THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Vision Systems for Autonomous Aircraft Guidance

Richard J. D. Moore

A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2012

Queensland Brain Institute

Abstract

Unmanned aerial vehicles (UAVs) are increasingly being used for a wide variety of civil and commercial applications such as infrastructure inspection and maintenance, search and rescue, mapping and cartography, as well as agricultural and environmental monitoring, to name just a few. Unmanned aircraft are suited to these roles because they can be smaller and lighter than manned aircraft and hence cheaper to operate, as well as being able to perform dull or repetitive tasks with greater precision than human operators, and dangerous tasks with greater safety. With the expanding set of roles for UAVs, there is an increasing need for them to be able to fly with a degree of low-level autonomy, thus freeing up their human controllers to concentrate on high-level decisions.

Modern UAVs control their position and orientation in space using technologies such as the Global Positioning System (GPS) and attitude and heading reference systems (AHRSs). They are unable to detect or avoid other objects or vehicles using these systems only, however, rendering them incapable of operating autonomously in near-Earth environments or around other moving vehicles. In these situations the aircraft must be able to monitor its surroundings continuously. Active proximity sensors, such as laser range-finders or radar, can be bulky, stealth-compromising, high-power, and low-bandwidth – limiting their utility for small-scale UAVs. There is considerable benefit to be gained, therefore, by designing guidance systems that use passive sensing, such as vision.

This thesis builds on recent research into biological vision-based flight control strategies to demonstrate that such bioinspired methods can offer dramatically improved sensing and control efficiencies over more complex computer vision-based approaches. Furthermore, this thesis establishes that wide-angle vision systems enable a broad array of sensing and guidance strategies, which can be implemented in parallel, in turn enabling complex flight behaviours that would traditionally require sensing and processing architectures incompatible with small-scale UAVs.

Two wide-angle vision systems are developed during the course of this research as well as a number of novel vision-based sensing and guidance algorithms, enabling detection of oncoming obstacles, estimation of attitude and altitude, long-term tracking of features, and interception of independently moving objects. Using these systems, complex capabilities such as terrain following at low altitude, aerobatics, landing in an uncontrolled environment, as well as tracking and interception of

an independently moving object are all demonstrated for the first time using only computing resources available on board a small-scale UAV and using only vision for all sensing and guidance.

The findings of this thesis contribute towards a greater understanding of the minimum requirements – in terms of sensing and guidance architectures – for complex UAV behaviours; and the design methodologies proposed herein represent an important step towards full autonomy for small-scale airborne platforms, thereby contributing towards exploitation of UAVs for civil and commercial applications and bringing autonomous UAVs a step closer to the remarkable capabilities of their biological counterparts.

Declaration by the Author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the *Copyright Act 1968*.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Refereed conference proceedings

[**Moore et al., 2012**] R. J. D. Moore, S. Thurrowgood, M. V. Srinivasan, "Vision-only estimation of the wind field strength and direction from an aerial platform". In *Proc. IEEE/RSJ International Conference on Intelligent Robots and System*, Vilamoura, Portugal, Oct. 2012.

[**Moore et al., 2011c**] R. J. D. Moore, S. Thurrowgood, D. Soccol, D. Bland, M. V. Srinivasan, "A method for the visual estimation and control of 3-DOF attitude for UAVs". In *Proc. ARAA Australasian Conference on Robotics and Automation*, Melbourne, Australia, Dec. 2011.

[**Moore et al., 2011a**] R. J. D. Moore, S. Thurrowgood, D. Bland, D. Soccol, M. V. Srinivasan, "A fast and adaptive method for estimating UAV attitude from the visual horizon". In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4935–4940, 2011.

[**Thurrowgood et al., 2010**] S. Thurrowgood, R. J. D. Moore, D. Bland, D. Soccol, M. V. Srinivasan, "UAV attitude control using the visual horizon". In *Proc. ARAA Australasian Conference on Robotics and Automation*, Brisbane, Australia, Dec. 2010.

[**Moore et al., 2010**] R. J. D. Moore, S. Thurrowgood, D. Bland, D. Soccol, M. V. Srinivasan, "UAV altitude and attitude stabilisation using a coaxial stereo vision system". In *Proc. IEEE/RAS International Conference on Robotics and Automation*, pages 29–34, 2010.

[**Moore et al., 2009**] R. J. D. Moore, S. Thurrowgood, D. Bland, D. Soccol, M. V. Srinivasan, "A stereo vision system for UAV guidance". In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3386–3391, 2009.

[**Thurrowgood et al., 2009**] S. Thurrowgood, D. Soccol, R. J. D. Moore, D. Bland, M. V. Srinivasan, "A vision based system for attitude estimation of UAVs". In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5725–5730, 2009.

Book chapters

[**Srinivasan et al., 2012**] M. V. Srinivasan, R. J. D. Moore, S. Thurrowgood, D. Soccol, D. Bland, "Insect vision and applications to robotics". In F. G. Barth, J. A. C. Humphrey, M. V. Srinivasan (Eds.), *Frontiers in Sensing: From Biology to Engineering*, pages 19–39, Springer-Verlag, Wien, 2012.

[**Moore et al., 2011b**] R. J. D. Moore, S. Thurrowgood, D. Soccol, D. Bland, M. V. Srinivasan, "A bio-inspired stereo vision system for guidance of autonomous aircraft". In A. Bhatti (Ed.), *Advances in Theory and Applications of Stereo Vision*, pages 305–326, InTech, Rijeka, 2011.

Publications included in this thesis

Moore et al. [2011b] incorporated as portions of Chapters 1–4.

Contributors: R. J. D. M., design (50%), implementation (70%), analysis (100%), article (90%); S. T., design (30%), implementation (20%), article (5%); D. B., implementation (5%) and other assistance; D. S., implementation (5%) and other assistance; M. V. S., design (20%), article (5%) and other assistance.

Moore et al. [2009] incorporated as portions of Chapters 3 & 4.

Contributors as above.

Moore et al. [2010] incorporated as portions of Chapters 3 & 4.

Contributors as above.

Moore et al. [2011a] incorporated as portions of Chapter 6.

Contributors: R. J. D. M., design (80%), implementation (80%), analysis (100%), article (90%); S. T., design (20%), implementation (10%), article (5%); D. B., implementation (5%) and other assistance; D. S., implementation (5%) and other assistance; M. V. S., article (5%) and other assistance.

Moore et al. [2011c] incorporated as portions of Chapter 6.

Contributors: R. J. D. M., design (100%), implementation (80%), analysis (100%), article (90%); S. T., implementation (10%), article (5%); D. B., implementation (5%) and other assistance; D. S., implementation (5%) and other assistance; M. V. S., article (5%) and other assistance.

Moore et al. [2012] incorporated as Section 6.6.1.

Contributors: R. J. D. M., design (100%), implementation (75%), analysis (100%), article (90%); S. T., implementation (25%), article (5%); M. V. S., article (5%) and other assistance.

Contributions by others to this thesis

Aside from the contributions by others to publications incorporated within this thesis (listed above), the research and ideas, as well as the design and development of algorithms, experimental design, analysis and interpretation of results, and writing contained within this thesis are all entirely the candidate's own work. Where this research relies on work conducted jointly with others, and no appropriate prior publication is available, the candidate's contribution is stated explicitly.

Statement of parts of this thesis submitted to qualify for the award of another degree

None.

Acknowledgements

First and foremost I would like to thank my principal advisor, Prof. Mandyam Srinivasan, for his expert advice and generous support throughout the duration of my PhD. Without his direction and tireless dedication none of this research would have been possible. I am also very grateful for the invaluable guidance and advice provided to me by my associate advisor, Prof. Gordon Wyeth, by my milestone assessors, Dr. Ben Upcroft and Dr. Jonathan Roberts, and by my milestone chair, Prof. Geoff Goodhill, all of whom I thank for their tremendous support, both professional and personal, during my candidature.

I would like to thank the members of the biorobotics laboratory at the Queensland Brain Institute for their invaluable contributions to this research – especially Mr. Saul Thurrowgood, whose knowledge and vision were instrumental in realising many of the systems developed during this work; Mr. Dean Soccol, whose experience and handiwork turned groundbreaking research back into a functioning aircraft on many occasions; as well as Mr. Daniel Bland, Mr. Michael Knight, and Mr. Samuel Baker for their technical assistance and piloting skills.

A big thankyou to my friends and colleagues at the Queensland Brain Institute, Young Visionaries at the ARC Centre of Excellence in Vision Science, and Robotics and Automation Society student members from around the world for making this PhD such an enjoyable experience. Finally, to my family and friends, thankyou for your endless support, patience, and encouragement – especially my parents Ian and Marilyn, who have inspired and supported my love of science since an early age.

During my candidature I was supported by scholarships and travel grants from the Queensland Brain Institute and the ARC Centre of Excellence in Vision Science, which assisted with attendance at various national and international conferences, workshops, and laboratories. This work was also supported partly by US Army Research Office MURI ARMY-W911NF041076, Technical Monitor Dr Tom Doligalksi; US ONR Award N00014-04-1-0334; ARC Centre of Excellence Grant CE0561903; and a Queensland Smart State Premier's Fellowship (M. V. Srinivasan).

Keywords

visual guidance, autonomous robotics, unmanned aerial vehicle (UAV), bioinspired systems

Australian and New Zealand Standard Research Classification (ANZSRC)

080106 Image Processing (50%); 080101 Adaptive Agents and Intelligent Robotics (30%); 080104 Computer Vision (20%).

Fields of Research (FoR) classification

0801 Artificial Intelligence and Image Processing (70%); 0906 Electrical and Electronic Engineering (20%); 0901 Aerospace Engineering (10%).

Contents

List of Figures

List of Tables

List of Acronyms

AGL above ground level. 38, 42, 44, 54, 60, 62, 63, 65, 66, 68–70, 75, 124, 135, 157, 158, 186

AHRS attitude and heading reference system. i, 2

CATD constant absolute target direction. 164, 165, 167, 181, 198

DGPS differential GPS. 68

DOF degrees of freedom. 7, 8, 28, 73, 74, 81, 85, 88–93, 95, 97, 98, 100–102, 106–113, 115–119, 121, 128, 132, 133, 135, 136, 139, 146, 147, 149–151, 153–158, 161, 172, 177, 189, 198, 204

EKF extended Kalman filter. 162

FOC focus of contraction. 81, 141

FOE focus of expansion. 81, 141

FOV field of view. 3, 4, 12, 19, 20, 22, 25–27, 33, 35, 45–47, 76, 79, 81–88, 96, 103, 104, 106, 116, 118, 119, 138, 139, 141, 146, 154, 160, 188, 198

GPS the Global Positioning System. i, 2, 3, 16, 18, 39, 61, 69–71, 112, 113, 120–122, 127, 129, 130, 133, 152, 155, 163, 177, 179–188, 191, 192

IMU inertial measurement unit. 15, 17, 18, 20, 54, 56, 57, 59–61, 63–68, 70, 71, 74, 86, 89, 106, 108–111, 113, 115–117, 120–122, 133, 136–138, 170

LOS line-of-sight. 152, 153, 155, 163, 164, 172, 177–180, 185–192, 195–198

LUT lookup table. 25, 45

MAV micro air vehicle. 43, 163, 200, 202

MRF Markov random field. 92

NED North-East-Down. 49, 157

PD proportional-derivative. 195

PID proportional-integral-derivative. 21, 58–61, 64, 68, 112, 113, 133, 135, 172, 175

PN proportional navigation. 165–167, 190, 193, 198

RANSAC RANdom SAmple Consensus. 53, 58, 69, 72

RGB red-green-blue. 86, 91, 92, 94, 139, 146–149, 159

RHS right-hand side. 169

RMSE root mean square error. 115, 117

ROI region of interest. 146, 148

SAD sum of absolute differences. 28, 52, 97, 105

SFM structure from motion. 17–19, 140, 141

SLAM simultaneous localisation and mapping. 17, 18, 160, 205

SNR signal-to-noise ratio. 54

SSE sum of squared errors. 130

SVM support vector machine. 92, 145

TTC time to contact. 152, 191–193

UA unmanned aircraft. 8

UAS unmanned aircraft systems. 8

UAV unmanned aerial vehicle. i, ii, 1–8, 11, 12, 14–16, 18–21, 38, 42, 43, 45–48, 56, 70, 73, 76, 79, 84, 85, 88–90, 102, 106, 113, 115, 119, 120, 131–133, 135, 138, 139, 145, 150, 160–163, 165, 172, 177, 183, 185, 188–190, 198–201, 203–206

UKF unscented Kalman filter. 162

VTOL vertical take-off and landing. 1

YUV luminance-chrominance. 91, 92, 94, 95, 98–100, 102, 118, 145

1

Vision Systems for Guidance of Autonomous Aircraft

1.1 Visual guidance for autonomous aircraft

Unmanned aerial vehicles (UAVs) have seen unprecedented levels of growth in both military and civilian applications since their inception during World War I. So much so, in fact, that the Joint Strike Fighter, currently under production, is predicted to be the last manned aircraft produced by the U.S. Armed Forces [Valavanis, 2007], and the U.S. military already trains twice as many ground operators for its UAVs as pilots for its military jets [Ross, 2011]. The first pilotless aircraft were intended for use as aerial torpedoes. Today, however, autonomous or semi-autonomous fixed-wing aircraft, airships, or helicopters and vertical take-off and landing (VTOL) rotorcraft (see Figure 1.1) are increasingly used for applications such as surveillance and reconnaissance, mapping and cartography, border patrol, infrastructure inspection and maintenance, military and defence missions, search and rescue operations, law enforcement, fire detection and fighting, agricultural and environmental imaging and monitoring, traffic monitoring, *ad hoc* communication networks, and extraterrestrial exploration, to name just a few.

The reason that UAVs are increasingly preferred for these roles is that they are able to operate in situations that are either too dangerous, too distant, too dull, or too difficult for manned aircraft [Valavanis, 2007]. Today's UAVs are typically flown remotely by a human pilot. However, the expanding set of roles means there is an increasing need for UAVs to be able to fly with a degree of low-level autonomy, thus freeing up their human controllers to concentrate on high-level decisions.

Figure 1.1: The many and varied designs for modern UAVs. From top-left to bottom-right: Rafael Skylite [Defense Update, 2005], AeroVironment Hornet [UAS Roadmap, 2005], AeronVironment Raven [UAS Roadmap, 2005], Aurora Flight Sciences Organic Air Vehicle [UAS Roadmap, 2005], Draganfly Innovations Inc. Stabilized Aerial Video System [Draganfly Innovations Inc., 2012], Elbit Systems Seagull [Defense Update, 2005], Sikorsky Aircraft Corp. Cypher II [Valavanis, 2007], Boeing X-50 Canard Rotor Wing [UAS Roadmap, 2005], ARA Battlefield Air Targeting Camera Micro Air Vehicle [UAS Roadmap, 2005], Honeywell Micro Aerial Vehicle [UAS Roadmap, 2005], Lockheed–Martin Force Protection Aerial Surveillance System [UAS Roadmap, 2005], University of Sydney T-wing [Stone, 1999], Aurora Flight Systems Corp. Golden Eye 100 [UAS Roadmap, 2005].

1.1.1 Short-range navigation

Modern UAVs are capable of controlling their position and orientation in space using technologies such as the Global Positioning System (GPS) and attitude and heading reference systems (AHRSs). This is sufficient when navigating over large distances at high altitude or in controlled airspaces. However, the expanding set of roles for UAVs increasingly calls for them to operate in near-Earth environments, and in environments containing 3D structures and obstacles. In such situations the UAV must know its position with respect to the surrounding environment accurately, which can be difficult using GPS alone due to occlusions and signal reflections from buildings and other objects. Moreover, the UAV must know *a priori* the 3D structure of the surrounding environment in order to avoid obstacles. Obviously such a scheme presents severe difficulties in situations where there is no foreknowledge of the 3D structure of the environment, or where this structure can change unpredictably.

A more efficient approach would be for the aircraft to monitor its surroundings continuously during flight, without relying on a previously constructed map. Active proximity sensors, such as laser range-finders (*e.g.* [Scherer et al., 2007]), or radar (*e.g.* [Viquerat et al., 2008]) have been considered for this purpose. However, such systems can be bulky, stealth-compromising, high-power, and low-bandwidth – limiting their utility for small-scale UAVs. There is considerable benefit to be gained, therefore, by designing guidance systems for UAVs that use passive sensing such as vision.

1.1.2 Visual guidance

The importance of vision for short range navigation was realised many decades ago. It was only recently, however, that vision-based guidance systems were demonstrated successfully on board real robots outside a controlled laboratory environment (for a review see DeSouza and Kak [2002]). The difficulty is that vision provides such a wealth of information about the surrounding environment and the self-motion of the vehicle, that it can be a laborious task to extract the information necessary for robot guidance.

For many insects, vision provides the primary sensory input for stabilisation of flight, detection of prey or predators, and interaction with other conspecifics – despite these animals possessing only very limited processing capabilities. Over the last two decades, a significant amount of research has shown that biological visual systems can inspire novel, vision-based solutions to some of the challenges facing autonomous aircraft guidance (see Floreano et al. [2009], Franceschini [2004], Srinivasan et al. [2009, 2012] for reviews).

1.2 Problem description

Recent research into bioinspired visual systems has resulted in significant advances to the field of UAV guidance. Miniature optic flow devices, for example, have been developed and used to sense aircraft egomotion and to detect obstacles, thus enabling autonomous flight in near-Earth environments with no reliance on external beacons such as GPS. While lightweight and efficient, these devices typically possess narrow fields of view (FOVs) and are single-purpose, permitting simple behaviours only. The aim of the current research is to investigate pairing of simple bioinspired sensing strategies with wide-angle vision systems. It is hypothesised that this synthesis will

allow for a much broader array of guidance algorithms, which can be executed in parallel to demonstrate flight behaviours more advanced than current state-of-the-art techniques can achieve.

By designing and using simple sensing and guidance algorithms, coupled with wide-FOV vision systems, this thesis will demonstrate complex UAV behaviours, such as landing in an uncontrolled environment and interception of an independently moving vehicle, using vision as the primary sensory input for all levels of control – from low-level flight control to high-level guidance. As a first step, this will involve design and implementation of a vision-based system enabling autonomous control of the fundamentals of UAV flight, such as stabilisation of attitude and altitude. In turn, this will facilitate investigation of more complex capabilities such as terrain-following and collision avoidance, landing, as well as visual tracking and interception. Traditionally such complex behaviours would require sensing and processing architectures incompatible with small-scale UAVs. The ultimate goal for this thesis is to better understand the minimum requirements for such high-level capabilities and to demonstrate these on board a small-scale UAV operating in the field.

1.3 Research contributions

This thesis investigates the use of vision as a sensory modality for providing guidance to an autonomous aircraft. A key outcome of this work is an increased control efficiency for an autonomously guided UAV using a bioinspired methodology. Contributing to this outcome are the findings that wide-angle vision is superior to sensors with a limited FOV; and parallel processing of visual information in a task-specific way increases autonomous capabilities. Two wide-angle vision systems are developed within this work as well as a number of novel vision-based sensing and guidance algorithms, enabling a range of complex flight behaviours to be demonstrated for the first time using only visual input and computing resources available on board a small-scale UAV.

Specifically, this thesis contributes:

- A stereo vision-based system enabling an aircraft to detect objects that penetrate a notional free-flight cylinder surrounding its projected flight path (Chapter 3); and to measure and control its attitude and altitude with respect to the ground plane (Chapter 4).

- An omnidirectional vision-based system enabling an aircraft to estimate and control its roll, pitch, and heading efficiently and robustly, based on position and orientation of the horizon and appearance of the sky (Chapter 6).
- An algorithm enabling an aircraft to visually track moving ground-based objects in an outdoor environment over the course of an entire flight and throughout occlusions, based on appearance of the object and estimated motion of the aircraft (Chapter 7).
- A guidance and control strategy enabling a fixed-wing aircraft to predict the future trajectory of a constant-velocity object and to intercept the object via the shortest possible 3D path, using only the apparent viewing direction to the object in the inertial reference frame (Chapter 8).
- A small-scale (2m wingspan) UAV capable of flying fully autonomously and completely independently in an outdoor environment and at low altitude in the presence of natural obstacles (Chapters 3 & 4); capable of performing aerobatic manoeuvres autonomously, and landing automatically in an uncontrolled environment (Chapter 6); and capable of tracking and intercepting an independently moving ground-based vehicle using only vision and on-board computing resources (Chapters 7 & 8).

The flight platform used in this work is pictured in Figure 1.2. A fixed-wing aircraft was selected over alternative platforms (*i.e.* a rotorcraft or flapping-wing vehicle) because of its superior efficiency, payload capacity, and endurance in the flight regime investigated in this work – an outdoor, non-urban environment. A fixed-wing platform is therefore likely to be better suited to many of the roles for small-scale autonomous aircraft discussed in Section 1.1. Additionally, the research platform used here facilitated a very quick development pathway, which allowed sensing and guidance algorithms to be implemented and tested without the prior need to develop complicated actuation and/or control systems.

1.4 Organisation of the thesis

Research presented within this thesis is organised into two parts. Each part covers design, development, and implementation of a range of sensing and guidance strategies and introduces a novel vision system designed specifically to complement those

Figure 1.2: The author conducting pre-flight checks on a small-scale UAV.

algorithms. Ordering of chapters loosely follows a chronology of the research that was conducted, and moves from development of low-level flight control strategies to higher-level guidance of complex behaviours. Each results chapter presents a distinct body of research on a particular guidance scheme, including a review of relevant prior research; a discussion of the design, development, and implementation of the method on board a small-scale UAV; an analysis of open- and closed-loop flight testing results; a discussion of the particular method's limitations and comparisons to other state-of-the-art techniques; examples of applications of the scheme to real-world scenarios for a small-scale UAV, where applicable; and finally, a concluding statement that summarises the contributions of that chapter within the context of the aims of this thesis.

Remaining chapters are summarised briefly below:

Part I Stereo Vision-based Guidance The first part of this thesis investigates a novel stereo vision system designed specifically for providing guidance to an autonomous UAV flying at low altitude in the presence of natural obstacles.

Chapter 2 Introduction to Stereo Vision-based Guidance An introductory chapter that provides some background research on visual guidance and contrasts the use of stereo vision with other visual guidance strategies for small-scale UAVs flying in near-Earth environments. This chapter also

gives an overview of the research presented in the remainder of Part I.

Chapter 3 A Stereo Vision System for Autonomous Aircraft Guidance This chapter discusses development and implementation of the stereo vision system used to investigate collision avoidance and terrain following behaviours in Part I. A technique for identifying obstacles penetrating a notional cylinder surrounding the projected flight path is described and implemented using this vision system. Autonomous collision avoidance is demonstrated by a small-scale UAV flying at low altitude in the presence of natural obstacles.

Chapter 4 Autonomous Stabilisation of Attitude and Altitude Two procedures for estimating and controlling altitude and 2-degrees of freedom (DOF) attitude (*i.e.* roll and pitch) of a low-flying aircraft with respect to the ground plane are described, both of which use the stereo vision system introduced in Chapter 3. The most computationally efficient approach is implemented within a closed-loop control system to demonstrate fast and low flight above uneven terrain with a small-scale UAV.

Part II Omnidirection Vision-based Guidance The second part of this thesis investigates several bioinspired sensing and guidance algorithms using a novel omnidirectional vision system designed specifically for estimating an aircraft's 3-DOF attitude and egomotion robustly in real-world conditions.

Chapter 5 Introduction to Omnidirectional Vision-based Guidance An introductory chapter that surveys the advantages of omnidirectional vision over stereo vision for the purpose of aircraft guidance. An omnidirectional vision system inspired partly by the honeybee is described, which serves as the platform for research conducted within Part II. This chapter also gives an overview of the remainder of the second part of this thesis.

Chapter 6 Visual Estimation of 3-DOF Attitude A novel procedure for estimating an aircraft's 3-DOF attitude (roll, pitch, and heading) from the visual horizon and appearance of the sky is described in this chapter. This approach is implemented within a closed-loop control system to demonstrate autonomous control of 3-DOF attitude during aerobatic manoeuvres and whilst performing an automatic landing.

Chapter 7 Visual Target Detection and Tracking Various vision-based approaches for detecting and tracking independently moving objects from a moving vehicle are reviewed. A computationally lightweight algorithm is presented, which enables an aircraft to visually track moving ground-based objects over extended frame sequences and throughout occlusions.

Chapter 8 Vision-based Interception of Moving Objects This chapter describes a novel guidance and control strategy, based on the predatory flights of dragonflies, which enables an aircraft to predict the future trajectory of a constant-velocity object and to intercept the object via the shortest possible 3D path, using only the apparent viewing direction to the target, obtained from the appearance-based tracker described in Chapter 7, and the 3-DOF attitude of the aircraft in the inertial reference frame, obtained from the visual method described in Chapter 6.

Chapter 9 General Conclusions and Outlook The final chapter of this thesis summarises the findings of Parts I & II; describes the outlook for ongoing and future research; and provides some discussion of the relevance of the outcomes of this thesis to the wider field of research.

1.4.1 Media attachments

Several media attachments are referenced throughout this thesis. If the attachments are not available with this document, all of them may be accessed at https://www.dropbox.com/sh/2ftorfp4tgf26vx/Yt3wb4D97j; some attachments are also available at http://www.youtube.com/qbibiorobotics and are labelled as such.

1.4.2 A note on terminology

Usage of the acronym UAV has recently been phased out by the Federal Aviation Administration in the United States as well as the Civil Aviation Safety Authority in Australia in favour of the term unmanned aircraft systems (UAS), which refers to the entire (ground-based and aerial) infrastructure supporting an unmanned aircraft (UA). The terms UAV and autonomous UAV are used interchangeably throughout this thesis for consistency with prior publications and the wider literature.

Part I

Stereo Vision-based Guidance

2

Introduction to Stereo Vision-based Guidance

2.1 Summary

Part I of this thesis investigates a stereo vision system designed specifically for providing guidance to an autonomous UAV flying at low altitude in the presence of natural obstacles. Stereo vision is useful for aircraft guidance because it enables a direct measure of range to obstacles in the environment irrespective of the motion of the aircraft. Stereo vision is contrasted with other visual guidance strategies in Section 2.2 and an overview of the remainder of Part I is presented in Section 2.3.

Portions of this chapter are contained within Moore et al. [2011b].

2.2 Motivations for stereo vision-based guidance

Biological vision systems have provided insight and inspiration for researchers in the field of visual guidance for decades. Visual processing systems of insects are particularly relevant for the design of visual guidance systems for small-scale UAVs because they are compact, robust, and computationally efficient. In this section, some approaches to visual guidance inspired by biology are discussed and compared with traditional computer vision-based approaches. Stereo vision addresses some limitations of both techniques and can provide an independent approach to visual guidance.

2.2.1 Biological vision

For many animals, vision provides the primary sensory input for stabilisation of flight, detection of prey or predators, and interaction with other conspecifics. Insects in particular provide a good study model because they have seemingly developed efficient and effective visual strategies to overcome many of the challenges facing autonomous guidance of small-scale UAVs. For instance, the humble housefly is often able to outwit even the most determined swatter, despite its tiny brain and relatively simple nervous system. In fact many flying insects have attained a level of skill, agility, autonomy, and circuit miniaturisation that greatly outperforms present-day aerial robots [Franceschini, 2004].

Unlike the eyes of vertebrates (including humans), insects' eyes are not independently moveable and have fixed-focus optics. Insects cannot, therefore, infer distances to objects in the environment using cues such as the gaze convergence or refractive power required to bring an object into focus on the retina. Furthermore, compared with human eyes, eyes of insects possess inferior spatial acuity and are positioned much closer together with limited overlapping FOVs. The precision with which insects could estimate range using stereopsis is thus limited to a few centimetres at most [Srinivasan et al., 1993].

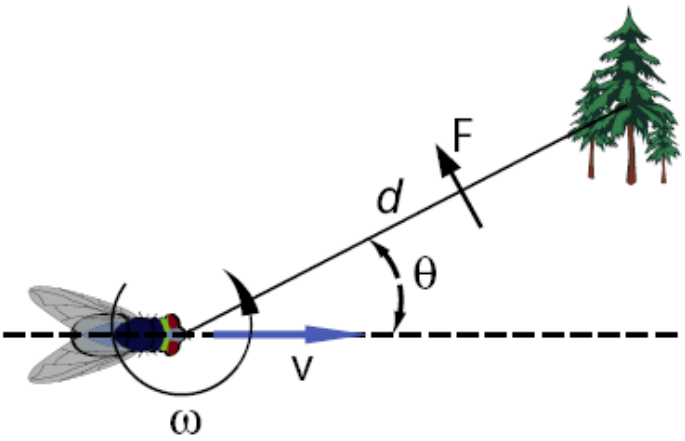


Figure 2.1: Optic flow, F, produced by an object as observed by an animal or vehicle in motion. Reproduced from Hrabar et al. [2005].

Not surprisingly then, insects have evolved alternative strategies for overcoming the challenges of visually guided flight [Srinivasan, 2011]. Many of these strategies rely on using image motion, or *optic flow*, generated by the insect's self-motion to infer distances to obstacles and to control various manoeuvres [Gibson, 1950, Nakayama

and Loomis, 1974, Srinivasan et al., 1993]. The relationship between optic flow and range to objects in the environment is remarkably simple: it depends only upon translational speed of the observer, v, distance to the obstacle, d, and θ, the angle to the object with respect to the translation direction of the observer (see Figure 2.1) [Nakayama and Loomis, 1974]. Optic flow generated by rotational motion of the observer, however, does not encode any information on range to objects and so must be discounted from the calculation. Alternatively, rotational movements of the vision system must be prevented and optic flow measured when the vision system is undergoing pure translation.

For an observer translating at a speed v, and rotating at angular velocity ω, the optic flow, F, generated by a stationary object at a distance d and angular bearing θ is given by

$$F = \frac{v \cdot \sin(\theta)}{d} - \omega. \tag{2.1}$$

To facilitate separation of rotation- and translation-induced optic flow, flying insects are known to stabilise their vision systems by performing compensatory head movements to cancel short-term rotations of the thorax during flight [Lindemann et al., 2005]. The rotational component of optic flow measured during longer-term movements, such as during turns, is minimised by performing short and fast head movements (*i.e. saccades*) interspersed with longer periods of fixation [Schilstra and van Hateren, 1999, van Hateren and Schilstra, 1999]. Exactly how insects are able to precisely compensate for rotations of the body remains an active area of study, although it is likely a combination of feed-forward control – generated from expected head motion during typical flight manoeuvres – and feed-back control from rotation rate measurements – obtained from inertial sensory organs and/or distilled from optic flow measurements by wide-field integrating neurons with sensitivities tuned for certain patterns of egomotion-induced optic flow [Franz et al., 2004].

In addition to their compound eyes, insects also possess a number of other sensory organs with which they can extract information from the environment relevant to controlling their flight. Flies use their *halteres* (small knobbed structures modified from hind wings in some two-winged insects) to measure rotation rates [Nalbach and Hengstenberg, 1994] and it is generally accepted that most flying insects gain rich inertial information from the numerous mechanosensors that are spread over most of their body. It is also widely believed that insects use their hair and antennae to feel air flow over their body and, in particular, estimate their airspeed [Dudley, 2000,

Taylor and Krapp, 2007]. Many insects also possess additional visual organs known as *ocelli* (literally "little eyes"). These simple eyes have very low resolution but are able to perceive light gradient and colour, as well as the polarisation pattern of the sky, thus facilitating stable and upright flight in an outdoor environment [Stange and Howard, 1979, Taylor, 1981, Wellington, 1974].

2.2.2 Bioinspired vision-based guidance systems

A significant amount of research over the past two decades has shown that biological vision systems can inspire novel, vision-based solutions to many of the challenges that must be overcome when designing guidance systems for autonomous aircraft (for reviews see Floreano et al. [2009], Franceschini [2004], Srinivasan [2011], Srinivasan et al. [2004, 2009]). It has been shown, for example, that honeybees use optic flow for negotiating narrow gaps and avoiding obstacles, regulating their flight speed and altitude, performing smooth landings, and for estimating their distance flown [Srinivasan and Zhang, 2004, Srinivasan et al., 2000]; an extensive list of studies has revealed that exquisite motion sensitivity in *dipteran* flies is fundamental to their flight control and ability to detect and chase other conspecifics (*e.g.* [Collett and Land, 1975, Schilstra and van Hateren, 1999, van Hateren and Schilstra, 1999]); and that locusts sense directly oncoming objects using flow divergence [Simmons and Rind, 1992]. A recent trend in biologically inspired vision systems for UAVs, therefore, has been to exploit optic flow information for collision avoidance, terrain and gorge following, as well as landing [Barrows and Neely, 2000, Barrows et al., 2003, Beyeler et al., 2009b, Conroy et al., 2009, Floreano et al., 2009, Kim and Brambley, 2007, Ruffier and Franceschini, 2005, Srinivasan et al., 2004, 2009].

The magnitude of the optic flow gives a measure of the ratio of the aircraft's ground speed to its distance from objects in the environment. Ruffier and Franceschini [2005] demonstrated that both forward speed and altitude could be regulated using a single optic flow detector (*e.g.* Figure 2.2) that was artificially maintained at a constant vertical orientation. Altitude control for cruise flight has also been demonstrated on board real UAV platforms [Barrows and Neely, 2000, Barrows et al., 2003, Chahl et al., 2004, Green et al., 2003, 2004, Oh et al., 2004] by regulating ventral longitudinal optic flow observed from the aircraft. While functional, these early systems were limited by their failure to take pitching motions of the aircraft into account (which also induce optic flow, but provide no information on range), and their passive or artificial

stabilisation of roll. Garratt and Chahl [2008] also controlled altitude via optic flow and additionally corrected for pitching motions of the aircraft using an inertial measurement unit (IMU), but did not take attitude of the aircraft into consideration. Neumann and Bulthoff [2001, 2002] used a similar strategy in simulation but regulated attitude using colour gradients present in the simulated test environment. Using similar principles, Moore et al. [2011a], Thurrowgood et al. [2009, 2010], Todorovic and Nechyba [2004] demonstrated methods for controlling UAV attitude based on the apparent orientation of the horizon. Likewise, Thakoor et al. [2002, 2003] described an attitude regulation scheme based on the operation of insect *ocelli*.

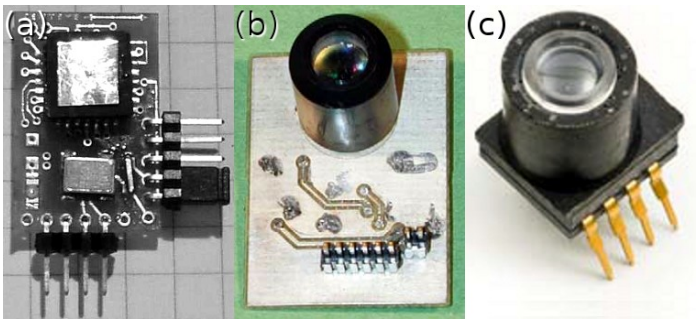


Figure 2.2: Examples of the specialised sensors used to measure optic flow from (a) Green et al. [2003], (b) Ruffier and Franceschini [2005], (c) and Beyeler et al. [2009b]. Sensor footprints are all in the order of 1cm^2.

It has been proposed that insects such as honeybees navigate through narrow openings and avoid obstacles by balancing optic flow observed on either side of their body [Srinivasan et al., 1991] and by turning away from regions of high optic flow [Srinivasan, 1993, Srinivasan and Lehrer, 1984, Srinivasan and Zhang, 1997]. Similar strategies have been employed by Conroy et al. [2009], Green [2007], Green et al. [2004], Hrabar and Sukhatme [2009], Oh et al. [2004], Zufferey and Floreano [2006], Zufferey et al. [2006] to demonstrate lateral obstacle avoidance in aircraft. Beyeler [2009], Beyeler et al. [2007, 2009b], Zufferey et al. [2008] steered their micro-UAV to avoid obstacles in three dimensions and additionally incorporated rate gyroscopes and an anemometer to account for rotational motions and airspeed of the aircraft respectively.

Studies of insect behaviour have also revealed novel strategies that may be used for controlling complex flight manoeuvres. It has been observed that when honeybees land, they tend to regulate their forward speed proportionally to their height such that optic flow produced by the landing surface remains constant [Srinivasan et al.,

2000]. As their height approaches zero, so does their forward speed, ensuring a safe, low speed at touch down for the bee. Similar strategies have been employed by Beyeler et al. [2009a], Chahl et al. [2004], Green et al. [2003, 2004], Oh et al. [2004] to demonstrate autonomous take-off and landing with small-scale UAVs. A purely visual approach for automatically landing a fixed-wing UAV in an uncontrolled environment that is based on this strategy is described in Section 6.6.2.

The successes of the approaches described above show clearly that measuring optic flow produced by an aircraft's motion through the environment is a viable means of providing guidance information for various manoeuvres. To retrieve accurate estimates of range from optic flow, however, the aircraft's ground speed must be measured independently so that this information can be combined with optic flow measurements to estimate the true ranges to objects in the environment. In practice this requires additional sensors, such as high-precision GPS or a Pitot tube[1]. In the case of the latter, however, the variable measured is actually airspeed, which would lead to incorrect range estimates for all but the case of low-altitude flight in still air. Furthermore, altitude perceived from a downward-facing optic flow sensor is not only dependent upon an aircraft's true altitude and ground speed, but also it's attitude. This is particularly relevant to fixed-wing aircraft in which relatively high roll and pitch angles are required to perform rapid manoeuvres. A method for overcoming these shortcomings is described by Beyeler et al. [2006], who interpolate images of the ground plane using measured flight data and compare with captured images to infer altitude and pitch angle simultaneously. Their technique is too restrictive to be implemented outdoors, however, as it does not permit rolling motions and requires ground speed to be estimated accurately.

Finally, and most significantly, systems that rely on optic flow for extracting range information need to discount the component of optic flow induced by *rotational* motion of the aircraft. This is because it is only the component of optic flow that is induced by *translational* motion of the observer that encodes information on range (see Equation 2.1). Discounting the rotation-induced component of optic flow requires either numerical estimates of the roll, pitch, and yaw rates of the aircraft, which are often noisy, or additional apparatus for their explicit measurement, such as a three-axis gyroscope. This approach can suffer from poor synchronisation between the multiple sensory modalities, but has been used to directly subtract the aircraft's rotational motion from the measured optic flow field [Zufferey and Floreano, 2006]

[1] A pressure measurement instrument used to compute fluid flow velocity.

and to predict visual motion caused by the rotational motion of the aircraft [Kendoul et al., 2009a].

Alternatively, if optic flow is measured from sufficiently varied viewing directions simultaneously, the rotational and translational components can be separated through application of wide-field motion templates [Dahmen et al., 2001, Franz and Krapp, 2000, Franz et al., 2004, Shoemaker et al., 2011]; by direct minimisation [Koenderink and Doorn, 1987]; or via an iterative, 'hypothesise and test' approach. These processes inherently require environmental structure and egomotion of the observer to be estimated simultaneously – known as obtaining structure from motion (SFM)[2]. This technique forms the basis for the classical computer vision approach to visual guidance and is discussed in greater detail below.

2.2.3 Computer vision

The SFM approach to short-range navigation dates back to works such as Harris and Pike [1988], Longuet-Higgins [1981]. The basis of this approach consists of estimating relative pose of the observer and 3D structure of the environment simultaneously from a set of camera images. For the monocular case, scene depth and camera motion are obtained from image correspondences between multiple camera frames as the vehicle moves through the environment. Image correspondences can be obtained from optic flow or feature-based techniques. In general, the problem of SFM is underdetermined and its solution requires a computationally intensive, 'hypothesise and test' approach. Monocular approaches compatible with aerial-based navigation in structured environments have been described (*e.g.* Mandelbaum et al. [1999], Merrell et al. [2004], Nister et al. [2004]), but they require recursive integration of measurements over time until reliable structure is obtained and they have not been implemented in real time on board aerial vehicles. To achieve autonomous flight, attitude control is typically performed on board the aircraft using IMU-based controllers and the visual information is transferred to a ground-based workstation, which computes SFM and relays flight commands back to the aircraft (*e.g.* Kendoul et al. [2009a]).

Simultaneous localisation and mapping (SLAM) is an extension of the SFM problem that aims to reduce errors accumulated in the estimated pose of the

[2] *Structure from motion* is a widely used term describing a process for estimating structure *and* egomotion simultaneously, despite the denotation of the term.

vehicle and structure of the environment, which occur over long periods of time (refer to Bailey and Durrant-Whyte [2006], Durrant-Whyte and Bailey [2006] for a detailed overview of SLAM). Efficient, vision-based SLAM implementations have been described for estimating pose of the observer and structure of the environment in real time [Klein and Murray, 2007]. Such approaches have been implemented recently on board rotorcraft to demonstrate autonomous hover and position control of small-scale UAVs in outdoor environments [Achtelik et al., 2011, Weiss et al., 2011]. Due to the relatively high latency and low update rates of visual SLAM algorithms, however, dynamics of the aircraft were neccessarily stabilised by IMU-based controllers.

In general, the problems of estimating environmental structure and camera egomotion are mutally dependent. Prior knowledge of vehicular motion allows scene structure to be computed directly through triangulation of corresponding image points. Vehicular self-motion may be constrained to simpler motions [Amidi et al., 1999, Chahl and Srinivasan, 1996, Stein et al., 2000] or measured by additional sensors such as an IMU [Amidi et al., 1999, Han and DeSouza, 2009, Hwangbo, 2009, Kendoul et al., 2009b, 2010], anemometer [Hwangbo, 2009], or GPS [Han and DeSouza, 2009]. Conversely, prior knowledge of scene structure enables direct computation of egomotion. This is exploited by Amidi et al. [1999], Stein et al. [2000], Suzuki and Kanade [1999], who constrain scenes to simple 2D environments. Alternatively, environmental structure can be measured, either *a priori* or online, using active sensors or stereo vision [Agrawal et al., 2005, Badino et al., 2006, Demirdjian and Horaud, 2000, Franke et al., 2005, 2008, Maimone et al., 2007, Mallet et al., 2000, Mandelbaum et al., 1999, Nister et al., 2004, Olson et al., 2003, Rabe et al., 2007, Talukder and Mathies, 2004]. The complexity of the SFM computation can thus be reduced significantly by measuring scene structure directly via stereo vision. The observer's egomotion can then be computed simply from feature correspondences between image frames, or optic flow. Such approaches have been implemented recently on board small-scale UAVs using stereo sensors such as the Microsoft Kinect[3] to demonstrate autonomous mapping and navigation within indoor environments [Heng et al., 2011, Huang et al., 2011, Lange et al., 2012], albeit under near-hover flight conditions.

[3] http://www.microsoft.com/en-us/kinectforwindows/

2.2.4 Stereo vision

SFM-based approaches can provide a detailed representation of vehicular egomotion as well as environmental structure, as discussed above. The computational complexity of this technique can be reduced if independent measurements of scene structure are available. Despite recent technological advance, however, such approaches have thus far only been able to demonstrate autonomous control of low-speed manoeuvres, such as hovering with rotorcraft, due to the computational constraints on board small-scale UAVs. Sensing and control schemes that are purely reactive may therefore offer a more efficient approach to visual guidance for UAVs performing specific tasks.

In Section 2.2.2, several previously published approaches were discussed in which optic flow was used directly to steer an aircraft around obstacles and to perform various manoeuvres. These approaches used specialised optic flow sensors (*i.e.* Figure 2.2) and represent the state of the art for visual guidance of small-scale fixed-wing UAVs, which are required to fly at high speed and low altitude during manoeuvres such as terrain following or, critically, during take-off and landing. Guidance systems based solely on optic flow, however, face the difficulty of having to 'de-rotate' the measured optic flow field in order to extract useful measurements of the range to objects in the environment. Additionally, the optic flow sensors pictured in Figure 2.2 possess narrow FOVs, which means that systems based on these devices must account for attitude and direction of translation of the aircraft in addition to rotational motion in order to compute valid estimates of altitude and range.

Stereo vision, on the other hand, enables the range to objects in the environment to be measured directly, irrespective of the rotational or translational motion of the aircraft. Section 2.2.3 showed how this property enables stereo vision to simplify the problem of computing a vehicle's *motion* from the *structure* of the environment via SFM. For high-speed flight at low altitude, however, stereo vision may be more useful when used independently as a reactive guidance system, since it facilitates direct computation of range to obstacles without the complicating factors experienced by optic flow-based approaches. Furthermore, for stereo systems, the search for image correspondences is constrained to a single dimension, thus reducing the complexity of the required computation. Stereo vision therefore provides an attractive approach to solving some of the problems facing the guidance of autonomous aircraft in low-altitude or cluttered environments.

2.2.5 Stereo vision-based guidance systems

Several stereo vision systems designed for reactive guidance of autonomous UAVs exist already. An altitude regulation scheme was presented by Roberts et al. [2002, 2003], who used a downward-facing stereo system; Hrabar and Sukhatme [2009], Hrabar et al. [2005] used a combined stereo and optic flow approach to navigate urban canyons and avoid frontal obstacles; and Andert and Adolf [2009], Andert et al. [2011] use stereo vision to detect obstacles and plan a safe flight path on-the-fly. All of these approaches, however, required attitude to be regulated with an IMU on board the aircraft.

In Chapter 3, a wide-angle stereo vision-based guidance system is introduced, which enables a small-scale UAV to detect and avoid oncoming obstacles efficiently. It is further shown in Chapter 4 that the wide-FOV of this new vision system permits attitude and altitude of an aircraft to be estimated accurately with respect to the ground plane. Behaviours specific to aircraft guidance such as terrain and gorge following, obstacle avoidance, and landing have not been demonstrated with existing wide-angle stereo vision systems [Thurrowgood et al., 2007, Tisse et al., 2007b]. In the first part of this thesis, a stereo vision system is described that is specifically designed to serve these requirements.

2.3 Outline of Part I

This research investigates whether simple sensing and guidance algorithms can be combined with wide-angle vision systems to demonstrate UAV behaviours that are beyond the capabilities of the state of the art for visual guidance of small-scale UAVs. Part I covers development and testing of a wide-angle stereo vision-based system designed to simplify computation of range and provide guidance for manoeuvres such as low-altitude terrain following and collision avoidance. The concept and design of the vision system is described in Chapter 3, along with a simple control strategy, which enables an aircraft to sense and react to obstacles lying close to its projected flight trajectory. Performance of the guidance system is evaluated in indoor and outdoor trials. A more advanced control strategy is introduced in Chapter 4, which permits specific control of the aircraft's altitude and attitude (roll and pitch). Flight tests demonstrate practicality and usefulness of this system for providing guidance to an aircraft performing low-altitude terrain following in the presence of obstacles.

3

A Stereo Vision System for Autonomous Aircraft Guidance

3.1 Summary

Stereo vision provides an attractive approach for overcoming some of the challenges facing guidance of autonomous aircraft in low-altitude environments, because accurate and dense range maps can be produced at a lower computational cost than equivalent monocular approaches. A wide-angle stereo vision system is introduced in this chapter that is tailored to the specific needs of aircraft guidance at low altitudes. The vision system was originally conceived to simplify computation of range from optic flow, but this design also makes it well suited to operation as a coaxial stereo system.

In Section 3.2, the concept as well as the final design and implementation of the stereo vision system are discussed in detail. The system enables radial distances from the optic axis to objects in the environment to be computed simply. It is proposed that this characteristic reduces the complexity of extracting appropriate guidance commands from visual input. It is shown through indoor testing that range estimates obtained with this system are capable of representing simple 3D structures accurately. A technique for identifying and segmenting obstacles that present a danger to the aircraft is presented in Section 3.3. When combined with a simple proportional-integral-derivative (PID)-based control scheme, this allows an aircraft to automatically detect and avoid obstacles that conflict with its projected flight path. Autonomous collision avoidance is demonstrated with closed-loop flight tests in which a small-scale UAV flies at low altitude amongst natural obstacles. In Section 3.4, limitations and capabilities of the proposed approach are contrasted with

alternative vision-based approaches and are found to compare favourably.

Portions of this chapter are contained within Moore et al. [2009, 2010, 2011b].

3.2 A new stereo vision system for UAV guidance

In this section, a novel wide-angle stereo vision system is introduced, which enables simple computation of the radial distances from the optic axis to objects in the environment. This information can be used directly to guide an autonomous aircraft around obstacles.

3.2.1 Conceptual design

The concept of the vision system is best described by considering an assembly in which a camera views a specially shaped reflective surface (a mirror). As well as increasing FOV of the camera, the profile of the mirror is designed such that equally spaced points on the ground plane, on a line parallel to the camera's optic axis, are imaged to points that are equally spaced in the camera's image plane. This has the effect of removing perspective distortion (and therefore distortion in image motion) that a camera experiences when viewing a horizontal plane stretching out to infinity in front of the aircraft. The mapping produced by the mirror (*terrain following* mapping) is illustrated in Figure 3.1. It is clear that equal distances along the ground, parallel to the optic axis of the system, are mapped to equal distances on the image plane, validating the design of the mirror. The concept of the terrain following mirror was first described in Srinivasan et al. [2006] along with a full derivation of its profile. The vision system described here, however, differs significantly from the original system and the outcomes of this research far exceed those of Srinivasan et al. [2006].

The special geometric remapping afforded by the mirror means that, for a given vehicle speed, the motion in the camera's imaging plane of the image of an object in the environment is inversely proportional to the radial distance of that object from the optic axis of the vision system. Surfaces of constant image motion reprojected into the environment, therefore, are cylindrical, as illustrated in Figure 3.2. This property makes the system particularly useful for aircraft guidance. For any given aircraft speed, the maximum image velocity observed in the remapped image specifies the radius of a cylinder of space in front of the aircraft through which collision-free flight can occur. Characterising the collision-free space in front of the aircraft as a

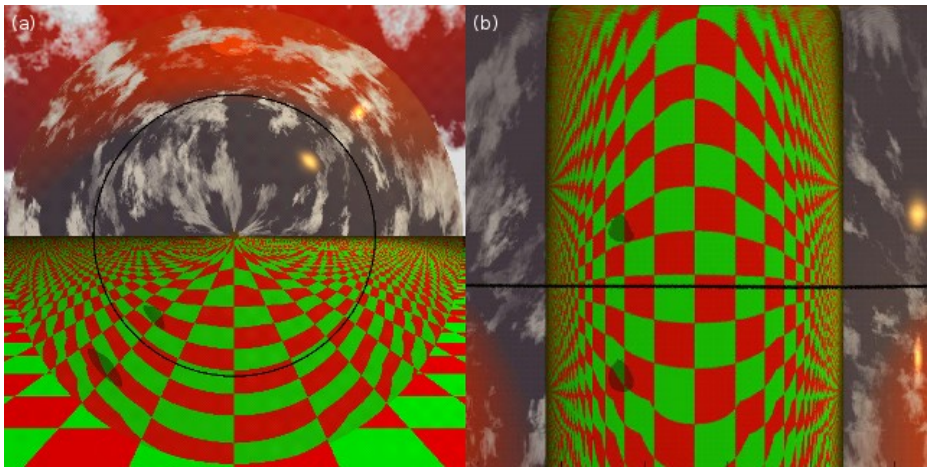


Figure 3.1: Visualisation of the imaging properties of the terrain following mirror showing (a) the raw image as viewed by the camera and (b) the remapped image. The remapped image is produced by mapping the polar coordinates of the pixels in the raw image to Cartesian coordinates in the remapped image. Only pixels that correspond to the reflection from the mirror surface in the raw image are remapped. The dark line indicates viewing directions at 90° to the camera's optic axis. Reproduced from Srinivasan et al. [2006].

virtual cylinder simplifies the problem of determining in advance whether an intended flight trajectory will be collision-free; and, therefore, of making any necessary course corrections to facilitate this.

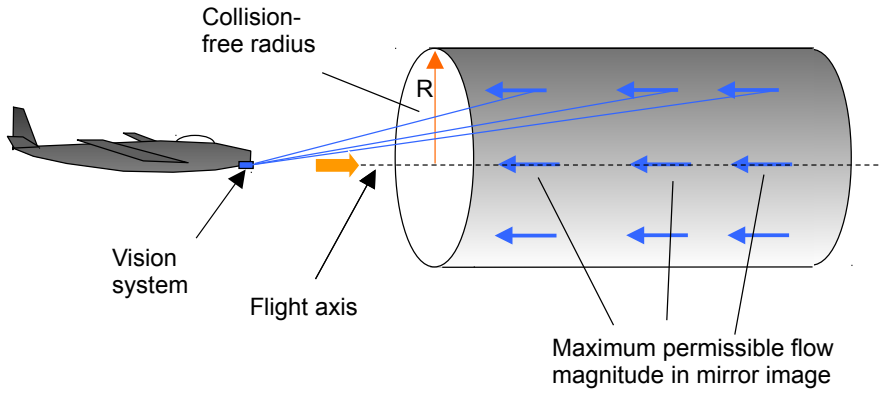


Figure 3.2: Illustration of the cylindrical clear-space mapping provided by the stereo vision system. Reproduced from Srinivasan et al. [2006].

Now consider a system in which two such camera-mirror assemblies are arranged coaxially, as illustrated in Figure 3.3. Each camera views the environment through a mirror that has the imaging properties described above. It follows that the pixel disparity, D_{pixel}, produced by a point imaged in both cameras is inversely proportional

to the radial distance, d_{radial}, of that point from the common optic axis of the two camera-mirror assemblies. The relationship is given by

$$D_{pixel} = \frac{d_{baseline} \cdot h_{image}}{r_{tot}} \cdot \frac{1}{d_{radial}}, \tag{3.1}$$

where $d_{baseline}$ is the stereo baseline; h_{image} is the vertical resolution of the remapped images; and r_{tot}, the forward viewing factor, is the ratio of the total visible distance along the ground to the height of the aircraft (*i.e.* the vertical angular extent of the remapped image).

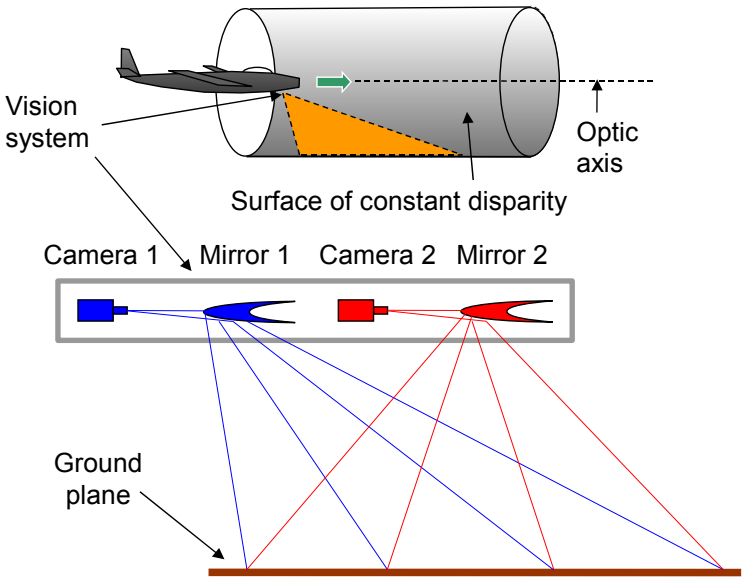


Figure 3.3: Schematic illustration of the conceptual stereo vision system, surface of constant disparity, and collision-free cylinder.

The first term in Equation 3.1 is simply a constant that depends on the configuration of the system. The maximum image disparity in a given stereo pair therefore directly defines the radius of a collision-free cylinder surrounding the optic axis, independent of the aircraft's speed. A simple control scheme may be implemented in which an aircraft is repelled from objects penetrating the notional flight cylinder required by the aircraft for collision-free flight. The image disparity will be one dimensional only, thereby reducing the complexity of the computation. This system is therefore well-suited to provide real-time information for visual guidance of an autonomous aircraft flying in the presence of obstacles.

	Arena test	Terrain following	Collision avoidance
Stereo baseline ($d_{baseline}$)	0.2m		
Remap image rows (h_{image}) / cols	384px / 128px		
Forward viewing factor (r_{tot})	2.5		3.5
Forward FOV from vertical	$20.6° - 70.8°$	$0° - 68.2°$	$0° - 74.0°$
Horizontal FOV from vertical	$-100°$ to $100°$		
Disparity search range (D_{pixel})	10px $\rightarrow$ 40px	0px $\rightarrow$ 15px	
Measurable distance range (d_{radial})	0.8m $\rightarrow$ 3.1m	2.0m $\rightarrow \infty$	1.5m $\rightarrow \infty$

Table 3.1: Configuration parameters for the stereo vision system and their values. Most of the parameters can be varied in software to suit different scenarios. The parameters are related to each other via Equation 3.1.

3.2.2 Implementation

The system parameters and the values used to obtain results presented in this chapter are listed in Table 3.1.

Software remapping

In recent implementations of the vision system [Moore et al., 2009, 2010, 2011b], the function of the specially shaped mirrors was simulated using software lookup tables (LUTs). This required calibrated camera-lens assemblies in order to generate the LUTs but reduced the physical bulk and cost of the system. It also avoided aberrations due to imperfections in the mirror surfaces. The LUTs were generated according to

$$\begin{bmatrix} p_x \\ p_y \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} v_z / \sqrt{v_x^2 + v_y^2} \\ \arctan\left(v_y / v_x\right) \end{bmatrix}, \tag{3.2}$$

where v_x, v_y, and v_z are the components of $\hat{\mathbf{v}}$, which is the unit view vector that is obtained from the camera-lens calibration for the i^{th} pixel in the raw image; p_x and p_y are the components of $\mathbf{p}$, which is the resulting position of the i^{th} pixel in the remapped image; and s_x and s_y are simply scaling parameters defining the number of pixels per unit in each of the axes in the remapped image. Put simply, this transformation maps pixels in the raw image to the inner surface of a unit cylinder surrounding the optic axis and extending towards infinity in front of the camera, as depicted in Figure 3.3. In the remapped image, the x-y coordinate axes are θ and d respectively, where θ is the angle (°) taken around the optic axis and d is the distance

(m) along the optic axis to the i^{th} pixel on the surface of the cylinder.

The software remapping process is illustrated in Figure 3.4. In this example, an image of a rendered scene is captured through a rectilinear lens with a 120° FOV. The shaded area of the raw image is transformed according to Equation 3.2 to produce the remapped image. A comparison with Figure 3.1 indicates that the image remapped in software shares the same properties with the image remapped in (simulated) hardware, as expected.

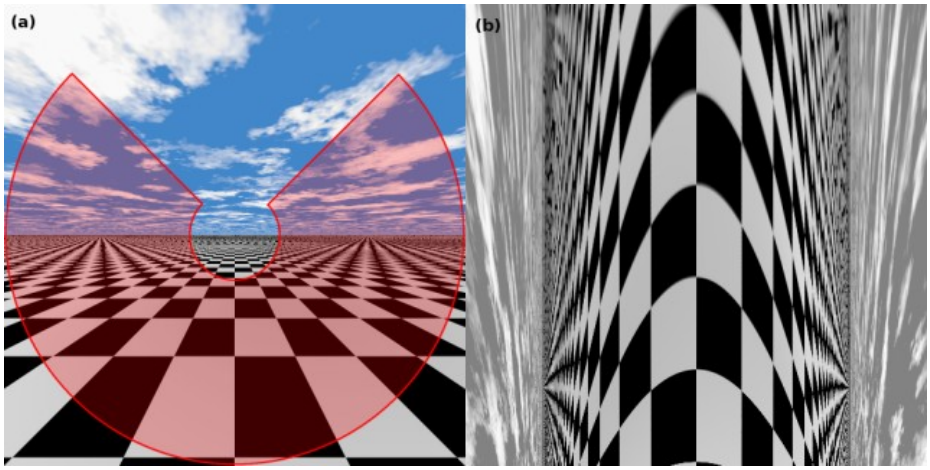


Figure 3.4: Illustration of the process for producing the terrain following mapping in software. The shaded area in the raw image (a) is remapped to (b) the terrain following representation by mapping the raw pixels to the surface of a unit cylinder surrounding the optic axis. The outer circumference of the shaded area in the raw image (a) corresponds to the bottom of the remapped image (b), and the inner circumference corresponds to the top of the remapped image.

The central region of the raw image (see Figure 3.4) is not remapped, because in this region equal distances along the ground plane project onto infinitesimally small distances on the cameras' image planes. If remapped, therefore, the resolution in the corresponding region of the remapped image would be very poor. For the configuration of the system presented in Table 3.1 for collision avoidance, the central unmapped region results in a cone-shaped 'blind zone' surrounding the optic axis with an opening angle of 32°. Because the cameras are situated coaxially, this region corresponds to the image of the fore camera on the imaging plane of the rear camera – hence this region would be unusable in any case. If physical mirrors were used, this region would still be unusable as it would be obscured by the self-reflection of each camera. The outer diameter of the remapped area is limited by the FOV of the camera-lens assemblies.

Camera-lens assemblies

The two cameras were mounted rigidly in a coaxial stereo configuration (Figure 3.5) to minimise measurement errors resulting from relative motion between the two camera-lens assemblies during flight. In initial work [Moore et al., 2009], high-resolution fisheye lenses (Fujinon FE185C057HA-1[1]) were used to provide good spatial resolution whilst maintaining a large FOV. In later work [Moore et al., 2010] it was found necessary to use lightweight, miniature fisheye lenses (Sunex DSL215[2]) to reduce vibration-induced motion of the lenses relative to the camera sensors without compromising the FOV. In both cases, high-resolution monochromatic video cameras (PointGrey GRAS-20S4M[3]) were used to capture raw images at $1040\text{px} \times 1040\text{px}$ and 25Hz. The stereo cameras were synchronised to within 125μs across the IEEE 1394b interface.

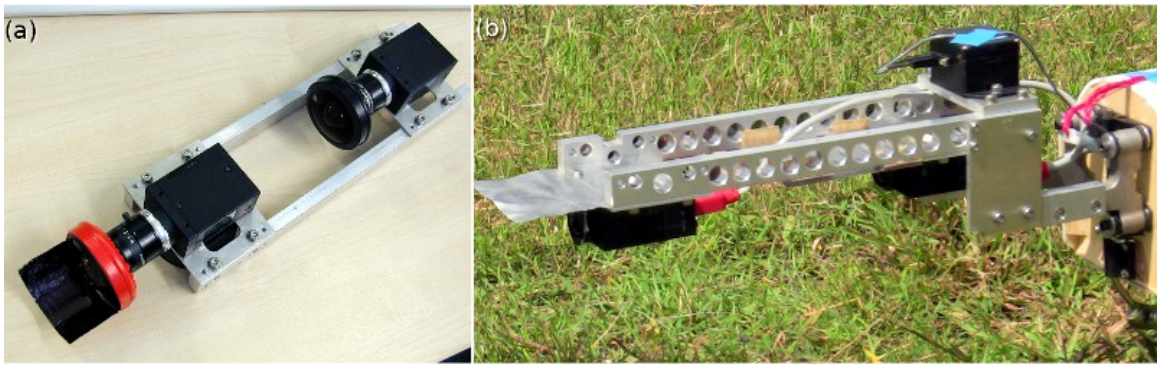


Figure 3.5: Stereo vision system (a) with high-resolution Fujinon lenses and (b) mounted on the aircraft with miniature Sunex lenses.

Each camera-lens assembly was calibrated according to the generic camera model described in Kannala and Brandt [2006] to correct for any aberrations due to imperfections in the imaging optics or positioning of the fisheye lenses with respect to the cameras. Additionally, a corrective rotational transformation was applied during the remapping process for the fore camera to account for any rotational misalignment between the two camera-lens assemblies. This stereo rotational calibration was performed as follows:

1. A stereo pair of images of a distant scene (*e.g.* cloudy sky) was captured with the cameras rigidly mounted in their flight configuration.

[1] http://www.fujifilmusa.com/shared/bin/FE185C046&057HA-1.pdf
[2] http://www.optics-online.com/OOL/DSL/DSL215.PDF
[3] http://www.ptgrey.com/products/grasshopper/Point_Grey_Grasshopper_datasheet.pdf

2. Each camera's raw image was rectified to a true fisheye image using the calibration parameters for each camera-lens pair.

3. A candidate corrective 3-DOF rotation was applied to the fore camera's rectified fisheye image.

4. The candidate rotation was scored by computing the sum of absolute differences (SAD) between the remapped images from the rear and (rotated) fore cameras.

The optimal 3-DOF corrective rotation was found by iteratively applying a non-linear optimisation routine on steps 3 & 4. The non-linear optimisation was performed using the NLopt library [Johnson, 2009] implementation of Powell [2009]'s BOBYQA algorithm.

Image disparity

Range information was extracted from the remapped images by computing image disparities between the stereo pairs. Disparity (in pixels) between each pixel in the remapped rear image and the corresponding pixel in the remapped fore image was computed using a block matching algorithm. To remove low-frequency image intensity gradients, which could confuse the matching algorithm, the remapped images were convolved with a Scharr filter kernel [Scharr, 2000] before the disparity was computed. Candidate disparities were scored by computing the SAD between a window of pixels centred on the i^{th} pixel in the remapped rear image and the corresponding window in the remapped fore image. Correlation scores for many windows were calculated quickly by pre-computing a running integral of absolute differences between pixels in the two remapped images. This process was then repeated as one image was increasingly offset with respect to the other. Disparity computed for the i^{th} pixel window was then simply the image offset at which the computed SAD score was a minimum.

To obtain sub-pixel disparity estimates, an equiangular fit [Shimizu and Okutomi, 2003] was applied to the minimum and neighbouring SAD scores for each window. Disparities were computed to an accuracy of $\frac{1}{8}$px and represented as an unsigned byte, giving a maximum possible search range of $\frac{255}{8} = 31.875$px. Incorrect matches were rejected by re-computing the disparity for the reverse image order and discarding disparities that differed by more than 2px from the original estimate. This bi-directional technique was effective at rejecting mismatches due to noise and stereo

discontinuities but doubled execution time of the algorithm. The matching algorithm was implemented using the Intel Performance Primitives[4] library. When tested on a 1.5GHz processor, the matching algorithm (with sub-pixel approximation and bi-directional search) executed in 20ms, generating 384px × 128px disparity images for a window size of 11px × 11px and a search range of 0px → 15px.

Stereo disparities were extracted from remapped images via the process described above. Radial distances from the optic axis of the system to objects in the environment were then calculated from the stereo disparities according to Equation 3.1. Values for the system parameters were selected arbitrarily to suit the desired operational range (altitude) envelope, the required resolution of the range estimates, and the computational time available for processing disparities. Typical values used to obtain the results presented in this chapter are listed in Table 3.1.

3.2.3 Range testing

Performance and accuracy of the stereo system were initially evaluated using an artificial arena of known geometry. A cropped image of the arena as viewed by the vision system's fore camera is displayed in Figure 3.6a. The texture used to line the arena comprised black circles of varying diameter (6.5cm → 15cm) on a white background. Dimensions of the arena were 3.20m × 2.35m × 1.15m. The stereo rig was positioned in the centre of the arena with the optic axis parallel to the longest dimension of the arena. The remapped view of the testing arena overlaid with the computed stereo disparities is displayed in Figure 3.6b. The 2D array of disparities is also shown in Figure 3.7, plotted as a 3D mesh for visualisation. It can be seen from Figures 3.6b & 3.7 that disparity vectors have equal magnitudes along an axis parallel with the long axis of the arena. This verifies the expected result – that image disparity between stereo pairs depends only on radial distance between the viewed points and the optic axis.

Estimates of radial distances to observed points were calculated from the measured stereo disparities via Equation 3.1. In order to quantify the accuracy of the stereo system, the relationship between the estimated radial distance to the arena and the viewing angle was plotted against the actual relationship (Figure 3.8). The actual relationship was calculated from the known geometry of the arena. Viewing angle in this case corresponds to horizontal elevation of the viewing ray – *i.e.* points that

[4] http://software.intel.com/en-us/articles/intel-ipp/

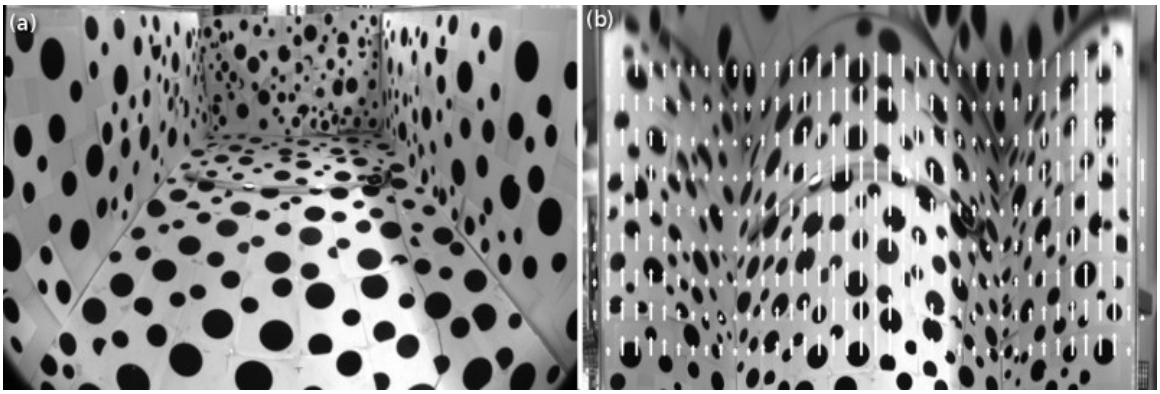


Figure 3.6: (a) Cropped image of the indoor arena used for range testing the stereo vision system as seen by the fore camera, and (b) the same view of the arena after remapping, overlaid with computed stereo disparities. Disparity vectors have been scaled to aid visualisation.

lie in the same column in the remapped image (Figure 3.6b) share the same viewing elevation. Estimated radial distance errors for each viewing angle in Figure 3.8 were thus computed from the variance of the multiple estimates at each viewing elevation.

It can be seen from Figure 3.8 that estimated radial distance errors are greatest for viewing elevations that correspond to the join between the walls and floor of the arena. This is partially a result of the non-zero size of the window used to compute stereo disparities. A window size larger than one pixel would be expected to cause an underestimation of radial distance to the corners of the arena, where surrounding pixels correspond to closer surfaces. Indeed, this is observed in Figure 3.8. Similarly, one would expect to observe a slight overestimation in the radial distance to the arena floor directly beneath the vision system, where surrounding pixels correspond to regions of the surface that are further away. Again, this is observed in Figure 3.8.

Data presented in Figure 3.8 was computed from a single typical stereo pair and is unfiltered. Small errors in the reprojected viewing angles may arise from inaccurate calibration of the camera-lens assemblies but are presumed to be negligible in this analysis. Total error in the reconstruction can therefore be specified as the error in radial distance to the arena at each viewing angle. The standard deviation of this error, measured from approximately 2.5×10^4 reprojected points, was computed to be $\sigma = 3.5 \times 10^{-2}$m. When represented as a percentage of the estimated radial distances at each viewing angle, the absolute (unsigned) reprojection error was calculated as having a mean of 1.2% and a maximum of 5.6%.

It can be seen from Figure 3.8 that the standard deviation of the range estimates

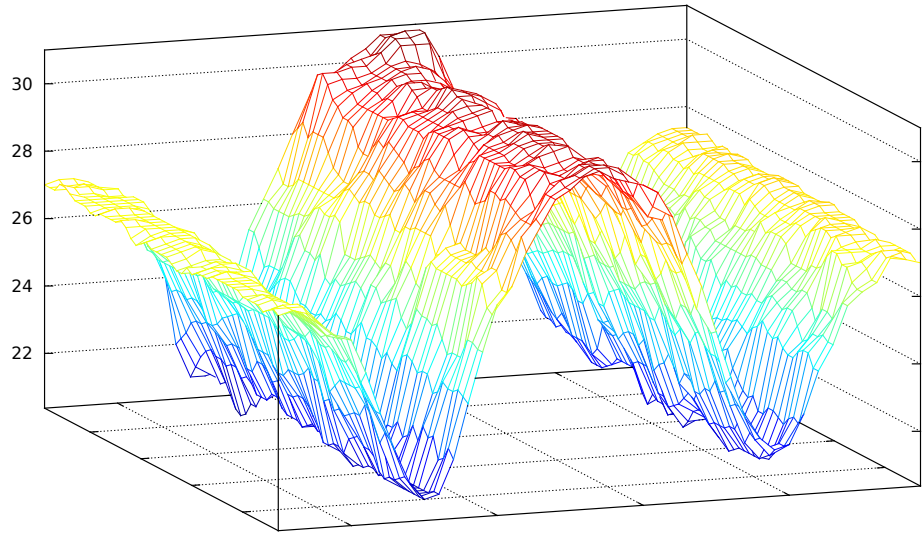


Figure 3.7: Disparities computed from the stereo pair of images taken inside the testing arena. The disparities are shown as a 3D mesh to aid visualisation. All axes have units of pixels. The labelled (vertical) axis shows the magnitude of the stereo disparities. The two unlabelled (horizontal) axes correspond to the image coordinates of the remapped image from the rear camera. The longitudinal axis of the arena has been compressed for display.

increases as a function of range. In fact the standard deviation of the range estimates increases approximately as the square of the true radial distance. This approximation is valid for distributions with a small *variability*, where the standard deviation is small with respect to the mean. It can be confirmed by comparing the statistics of a distribution, x, with the statistics of its inverse, x'. For a normal distribution with a small spread of data, the statistics of the original distribution (median, $\tilde{x}$, and standard deviation, σ_x) can be recovered approximately from the statistics of the inverse distribution via

$$\tilde{x} \approx \frac{\tilde{x'}}{(\tilde{x'})^2} \approx \frac{1}{\tilde{x'}} \quad \text{and} \quad \sigma_x \approx \frac{\sigma_{x'}}{(\tilde{x'})^2}.$$

In this case, estimated radial distance is not exactly the inverse of measured disparity, rather it is proportional to the inverse of measured disparity (*e.g.* Equation 3.1) and the variability of measured disparity data is small for each viewing angle. Hence the standard deviation of range estimates is approximately proportional to the square of radial distance multiplied by the standard deviation of measured disparities over the ranges at which this test was conducted. This finding agrees with the accepted theory of stereo range resolution for conventional stereo

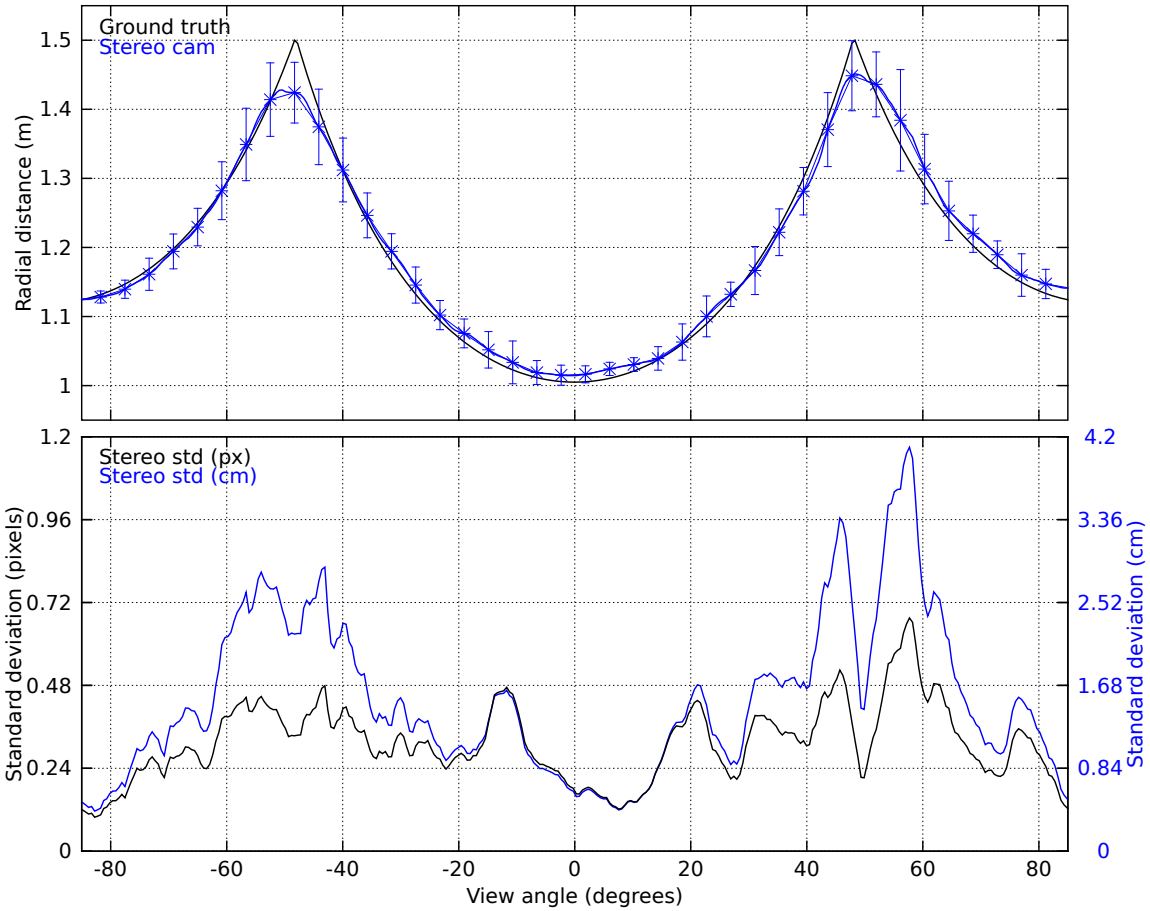


Figure 3.8: Profile of the estimated radial distances to the arena wall and floor (top, blue) shown alongside the actual radial distances (top, black) at each viewing angle. Error bars represent $\pm 2\sigma$ at each viewing angle. The profile of the standard deviations of the computed radial distances is also shown (bottom, right axis, blue) alongside the standard deviations for the measured stereo disparities (bottom, left axis, black). Note that the variance of the measured stereo disparities is approximately constant, but the variance of the computed radial distances increases with range. The two y-axes in the lower plot have been drawn to highlight the approximately squared dependence on range for the standard deviation of the computed radial distances.

systems [Matthies and Shafer, 1987]. Accuracy of the stereo system is therefore more properly described in terms of pixel disparity and is approximately $\sigma = 2.5 \times 10^{-1}$px. Measurement error (in pixels) will therefore be approximately constant for any detection range, but accuracy of metric range estimates will drop exponentially with detection range. This is not a significant problem in this case, however, as collision avoidance only requires reliable detection of nearby obstacles. Given the average flight speed ($\sim 20\text{ms}^{-1}$) and manoeuvrability of the fixed-wing platform, successful collision avoidance will require detection of obstacles at ranges in the order of 10m. At this range, the accuracy of the stereo vision system is expected to be sufficient for reliable obstacle detection.

The predicted growth of the standard deviation of radial distance estimates as a function of radial distance is plotted in Figure 3.9 for the stereo vision system proposed here (using the configuration presented in Table 3.1 for testing) and a conventional stereo vision system[5]. Growth of radial σ is plotted for the conventional stereo system at three different viewing angles in the horizontal plane (*i.e.* for 'oncoming' obstacles at three different angles with respect to the forward viewing axis of the stereo system). It can be seen that growth of radial range error is equivalent for the two systems for oncoming obstacles at an angle of approximately 36.5° from the optic (forward) axis. Equation 3.1 predicts that radial range error should be independent of viewing angle for the proposed stereo system, which is not the case for conventional stereo systems. Obstacles close to the forward viewing axis (*i.e.* obstacles closely coincident with the projected flight path) should therefore be more accurately localised by the proposed stereo system than a conventional stereo system with the same coaxial camera configuration. A conventional stereo system with a side-by-side forward-looking configuration would be expected to provide more accurate range estimates for oncoming obstacles.

Radial distances computed from the stereo images of the test arena (seen in Figure 3.6) were used to reconstruct the arena in 3D space (Figure 3.10). It can be seen that the relatively small errors in the estimated radial distances lead to an accurate 3D representation of this simple test environment.

[5]The conventional stereo system considered for the purpose of this analysis differs from the proposed stereo system only by the fact that the terrain following mapping described in Section 3.2 is not performed – *i.e.* the baseline of the system as well as the resolution, FOV, and positioning of the cameras is identical.

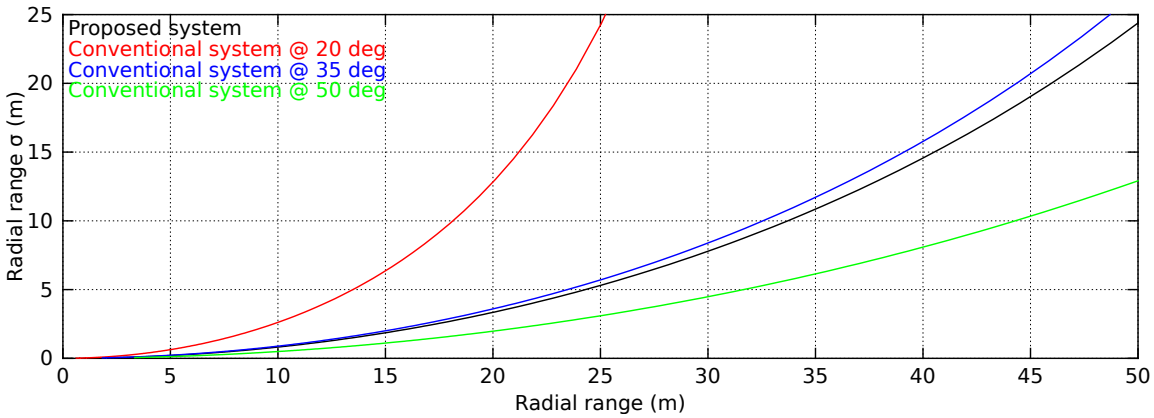


Figure 3.9: Predicted growth of the standard deviation of radial distance estimates as a function of radial distance for (black) the proposed stereo vision system and (red, blue, green) a conventional stereo vision system for viewing angles of 20°, 35°, and 50° from the forward viewing axis respectively. Details in text.

3.3 Stereo vision-based collision avoidance

Once the radius of a collision-free cylinder surrounding the optic axis is defined, this information can be used directly in a closed-loop feedback control scheme to steer an aircraft away from objects that penetrate the notional flight cylinder required by the aircraft for collision-free flight.

3.3.1 Obstacle detection

For a guidance system to be effective, it must be able to repel the aircraft from objects that are detected penetrating the notional flight cylinder. In order to achieve this, the closest object to the projected flight path was localised by performing the following steps in real time on board the aircraft:

- A stereo pair of images were captured from the cameras at 25Hz and remapped to the terrain following model. A dense disparity map was then computed according to the method described in Section 3.2.2 and using the parameters defined in Table 3.1 for collision avoidance.

- Morphological and filtering operations were performed on the disparity map to eliminate noise caused by incorrect stereo correspondences. The operations, performed in order, were erosion and dilation (both 3px × 3px window), followed by blurring (2px × 4px window). The most critical obstacle was then located

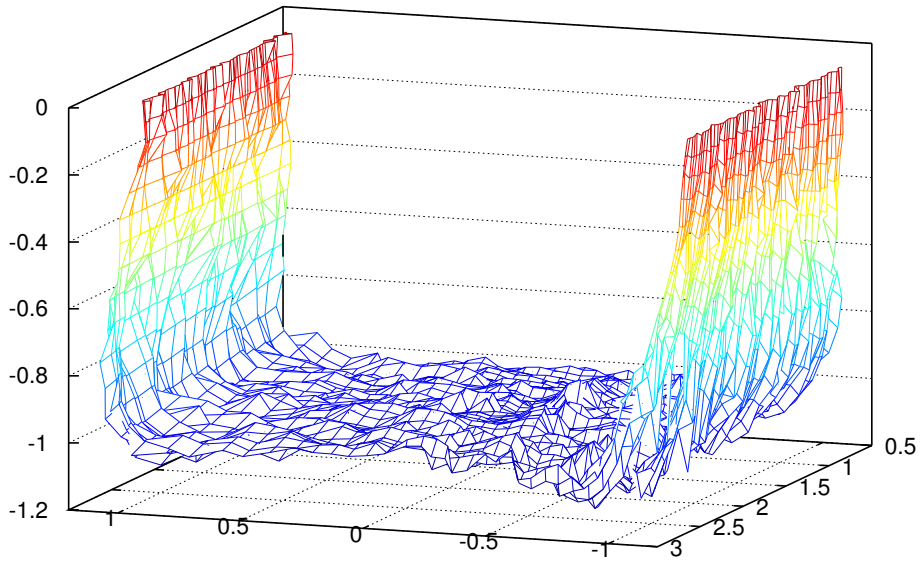


Figure 3.10: 3D reconstruction of the arena used to evaluate performance of the stereo vision system. Measurements are given in metres relative to the nodal point of the vision system. True dimensions of the arena are 3.20m × 2.35m × 1.15m.

by searching for the highest average disparity within a 10px × 20px region in the disparity map.

- Radial distance from the closest visible obstacle to the optic axis of the vision system was computed according to Equation 3.1. Roll attitude of the aircraft with respect to the position of the obstacle was determined from the horizontal position of the detected obstacle in the disparity map (horizontal FOV of the disparity map is 200° – see Table 3.1).

- If the computed radial distance to the obstacle was less than the minimum radius required for collision-free flight, an evasive manoeuvre was performed.

The obstacle detection stage is illustrated in Figure 3.11.

3.3.2 Closed-loop control

In the intended control system, roll and pitch of the aircraft would be controlled independently. An aileron command would be generated to centre the position of the nearest detected obstacle beneath the aircraft and an elevator command would be generated to maintain the range to the closest detected obstacle equal to the minimum radius of the notional free-flight cylinder. Using this scheme, the aircraft

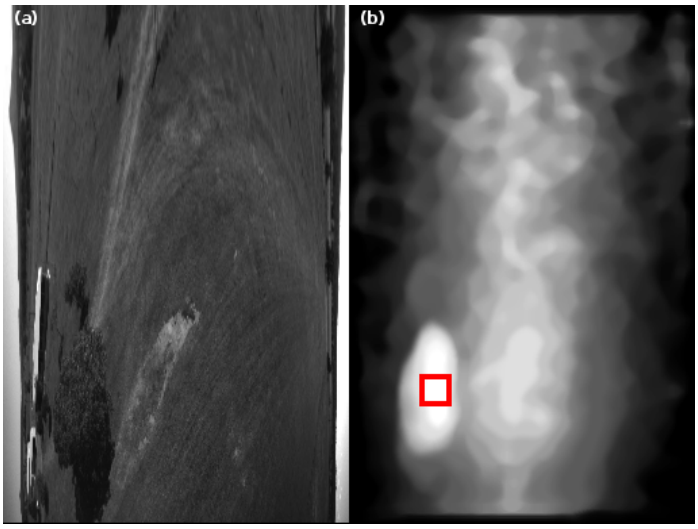


Figure 3.11: Illustration of the obstacle detection stage showing (a) the remapped image from the rear camera, and (b) the corresponding disparity map. The region of the disparity map with the highest average disparity is highlighted in red. This corresponds to the obstacle that lies closest to the projected flight trajectory of the aircraft.

should roll and 'pull up' to avoid oncoming obstacles and then return to level after the obstacle is passed. During flight testing, however, it was found that directly controlling the elevator from the error between the detected obstacle range and the minimum free-flight radius resulted in a control system that was unstable during terrain following. This was due to the large changes in radial distances to objects in the environment that could be effected by relatively small changes in pitch attitude. To facilitate stable terrain following as well as effective collision avoidance, a mixed control system was devised:

- Whilst no obstacles were detected within the notional free-flight radius, the aircraft was commanded to hold a constant altitude and attitude with respect to the ground plane using the approach prescribed for terrain following in Chapter 4, and specifically, the control system described in Section 4.3.

- If an obstacle was detected within the radius of the free-flight cylinder, an evasive manoeuvre was initiated by mixing the roll angle and elevator commands within the terrain following control system with evasive roll angle and elevator commands, which were computed from the position of the obstacle within the disparity map and its estimated radial distance respectively.

  The mixing factor, α_{avoid}, was a linear function of the error between the notional

radius of the free-flight cylinder, $r_{cylinder}$, and the radial distance measured from the closest obstacle to the optic axis, d_{radial},

$$\alpha_{avoid} = (r_{cylinder} - d_{radial})/4, \quad \alpha_{avoid} \in [0, 1],$$

with all parameters specified in metres. For the closest detected obstacle, if $d_{radial} \geq r_{cylinder}$ then the control system was purely terrain following; if $(r_{cylinder} - 4) < d_{radial} < r_{cylinder}$ then the control system was a weighted combination of terrain following and collision avoidance; and if $d_{radial} \leq (r_{cylinder} - 4)$ then the control system was purely collision avoidance.

The elevator command generated for collision avoidance, e^*_{avoid}, was a linear function of the extent to which the closest obstacle penetrated the free-flight cylinder and is identical to the mixing function,

$$e^*_{avoid} = \alpha_{avoid}, \quad e^*_{avoid} \in [0, 1].$$

The roll angle command, ϕ^*_{avoid}, was set to nullify the apparent roll angle of the aircraft with respect to the closest detected obstacle, $\phi_{obstacle}$,

$$\phi^*_{avoid} = -\phi_{obstacle},$$

where $\phi_{obstacle}$ corresponds to the horizontal elevation, or equivalently, the horizontal position, of the detected obstacle in the disparity map.

The mixed elevator, e^*_{mix}, and roll angle, ϕ^*_{mix}, set-points were therefore defined by

$$e^*_{mix} = \alpha_{avoid} \cdot e^*_{avoid} + (1 - \alpha_{avoid}) \cdot e^*, \tag{3.3a}$$

$$\phi^*_{mix} = \alpha_{avoid} \cdot \phi^*_{avoid} + (1 - \alpha_{avoid}) \cdot \phi^*, \tag{3.3b}$$

where e^* and ϕ^* are the elevator and roll angle set-points respectively generated by the terrain following closed-loop control system described in Section 4.3. Autonomous collision avoidance was achieved by replacing the elevator and roll angle set-points in the terrain following closed-loop control system by their counterparts defined in Equation 3.3. The remainder of the control system was identical to that described for terrain following in Section 4.3.

3.3.3 Flight test results

To validate performance of the stereo vision system, a series of flight tests was conducted in which the vision system provided guidance to a small-scale UAV flying in close proximity to the ground and various other natural obstacles. The aircraft used for flight testing was a TWM Super Frontier Senior-46[6] with a wingspan of 2.04m and modified to carry the engine and propeller assembly above the wing (pictured in Figure 3.12). All processing was completed on board the aircraft in real time via a PC-104 form-factor Digital-Logic MSM945P/SMX945B-L7400[7] computer incorporating an Intel Core 2 Duo 1.5GHz processor.

Figure 3.12: The aircraft used for collision avoidance and terrain following flight tests with the stereo vision system mounted on the nose.

Guidance commands were computed continuously by the stereo vision system during the flight test. A switch on the ground-based controller allowed a human pilot to select between manual or autonomous control of the aircraft. During flight testing, the aircraft was piloted manually in a rough racetrack pattern and autonomous control was engaged for segments of flight, during which the stereo vision system had full control of the aircraft. The flight plan is illustrated in Figure 3.13.

Seventeen autonomous passes were conducted in which the aircraft was required to maintain a constant roll angle (level with the ground plane) and altitude (6m above ground level (AGL)). For 10 of the passes, the aircraft's trajectory brought it in close proximity to a stand of trees, which were approximately the same height as the altitude of the aircraft and hence required the aircraft to perform an evasive manoeuvre to avoid a collision. Flight data from two autonomous passes during

[6]http://www.theworldmodels.com/para/instruction/catalog/2009/pdf/35.pdf
[7]http://www.adl-usa.com/products/cpu/datapage.php?pid=MSM945

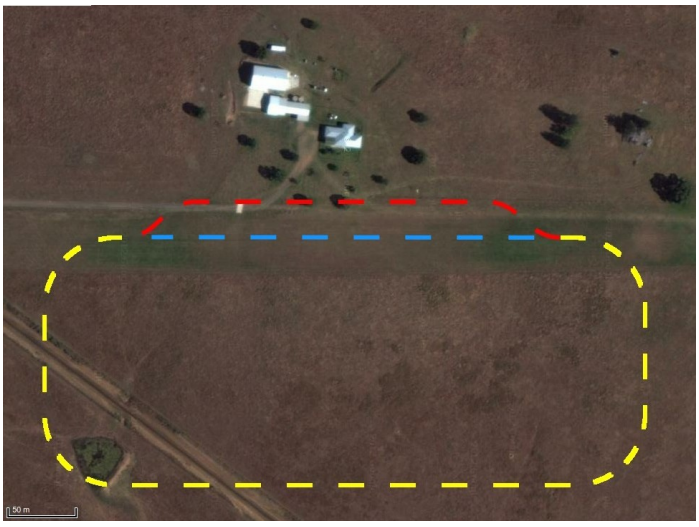


Figure 3.13: Bird's-eye view of the area in which collision avoidance flight testing was performed, indicating the approximate flight path during manual flight (yellow), autonomous terrain following (blue), and autonomous collision avoidance (red). Unfortunately the GPS path of the aircraft was not available during these flight tests. The scale indicates 50m. Satellite imagery ©2012 Google, DigitalGlobe, GeoEye, Cnes/Spot Image.

the flight test are shown in Figure 3.14. These passes represent typical segments of autonomous flight from the *terrain following* and *collision avoidance* segments of the flight test, as indicated in Figure 3.13.

During flight testing, a radius of $r_{cylinder} = 6$m was used for the notional cylinder required for collision-free flight. It can be seen clearly from the flight data presented in Figure 3.14 that once an obstacle was detected violating the collision-free flight cylinder, the guidance system was able to generate appropriate elevator and roll angle commands (defined in Equation 3.3), thus enabling the aircraft to avoid the obstacle effectively. A single autonomous collision avoidance pass is shown in the accompanying video[8]. A segment of flight from this pass, corresponding to 415s $\sim$ 416s in Figure 3.14 and in which the aircraft avoided a tree successfully, is analysed in greater detail in Figure 3.15. The sequence of frames taken from this segment of the flight test show initial detection of the tree as an obstacle and subsequent evasive response of the aircraft.

From Figures 3.14 & 3.15 it can be seen that the obstacle was detected initially by the vision system at $\sim$ 415s; at $\sim$ 415.25s the evasive manoeuvre was intitiated; by

[8]Attachment AV01. Note that colours overlaid on the on-board imagery in this video represent objects that are measured to be either close to (green) or penetrating (red) the free-flight cylinder.

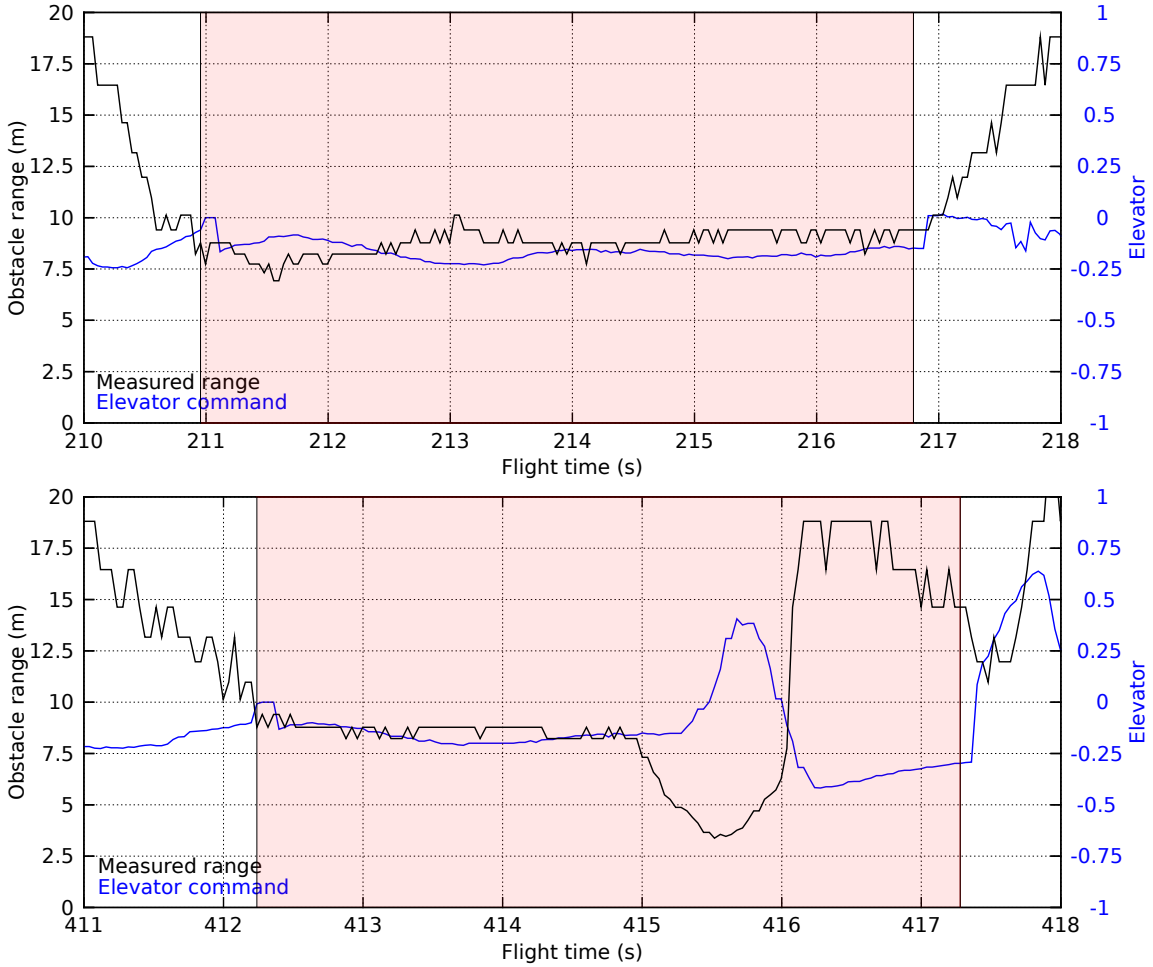


Figure 3.14: Two closed-loop passes during the collision avoidance flight test demonstrating terrain following above flat ground with no obstacles (top) and with an obstacle in the flight path (bottom). Radial distance to the closest obstacle as measured from the disparity map (black) is plotted alongside normalised elevator command (blue). During autonomous flight (shaded red), the aircraft was commanded to hold a radial distance of approximately 6m to the closest obstacle, and if an obstacle violated this constraint, an evasive manoeuvre was performed.

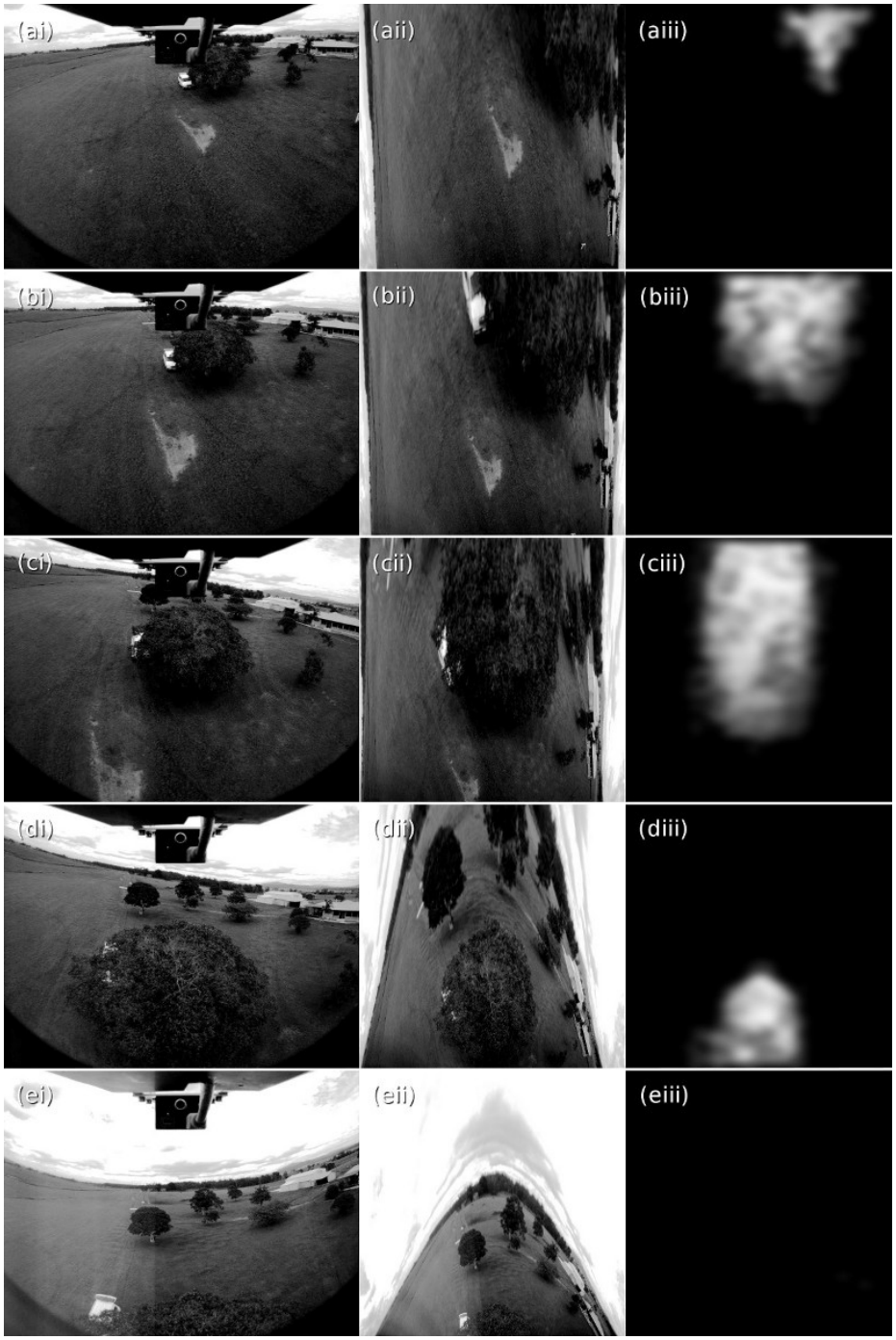


Figure 3.15: Sequence of images (rows a-e) taken on board the aircraft during a segment of closed-loop flight. The images are: a crop from the rear camera (column *i*); a terrain following remapping of the rear image (column *ii*); and an obstacle map (column *iii*) that shows the measured stereo disparities corresponding to regions of the remapped rear image detected violating the free-flight cylinder. The five frames were captured at intervals of approximately 240ms and cover a segment of the flight test from 415s $\sim$ 416s. Further details in text.

$\sim$ 415.5s the aircraft had begun to respond to the avoidance commands – the aircraft had rolled to centre the obstacle beneath it and pitched up; the aircraft continued to respond at $\sim$ 415.75s, and by $\sim$ 416s the avoidance manoeuvre had been completed and the aircraft returned to terrain following guidance.

When no obstacle was present in the flight path, the aircraft was able to maintain a stable altitude and hold level with the ground. It can be seen from Figure 3.14 however that there was a small positive bias to the radial distance measured to the closest point on the ground during terrain following. It can also be seen that the aircraft-ground range remained constant even though a negative elevator position was commanded. This indicates that the aircraft had a significant and negative trim position for the pitch axis, which the control system was unable to overcome completely. Had the elevator trim point been set appropriately, it is expected that the aircraft would have been able to maintain the predefined altitude of 6m AGL more precisely. The performance of the terrain following system is analysed in greater detail in Chapter 4.

3.4 Discussion

Performance of the stereo vision-based guidance system described above is analysed in this section. The ability of the proposed system to guide a small-scale UAV around obstacles was evaluated by comparing its performance with other vision-based systems.

3.4.1 Performance

Flight data presented above (Section 3.3.3) demonstrated the ability of this stereo vision system to provide effective guidance to a small-scale aircraft flying at low altitude in the presence of obstacles. It is difficult, unfortunately, to obtain a quantitative measure of the efficacy and robustness of this system, as no measurement of the position of the aircraft was available at the time of conducting these flight tests. During the 10 autonomous segments of flight in which the aircraft's trajectory would have resulted in a collision or near-miss[9] had the aircraft not reacted, however, the vision system detected the obstacle successfully and was able to increase the miss distance in each case. In the 7 segments of autonomous flight where there was no

[9] A near-miss is defined here as an obstacle passing within a wingspan of the body of the aircraft.

obstacle other than the ground present, the aircraft maintained stable altitude and attitude.

The performance of the proposed approach somewhat exceeds that of other state-of-the-art vision-based collision avoidance approaches. Green and Oh [2008] use two 1D optic flow sensors oriented at 45° from the forward axis to provide a collision avoidance capability for their fixed-wing micro air vehicle (MAV). They provide qualitative evidence to demonstrate the effectiveness of their approach, although they caution that their optic flow sensors are ineffective for detecting the presence of small obstacles directly in the flight path of the aircraft. Beyeler et al. [2009b], Zufferey et al. [2010] describe a similar but more capable system and present more comprehensive results from flight tests conducted outdoors. They fly a small-scale, fixed-wing UAV along a predefined flight path that brings the aircraft in close proximity to trees, which are approximately the same height as the altitude of the aircraft during its approach. They demonstrate that their system is able to steer the aircraft around and/or over the obstacles robustly, but their approach is likely to suffer from the same susceptibility to directly oncoming small obstacles as Green and Oh [2008].

To overcome the limitations of purely optic flow-based systems for detecting oncoming obstacles, [Hrabar and Sukhatme, 2009, Hrabar et al., 2005] complement their optic flow-based guidance system with forward-looking stereo and demonstrate the effectiveness of their system operating on board a small-scale helicopter by avoiding a large tree and buildings. Forward-looking stereo is also used by Byrne et al. [2006] to detect oncoming small obstacles from a small-scale helicopter, although they do not demonstrate closed-loop object avoidance.

The stereo vision-based guidance system presented in this chapter demonstrated a capability for object detection and avoidance comparable to the approaches described above – it was capable of steering the aircraft around obstacles that were approximately the same height as the altitude of the aircraft and extended to the ground. The coaxial design of this stereo vision system means, however, that it would be unable to detect the presence of small obstacles that are directly coincident with the flight trajectory, similar to the purely optic flow-based approaches discussed here. Limitations of the stereo vision system are discussed further in Section 3.4.2. Despite these limitations, autonomous collision avoidance was demonstrated successfully using the proposed stereo vision-based guidance system at flight speeds of approximately $20\text{ms}^{-1} \sim 30\text{ms}^{-1}$, which is roughly twice that of the other published approaches discussed in this section.

3.4.2 Limitations

The coaxial design of the stereo vision system and the terrain following remapping allow radial distance from obstacles in the environment to the optic axis to be computed directly. By projecting the instantaneous flight trajectory along the optic axis, this approach permits the aircraft to identify potentially dangerous obstacles effectively and to correct its course accordingly. A disadvantage of this approach, however, is that the coaxial positioning of the stereo cameras results in a conical blind zone surrounding the optic axis. The configuration of the stereo system used during flight testing (see Table 3.1 for the configuration parameters) produces a maximum forward viewing factor of $r_{tot} = 3.5$, which results in a conical blind zone with an opening angle of 32°. This means that, for level flight above an infinite ground plane, the vision system can detect an oncoming obstacle at a maximum range of 3.5 times the height of the aircraft, for an obstacle at ground level. Maximum detection range is lower for objects above ground level. It was demonstrated in Section 3.3.3 that this was sufficient for the proposed visual guidance system to detect and react to oncoming trees, which intersect the ground plane. Collision avoidance was tested at a minimum altitude of 6m AGL. At lower altitudes a maximum forward viewing factor of $r_{tot} = 3.5$ would preclude detection of obstacles further than 20m in front of the aircraft. Given the practical limitations of the obstacle detection system and aircraft manoeuvrability, 6m probably represents the minimum altitude at which this system could provide useful collision avoidance information to this aircraft. The proposed approach would be less effective if the aircraft were confronted with smaller airborne obstacles.

To overcome this limitation, the stereo cameras could be positioned laterally relative to each other, instead of coaxially. In this configuration, a forward-looking stereo system would provide information on directly oncoming small obstacles, as demonstrated by Hrabar and Sukhatme [2009], Hrabar et al. [2005], but would not be able to make use of the terrain following mapping nor directly extract radial distances from obstacles to the projected flight trajectory. Once the position of objects in the environment had been triangulated, however, their radial distances from the optic axis could be computed simply via trigonometry. It is expected that a manipulation of the computed obstacle ranges would produce estimates of the radial distances to obstacles that would be less accurate than estimates obtained directly from stereo images remapped to the terrain following model (described earlier in this chapter).

This is due partly to the asymmetry of the error distribution associated with the reprojected ranges (see Section 3.2.3 for an analysis of the reprojection error).

An advantage of using stereo vision over alternative optic flow-based approaches is that stereo vision does not require rotational motion or ground speed of the aircraft to be estimated or measured in order to extract accurate estimates of range to obstacles in the environment. This advantage comes at the cost of a limited range of distances over which measurements can be made precisely, due to the fixed baselines of stereo vision systems. This is an issue that affects most stereo vision-based systems (it is possible to construct a stereo system with a variable baseline).

An advantage of using high-resolution cameras and fisheye lenses is that smaller obstacles may be detected over a wider FOV, compared to the approaches described by Beyeler et al. [2009b], Green and Oh [2008], Zufferey et al. [2010], which all rely on specialised optic flow sensors (see Figure 2.2 for examples of such sensors). The stereo cameras must be calibrated very accurately, however, to realise the increased resolution. The calibration scheme discussed in Section 3.2.2 proved sufficient for detecting and avoiding tree-sized obstacles in the real world. It was necessary, however, to post-process the computed disparity maps to remove invalid stereo correspondences in order for obstacles to be detected reliably (the processing steps are discussed in Section 3.3.1). This had the effect of reducing the practical resolution of the vision system. To offset this, it is expected that the effective resolution of the system could be enhanced greatly by generating a per-pixel stereo calibration LUT that could be applied during the remapping process, rather than performing a rotational alignment of the two camera models only. Such a LUT could be generated simply by finding the average 2D correspondences between the fore and rear remapped images over an extended period of flight, during which the aircraft was at an altitude high enough to be beyond the effective range of the stereo system.

3.5 Conclusion

This chapter introduced and described a novel wide-angle stereo vision system that was designed for providing guidance to a small-scale UAV flying fully autonomously at high speed and low altitude. The design of the vision system was intended to simplify the process of extracting guidance commands from visual input by allowing an aircraft to measure radial distances from the optic axis to objects in the environment directly. The maximum disparity measured by this system thus defines the radius of

a collision-free cylinder surrounding the optic axis, through which the aircraft can fly unobstructed. Oncoming obstacles that violate the minimum radius of the notional flight cylinder can therefore be detected easily and the trajectory of the aircraft can be modified appropriately to avoid a collision. This system is therefore well suited to providing visual guidance for an aircraft flying autonomously at low altitude.

Results of open-loop and closed-loop flight tests demonstrate the accuracy of this stereo vision system and the capabilities of the guidance system. This vision system was proven to be effective for guiding a small-scale UAV around natural obstacles when flying at high speed in a low-altitude environment and would usefully complement a specific terrain following guidance system. Such a guidance scheme is the subject of Chapter 4.

Within the context of the aims of this thesis, this chapter has introduced a novel vision system that is as capable as, if not more so, other state-of-the-art visual approaches for providing guidance to an autonomous aircraft flying in a low-altitude environment. Additionally, the design of this vision system provides greater scope for complementary guidance algorithms to be implemented alongside the basic guidance scheme presented in this chapter. Such additional capacities would be difficult to achieve using front-end visual sensors dedicated exclusively to computing optic flow. This is an important advantage of the proposed approach. The ability to combine multiple, simple guidance algorithms allows for a broader range of capabilities than that currently achievable using pre-existing approaches. Chapter 4 shows how a wide-FOV stereo system facilitates precise control of attitude and altitude of a small-scale UAV.

4

Autonomous Stabilisation of Attitude and Altitude

4.1 Summary

Chapter 3 introduced a coaxial stereo vision system that simplified the task of extracting guidance commands from visual input for a UAV flying at low altitude. The simple guidance scheme described there allowed an aircraft to autonomously avoid colliding with obstacles that penetrated a notional flight cylinder surrounding the optic axis of the vision system. A complementary guidance scheme is now presented that uses the wide-FOV stereo system to control aircraft attitude and altitude accurately.

A planar model is assumed for the ground directly beneath, and in front of, the aircraft. Two methods for estimating attitude and altitude with respect to the ground plane are described and their relative strengths and weaknesses are analysed using results obtained from ground-based testing and open-loop flight tests (Section 4.2). A control scheme is described in Section 4.3 and results from closed-loop flight tests demonstrate the usefulness of this approach for controlling attitude and altitude of an autonomous UAV flying at low altitude. The proposed approach is compared with alternative vision-based approaches and found to provide more accurate and precise control of attitude and altitude than other similar schemes (Section 4.4). Additionally, the guidance system presented here is found to facilitate terrain following flight at a faster ground speed than previously published optic flow-based approaches and at a higher altitude than previously published stereo vision-based approaches. The possibility of combining the stereo system with measurements of optic flow is

investigated briefly in Section 4.5 as a means of overcoming one of the limitations of the proposed system.

Portions of this chapter are contained within Moore et al. [2009, 2010, 2011b].

4.2 Attitude and altitude stabilisation

In section 3.3.2, a stereo vision-based guidance system was described in which an aircraft is repelled from objects penetrating a notional collision-free cylinder surrounding the projected flight trajectory. This control scheme provided an effective collision avoidance strategy for an autonomous UAV flying at low altitude. In this section it is shown that aircraft attitude and altitude with respect to the ground can be measured accurately using the same stereo vision system. This enables more precise attitude and altitude control during low-altitude terrain following, and also allows for other manoeuvres such as constant altitude turns and landing.

4.2.1 Estimating attitude and altitude

If it is assumed that the ground directly beneath and in front of a low-flying aircraft can be modelled as a planar surface, then the aircraft's attitude can be measured with respect to the normal to the plane. Furthermore, the aircraft's altitude can be specified as the distance from the nodal point of the vision system to the ground plane, measured parallel to the plane normal. Attitude and altitude can thus be measured from the parameters of a planar model fitted to the observed disparity points.

Two approaches for fitting a model ground plane to observed disparities were considered in this thesis. Initially, the ground plane was modelled in disparity space [Moore et al., 2009], which is a direct approach and results in precise and robust estimates of an aircraft's attitude and altitude. For closed-loop flight testing, however, it was found necessary to fit the ground plane in 3D space [Moore et al., 2010], which is computationally more efficient. The two approaches are described in detail below.

The three frames of reference (camera, body, and inertial) that are used in the following discussion and throughout this thesis are illustrated in Figure 4.1.

Modelling the disparity surface

Given calibration parameters of the camera-lens assemblies, as well as attitude and altitude of the aircraft carrying the vision system, the lengths and directions of the

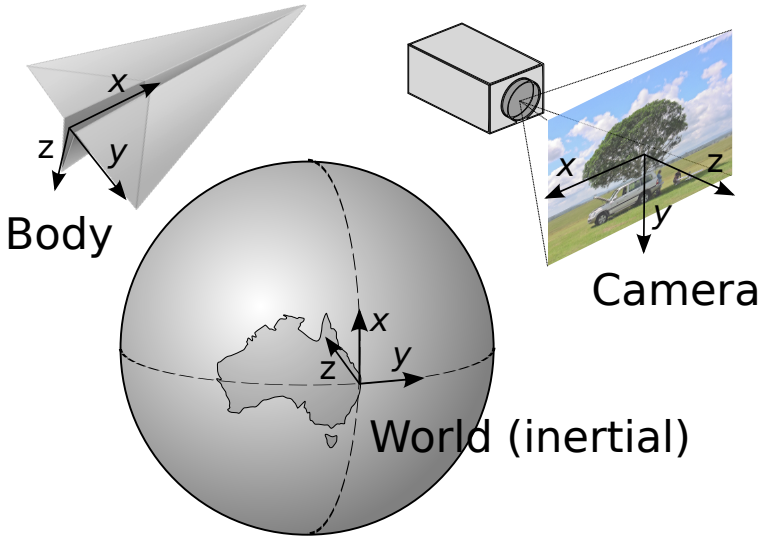


Figure 4.1: Illustration of the coordinate frames referenced in this thesis. The world, or inertial, frame is defined by the standard geographic NED coordinate frame, where x and y are aligned with North and East respectively and z is aligned with the direction of gravity. The body axes are defined by a body-fixed coordinate frame, where positive rotations about each of the axes correspond to a rolling motion to the right; pitching up; and a clockwise change in heading direction respectively. The camera frame x and y axes are given by the direction of increasing pixel coordinates for image columns and image rows respectively, and the z axis is defined to complete a right-handed coordinate frame. Paper plane image credit: http://www.graphicsfuel.com.

view rays that emanate from the nodal point of the vision system and intersect with an ideal, infinite ground plane can be computed. Additionally, by reformulating the ray distances as radial distances from the optic axis of the vision system, the ideal disparities computed by the stereo vision system may be calculated via Equation 3.1. The disparity surface that should be measured by the vision system at a given attitude and altitude above an infinite ground plane can thus be predicted. Conversely, given the measured disparity surface, the roll, pitch, and height of the aircraft with respect to the ideal ground plane can be estimated by iteratively fitting the modelled disparity surface to the measurements. This is a robust method for estimating the attitude and altitude of the aircraft because the disparity data is used directly, hence the data points and average error will be distributed evenly over the fitted surface.

In order to fit the modelled disparity surface to the observed data, the disparity model must be parametrised using roll, pitch, and height of the aircraft with respect to the ground plane. First, the intersection points between the view vectors and the

ideal plane are calculated. A point on a line can be parametrised as $\mathbf{p} = t\hat{\mathbf{v}}$, where in this case $\mathbf{v}$ is a view vector (*i.e.* a vector connecting the nodal point of the vision system with a point in 3D space) and $\hat{\mathbf{v}}$ is its normalised counterpart, and t is the distance to the point of intersection with the ground plane, $\mathbf{p}$, from the origin (*i.e.* the nodal point of the vision system). A plane can be defined as $\mathbf{p} \cdot \hat{\mathbf{n}} + d = 0$, where $\mathbf{p}$ is the same as above, $\hat{\mathbf{n}}$ is the normal to the plane, and d is the distance from the plane to the origin. Solving for t gives

$$t = \frac{-d \cdot |\mathbf{v}|}{\mathbf{v} \cdot \hat{\mathbf{n}}}. \tag{4.1}$$

Now, in the inertial frame, the ideal plane will remain stationary and the aircraft will rotate, so the plane normal can be defined $\hat{\mathbf{n}} = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}$. Therefore, $d = d_{height}$ is the distance from the ideal plane to the origin and, equivalently, the height of the aircraft above the ground plane. So, making the substitutions,

$$t = \frac{-d_{height} \cdot |\mathbf{v}|}{\mathbf{v} \cdot \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}} = \frac{d_{height} \cdot |\mathbf{v}|}{\mathbf{v}_z}, \tag{4.2}$$

thus, only the z component of the view vector in the inertial frame must be found.

In the camera frame, a view vector can be defined by the viewing angle, ν, taken about the positive z axis clockwise from the positive x axis; and the forward viewing ratio, r (see Section 3.2.1 for a description of the forward viewing ratio). Thus, in the coordinate frame of the camera (denoted by the superscript *cam*) a view vector can be written as $\mathbf{v}^{\text{cam}} = \begin{bmatrix} \cos\nu & \sin\nu & r \end{bmatrix}$. To find the same view vector in the inertial coordinate frame, $\mathbf{v}^{\text{cam}}$ is first transposed to the body frame, $\mathbf{v}^{\text{body}} = \begin{bmatrix} r & -\cos\nu & -\sin\nu \end{bmatrix}$ (note that here the cameras are mounted to the aircraft upside down), and then rotated from the body frame to the inertial frame (neglecting yaw, since a heading direction cannot be defined with respect to an infinite ground plane).

A rolling motion, ϕ, about the body frame x axis, and a pitching motion, θ, about

the body frame y axis are defined by $\mathbf{R}_x(\phi)$ and $\mathbf{R}_y(\theta)$ respectively, where

$$\mathbf{R}_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix},$$

$$\mathbf{R}_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}.$$

Therefore,

$$\mathbf{R}^{\text{body} \rightarrow \text{world}}(\phi, \theta) = \mathbf{R}_y(\theta) \cdot \mathbf{R}_x(\phi) = \begin{bmatrix} \cos\theta & \sin\theta\sin\phi & \sin\theta\cos\phi \\ 0 & \cos\phi & -\sin\phi \\ -\sin\theta & \cos\theta\sin\phi & \cos\theta\cos\phi \end{bmatrix},$$

and

$$\mathbf{v}^{\text{world}} = \mathbf{R}^{\text{body} \rightarrow \text{world}}(\phi, \theta) \cdot \mathbf{v}^{\text{body}}. \tag{4.3}$$

Now, only $\mathbf{v}_z$, the z component of the inertial frame view vector, $\mathbf{v}^{\text{world}}$, is required. Multiplying out Equation 4.3 gives

$$\mathbf{v}_z^i = -\cos(\theta) \cdot \sin(\nu^i + \phi) - r^i \cdot \sin(\theta), \tag{4.4}$$

where the superscript i is included to indicate that this is the view vector corresponding to the i^{th} pixel in the remapped image. Substituting Equation 4.4 back into Equation 4.2 gives

$$t^i = \frac{d_{height} \cdot |\mathbf{v}^i|}{-\cos(\theta) \cdot \sin(\nu^i + \phi) - r^i \cdot \sin(\theta)}, \tag{4.5}$$

where t^i is the direct ray distance to the ideal ground plane along the view vector, $\mathbf{v}^i$, for the i^{th} pixel. Now, the stereo vision system actually measures radial distance to objects from the optic axis. Therefore to convert t in Equation 4.5 from a ray distance to a radial distance, the scale factor $|\mathbf{v}|$ is dropped. Substituting Equation 4.5 back into Equation 3.1, the expected disparity surface measured by the stereo vision system

for a particular attitude and altitude above an ideal ground plane is obtained,

$$D^i_{pixel} = \frac{d_{baseline} \cdot h_{image}}{r_{tot}} \cdot \frac{1}{d_{height}} \cdot \left[-\cos(\theta) \cdot \sin(\nu^i + \phi) - r^i \cdot \sin(\theta) \right], \tag{4.6}$$

where the first term is a system constant as described in Section 3.2.1; and radial distance has been replaced by d_{height}, the vertical height (in the inertial frame) of the aircraft above the ideal ground plane. The bracketed term describes the topology of the disparity surface and depends on the roll, ϕ, and pitch, θ, angles of the aircraft as well as two parameters ν^i and r^i, that determine the viewing angles in the x and z (camera frame) planes respectively for the i^{th} pixel in the remapped image.

To retrieve the aircraft's roll, pitch, and height, the SAD between the model disparities predicted by Equation 4.6 and the stereo disparities measured by the vision system was minimised using a non-linear optimisation algorithm. The non-linear optimisation was performed using the NLopt library [Johnson, 2009] implementation of Powell [2009]'s BOBYQA algorithm. Prior to fitting the disparity surface, the disparity maps were reduced to 64px × 96px to limit the number of operations performed during the minimisation. This implementation typically gave minimisation times in the order of 10ms, using $\sim 6 \times 10^3$ disparities on a 1.5GHz processor. This is faster than the time required to transfer image data from the cameras to the on-board computer ($\sim$ 40ms), and is thus sufficient for closed-loop control.

The advantage to this approach for estimating the attitude and altitude of an aircraft, is that measured stereo disparities are fitted to the ideal ground plane in disparity space. In Section 3.2.3, it was shown that the standard deviation of computed stereo disparities was independent of range. Fitting the ideal ground plane in disparity space, therefore, ensures that data points and average error are both distributed evenly over the fitted surface.

Reprojecting to 3D coordinates

The second approach for determining the attitude and altitude of an aircraft with respect to an ideal ground plane is to reproject measured stereo disparities into 3D coordinates relative to the nodal point of the vision system, and to fit the ideal ground plane in 3D space. While this procedure does not sample data points uniformly over the plane, it leads to a single-step, non-iterative optimisation that offers the advantage of low computational overheads and reliable real-time operation.

Radial distances, computed directly from stereo disparities via Equation 3.1, are

used to reproject disparity points into 3D space according to

$$\mathbf{p}^i = \frac{d^i_{radial}}{\sin \alpha^i} \cdot \hat{\mathbf{u}}^i, \tag{4.7}$$

where $\mathbf{p}^i$ is the reprojected location of the i^{th} pixel in 3D coordinates relative to the nodal point of the vision system; $\hat{\mathbf{u}}^i$ is the unit view vector for the i^{th} pixel (derived from the calibration parameters of the cameras); and α^i is the angle between the i^{th} view vector and the optic axis.

The ground plane, $\mathbf{p} \cdot \hat{\mathbf{n}} + d = 0$, is defined as above and the parameters for the plane, $\hat{\mathbf{n}}$ and d, are computed via an orthogonal, least-squares (non-iterative) regression. The regression was performed using the Geometric Tools [2010] mathematics library. Attitude and altitude of the aircraft are then extracted from the parameters for the fitted plane according to

$$\begin{aligned} \phi &= -\arctan\left(\frac{n_y}{n_z}\right), \\ \theta &= \arcsin\left(n_x\right), \\ d_{height} &= d, \end{aligned}$$

where ϕ and θ are the roll angle and pitch angle of the aircraft respectively; n_x, n_y, and n_z are the components of the plane normal, $\hat{\mathbf{n}}$, given in the body frame; and d_{height} and d are the altitude of the aircraft and the distance from the plane to the origin respectively.

Prior to fitting the ground plane, disparity maps are reduced to 64px × 96px to limit the maximum number of reprojected points used when applying the fit. This implementation typically gives computation times $<$ 1ms, using $\sim 6 \times 10^3$ reprojected points on a 1.5GHz processor. Hence, applying the planar fit in 3D space offers lower computational overheads at the cost of reduced precision for the estimates of the aircraft's attitude and altitude. Due to its computational efficiency, however, the least-squares regression may be implemented within an iterative framework (*i.e.* RANdom SAmple Consensus (RANSAC)) to reject outliers and improve the robustness of the state estimation.

4.2.2 Range testing

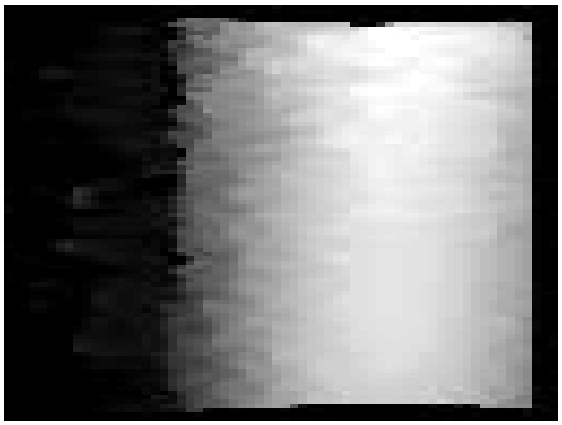
An outdoor test was conducted to analyse performance of the two approaches for estimating aircraft attitude and altitude. In the test, an aircraft was supported above the ground at a flat and grass-covered area. A single frame from the test is displayed in Figure 4.2. During the test, the aircraft was moved through a range of orientations and altitudes. The results of the outdoor test are shown in Figure 4.3. An IMU (MicroStrain 3DM-GX2[1]) was rigidly affixed to the vision system during the test to provide comparative measurements of the attitude of the aircraft, and an ultrasonic altimeter (MicroPilot AGL[2]) was attached to the wing of the aircraft to provide comparative measurements of altitude.

It can be seen from Figure 4.3 that both visual methods for computing aircraft attitude and altitude produced estimates that correlated well with the measurements used for comparison, over the range of altitudes tested. At higher altitudes, however, the signal-to-noise ratio (SNR) of the measured stereo disparities would decrease and the spread of the error in the reprojected 3D points would increase exponentially, while the spread of the error in the measured stereo disparities would remain relatively constant (see Section 3.2.3 for an analysis of the reprojection error). Estimates of attitude and altitude obtained from fitting the ideal ground plane in 3D space will therefore not be well-constrained as the aircraft approaches the altitude ceiling for the configuration of the stereo vision system used during flight testing (see Table 3.1 for the system configuration parameters and their values). This phenomenon is evident in the flight data presented in Figure 4.4, which was recorded as the aircraft was taking off under manual control.

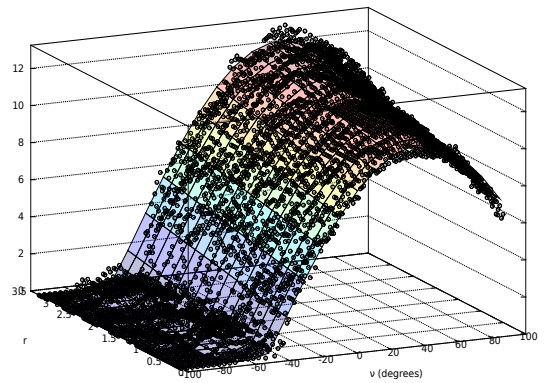
It can be seen from Figure 4.4 that attitude and altitude estimates obtained from fitting the ground plane in 3D space and in disparity space correlate well with each other, and that the visual attitude estimates also correlate well with the comparative measurements from the IMU. However, at heights > 20m AGL, attitude and altitude estimates obtained from fitting the disparity surface show less deterioration than those obtained from fitting the 3D point cloud. It was planned to use measurements from the ultrasonic altimeter as a reference for the altitude of the aircraft during the flight test. It was found, however, that the altimeter was unable to provide useful measurements above ~ 5m AGL.

[1] http://www.microstrain.com/inertial/3DM-GX2
[2] http://www.micropilot.com/products-accessories.htm

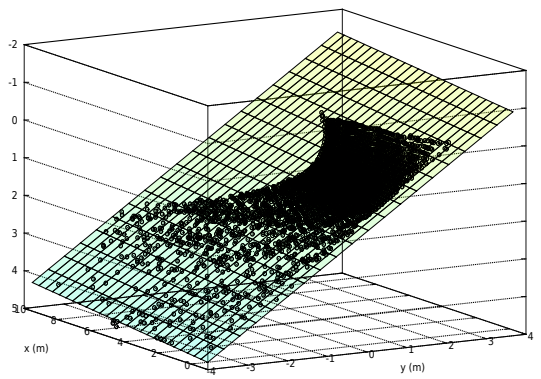
(a) Rear image crop.



(b) Terrain following disparity map.



(c) Fitted disparity surface. Horizontal axes define the viewing directions of the corresponding pixels in the disparity map. Vertical axis corresponds to stereo disparity in pixels.



(d) Fitted 3D plane. All axes are given in metres in the *body* coordinate frame (y is camera-right, z is camera-down, and x is parallel with the optic axis.

Figure 4.2: Single frame from the ground-based test showing the two approaches for computing attitude and altitude on board the aircraft. Pictured (a) is a crop of the raw image from the rear camera; (b) disparity map computed from the stereo pair of remapped images; (c) stereo disparities plotted with the best fitting disparity surface (defined in Equation 4.6); and (d) stereo disparities reprojected into 3D space and plotted with the best fitting planar surface.

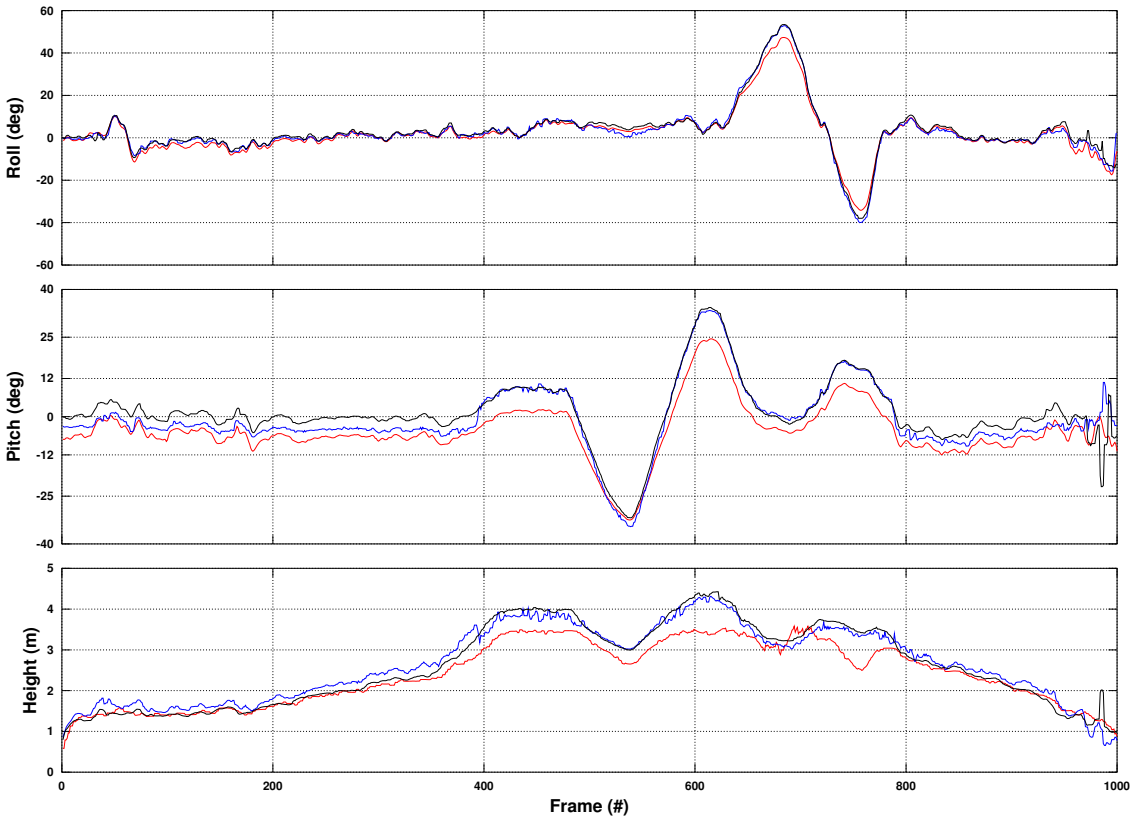


Figure 4.3: Attitude and altitude of the aircraft during an outdoor test as estimated via fitting the disparity surface (black) and fitting the 3D plane (blue). Also shown for comparison (red) are the attitude and altitude reported by an IMU and an ultrasonic altimeter respectively, which were rigidly affixed to the aircraft during the test. Frames were captured at approximately 12Hz.

4.3 Stereo vision-based terrain following

The stereo vision system can be used to estimate aircraft attitude and altitude accurately (Section 4.2). This information can then be used in combination with a closed-loop control system to enable an aircraft to perform low-altitude terrain following autonomously. The flight testing setup and control system are described below. Flight test results demonstrate the capabilities of this vision system for providing real-time guidance to a small-scale UAV.

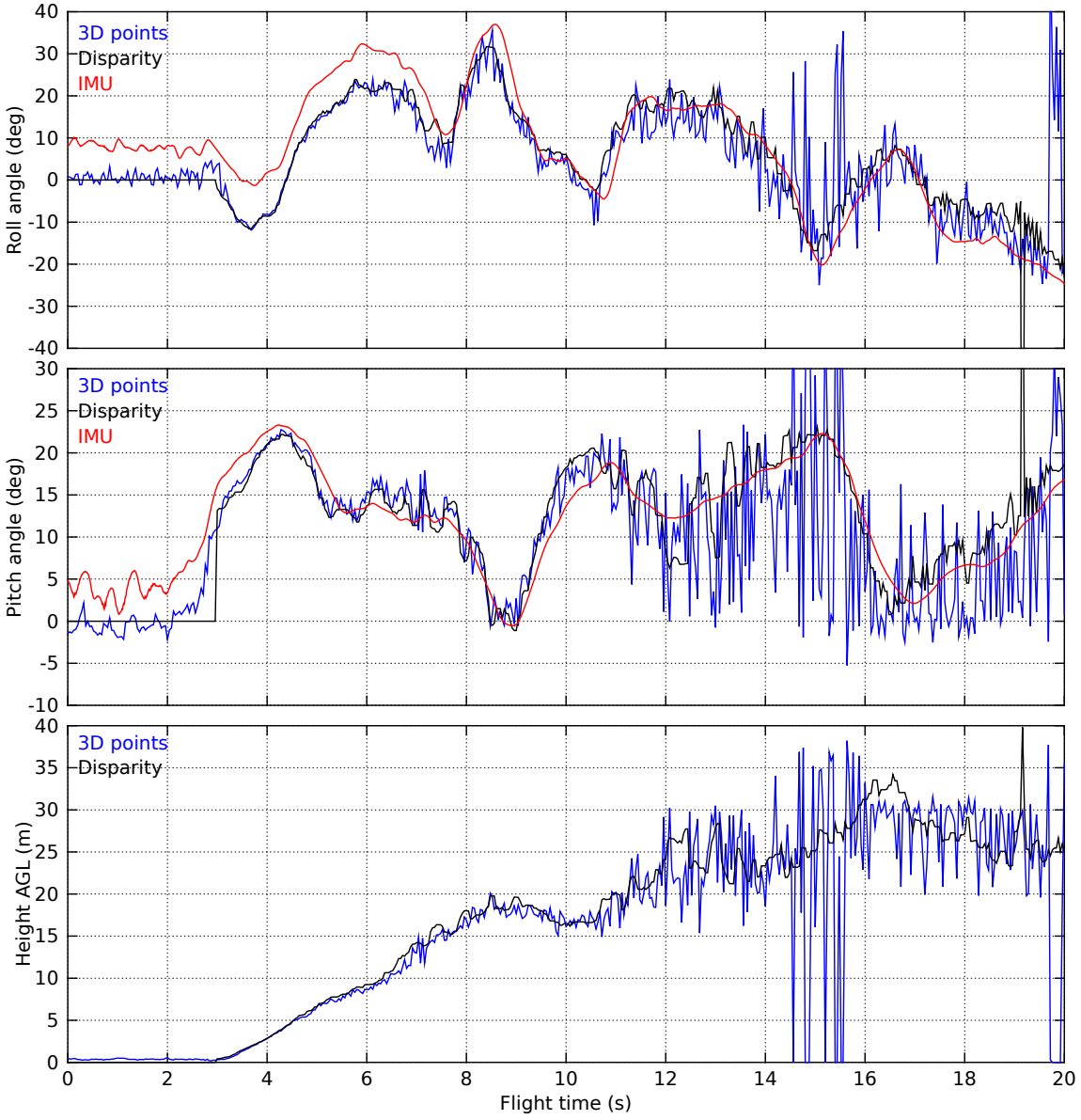


Figure 4.4: Attitude and altitude of the aircraft during a segment of manual flight in which the aircraft was performing a take-off. Attitude and altitude estimates obtained by fitting the disparity surface (black) are plotted with estimates obtained by fitting the 3D point cloud (blue). Also shown for comparison (red) is attitude reported by an IMU on board the aircraft.

4.3.1 Flight testing

The stereo vision system's ability to stabilise the attitude and altitude of an aircraft performing low-altitude terrain following was evaluated in a series of closed-loop flight tests. During the flight tests, the aircraft was required to recover from a variety of irregular orientations and trajectories quickly, and to restore a predefined attitude and altitude. The aircraft and on-board hardware were identical to the setup described for the collision avoidance flight tests in Section 3.3.3.

It was found in Section 4.2.2 that modelling the ground plane in disparity space provided estimates of the aircraft's attitude and altitude that were more precise and robust than those obtained via fitting the ground plane in 3D space. Figures 4.3 & 4.4 show, however, that modelling the ground plane in 3D space is still a viable means of estimating aircraft orientation and height with respect to the ground, as long as the aircraft is well below the operational altitude ceiling of the vision system. Furthermore, modelling the ground plane in 3D space rather than disparity space results in an optimisation that is at least one order of magnitude faster. During closed-loop flight testing, where the aircraft's proximity to the ground and other obstacles required minimal control lag, the ideal ground plane was therefore modelled in 3D space. To increase the robustness of the visual estimates of the aircraft's attitude and altitude to outlying stereo correspondences, the least-squares regression applied to the 3D point cloud was implemented within a RANSAC framework.

4.3.2 Closed-loop control

During closed-loop flight testing, PID feedback control loops were used to generate flight commands whilst attempting to minimise the errors between the visually estimated attitude and altitude of the aircraft and their respective set-points. The closed-loop control scheme is depicted in Figure 4.5. The aircraft's lateral and longitudinal dynamics were treated separately, so full autonomous control was achieved using two feedback control subsystems. Additionally, within each control subsystem, multiple control layers were cascaded to improve the stability of the system.

The first control subsystem comprised two cascaded PID controllers for stabilising the roll dynamics of the aircraft. The highest-level controller measured error between the roll angle set-point and the estimated roll angle and delivered an appropriate roll rate command to the lower-level controller, which implemented the desired

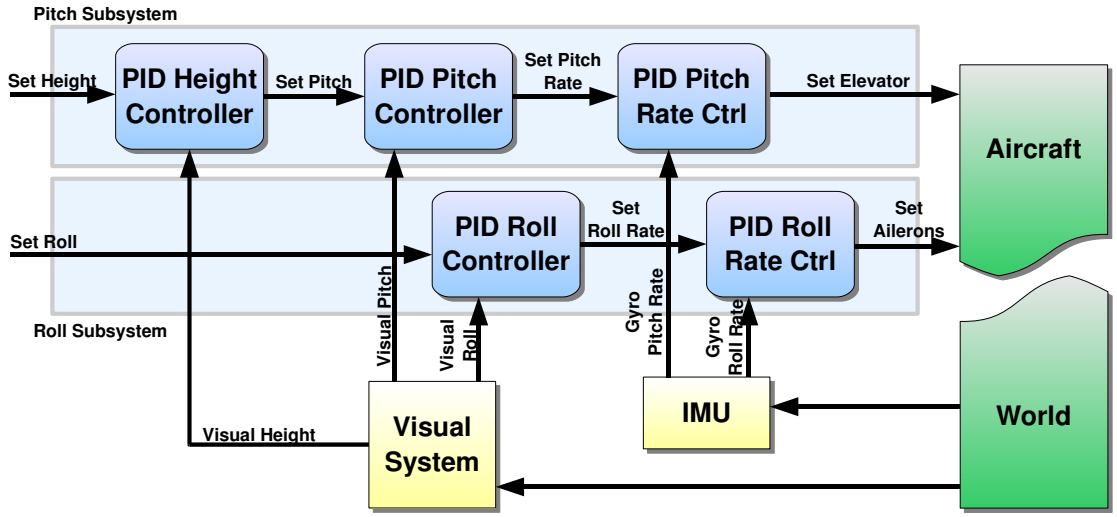


Figure 4.5: Block diagram illustrating the closed-loop control scheme used with the stereo vision system for autonomous terrain following.

roll rate. The aircraft's pitching dynamics were controlled in a manner similar to the roll dynamics. However, the pitch control subsystem included an additional cascaded PID controller to incorporate altitude stabilisation. As shown in Figure 4.5, error between the altitude set-point and the estimated altitude was measured by the highest-level controller, which generated an appropriate pitch angle command, and the remainder of the pitch control subsystem functioned identically to the roll subsystem. Measurements of absolute attitude and altitude of the aircraft were made by the stereo vision system and were used to drive all other elements of the closed-loop control system. Low-level control feedback for the roll rate and pitch rate commands was provided by an IMU (MicroStrain 3DM-GX2), which was affixed rigidly to the vision system during the flight test.

Setting dominant proportional gains for roll rate and pitch rate PID controllers means that the control surfaces of the aircraft are effectively being treated as accelerating inputs – where ailerons accelerate rolling motion and elevators accelerate pitching motion. During flight testing, this assumption was found to be true only for very high frequency motions, because the air flow field surrounding the aircraft quickly readjusts following a control surface deflection and acts to dampen rotation of the aircraft. For large changes in orientation, therefore, the aircraft's control surfaces could more accurately be approximated as velocity controls. A more responsive control system could thus be produced by collapsing the absolute angle and rate

controllers in Figure 4.5 into a single PID controller for each subsystem (in this case, the rate measurements from the IMU could be used as inputs for derivative control). Closed-loop flight data presented in Section 4.3.3, however, was collected using the control system illustrated in Figure 4.5. Removing the rotation rate feedback provided by the IMU altogether would not necessarily affect the overall function of the system, although the proportional feedback gains for the roll and pitch angle of the aircraft would have to be reduced to avoid instabilities, which would lead in turn to reduced responsiveness.

4.3.3 Autonomous terrain following

Two closed-loop flight tests were performed to evaluate the stereo vision system's ability to provide real-time guidance to an aircraft performing low-altitude terrain following. The approximate flight plan for each terrain following flight test is illustrated in Figure 4.6.

Stabilising roll angle and altitude

For the first test, the aircraft was piloted manually around a circuit. On each pass, the aircraft was positioned with an irregular attitude and altitude and then autonomous control was engaged for a period of 5s $\sim$ 10s. A quantitive measure of the performance of the system was obtained by computing the response time and steady-state precision of the guidance system during autonomous flight.

Altitude of the aircraft during the first flight test, estimated via the vision-based approach proposed here, is displayed in Figure 4.7. During the periods of autonomous control, the aircraft was programmed to hold a roll angle of 0° with respect to the ground and an altitude of 10m AGL. A typical segment of flight is analysed in Figure 4.8, during which the aircraft made two autonomous passes (corresponding to 375s $-$ 410s in Figure 4.7 and shown in the accompanying video[3]).

Figure 4.7 shows that the aircraft was able to reduce the error between its initial altitude and the set-point in every autonomous pass during the flight test, despite initial altitudes varying between 5m $\sim$ 25m AGL. From Figure 4.8 it can be seen that the aircraft was also able to stabilise its attitude effectively during the autonomous passes – *e.g.* recovering quickly from a roll angle error of $> 50°$ at 397s.

[3]Attachment AV02 (available at http://youtu.be/qW78288a3Cs).

Figure 4.6: Bird's-eye view of the area in which the first (a) and second (b) terrain following flight tests were performed, indicating the approximate flight path during manual flight (yellow) and autonomous terrain following (blue). Green arrows (b) indicate the approximate direction of the downhill slope in the vicinity of the flight testing region. Unfortunately the GPS path of the aircraft was not available during these flight tests. The scale indicates 50m. Satellite imagery ©2012 Google, DigitalGlobe, GeoEye, Cnes/Spot Image.

Temporary deviations between visually estimated roll and pitch angles and values reported by the IMU are to be expected, due to the inherent difference between the measurements performed by the stereo vision system, which measures attitude with respect to the local orientation of the ground plane, and those performed by the IMU, which measures attitude with respect to gravity. It can be seen from Figure 4.8, however, that the visual estimate of the aircraft's pitch angle was offset negatively with respect to the measurement obtained from the IMU for extended periods of time. This was a factor throughout the first flight test, as can be seen clearly from Figure 4.9, which displays the distributions of the differences between the roll angles and pitch angles measured by the stereo vision system and IMU during the autonomous segments of flight. The histograms do not include data from the initial response phase of each autonomous segment.

Following the flight test it was found that the gyroscopic pitch rate reported by the IMU had a significant positive bias, which – when integrated by the IMU to obtain attitude – would have contributed to the discrepancy between the pitch angles reported by the IMU and stereo vision system during the flight test. Figure 4.7 shows, however, that the aircraft tended to hold an altitude that was up to one metre above the set-point during the autonomous passes, suggesting that either a negative angle-of-attack was required for level flight (which wasn't accommodated by the height PID

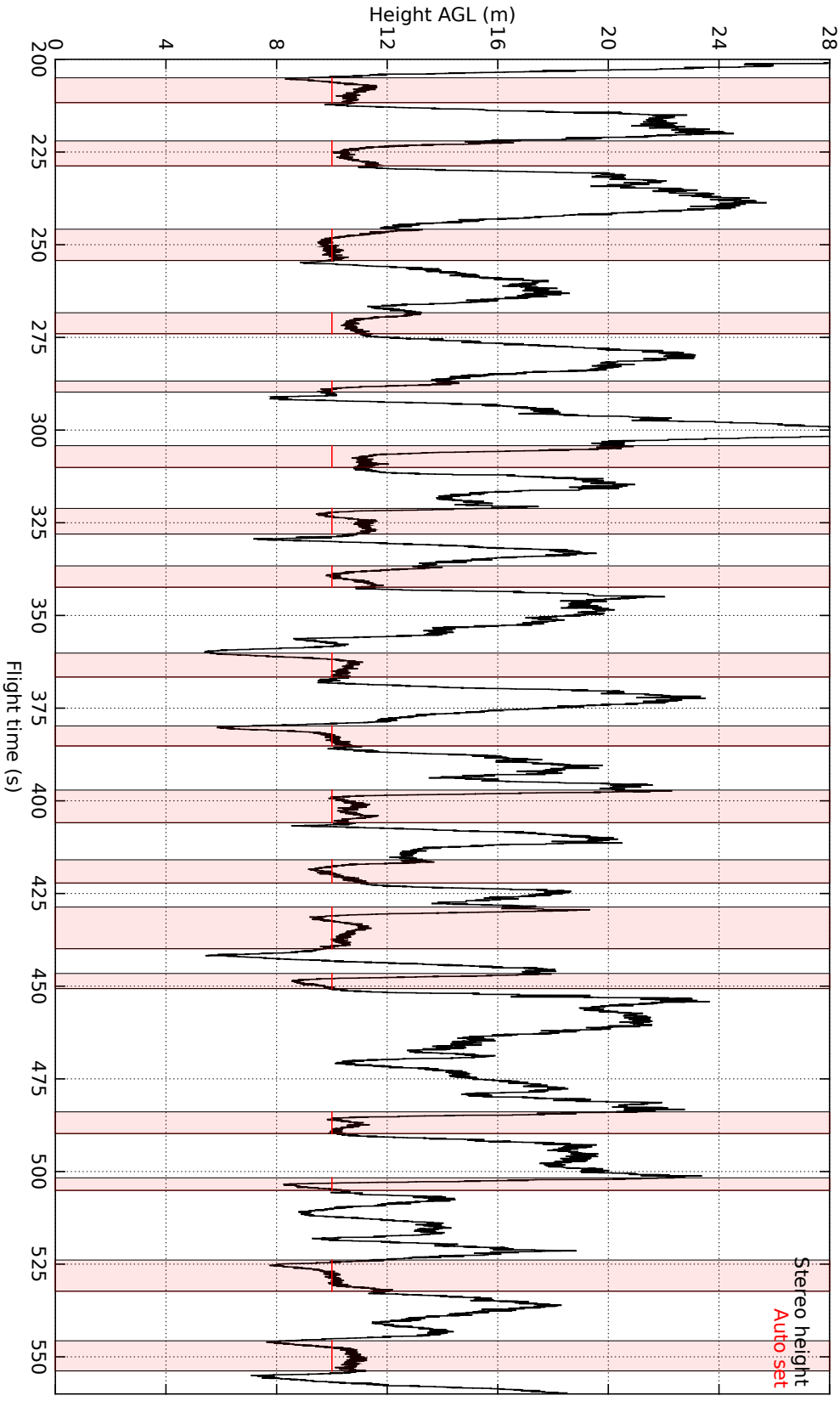


Figure 4.7: Altitude of the aircraft estimated by the stereo vision system (black) during the first terrain following flight test. During the periods of autonomous control (shaded red) the aircraft was programmed to hold an altitude of 10m AGL (red).

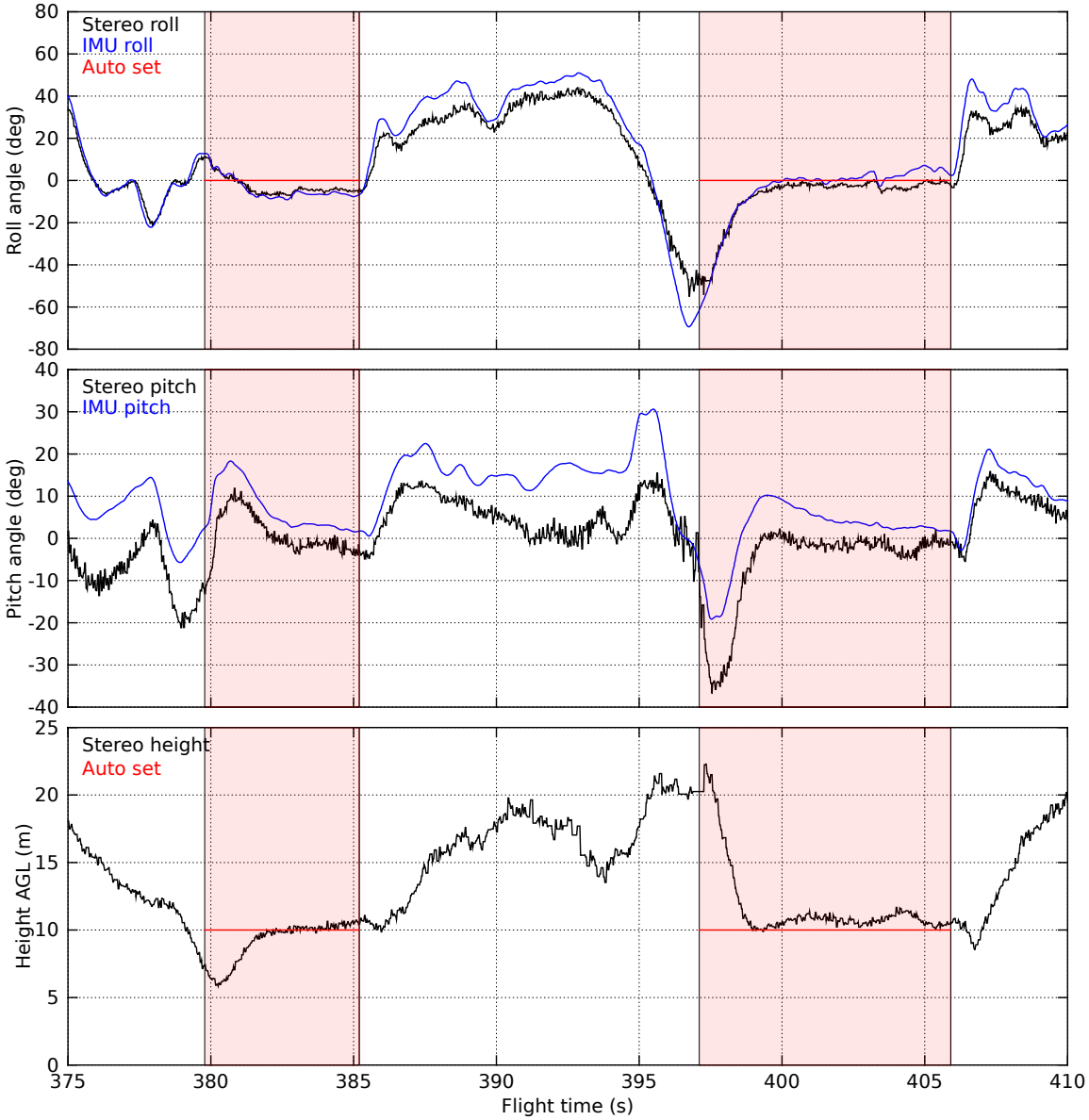


Figure 4.8: Roll angle (top), pitch angle (centre), and altitude (bottom) of the aircraft during a segment of the first terrain following flight test, as estimated by the stereo vision system (black), and reported by the IMU on board the aircraft (blue). Also shown are the periods of autonomous control (shaded red), during which the aircraft was programmed to hold a roll angle of 0° (red) with respect to the ground and an altitude of 10m AGL (red).

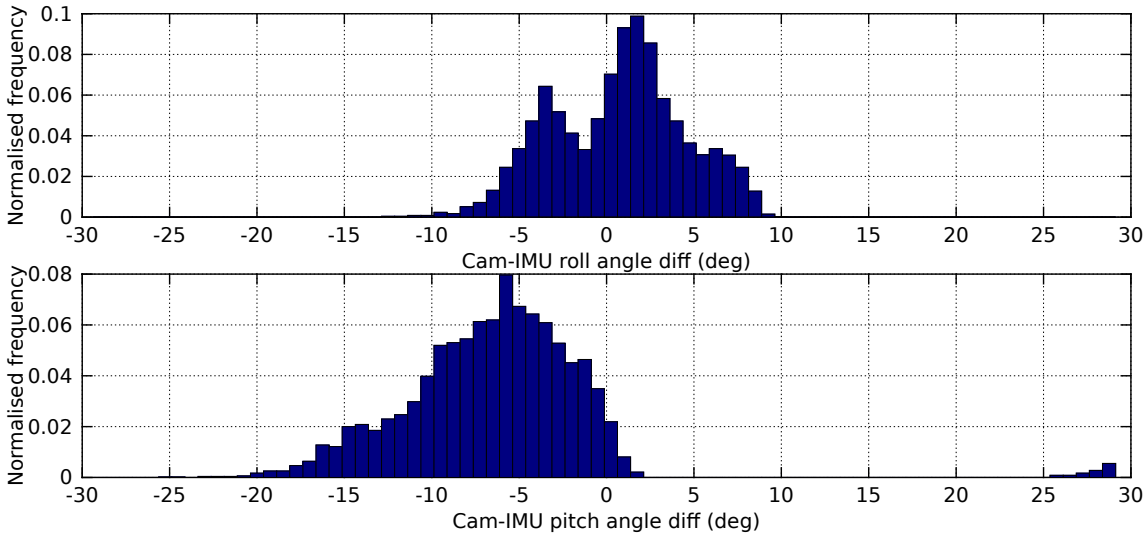


Figure 4.9: Distributions of the differences between roll angles (top) and pitch angles (bottom) measured by the stereo vision system and IMU on board the aircraft during the first terrain following flight test. Measurements were made only during autonomous passes once the aircraft had levelled out. Histograms have been normalised such that the sum of all bin frequencies is equal to one.

in the control system), or that the vision system was underestimating the aircraft's pitch angle. Discrepancy between pitch angles reported by the vision system and the IMU during the first flight test was thus likely to be due to a combination of factors.

It can be seen from Figure 4.8 that during the autonomous pass from 380s $\sim$ 385s the control system was unable to level the roll angle of the aircraft fully but was able to do so for the autonomous pass that commenced at 397s. In fact, for several of the autonomous passes during the first flight test, the steady-state roll angle showed a negative bias with respect to the set-point. This suggests that the aircraft had a negative aileron trim, which the closed-loop control system was unable to overcome fully. This was attributed to the inclusion of the roll rate and pitch rate PID controllers, which may have introduced a damping effect that caused the system to be unresponsive to small attitude errors, as discussed in Section 4.3.2.

Interestingly, the histogram of the differences between the roll angles estimated by the vision system and those measured by the IMU during the first flight test (plotted in Figure 4.9) indicates a bimodal distribution, centred around zero. This type of distribution might be expected if the area in which the straight segments of terrain following flight were conducted was sloping slightly in a direction perpendicular to the flight path (as was the case for these flight tests). Under this circumstance, a

gravity-referenced roll angle (such as that reported by an IMU) would be expected to show a positive or negative bias with respect to a local ground plane-referenced roll angle, depending on the direction of flight. Despite the irregularities with the attitude measurements, however, the aircraft recovered from a range of dangerous situations successfully and maintained a stable attitude and altitude during terrain following.

Performance of the guidance system was quantified by considering two metrics: the time that elapsed between the start of each autonomous segment and the aircraft first passing within one metre of the altitude set-point; and the average attitude and altitude of the aircraft during the remainder of each autonomous segment (*i.e.* not including the initial response phase) compared to the respective set-points. These metrics were used to obtain a measure of the response time and steady-state precision of the system respectively. From the data presented in Figure 4.7, the average response time of the system was calculated to be $1.45\text{s} \pm 1.3\text{s}$, where error bounds represent $\pm 2\sigma$ from the 18 closed-loop trials. The relatively high variance of the average response time was due to the large range of initial altitudes. Using the second metric defined above, the average unsigned roll angle and altitude errors were calculated to be 3.9° and $6.4 \times 10^{-1}\text{m}$ respectively from approximately 92s of continuous segments of autonomous terrain following. These performance metrics indicate that the closed-loop system is able to respond to sharp adjustments in altitude quickly and also that the system is able to hold a set attitude and altitude precisely, thereby validating its use for autonomous terrain following.

Maintaining a constant-height turn over sloping terrain

For the second closed-loop flight test, both the stereo vision system's ability to handle uneven terrain and the closed-loop control system's ability to handle external disturbances were evaluated by commanding the aircraft to perform a circular terrain following flight above sloped and uneven terrain (the flight plan is illustrated in Figure 4.6b). During periods of autonomous control the aircraft was programmed to hold an altitude of 15m AGL and a roll angle of -45° with respect to the immediate ground plane. A typical segment of flight (shown in the accompanying video[4]) during which the aircraft performed a single autonomous pass is analysed in Figure 4.10. Attitude and altitude estimated by the vision system are plotted alongside attitude measured by the IMU on board the aircraft for comparison.

[4]Attachment AV03 (available at http://youtu.be/wLbwgo1oX5Y).

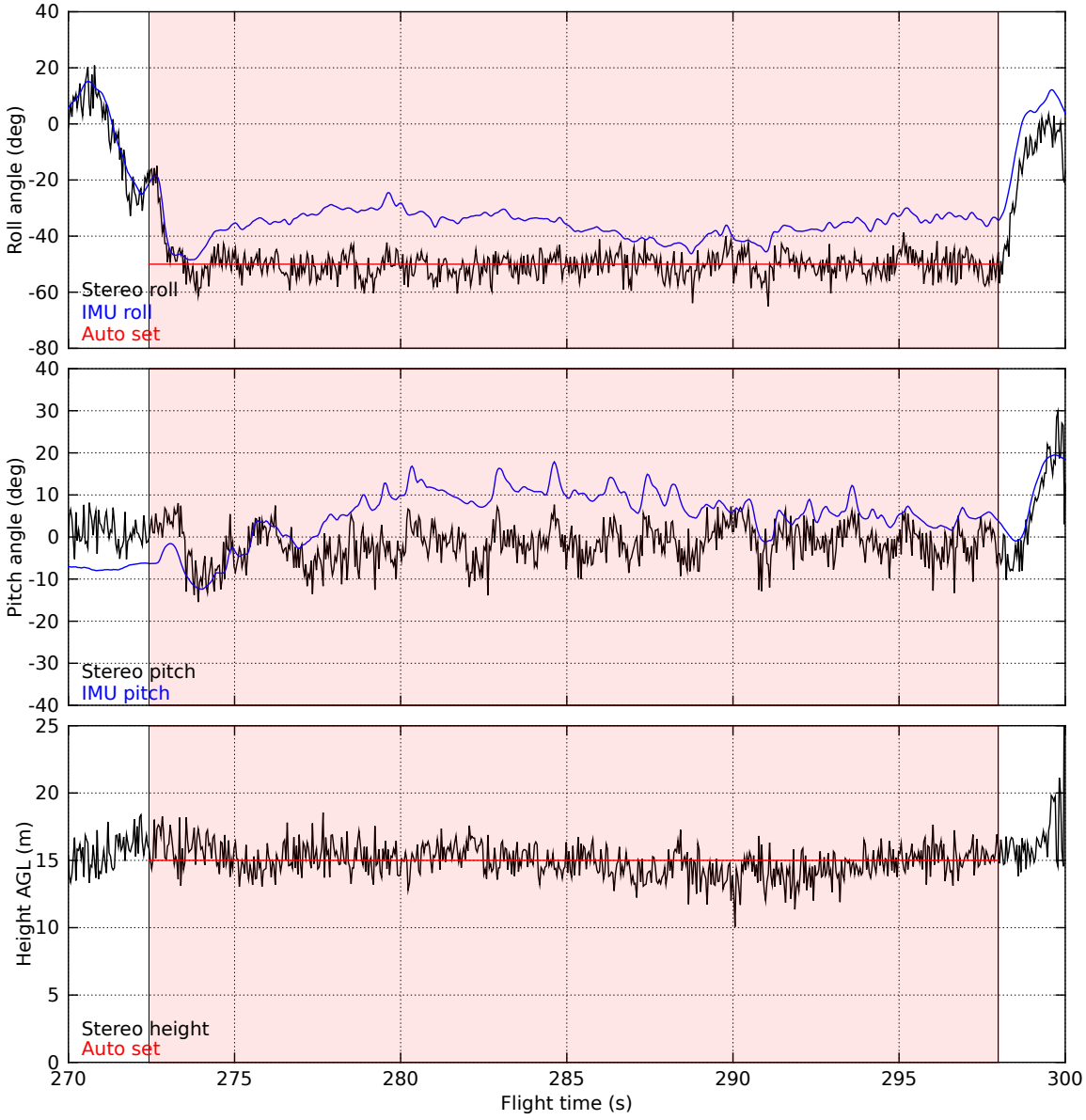


Figure 4.10: Roll angle (top), pitch angle (centre), and altitude (bottom) of the aircraft during the second terrain following flight test, as estimated by the stereo vision system (black) and reported by the IMU on board the aircraft (blue). Also shown is a period of autonomous control (shaded red), during which the aircraft was programmed to hold a roll angle of $-45°$ (red) with respect to the ground and an altitude of 15m AGL (red).

It can be seen that the stereo vision system reported that the aircraft was able to maintain the commanded roll angle and altitude successfully, despite the unevenness and slope of the terrain. The average unsigned errors between the visually estimated roll angle and altitude and their respective set-points were 3.7° and 9.5×10^{-1}m respectively. The attitude reported by the IMU, however, did not seem to correlate particularly well with the visual estimates. This was attributed to two factors, discussed below.

Firstly, the aircraft flew approximately 1.5 laps of a circular path above a significantly sloped hillside during the autonomous segment of flight. The error between a gravity-referenced attitude (such as that obtained from an IMU) and a local ground plane-referenced attitude would therefore be expected to vary sinusoidally, with a period equal to the period of each of the aircraft's laps and a 90° phase shift between the roll angle error and pitch angle error. It can in fact be seen from Figure 4.10 that the differences between the visual estimates of roll angle and pitch angle and those from the IMU vary sinusoidally and are approximately one quarter of a lap out of phase.

Secondly, the centripetal force exerted by the aircraft in order to maintain its circular flight path would have confounded the IMU's estimate of the gravity vector – leading it to report a more upright roll attitude than was actually the case. Furthermore, because the fixed-wing aircraft was required to apply a continuous body-frame pitch rate in order to maintain a constant turn rate and the IMU on board the aircraft was underestimating the roll angle, this would have lead the IMU to overestimate the pitch angle of the aircraft through integration of the positive pitch rate. Both of these phenomena are visible in the data presented in Figure 4.10 and they serve to highlight the fundamental difference between the measurements performed by the stereo vision system and an IMU, and the superior performance of the vision system under these conditions.

4.4 Discussion

The closed-loop flight testing results presented in Section 4.3 have been analysed to quantitatively compare the performance of the proposed stereo vision system with other state-of-the-art approaches.

4.4.1 Performance

Section 4.2 described two methods for determining aircraft attitude and altitude with respect to the local ground plane using the stereo vision system introduced in Chapter 3. This vision system has been coupled with a simple PID-based closed-loop control system to demonstrate low-altitude terrain following and attitude control during two autonomous flight tests (Section 4.3). In the first flight test it was found that the guidance system was able to recover the aircraft from a range of initial conditions quickly and to maintain a predetermined attitude and altitude precisely. The average time taken for the aircraft to recover was 1.4s and the mean unsigned deviation of the attitude and altitude estimates from the set-points were 3.9° and 6.4×10^{-1}m respectively, once the aircraft had levelled out at the predefined altitude of 10m.

During the second flight test it was found that the guidance system was able to maintain the attitude and altitude set-points precisely even as the aircraft traversed uneven terrain, giving average unsigned errors of 3.7° and 9.5×10^{-1}m respectively. The slightly larger average altitude deviation during the second flight test was attributed to several factors: the aircraft was flying at an altitude of 15m AGL, which was slightly higher than during the first flight test, hence the stereo vision-based range estimates would have been less accurate; the average roll angle ($\sim 40°$) during the second flight test would have reduced the effectiveness of the elevators at controlling the height of the aircraft; and the terrain over which the aircraft flew was uneven and sloped.

To quantitatively compare the performance of the stereo vision system with other previously published vision-based guidance systems, ground-truth measurements of the aircraft's attitude and altitude during flight testing are neccessary. Such measurements are difficult to obtain, or would require expensive equipment such as a high-precision IMU and a differential GPS (DGPS). Here, a relatively inexpensive IMU was used to provide comparitive measurements of the aircraft's attitude, but unfortunately no such measurements were available for the height of the aircraft. In Section 3.2.3, however, it was found that the vision system measured stereo disparities with an accuracy of approximately $\sigma = 2.5 \times 10^{-1}$px, corresponding to a mean unsigned range error of approximately $\sigma = 3.5 \times 10^{-2}$m over the ranges at which that indoor test was conducted.

Using the computed variance of the disparity estimates, the configuration of the

system during the terrain following flights (Table 3.1), and the relationship between disparity and radial distance (Equation 3.1), it can be predicted that radial distance error at an altitide of 10m AGL is in the order of $\sigma = 8 \times 10^{-1}$m for each range measurement. In fact, during the second terrain following flight test, average unsigned distance error measured orthogonally from each reprojected 3D inlier to the fitted ground plane was approximately 5×10^{-1}m, with an average of approximately 500 inliers per frame. From Figures 4.4 & 4.8 it can be seen that the standard deviation of estimates of the aircraft's height at 10m AGL are in fact significantly less than the predicted variance of the range measurements. This is a result of the least-squares regression used to fit the ground plane in 3D space and the RANSAC framework in which the fit is applied. It is assumed, therefore, that estimates for attitude and altitude of the aircraft provided by the stereo vision system are reliable for an aircraft flying at low altitude.

Previous authors have described various techniques for performing vision-based terrain following autonomously. For example, Herisse et al. [2010] describe an approach in which the distance from the aircraft to the ground plane is maintained by controlling the height of the aircraft such that the measured optic flow corresponds to a predefined value. The optic flow set-point is computed from the ground speed of the aircraft, which is regulated externally. The system is implemented on board a quadrotor and its performance is demonstrated indoors by following and avoiding angled planar surfaces at a commanded distance of 8×10^{-1}m and a speed of 5×10^{-1}ms^{-1}. No quantitative analysis of measurement accuracy with respect to a ground truth is provided, however the control system is able to maintain the vehicle's height to within approximately $\pm 4 \times 10^{-1}$m of the set-point. The orientation of the quadrotor is controlled on board via an inertial system but the optic flow-based height guidance is relayed from a ground-based workstation.

Garratt and Chahl [2008] describe a similar approach, in which the height of a helicopter above the ground plane is regulated using optic-flow measurements obtained from a downward facing camera on board the aircraft. Extended periods of terrain following are demonstrated in an outdoor environment in which the pitch of the helicopter is controlled manually and the thrust and roll angle are controlled autonomously using the magnitude and direction of the measured optic flow. The ground speed of the aircraft is measured by GPS and used by the control system on board the helicopter to regulate its height at 1.27 ± 0.36m AGL, where the error bounds represent $\pm 2\sigma$ from the optic flow-based height estimates. The accuracy

of the height estimates is analysed by comparing them with measurements from a laser range-finder on board the aircraft. The maximum error range is indicated as approximately $\pm 2 \times 10^{-1}$m at a forward speed of 5ms^{-1}.

Beyeler [2009], Zufferey et al. [2010] also describe an optic flow-based guidance scheme for a small-scale fixed-wing UAV that consists of a set of miniature optic flow sensors arrayed at viewing directions 45° from the forward axis of the aircraft. The roll angle of the aircraft is commanded to counteract asymmetric distributions of optic flow across the sensors and the height of the aircraft is regulated using the magnitude of the measured optic flow and an assumed ground speed. No quantitative measurements of the accuracy of the system are given but its robustness is demonstrated through extended flights in outdoor environments. Measurements from a GPS and an IMU on board the aircraft during the flight tests indicate that the guidance system is able to maintain a roll angle of $\sim 0 \pm 4°$ and a height of $\sim 9 \pm 2$m AGL, where the error bounds correspond to $\sim 2\sigma$ for a flight speed of 12ms^{-1}.

Stereo vision-based approaches for controlling the height of an aircraft flying at low altitude have also been described previously. For example, Roberts et al. [2002, 2003] utilise two downward-facing cameras and compute feature-based stereo correspondences to estimate the height of a small-scale helicopter flying in an outdoor environment. It is shown that the height estimates generated by the system correspond to measurements reported by an ultrasonic altimeter over a small range of altitudes at which a comparison was possible [Corke et al., 2001] and that the stereo vision-based height estimates have a standard deviation of $\sigma \approx 2.5 \times 10^{-1}$px at an altitude of approximately 5m during manual flight outdoors [Roberts et al., 2002, 2003]. However, no quantitative analysis is given for the accuracy of the approach.

Alternatively, Eynard et al. [2012] describe a hybrid fisheye-perspective stereo system with which they use a plane-sweeping [Collins, 1996] approach to estimate the altitude of a small-scale UAV. They compare the stereo vision-based height – estimated during a manual take-off manoeuvre – with values reported from a laser range-finder on board the aircraft and report the mean error to be 2.41%, or approximately 3.6×10^{-2}m over the range of altitudes covered in the test (0.6m−2.2m). The attitude of the aircraft is estimated visually by identifying vanishing points [Bazin et al., 2010, Demonceaux et al., 2007] and is quoted as providing roll angle and pitch angle estimates with mean absolute errors of 1.32° and 1.96° respectively when compared with measurements provided by an IMU. The system operates in real time

on a ground-based workstation, although no closed-loop flight data is presented.

A stereo vision-based system is also described by Campoy et al. [2009], Mondragón et al. [2010b] for the purpose of estimating the height of a small-scale helicopter above the ground plane. Results from outdoor flight testing are presented in which the helicopter was programmed to hold a certain trajectory autonomously using state information computed on board the helicopter through the integration of multiple sensory inputs (GPS, IMU, magnetometer, vision). The altitude of the aircraft computed via stereo vision is compared with the filtered state estimate to produce a measure of the mean squared-error of the height estimates of 1.73×10^{-1}m at a height of approximately 1.6m.

It can be seen from Table 4.1 that the stereo vision-based guidance system investigated in Part I of this thesis compares favourably with the state-of-the-art approaches described above, where comparison is possible. The spread of range estimates obtained using the stereo vision system is more accurate (in some cases an order of magnitude more accurate) than that for systems with ground truth comparisons discussed above. In fact, the comparative ranging accuracies listed in Table 4.1 are derived from estimates of the heights of the various aircraft with respect to the ground truths, whilst the value listed for the proposed approach is derived from individual range estimates measured during testing (Section 3.2.3). The static altitude estimates obtained from the stereo vision system would therefore be expected to be more accurate than the value listed (at the same range of 1.2m) due to the error rejection properties of the plane fitting procedure.

Accuracy of the proposed approach enables the guidance system to control aircraft attitude and altitude more precisely and at higher altitude than the previously published approaches discussed in this section. Additionally, the system operates in real time and completely independently of any ground-based workstation, which enables vision-based guidance at a faster flight speed than has been described previously. For altitudes $>$ 20m, however, the accuracy of this approach degrades quickly due to limitations of stereo vision and the configuation of the system. The limitations of this approach are discussed in greater detail below.

4.4.2 Limitations

General limitations of the stereo vision system were discussed in Section 3.4.2. Further limitations can also be identified that are specific to the methods proposed

	Terrain following precision		Ranging accuracy	
	height$\pm 2\sigma$(m)	mean error (%)	range$\pm 2\sigma$(m)	mean error (%)
Herisse et al. [2010]	0.8 ± 0.4	20	-	-
Garratt and Chahl [2008]	1.27 ± 0.36	14	1.27 ± 0.2	6.3
Beyeler [2009]	9 ± 2	8.8	-	-
Eynard et al. [2012]	-	-	1.4 ± 0.085	2.4
Campoy et al. [2009]	-	-	1.6 ± 0.83	21
Proposed	$\mathbf{10 \pm 0.64}$	**2.5**	$\mathbf{1.2 \pm 0.035}$	**2.3**

Table 4.1: Comparison of the errors indicated for the various approaches discussed in Section 4.4.1. Terrain following precision represents the closed-loop performance for each approach with respect to the mean altitude estimate or set-point, and ranging accuracy represents the accuracy of the range or altitude estimates with respect to a ground truth measure. Two values are given for each metric: the 95% confidence bounds ($\pm 2\sigma$) for the visual estimates and the height/range at which the measurements were made; and the average unsigned error (mean error), given as a percentage of the height/range. Standard deviations have been converted to approximate mean errors by multiplying by $\sqrt{\frac{2}{\pi}} \approx 0.798$, which assumes normally distributed data.

in this chapter for estimating and controlling the attitude and altitude of an aircraft performing low-altitude terrain following. Inherent for stereo vision-based systems is the limitation that range can be measured accurately over a finite spectrum of distances only, due to the fixed baseline of the stereo system. For the vision system investigated here, the fixed baseline results in a practical altitude ceiling, where error bounds associated with each range measurement become too large for the measurement itself to be meaningful. For example, the accuracy of the stereo vision system has been stated as $\sigma = 2.5 \times 10^{-1}$px; therefore, according to Table 3.1 and Equation 3.1, the disparity measured by the stereo system for an object at $d_{radial} = 30$m would be expected to be $D_{pixel} = 1 \pm 0.5$px, for a confidence level of 2σ. This corresponds to a 95% confidence bound of $20\text{m} - 61\text{m}$, which is significant compared to the true range of $d_{radial} = 30$m.

Modelling the ground as a planar surface allows partial information from many simultaneous measurements to be collated, thereby reducing the uncertainty in the final measurement. From Figure 4.4 it can be seen that this approach allows the attitude and altitude of the aircraft to be estimated accurately up to an altitude of ~ 20m. The susceptibilty of the system to noisy range measurements for altitudes > 20m can be reduced further by fitting the 3D ground plane model within a RANSAC

framework, which increases the range over which valid measurements of attitude and altitude can be made to $\sim$ 30m (*e.g.* Figures 4.7 & 4.8). Alternatively, the ground plane can be modelled in disparity space, which results in estimates of attitude and altitude that are more accurate and more robust $>$ 30m (*e.g.* Figure 4.4), because very small disparities ($|D_{pixel}| \ll$ 1px) can be included in the fit without injecting substantial amounts of noise.

Another limitation of vision systems that rely solely on stereo-based range measurements is that the motion of the aircraft is not measured directly. Using the approach described within this chapter, the roll rate, pitch rate, and height rate of the aircraft may be derived numerically from the raw attitude and altitude measurements. Translational motion of the aircraft, however, must be measured from image correspondences between successive frames using either optic flow or feature-based methods. This topic is addressed briefly below.

4.5 Applications

Sections 3.3 & 4.3 have demonstrated the ability of this stereo vision system to provide guidance for terrain following and collision avoidance to a small-scale UAV flying autonomously at low altitude. Tasks such as mapping or visual odometry, however, require that aircraft egomotion (rotation and translation) is measured or estimated in addition to attitude and altitude. To retrieve the translational component of egomotion, image correspondences must be made between successive frames using either optic flow or feature-based methods. In Section 4.5.1, the results of a brief investigation are presented in which the stereo vision system was augmented with measurements of optic flow.

4.5.1 Combining stereo disparity and optic flow

Optic flow allows the motion of the aircraft to be measured. However, to facilitate computation of the 3-DOF translation of the aircraft, the rotational component of the optic flow must be subtracted from the raw measurements. This 'de-rotation' step is difficult to perform using visual information only, but is greatly simplified when optic flow vectors are available from opposing hemispheres of the view sphere [Koenderink and Doorn, 1987]. A vision system that is designed specifically for this task is introduced in Chapter 5. For the present investigation, however, rotational rates

reported by an IMU on board the aircraft were used to isolate just the translational components of the raw optic flow vectors (*i.e.* Figure 4.11). The distance and orientation of the ground plane with respect to the aircraft were obtained from the stereo vision system. Individual estimates of 3-DOF translation of the aircraft were thus computed from each flow vector by intersecting view vectors corresponding to each motion start and end point with the modelled ground plane.

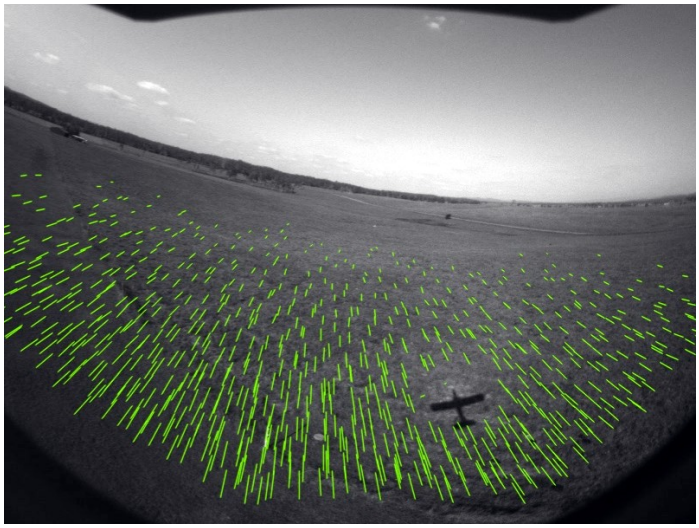


Figure 4.11: Crop of a raw image from the fore camera during a manual take-off overlaid with the estimated translational component of the measured optic flow vectors (green). Note that flow vectors that did not intersect the modelled ground plane were discarded. The 'hole' in the optic flow field surrounding the shadow of the aircraft occurs because the shadow generates no parallax and hence the only observed motion in this region is due to rotation of the aircraft, and is thus subtracted by the 'de-rotation' step.

This approach was implemented offline and used to compute attitude, altitude, and ground speed of the aircraft from a recorded segment of manual flight during which the aircraft was performing a take-off (shown in the accompanying video[5]). The flight data is plotted in Figure 4.12. Raw image pairs were captured from the stereo cameras at 25Hz and used to compute attitude and altitude of the aircraft via the disparity surface approach described in Section 4.2. Flow vectors were computed between successive raw images from the fore camera at a regular grid of pixel locations using a hierarchical block matching algorithm developed by Mr. Saul Thurrowgood and then de-rotated using the rotational rates recorded from an IMU on board the aircraft during the test. Ground speed of the aircraft was estimated from the median

[5]Attachment AV04 (available at http://youtu.be/WhhDeXCQpwM). Note that translational components of measured optic flow vectors have been drawn on the on-board imagery in this video.

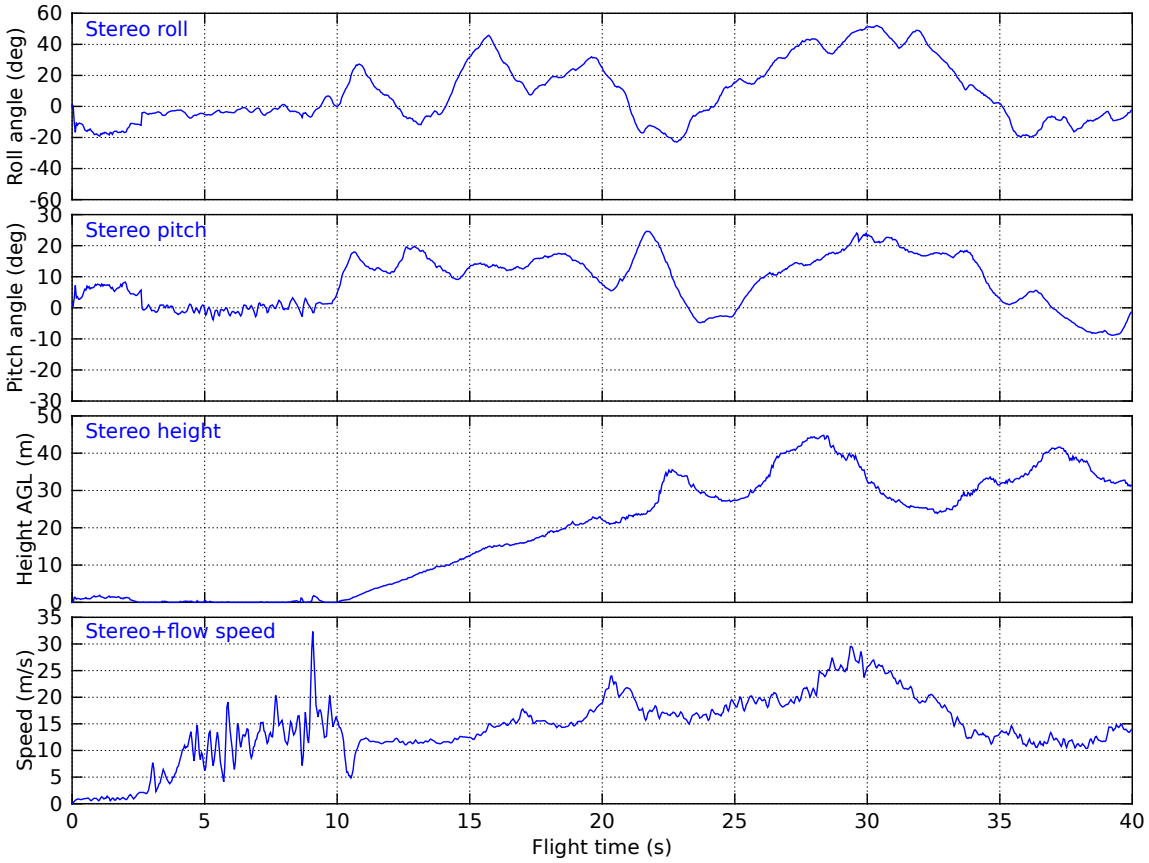


Figure 4.12: Roll angle (top panel), pitch angle (2nd panel), and height (3rd panel) of the aircraft with respect to the ground plane, estimated by the stereo vision system; and the ground speed (bottom panel) of the aircraft, estimated by augmenting the stereo vision system with measurements of optic flow between successive frames from the fore camera. Details in text.

of the magnitudes of the scaled translational flow vectors, where the flow vectors were scaled by the distance to the modelled ground plane along each view vector.

It can be seen from Figure 4.12 that the flight data estimated by the vision system during the flight test is internally consistent (increases in pitch attitude correspond to altitude gains and *vice versa*), although no ground truth data was available to validate the estimated altitude or ground speed. At the beginning of the flight segment the aircraft was moving but still in contact with the ground. Hence, much of the ground plane was too close for the vision system to measure stereo disparity or optic flow reliably, which lead to noisy state estimates < 10s. After leaving the ground, the vision system was able to provide valid estimates of the attitude, altitude, and ground speed of the aircraft despite reaching altitudes > 45m AGL.

4.6 Conclusion

A wide-angle stereo vision system designed specifically for providing guidance to a small-scale UAV flying in a low-altitude environment was introduced in Chapter 3. A simple guidance scheme was also described that allows an aircraft to autonomously avoid colliding with obstacles that penetrate a notional flight cylinder surrounding the projected flight trajectory. A complementary guidance scheme that allows an aircraft to estimate its attitude and altitude with respect to the local ground plane was described in this chapter. The system has been shown to operate in real time and is hence well suited to providing guidance for an aircraft performing tasks such as terrain following or landing, where the aircraft is in close proximity to the ground and must control its attitude and altitude precisely.

The results of open-loop and closed-loop flight tests have demonstrated the capabilities of this approach. The proposed vision system provides more precise control of attitude and altitude than other similar vision-based systems. Additionally, terrain following flight was demonstrated at a faster ground speed than previously published optic flow-based approaches and at a higher altitude than previously published stereo vision-based approaches. Coupled with the collision avoidance scheme presented in Chapter 3, this vision system provides the neccessary guidance for an aircraft to fly fast and low to the ground safely, even in the presence of obstacles.

The aim of this thesis is to investigate whether simple, vision-based sensing and guidance algorithms can be combined with wide-angle vision systems in order to demonstrate complex behaviours that are beyond the capabilities of the current state-of-the-art visual guidance systems. In Part I, two related behaviours were investigated – collision avoidance and terrain following – both of which are crucial for small-scale UAVs operating in near-Earth environments. For most metrics it was found that the proposed vision system surpassed the performance of other similar approaches. Additionally, the wide-FOV of the stereo system permits full control of the attitude and altitude of the aircraft, which many other leading approaches have not been able to demonstrate using vision only. Despite the success of the proposed approach, several limitations have been identified that restrict its utility for guidance of UAVs in other flight regimes. In Chapter 5 a more general purpose vision system is described that addresses these limitations and serves as the platform for the investigations in Part II.

Part II

Omnidirectional Vision-based Guidance

5

Introduction to Omnidirectional Vision-based Guidance

5.1 Summary

The first part of this thesis described a stereo vision system designed specifically for providing guidance to a small-scale aircraft flying at low altitude. That vision system performed well for its intended purpose, but several limitations were identified in Chapters 3 & 4 that restrict its usefulness for more general UAV guidance.

A more general-purpose vision system, inspired partly by the honeybee, is now introduced. This serves as the platform for the research presented in Part II. The improved vision system is designed to expand on the capabilities of the stereo vision system discussed in Part I, primarily by simplifying the problem of computing egomotion and attitude robustly, whilst retaining the low-altitude capabilities of the stereo system. This new vision system possesses an *omnidirectional* FOV. Motivations for using omnidirectional vision are discussed in Section 5.2 and the design and implementation of the new vision system are presented in Section 5.3. An overview of the aims of Part II is given in Section 5.4 along with an outline of the remainder of the second part of the thesis.

5.2 Motivations for omnidirectional vision-based guidance

Omnidirectional vision is important for estimating attitude and egomotion robustly. FOV is therefore an important consideration in the design of visual guidance systems for UAVs. Not coincidentally, many flying insects possess omnidirectional visual

fields. Biological vision systems can therefore influence the design of visual guidance systems and sensing algorithms.

5.2.1 Advantages and limitations of stereo vision

The visual guidance system described in Part I was based on stereo vision to overcome some of the difficulties faced by the optic flow-based visual guidance systems described in Chapter 2. Specifically:

- Stereo vision-based systems avoid the need to separate rotational and translational components of measured optic flow to extract estimates of range.
- The translational component of optic flow encodes range information but is also proportional to the speed of the observer. Stereo vision-based systems do not require measurement of the aircraft's ground speed to compute range accurately.
- Computing range via stereo is more computationally efficient than computing range from optic flow, because (even disregarding the need to isolate the translational component of optic flow) optic flow fundamentally requires a 2D search for corresponding image regions, whilst stereo requires a 1D search only. In practice, this means it is possible to compute stereo range maps more densely than their monocular equivalent, which facilitates detection of smaller obstacles.

These factors all motivated use of stereo vision for a system designed to provide real-time guidance to an aircraft flying in a near-Earth environment. However, stereo vision suffers from its own inherent limitations, identified in Chapters 3 & 4. Namely:

- The two cameras must be calibrated precisely, both individually and as a stereo pair, to compute dense and accurate range maps.
- The spectrum of ranges that the stereo vision system is able to measure precisely depends on the configuration of the system, primarily the baseline between the two cameras. In practice, this means that the stereo vision system is only able to compute accurate range to objects that fall within the finite spectrum of ranges that the system was designed for, and very quickly degenerates to a monocular system outside this spectrum.
- Systems based on stereo vision only do not directly measure motion of the aircraft. A system that allows aircraft attitude and altitude to be estimated

with respect to the ground plane was described in Part I. Estimates of rotational and translational motion of the aircraft, however, must be either computed numerically; measured from image correspondences between successive frames using either optic flow or feature-based methods; or by using additional sensors.

To overcome some of the individual limitations of stereo vision and optic flow, the guidance system presented in Part I could be augmented such that optic flow is computed between successive frames from the fore camera, whilst stereo range is computed between instantaneous stereo image pairs from both cameras, as described in Section 4.5. The advantage of this approach would be that full 6-DOF motion of the aircraft could be obtained when the aircraft was close to the ground, such as during critical take-off and landing phases of flight. This configuration would still suffer, however, from an inability to compute stereo range when the aircraft was not close to the ground and from the difficulty of separating rotational and translational components of measured optic flow.

5.2.2 Omnidirectional vision

It has been shown by others previously that a very wide FOV is neccessary for separating rotational and translational components of measured optic flow robustly [Adiv, 1989, Fermuller and Aloimonos, 1998, Gluckman and Nayar, 1998, Koenderink and Doorn, 1987]. This is because the apparent motion field, or optic flow field, contains global patterns that do not always manifest themselves in a restricted FOV ($\ll 180°$). In particular, the focus of expansion (FOE) and focus of contraction (FOC) are likely to exist outside the FOV for traditional rectilinear vision systems, thereby increasing the sensitivity of egomotion estimates to noise that is inherent in measurements of optic flow. In contrast, for wide-angle vision systems with hemispherical or greater FOVs (*e.g.* omnidirectional vision systems), either the FOE or FOC or both will always be visible. Translational and rotational motions can produce optic flow fields that appear similar in some regions of the view sphere (*e.g.* upward translation parallel to the vertical axis and clockwise rotation about the forward horizontal axis both produce generally downward optic flow to the right of the observer, as shown in Figure 5.1), and are therefore difficult to disambiguate with a limited FOV in the presence of noise. For large FOVs ($> 180°$), however, the motion fields are distinct even in the presence of significant amounts of noise [Gluckman and Nayar, 1998].

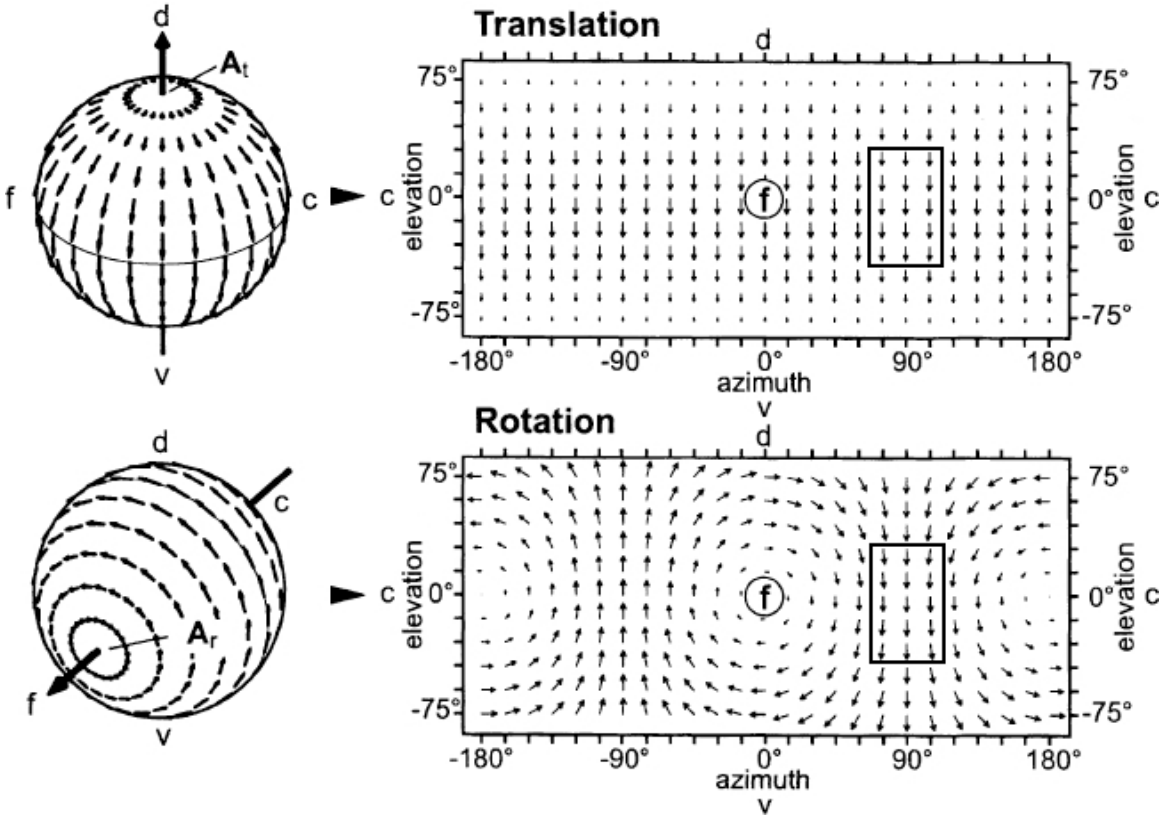


Figure 5.1: Comparison of global motion fields produced by translation along the vertical axis (A_t, top) and rotation around a horizontal axis (A_r, bottom). The two motions are represented by both an external view of a visual hemisphere (left) and by a Mercator map of the entire visual space (right). Frontal (f), caudal (c), dorsal (d), and ventral (v) positions are marked in both representations. Globally, the translational and rotational flow fields can be distinguished easily, but they may appear locally similar (*e.g.* marked frame) and can be difficult to disambiguate with a small FOV in the presence of noise. Modified from Krapp and Hengstenberg [1996].

5.2.3 Omnidirectional vision in insects

Many flying insects depend on motion cues such as optic flow to perform various complex flight manoeuvres, as discussed in Chapter 2. To judge range to obstacles and to compute their own flight speed and distance flown, insects must first resolve their observed visual self-motion into translational and rotational components. Perhaps not coincidentally then, the compound eyes of many flying insects endow the animal with omnidirectional motion sensitivity (*e.g.* the honeybee's visual field, illustrated in Figure 5.2). Krapp and Hengstenberg [1996], Krapp et al. [1998] have shown, in fact, that specialised visual neurons in the blowfly possess huge receptive fields, which

span a large fraction of the visual field. They found that the motion response fields for those neurons were remarkably similar to optic flow fields produced by pure rotational motion around various axes, suggesting that flying insects possess the ability to specifically extract the rotational component of optic flow around a particular axis. Once rotation has been subtracted or nullified, the residual translational component of optic flow can be used directly to estimate range to obstacles in the environment according to Equation 2.1.

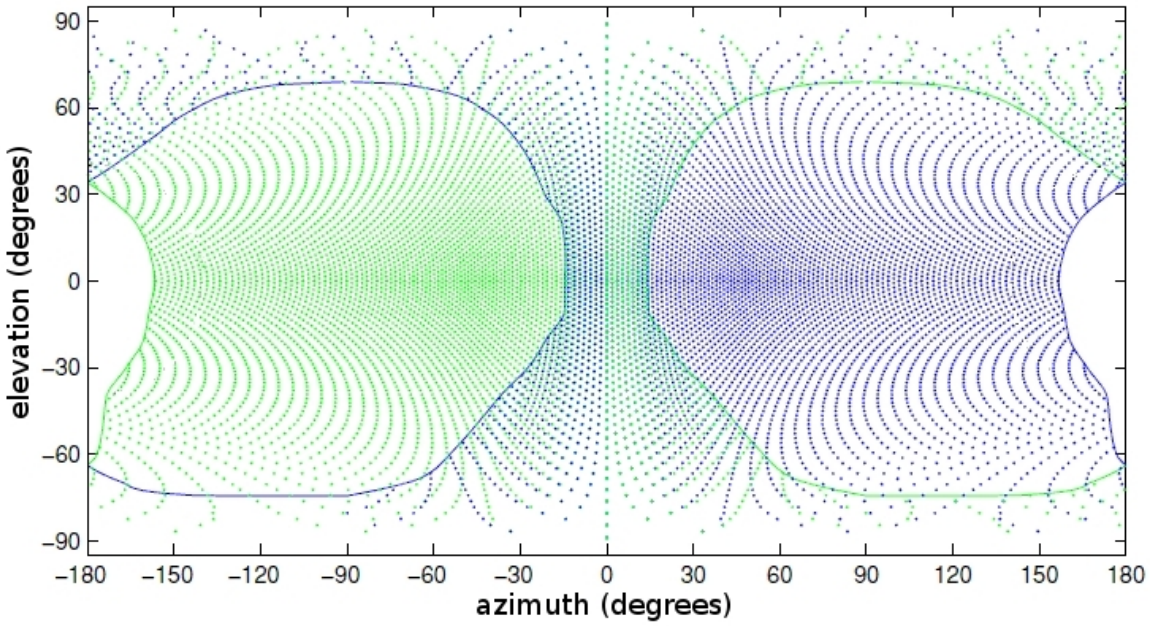


Figure 5.2: Viewing directions of the ommatidia (points) and extents of the visual fields (solid lines) of both eyes in the honeybee. Except for a small region that is obscured by the body of the animal, the honeybee's visual field is omnidirectional. Note that there is significant overlap between the two visual fields, affording the animal stereoscopic vision in a strip approximately $40° \times 240°$ that spans the ventral, frontal, and dorsal regions of the view sphere [Seidl and Kaiser, 1981]. The limited separation between the eyes in most insects, however, necessarily restricts the range over which their stereopsis would be effective. Modified from Sturzl et al. [2010].

In addition to their compound eyes, many insects also possess dorsal *ocelli*, with which they perceive wide-field spectral, intensity, and polarisation gradients [Ribi et al., 2011, Schuppe and Hengstenberg, 1993]. Due to the low spatial resolution and very high FOV of the *ocelli*, it is thought that their primary functions are maintenance of flight stability and navigation. Behavioural studies have shown that stimulating insects' *ocelli* with appropriate visual cues can, in fact, effect compensatory head movements and flight steering actions in locusts [Taylor, 1981], dragonflies [Stange

and Howard, 1979], bumblebees [Wellington, 1974], flies [Wellington, 1953], and ants [Fent and Wehner, 1985, Mote and Wehner, 1980, Schwarz et al., 2011]. It is likely, therefore, that many insects use wide-field visual cues (*i.e.* the position of the horizon) to monitor and stabilise their attitude. An algorithm for automatically stabilising the attitude of a UAV using the visual horizon has been developed and is described in Chapter 6. Stabilisation of attitude is very important for insects, not only for flight stability, but also for head stability – head rotations induce optic flow patterns that contaminate estimates of ranges to objects in the environment.

5.2.4 Omnidirectional vision-based guidance systems

Estimating attitude and egomotion (*i.e.* translational and rotational components of self-motion) robustly is of particular relevance to visual guidance of autonomous robots. The importance of omnidirectional vision for this task is widely understood (see Benosman and Kang [2001], Yasushi [1999] for overviews of the designs and applications of omnidirectional vision systems), but in practice it is difficult to achieve a truly spherical FOV. Typically, combinations of reflective and refractive elements (*catadioptric* systems) are used to achieve a panoramic view of the environment (*e.g.* Chahl and Srinivasan [1997], Nayar [1997]), or views from multiple cameras are stitched together to achieve an arbitrarily wide FOV (*e.g.* PointGrey Ladybug[1]).

Catadioptric systems have been designed previously for the guidance of ground-based robots [Chang and Hebert, 1998, Scaramuzza and Siegwart, 2008, Winters et al., 2000] and also UAVs [Demonceaux et al., 2006, 2007, Hrabar and Sukhatme, 2003, Mondragón et al., 2010a, Soccol et al., 2007, Srinivasan et al., 2006]. The size of such systems and the number of their optical elements means, however, that they are not ideally suited for use on board small-scale UAVs, unlike the miniature sensor described by Tisse et al. [2007a]. Multi-camera omnidirectional vision systems have also been used to estimate egomotion of both ground-based robots [Schill et al., 2008, Tardif et al., 2008] and aerial robots [Hrabar and Sukhatme, 2009]. One advantage of such systems is that the multiple sensors can be oriented with overlapping FOVs to achieve wide-angle stereo vision [Moore et al., 2009, 2010, 2011b, Spacek and Burbridge, 2007, Thurrowgood et al., 2007, Tisse et al., 2007b].

Although omnidirectional vision may contribute significantly to the ability of many flying insects to compute their egomotion efficiently and robustly, few of the

[1] http://www.ptgrey.com/products/ladybug3/ladybug3_360_video_camera.asp

designs discussed in this section show much similarity to their biological counterparts. Exceptions to this observation are electronic *ocelli* [Barrows et al., 2003, Thakoor et al., 2003], which were designed to provide attitude stabilisation for small-scale UAVs flying outdoors, and the miniature *catadioptric* sensor developed by Sturzl et al. [2010], which possesses a visual field matching that of many flying insects.

A new omnidirectional vision system possessing monocular and stereo visual fields comparable to that of the honeybee is described in Section 5.3. The system has a simple design comprising few optical elements, yet the near-spherical monocular FOV and region of stereo overlap it provides enables robust and efficient computation of 3-DOF egomotion and 3-DOF attitude. This vision system therefore provides an ideal research platform with which to investigate the capabilities of visual guidance for small-scale UAVs.

5.3 A new omnidirectional vision system for UAV guidance

A novel omnidirectional vision system was designed for the purposes of computing visual egomotion and odometry robustly. The omnidirectional FOV of this vision system simplifies the problem of estimating 3-DOF egomotion from the measured optic flow field; it also allows 3-DOF attitude to be computed robustly while permitting a region where stereo range can be computed. This vision system was designed and developed primarily by Mr. Saul Thurrowgood (with contributions by the author). The investigations presented in Part II are conducted using this platform, however, and as no comprehensive description is published elsewhere, its design and characteristics are given here.

The omnidirectional vision system, or *iEye* (meaning "two eyes"), comprises two colour firewire cameras (PointGrey Firefly MV[2]), which are equipped with miniature fisheye lenses (Sunex DSL216[3]). Each camera-lens assembly has a FOV of approximately 187°. The two assemblies are positioned back-to-back such that the complete vision system has a near-spherical monocular FOV. Additionally, each camera is oriented with a 10° camber relative to the baseline between the two assemblies to produce a wedge-shaped region of the view sphere imaged by both cameras. The baseline between the two camera-lens assemblies is ≈ 9cm. This design thus enables measurement of forward-looking stereo disparity in the region where

[2]http://www.ptgrey.com/products/fireflymv/fireflymv_usb_firewire_cmos_camera.asp
[3]http://www.optics-online.com/OOL/DSL/DSL216.PDF

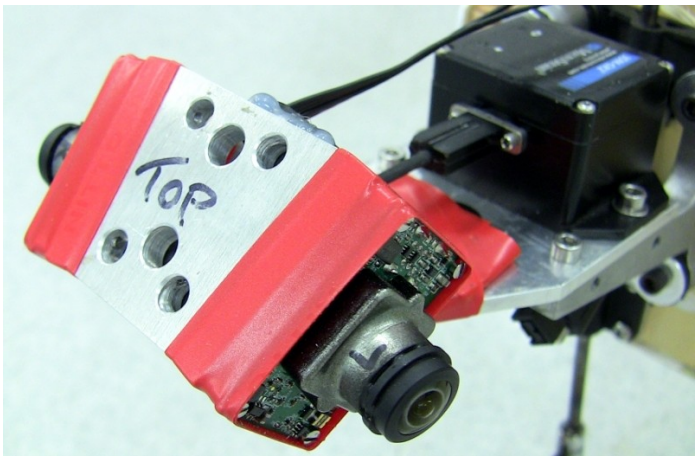


Figure 5.3: The *iEye* vision system, shown here attached on the nose of the test aircraft. The two cameras (details in text) are positioned back-to-back and angled to give the vision system a near-spherical monocular FOV and a strip of stereo overlap in front of and beneath the aircraft. Also visible is a MicroStrain 3DM-GX2 IMU, used for comparative measurements of rotational rate and orientation.

looming optic flow signals are weakest for a forward-moving observer. The vision system is mounted on the aircraft facing 45° downwards such that, on average, the ground directly in front of and beneath the aircraft is visible by both cameras. The *iEye* vision system is displayed in Figure 5.3 and the visual field of the system is shown in Figure 5.4.

Each camera-lens assembly is calibrated individually using a generic, 6-parameter calibration procedure developed by Mr. Saul Thurrowgood, although any procedure where image pixels are mapped smoothly to unit vectors on the view sphere may be used. The two camera models are rotationally aligned with respect to each other by hand. Raw 608px × 480px images are captured at 25Hz from each camera and typically stitched and mapped to a 360px × 220px equirectangular or equal-area cylindrical projection (*e.g.* Figure 5.4) for higher-level processing. The image pairs are synchronised in hardware by triggering the start of each camera's exposure with an external signal. Currently, a dedicated microcontroller is used for this purpose. Exposure is manually controlled to maintain 80% of intensity values at $\leq$ 50% of the dynamic range and a nonlinear histogram adjustment (similar to Larson et al. [1997]) is performed on the 10-bit raw pixel data prior to compression to maximise dynamic range of the 8-bit red-green-blue (RGB) pixel data. The resulting imagery is highly invariant to external brightness changes and internal shutter variation, whilst providing scene texture in direct sunlight and in shadow simultaneously.

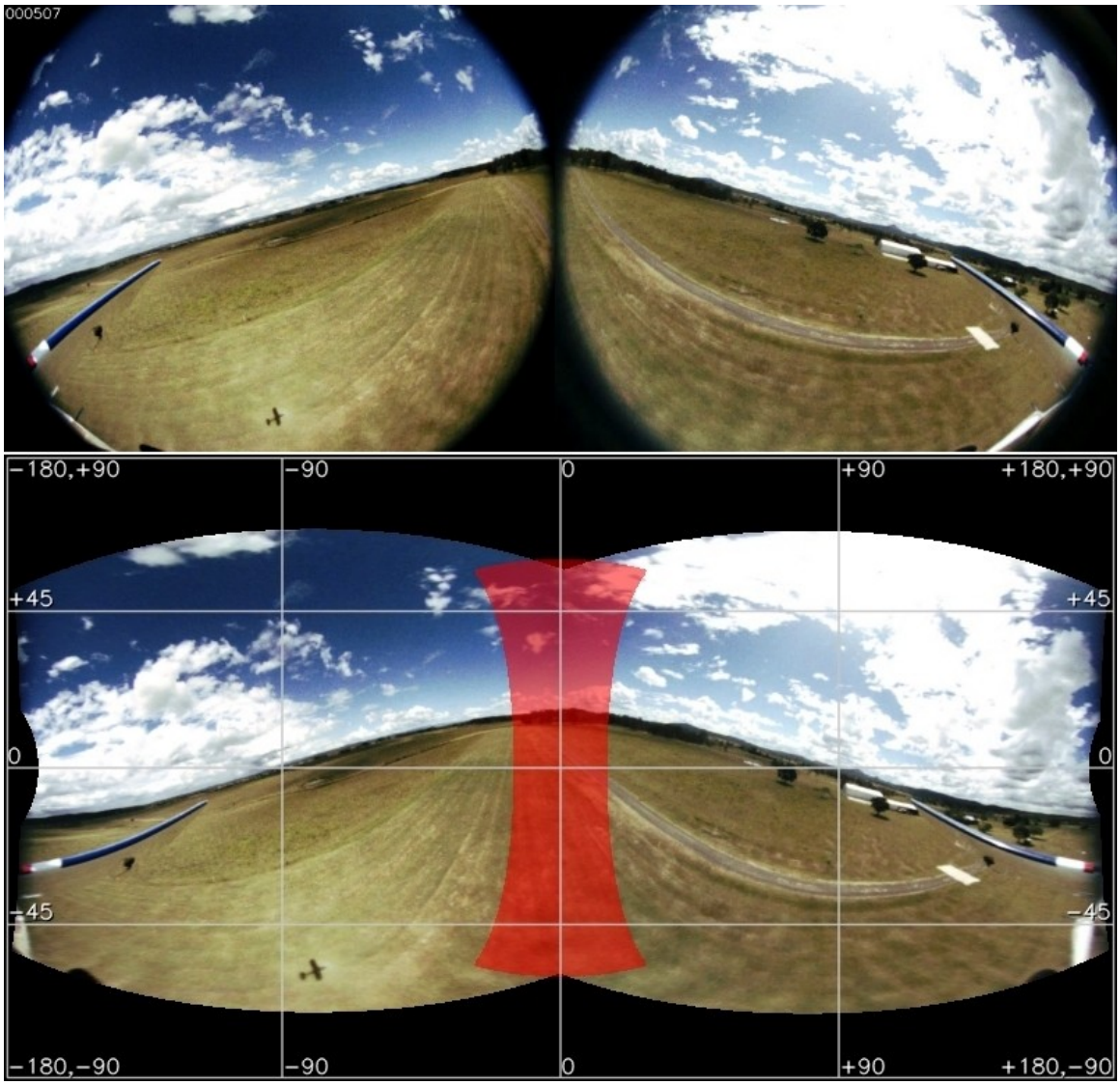


Figure 5.4: Raw fisheye images captured by the *iEye* vision system from the left and right cameras (top), and an equirectangular projection of the stitched omnidirectional image (bottom) showing lines of azimuth and elevation (grey) and extents of the area remapped to compute stereo disparities (red). The equirectangular projection covers the complete view sphere and pixel columns and rows are proportional to lines of constant azimuth and elevation respectively. Area on the view sphere appears distorted under this projection, however, and the proportion of the total view sphere imaged by the *iEye* vision system is actually $\approx 88\%$. Regions that are not imaged correspond to either segments of the image circles clipped by the imaging sensors, or a small wedge of the view sphere behind the vision system (azimuth $\approx \pm 180°$), which is not imaged due to the relative orientation of the two cameras. The cameras have been oriented to produce a strip of stereo overlap (approximately $30° \times 120°$) in front of and beneath the aircraft. The FOV provided by each fisheye lens is approximately 187°.

5.4 Outline of Part II

This thesis investigates wide-angle vision systems as a means for providing guidance to small-scale UAVs. It is proposed that a wide FOV enables simple and bioinspired sensing and guidance algorithms, which can be combined to enable complex and autonomous behaviours. A new omnidirectional vision system, the *iEye*, was described in this chapter. A novel method for obtaining 3-DOF attitude robustly and efficiently using the *iEye* vision system is introduced and described in Chapter 6. This system then forms the basis for investigations of more complex behaviours in Part II.

Various approaches for detecting and tracking an independently moving target using vision are reviewed in Chapter 7 and a computationally lightweight appearance-based tracker is presented. A bioinspired strategy that facilitates the time-optimal interception of a moving target is described in Chapter 8. A novel interception control algorithm, based on this strategy, is also presented. Performance of the interception guidance scheme is assessed and confirmed through a series of simulated and real-world flight tests. Finally, by combining the algorithms presented in Part II, the capability for a small-scale UAV to intercept a moving target using only visual input from the *iEye* vision system is demonstrated.

6

Visual Estimation of 3-DOF Attitude

6.1 Summary

An omnidirectional vision system, the *iEye*, designed specifically to simplify computation of 3-DOF egomotion and 3-DOF attitude from visual input was described in Chapter 5. A novel method for automatically obtaining an aircraft's 3-DOF attitude using this vision system is now described in this chapter. An overview of the approach is given in Section 6.2.

As input, the system takes the near-spherical view of the environment captured by the *iEye* vision system and classifies the image into fuzzy sky and ground regions, based on the spectral and intensity properties of the pixels. A novel approach to obtaining 2-DOF attitude from the classified image is used to generate a stabilised panoramic image of the horizon and sky. This 'visual compass' is then used to detemine heading direction. This approach is based partly on the *ocellar* system of flying insects, which senses the panoramic profile of the horizon to estimate and stabilise attitude. The full details of the approach and implementation are given in Section 6.3. Results from open-loop and closed-loop flight tests demonstrate the abilty of this system to outperform an inexpensive IMU and to provide real-time guidance for a small-scale UAV (Section 6.4). Through comparisons with alternative vision-based approaches in Section 6.5, it is shown that this system provides estimates for the 3-DOF attitude of an aircraft that are more accurate, more robust, and less computationally expensive than present state-of-the-art approaches. Results from additional investigations demonstrate application of this visual guidance scheme to real-world tasks such as landing automatically in an uncontrolled environment (Section 6.6).

Portions of this chapter are contained within Moore et al. [2011a,c].

6.2 Visual attitude

For a UAV, accurate estimation of 3-DOF attitude (roll angle, pitch angle, and heading direction) is crucial for applications such as mapping, landmark localisation, or augmented reality, where even small attitude errors can lead to failures in feature matching or misalignments during reprojection. Traditionally, UAV attitude is sensed by integrating the outputs from rate gyroscopes, but this method is susceptible to noise-induced drift [Rohac, 2005]. This shortcoming can be mitigated by augmenting the system with a direct measure of absolute orientation. Triaxial accelerometers or magnetometers, for example, may be used to sense the direction of gravity or the local magnetic field respectively. Accelerometers also measure the linear and centripetal accelerations of the aircraft, however, which confound the process of estimating the gravity vector when the aircraft is manoeuvring [Rohac, 2005]. Furthermore, magnetometers will not sense rotations about an axis parallel to the direction of the local magnetic field [Merhav, 1996]; the local magnetic field is also subject to environmental disturbances. The limitations of these approaches mean that they are not ideal for applications where aircraft attitude must be known precisely.

A vision-based method for sensing the absolute 3-DOF orientation of an aircraft is described in Section 6.3. This approach involves capturing a very wide-angle view of the environment, including the horizon. An adaptive classifier is used to segment the scene into sky and ground regions, using the spectral and intensity properties of the input pixels. The position and orientation of the horizon are then obtained and used to infer the aircraft's roll angle and pitch angle. A stabilised panoramic image of the horizon profile and sky is then generated using estimated 2-DOF attitude. The stabilised horizon image is independent of the roll and pitch of the aircraft, but the image undergoes a horizontal shift as the aircraft's heading direction changes. This so-called 'visual compass' can therefore be used to determine the heading direction of the aircraft, relative to any prior point in time.

Methods for obtaining the 3-DOF attitude of an aircraft using vision have been described previously (*e.g.* [Mondragón et al., 2010a]). The novelty of the proposed approach, however, is that it continuously adapts to changing environmental conditions and therefore does not require training off-line prior to use. Additionally, it is shown in Section 6.5 that this method is both robust and relatively computationally

efficient to implement, compared with previously described vision-based approaches.

6.3 Visually estimating 3-DOF attitude

In the proposed approach, full 3-DOF attitude is estimated in two steps. Firstly, the distributions of the sky and ground regions in the input image are used to compute the position and orientation of the horizon, from which the roll angle and pitch angle of the aircraft are inferred (Section 6.3.1). Then, a stabilised 'visual compass' is produced using the estimated 2-DOF attitude, from which the heading direction of the aircraft is obtained (Section 6.3.2).

6.3.1 Estimating roll and pitch

Automatic horizon detection schemes typically (either explicitly or implicitly) attempt to find an n-dimensional contour in colour, intensity, and/or texture space that separates the input scene into two distinct regions – one corresponding to the ground, and the other to the sky. The visual horizon corresponds to the projection of the n-dimensional contour into image space. Attitude of the aircraft can be inferred from the orientation and position of the horizon contour. Separation of the input pixels into two classes is, therefore, arguably the most crucial step in automatic visual horizon estimation and this problem has been tackled previously in various ways.

Cornall and Egan [2004], Cornall et al. [2006] use a static transformation of the RGB colour space to enhance the contrast between the sky and ground regions before applying a threshold to distinguish the two. Mondragón et al. [2010a], Thurrowgood et al. [2009, 2010] also apply a static transform to the RGB or luminance-chrominance (YUV) colour space but additionally define adaptive thresholding techniques that allow some limited variability between the appearance of the sky and ground and the static representations. These approaches can be computationally efficient and can also inherently handle cases where the sky and ground are inverted (*i.e.* when the aircraft is upside down [Thurrowgood et al., 2010]), or not present in the input image. The representations of sky and ground are largely predetermined and fixed, however, so these approaches can fail when the appearance of the sky or ground in the input image does not closely match the global average representation of sky or ground for all scenes.

Using an alternative approach, Dusha et al. [2007] do not explicitly label sky

and ground regions but instead directly search for the horizon contour in RGB colour space by employing a Hough transform (as in Ballard [1981], O'Gorman and Clowes [1976]) on an edge image. Similarly, Ettinger et al. [2002a,b] search exhaustively for the parameterised horizon contour that best separates the input image into two classes by minimising the *intra-class* RGB covariance of the input pixels. Todorovic and Nechyba [2004], Todorovic et al. [2003] perform an equivalent search, but instead maximise the *inter-class* RGB Mahalanobis distance for a number of atomic sub-images, while McGee et al. [2005] use a support vector machine (SVM), and Demonceaux et al. [2006] use Markov random fields (MRFs) to achieve the same result. While these approaches are robust to dramatic variations in the appearance of the sky and/or the ground, they can be computationally intensive to execute and they do not explicitly label the sky and ground classes. Thus, Ettinger et al. [2002b] also maintain a representation of the sky and ground classes to distinguish the two, and additionally detect cases where the horizon detection failed, while Todorovic and Nechyba [2004], Todorovic et al. [2003] use standalone classifiers, trained offline, to label sky and ground regions.

An *adaptive* approach to the classification of sky and ground regions in the input image is presented here. The classifier maintains a record of the spectral and intensity properties of the sky and ground regions, which is updated online continuously. This does not require a fixed, predetermined representation of the sky and ground classes and hence does not require the classifier to be trained offline prior to use. Additionally, the implementation is computationally efficient and hence fast to execute.

The proposed horizon detection scheme involves four stages (illustrated in Figure 6.1), which are performed online:

1. *Classification* – Spectral and intensity properties (i.e. the YUV values) are used by the classifier to assign a weight to each pixel in the colour input image. The weight represents likelihood that the corresponding input pixel belongs to either the *sky* or *ground* class, or a combination of the two. Together, the weights form the classified image.

2. *Matching* – The classified image is reduced to two 1D arrays (referred to as a kernel), which hold the row and column averages of the classified image. The classified image kernel is then matched against a database of reference kernels, which represent every possible combination of roll and pitch to find the current 2-DOF attitude.

3. *Masking* – A sky/ground mask is generated using the current estimate of roll and pitch. The mask is used to compute the desired spectral and intensity weights within the classifier.

4. *Training* – The weights in the classifier are updated using the sky/ground mask and an online reinforcement strategy.

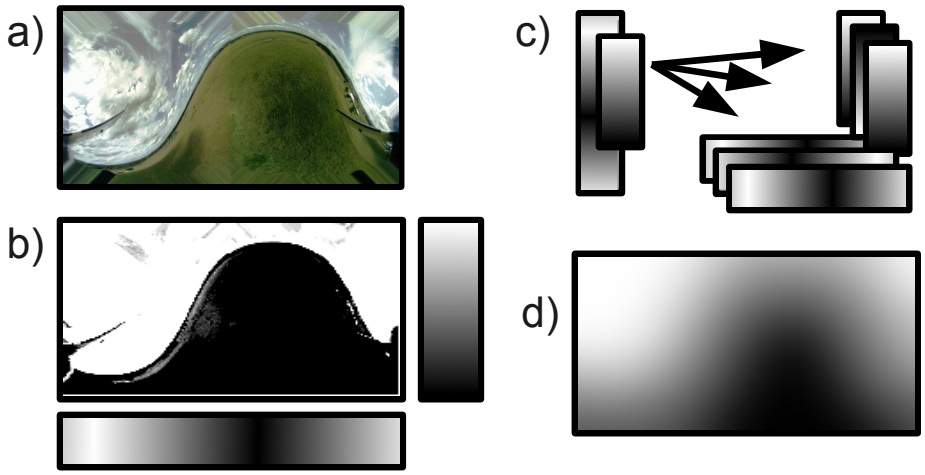


Figure 6.1: Illustration of the process for estimating an aircraft's 2-DOF attitude from the visual horizon. Firstly, (a) the input image is processed to give (b) the classified image. The classified image is then reduced to two 1D arrays (a kernel) which contain the row and column averages. The input arrays are matched to a pre-computed database of attitude kernels (c) to retrieve the current roll and pitch angles. Finally, a sky/ground mask (d) is generated, using estimated 2-DOF attitude, and used to update the weights in the classifier.

Each of these steps is performed in turn on each input image, so that, over time, the classifier builds up a representation of the ground and sky classes. The classifier is trained continuously, so it is able to adapt to the changing appearance of the ground and sky regions within the input images. Each stage is described in greater detail as follows:

Classifying sky and ground

It has been shown that using spectral information alongside image brightness allows a better segmentation between sky and ground than can be achieved using either alone [Thurrowgood et al., 2009]. Hence, here the Y, U, and V channels are used to preserve all visual information for image classification. The colour space transformation from

RGB to YUV used here is given by:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.3 & 0.586 & 0.114 \\ -0.168 & -0.332 & 0.5 \\ 0.5 & -0.418 & -0.082 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 127.5 \\ 127.5 \end{bmatrix}$$

Following the transformation to YUV colour space, the colour planes (U & V) are expanded by a factor of 3.0 and clamped to $[0, 255]$. This facilitates a better separation of the spectral and intensity properties of the input pixels. It is undesirable to saturate either of the U or V channels, however, as this will alter the hue of the pixel. The scale factor (3.0) was chosen, therefore, such that $|4 \cdot \sigma_{u,v}| < 128$ (*i.e.* 99.99% of the U & V values lie within the range $[0, 255]$), where $\sigma_{u,v}$ represents the standard deviation of either the U or V channel, for the typical range of values sampled with the vision systems investigated in this study.

The role of the classification stage is to label the sky and ground regions in the input image. The sky and ground classes here are fuzzy subsets (first described by Zadeh [1965]) of the YUV colour space. Each pixel in the input image, therefore, may be labelled as definitely *sky* or *ground*, or some uncertain combination of the two. To represent the range of possible classifications, each input pixel, i, is assigned a weight, $w_i(y, u, v)$, which is continuous and defined over the range $[-1, +1]$. The weight is a function of Y, U, and V, and is a measure of the likelihood that a particular combination of spectral and intensity properties represents a sky region ($w = +1$), or a ground region ($w = -1$). The relationship between the weight function and the membership functions for the sky and ground classes is illustrated in Figure 6.2.

The weights, $w_i(y, u, v)$, assigned to each input pixel, i, during classification are derived directly from a learned weight for each Y, U, and V combination,

$$w_i(y, u, v) = \begin{cases} +1 & \text{if } |w_{y,u,v}| \geq +1, \\ -1 & \text{if } |w_{y,u,v}| \leq -1, \\ w_{y,u,v} & \text{otherwise,} \end{cases} \tag{6.1}$$

where $w_{y,u,v}$ is defined over the range $[-a, +a]$, $a \geq 1$ (the reasoning for this is explained in the subsection *Training the classifier*).

In practice, the classifier is simply a three dimensional structure in YUV space, where each element, $w_{y,u,v}$, within the structure stores a classification weight for a particular YUV combination. To reduce the amount of data storage required, the

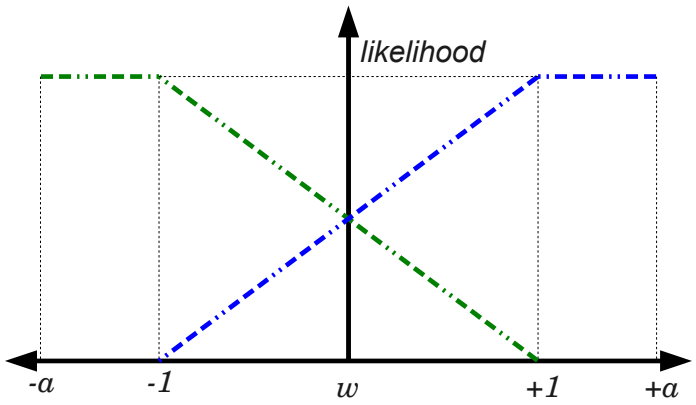


Figure 6.2: Relationship between the weight function and the membership functions for the sky (blue) and ground (green) classes. YUV combinations with a weight $w \geq +1$ are maximally likely to belong to the sky class, or ground class for $w \leq -1$. YUV combinations with a weight $|w| < 1$ are, to some degree, likely to belong to both the sky and ground classes.

dimensions of the classifier are reduced from 255^3 to 16^3. Each element within the classifier thus holds one classification weight for a cluster of similar YUV combinations. It was found through experimentation that these particular dimensions offer a suitable trade-off between achievable spectral and intensity resolution and ability to generalise classifications to neighbouring YUV combinations.

To classify an input image at run-time, the Y, U, and V values for each input pixel are used simply to address an element within the classifier and retrieve the appropriate weight according to Equation 6.1. The weights within the classifier are initially $w_{y,u,v} = 0$, for all Y, U, and V. As training examples are presented to the classifier, however, it quickly learns to separate the sky and ground classes and ideally $|w_{y,u,v}| \geq 1$ for all Y, U, and V after a few seconds.

Some example input images and the corresponding classified images are displayed in Figure 6.3.

Matching against stored attitude kernels

The classified image contains values in the range $[-1, +1]$, which correspond to the weights assigned to each input pixel during classification (see Figure 6.3). Ideally, $|w_i(y, u, v)| = 1$ for all pixels, i. To ascertain the current roll angle and pitch angle of the aircraft, the classified input image is matched exhaustively against a database of ideal image classifications representing all possible 2-DOF attitude combinations.

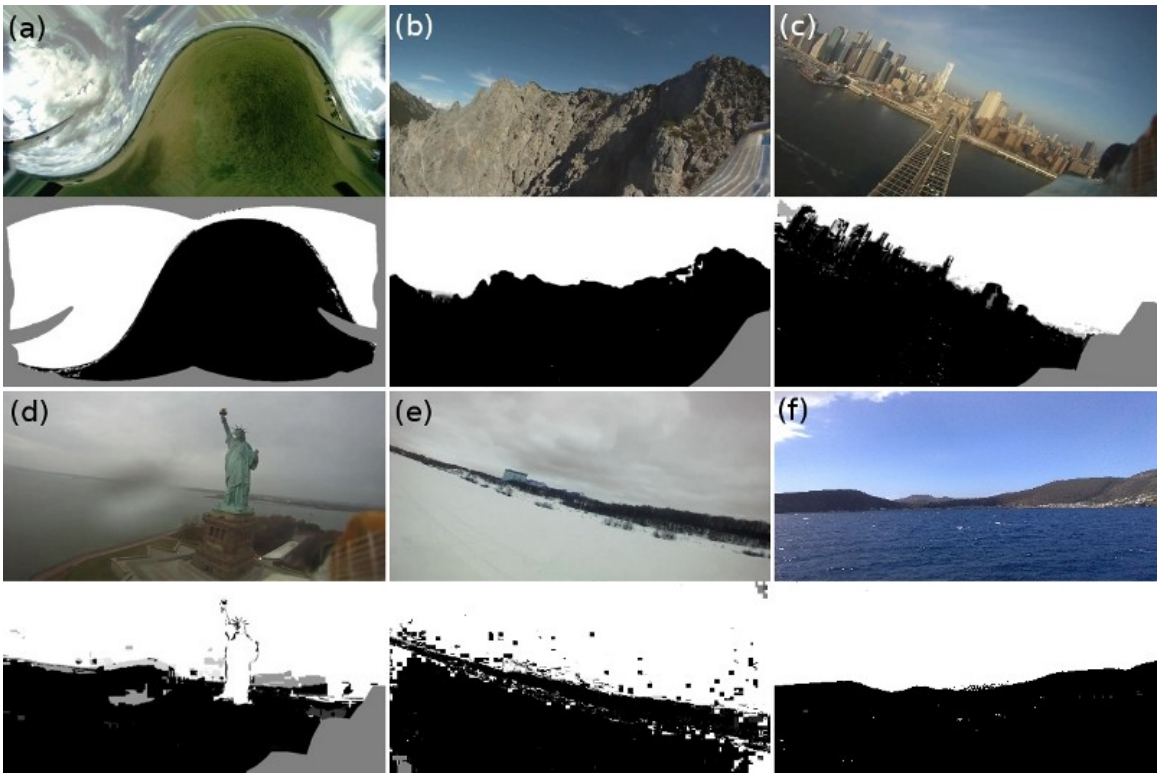


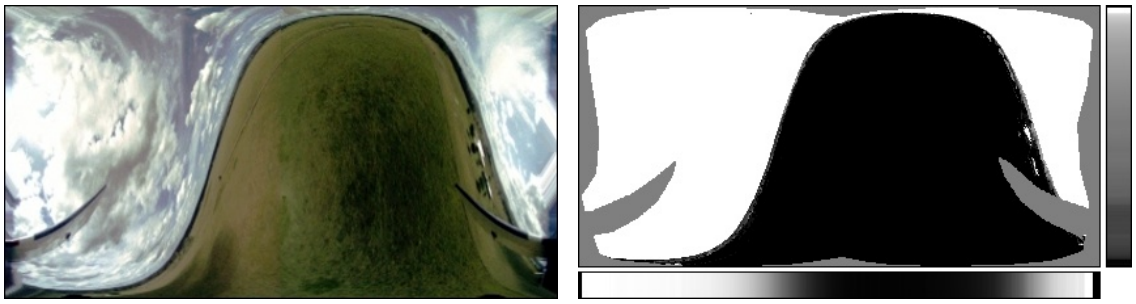
Figure 6.3: Example input images (a–f) and corresponding classifications. A range of environmental conditions, FOVs, image resolutions, and compression ratios are represented by the input images. To produce each classified image, the corresponding input image was presented repeatedly until the weights within the classifier converged to stable values. Contiguous grey regions adjoining the borders of classified images represent regions masked from the training phase. Note that high-resolution classified images are shown here to aid illustration, in practice a resolution of 80px × 40px is used. Image credits: (b–d) http://www.youtube.com/nastycop420; (f) http://commons.wikimedia.org/wiki/User:Dedda71.

The database is pre-computed offline by generating the ideal classification images for an observer above an infinite ground plane. Under this approximation the ideal attitude kernels are independent of the observer's altitude above the ground plane. In practice, buildings, mountains, trees, or other structures may protrude through the horizon at low altitudes and affect the accuracy of the matching process. However, the very-wide FOV of the vision system and low resolution of the classified images means that this approach is very robust to non-planar structure.

To reduce the complexity of the matching process, and to reduce the amount of data that is required to store the database of ideal image classifications, each classified image is reduced to two 1D arrays (a kernel), which hold the row and column averages of the classified image. By storing and matching just the row and

column average kernels, the attitude can be determined uniquely, whilst the number of required difference operations (and stored memory elements) is reduced from $n \times (r \times c)$ to $n \times (r + c)$, where n represents the number of candidate attitudes (*e.g.* for this study $n = [\frac{360^\circ}{5^\circ} + 1] \times [\frac{180^\circ}{5^\circ} + 1] = 2701$), and r and c represent the height and width of the input images respectively (in this study $r = 40$px, $c = 80$px).

The classification and kernel matching process is illustrated in Figure 6.4. An error score for each candidate attitude is computed from the SAD between the classified input kernel and each of the stored reference kernels. In practice, the reference kernels are computed at some finite number of attitudes (in this study an angular resolution of 5° is used), and hence the true current attitude is estimated by applying an equiangular fit [Shimizu and Okutomi, 2003] to the minimum and neighbouring SAD scores.



(a) Input image.

(b) Classified image and row/column kernel.

(c) Instantaneous row/column kernel and an illustration of the stored database of attitude kernels.

Figure 6.4: Classification and 2-DOF attitude matching process, showing (a) the raw input image; (b) the classified input image with the 1D column and row averages (kernel) shown below and to the right respectively; and (c) the same kernel from (b) alongside the pre-computed database of all possible 2-DOF attitude kernels. An angular resolution of 8° was used to compute the kernel database, so there are $n = [\frac{360^\circ}{8^\circ} + 1] \times [\frac{180^\circ}{8^\circ} + 1] = 1058$ unique kernels in the database. Each stored kernel is represented here as an individual column of pixels in (c). The database is stored in 'pitch-major' order, so scanning through the database from left to right corresponds to $\frac{180^\circ}{8^\circ} + 1 = 23$ increments in pitch angle within the range $[-90^\circ, 90^\circ]$, and there are $\frac{360^\circ}{8^\circ} + 1 = 46$ increments in roll angle within the range $[-180^\circ, 180^\circ]$ for each pitch angle. The best matching kernel for this input image corresponds to a roll angle of 61° and a pitch angle of −41°; the position of this attitude within the database is marked (blue ticks).

Generating the sky/ground mask

The classifier weights, $w_{y,u,v}$, are updated online using a reinforcement strategy similar to the well-known delta rule [Rumelhart, 1985],

$$\Delta w_{y,u,v} = \alpha \cdot \delta_{y,u,v}, \tag{6.2}$$

where, $\Delta w_{y,u,v}$ is the change applied to a weight within the classifier, α is the global learning rate (here $\alpha = 0.05$, which corresponds to $\tau_d \sim 0.8\text{s} \,@\, 25\text{Hz}$), and $\delta_{y,u,v}$ is the difference between the desired output and the current output for each element. Here it is desired only to maximise the difference between the classifier weights for the sky and ground regions. Hence, $\delta_{y,u,v}$ may be chosen arbitrarily, such that it acts as a velocity term that drives the corresponding weight towards the desired class ($w_{y,u,v} \to +a$ for sky regions and $w_{y,u,v} \to -a$ for ground regions).

To identify the desired class (*i.e sky* or *ground*) for each YUV combination, the desired class for each pixel, i, in the input image is first identified:

$$m_i = \hat{\mathbf{v}}_i \cdot \hat{\mathbf{n}}, \tag{6.3}$$

where m_i and $\hat{\mathbf{v}}_i$ are the mask value and unit view vector (obtained from the camera calibration) respectively for the i^{th} pixel in the input image, and $\hat{\mathbf{n}}$ is the unit vector direction of *up*[1] given in the camera frame, which is obtained from the current 2-DOF attitude estimate. m_i is permitted to vary continuously in the range $[-1, +1]$ to account for uncertainty inherent in determining the desired class for pixels lying close to the horizon. This uncertainty stems from the fact that the visual horizon may be uneven, or the estimated attitude may be inaccurate.

Note that generating the sky/ground training mask from estimated 2-DOF attitude creates a feedback loop, such that the proposed algorithm *tracks* the appearance of the sky and ground regions through YUV space. In some cases, it may be beneficial to reduce the *gain* of this feedback loop by biasing the generated sky/ground training mask towards an upright and wings-level attitude. This effect is discussed in more detail in Section 6.5.1.

[1] In this study, *up* is defined as a vector that is perpendicular to the surface of the Earth.

Training the classifier

Once the desired class for each pixel in the input image has been identified using Equation 6.3, the desired class for each YUV element within the classifier is computed according to

$$\delta_{y,u,v} = \frac{1}{k} \sum_{i \in \mathbb{K}} m_i, \tag{6.4}$$

where $\mathbb{K}$ is the subset of pixels in the input image whose Y, U, and V values correspond to the appropriate YUV element within the classifier, and k is the cardinality of $\mathbb{K}$.

Then, the classifier weights may be updated,

$$w^{t+1}_{y,u,v} = w^{t}_{y,u,v} + \Delta w_{y,u,v}, \tag{6.5}$$

where $w^{t+1}_{y,u,v}$ is the new classifier weight for the element YUV and is constrained by $|w^{t+1}_{y,u,v}| \leq a$, and $\Delta w_{y,u,v}$ is defined above in Equation 6.2.

According to Equations 6.2 – 6.5, YUV elements corresponding to image regions close to the estimated directions of *up* or *down* will saturate to either *sky* ($w_{y,u,v} \geq +1$) or *ground* ($w_{y,u,v} \leq -1$) more quickly than those elements corresponding to image regions close to the horizon. However, the update heuristic (Equation 6.5) affects the *rate of change* of the learned weights (as defined by Equation 6.2), which depends on the desired output for each element, so the two classes (*sky* and *ground*) will eventually be driven apart and the classifier will approach a stable state ($|w_{y,u,v}| \geq 1$ for all Y, U, and V). This method of updating the classification weights is used in place of a simpler probability-based heuristic to maximise the inter-class difference. A probability-based weight would likely result in many YUV elements with $|w_{y,u,v}| < 1$, which would bias the output of the attitude matching stage because the attitude kernels have been computed assuming a perfect separation of sky and ground classes ($|w_{y,u,v}| \geq 1$ for all Y, U, and V, or equivalently, $|w_i(y,u,v)| = 1$ for all pixels, i, in the classified images).

Allowing the maximum magnitude of the classifier weights to be $a \geq 1$, allows for some hysteresis in the classified image, due to Equation 6.1. Put simply, YUV elements that definitely belong to one class or the other ($|w_{y,u,v}| \geq 1$) will tend to retain their classification ($|w_i(y,u,v)| = 1$) even in the presence of noise. This is beneficial because the attitude kernels are generated assuming $|w_i(y,u,v)| = 1$ for all pixels, i. In this study $a = 2$.

Following the classifier update stage, a 1D Gaussian low-pass filter ($\sigma = \frac{1}{3}$), is applied to each of the axes in the classifier. This mitigates any quantisation effects resulting from the arbitrarily chosen dimensions for the classifier. The application of this filter has the additional effect of diffusing weights to neighbouring elements within the classifier, which assists with the classification of previously unseen YUV combinations. Prior to applying the Gaussian filter, the total sum of the weights for both classes within the classifier is equalised by scaling the weights of the class with the lowest absolute sum. This ensures that the application of the Guassian filter does not bias the weights within the classifier by favouring a dominant class.

Initialising the 2-DOF attitude estimate

At instantiation, the classifier is completely untrained and the weights within the classifier are zero (*i.e.* $w_{y,u,v} = 0$ for all Y, U, and V). To initialise the classification-training cycle, therefore, a single 2-DOF attitude estimate must be provided, or alternatively a reasonable set of weights must be loaded into the classifier. This is to ensure the classifier initially labels the sky and ground regions correctly. Any reasonable attitude estimate can be used. In this study it is assumed simply that the aircraft is initially upright and wings-level. The effect of this assumption on the convergence of the 2-DOF attitude estimate is analysed in Figure 6.5. To perform the analysis, the attitude estimation algorithm was allowed to execute on a single input image (shown in Figure 6.4a) for 500 frames (*i.e.* this image was presented to the algorithm 500 times). This process was repeated for 1000 different initialisation attitudes, which were distributed evenly across the spherical surface of possible attitudes. The results of the convergence analysis are plotted in Figure 6.5.

It can be seen from Figure 6.5 that the convergence plot is approximately symmetrical about an angular initialisation error of 90°. This is a result of the fact that the *ground* and *sky* distributions within the input image are not recognised explicitly upon initialisation. Hence, instances of the algorithm initialised with an estimate of the *up* direction $< 90°$ from the true *up* direction tended to converge to the correct attitude, while instances that were initialised with an estimate $> 90°$ from the true *up* direction tended to converge to the opposite solution (*i.e.* the ground and sky classes ended up reversed within the classifier and the aircraft was computed to be exactly upside down).

It can also be seen from Figure 6.5 that if the initial estimate of *up* is $\sim 90°$ from

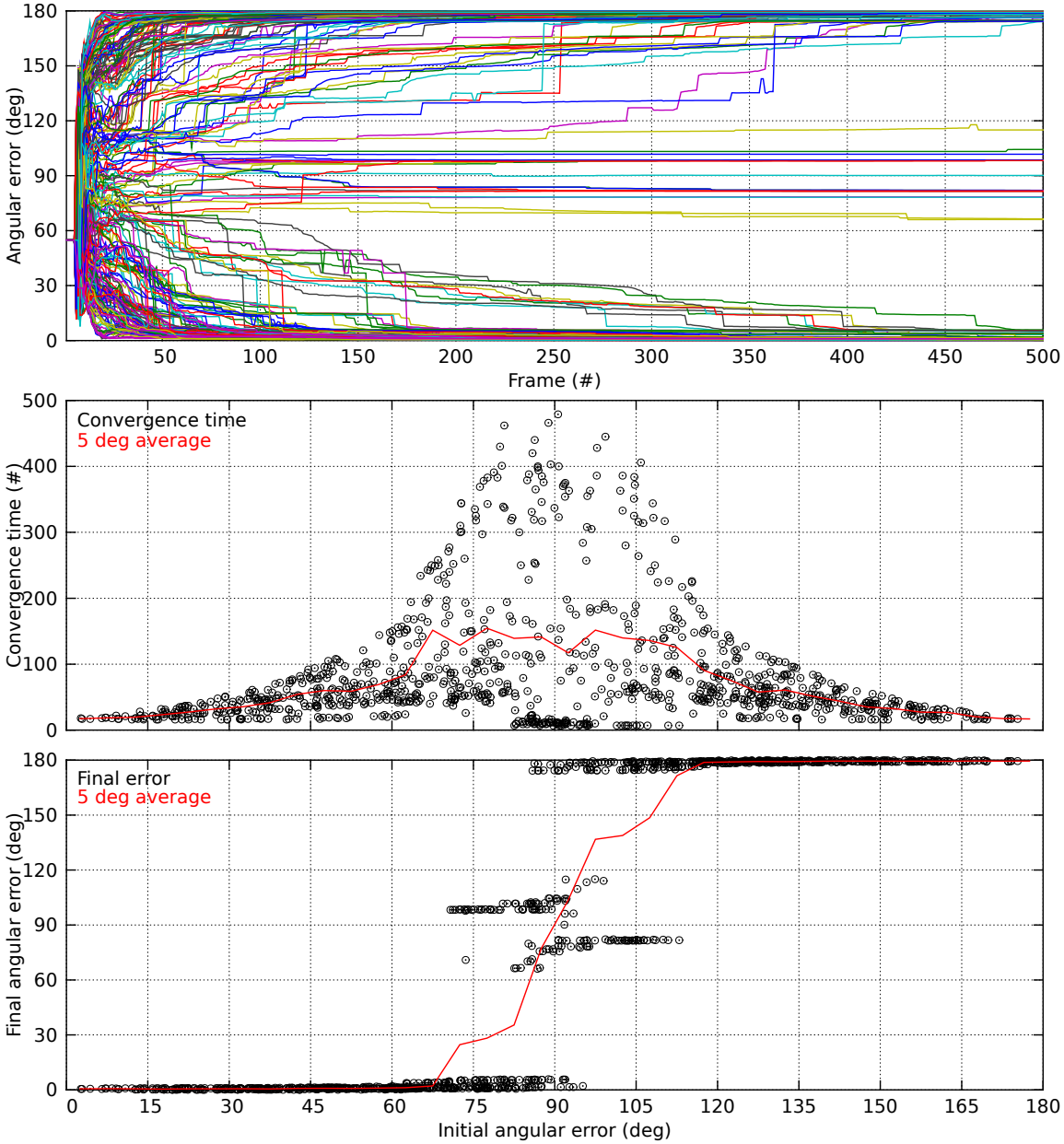


Figure 6.5: 2-DOF attitude convergence analysis data from 1000 trials. Plotted is angular error between computed attitude and true attitude (top) for each trial for the 500 frames following initialisation; number of frames taken to converge to within 5° of the final value (centre) plotted against initialisation error for each trial (black points); and angular error after 500 frames had elapsed (bottom) plotted against initialisation error for each trial (black points). Also plotted is average convergence time and average final angular error for initial angular errors within a 5° window (red lines).

the true direction, the classifier takes longer to converge or can become 'stuck', where the algorithm converges to an incorrect attitude. This is a result of approximately equal distributions of the same YUV combinations being attributed to both the ground and sky classes during the training phase, which stagnates the learning process. This situation is unlikely to occur for real image sequences, however, as even small variations in the distribution of the *ground* and *sky* YUV combinations (*e.g.* small changes in aircraft attitude between frames in an image sequence) will cause the classifier to 'pick a side' and converge to either the correct solution or the opposite solution.

In any case, for every trial where the initial *up* guess was $< 65°$ from the true direction of *up*, the attitude estimate converged to the correct solution within an average time of ~ 100 frames (or ~ 4s @ 25Hz) and much more rapidly for initialisation errors $\ll 65°$. Hence, Figure 6.5 indicates that assuming a level attitude initially is valid even during (non-aerobatic) flight. The algorithm is typically initialised prior to take-off when the aircraft is approximatley upright and wings-level. The 2-DOF attitude estimate will thus converge rapidly and no training offline prior to use is required.

6.3.2 Estimating heading direction

The 'visual compass' has been described previously for the purpose of extracting relative heading direction from a visual input [Labrosse, 2006, Mondragón et al., 2010a, Scaramuzza and Siegwart, 2008]. This approach typically operates on the principle that a rotation around the yaw axis of the vehicle corresponds to a simple left or right column shifting of a 360° panoramic image taken around the same axis.

Labrosse [2006], Scaramuzza and Siegwart [2008] describe implementations of the visual compass for ground vehicles, where the yaw axis of the vehicle is in general well aligned with the inertial frame *up* vector, hence they neglect the rolling and pitching motion of the vehicle. Mondragón et al. [2010a] extend the approach to include aerial vehicles, but also neglect roll and pitch. Their approach is limited, therefore, to estimating change in visual heading direction between consecutive frames, between which the roll angle and pitch angle of the vehicle are similar. Absolute heading direction is then obtained through integrating the estimated changes in heading direction.

A method is described here that enables a UAV to measure its heading direction

using a visual compass that accounts for the roll angle and pitch angle of the aircraft. This approach should therefore be more accurate than other previously described approaches and, additionally, the visual compass should be valid over much longer periods of time. Absolute heading can therefore be computed directly (relative to any prior heading), rather than by integrating the intervening changes in heading direction.

The proposed approach involves two stages. Firstly, the orientation and position of the horizon (estimated in Section 6.3.1) are used to extract a stabilised horizon image. Relative heading direction is then estimated by left- or right-column shifting the stabilised image (or equivalently rotating the panoramic image around the inertial frame *up* vector) to match a stored reference image. These two steps are illustrated in Figure 6.6 and described in more detail below.

Generating the visual compass

The visual compass is a stabilised panoramic image of the horizon and sky. The panorama extends 360° around the inertial frame *up* vector in the horizontal image axis and from 5° below the horizon to 85° above the horizon in the vertical image axis. Image regions well below the horizon plane are ignored because they are likely to contain objects that are close by, and hence transient in the FOV as the aircraft translates. Here, the visual compass is represented by a greyscale image with dimensions 80px × 20px.

To produce the visual compass, the estimated orientation and position of the horizon is used to generate a transform that remaps pixels from the input image to the stabilised horizon image. The transform is given by

$$\begin{bmatrix} s_x^i \\ s_y^i \\ s_z^i \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \cdot \cos\theta & \sin\phi \cdot \sin\theta \\ -\sin\phi & \cos\phi \cdot \cos\theta & \cos\phi \cdot \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r_x^i \\ r_y^i \\ r_z^i \end{bmatrix},$$

where $\mathbf{s}^i$ and $\mathbf{r}^i$ are view vectors in the source and remapped images respectively that correspond to the i^{th} pixel in the remapped image, and ϕ and θ are the estimated roll angle and pitch angle of the aircraft respectively. The relationship between view vector and pixel coordinates for a particular pixel is determined from the camera calibration parameters.

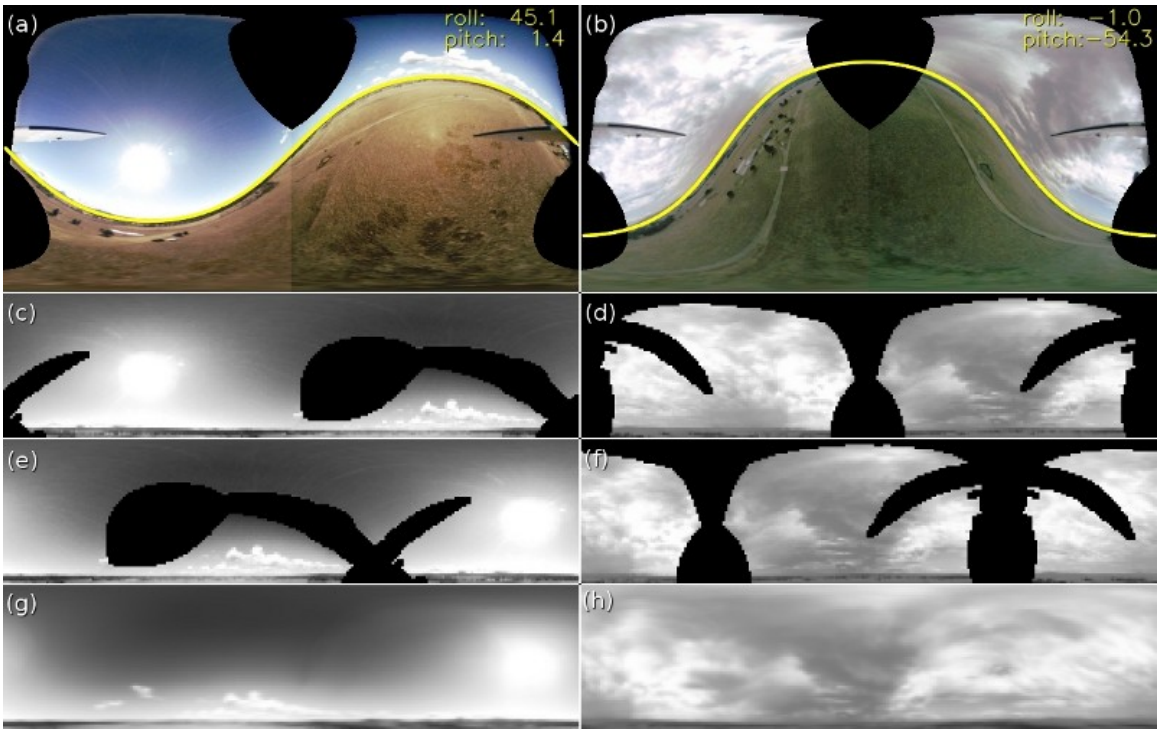


Figure 6.6: Illustration of how the visual compass is generated and used to estimate heading direction. Panels (a) and (b) show stitched images captured by the vision system during two different flight tests, overlaid with the estimated position and orientation of the horizon (yellow). Panels (c) and (d) show the stabilised horizon images, extracted from (a) and (b) respectively. The accumulated reference horizon images are shown in panels (g) and (h). At each frame the stabilised horizon image is shifted and matched against the reference image to obtain the corresponding difference in heading direction between the two. The best matching shifted horizon images for these two frames are shown in panels (e) and (f). The FOV of the stitched images is $360° \times 180°$, and the FOV of the remapped horizon images is $360° \times 90°$. Black regions in panels (a) and (b) represent regions of the view sphere not imaged by the vision system, and additionally regions of the view sphere that correspond to the aircraft in the other panels. Note that high-resolution horizon images are shown here to aid understanding, in practice a much lower resolution of 80px × 20px is used.

Determining heading direction

Removing the influence of the aircraft's rolling and pitching motions on the visual compass (*i.e.* by stabilising the horizon image) means that the only remaining degree of freedom is a 1D shift along the horizontal image axis, which corresponds to change in the aircraft's heading direction. Defining a reference horizon image, therefore, allows the relative heading direction of the aircraft to be determined by shifting the instantaneous horizon image to match the reference image.

An error score for each candidate heading direction is computed from the SAD between the instantaneous horizon image and the stored reference image. In practice, a discrete number of candidate heading directions are tested (an angular resolution of 4° is used in this study) and the true heading direction is estimated by performing an equiangular fit [Shimizu and Okutomi, 2003] on the minimum and neighbouring SAD scores.

In the proposed approach, the reference horizon image is not static but is accumulated from each matched horizon image according to

$$d_i^t = \alpha_d \cdot c_i^t + (1 - \alpha_d) \cdot d_i^{t-1},$$

where d_i and c_i are the i^{th} pixels in the accumulating reference image and the instantaneous matched horizon image respectively, and α_d is the accumulation rate of the reference image (here $\alpha_d = 0.01$, which corresponds to $\tau_d \sim 4\text{s} \,@\, 25\text{Hz}$). This procedure generates an exponentially decaying weighting of past reference images. Regions in the matched horizon image that correspond to the body of the aircraft or areas not imaged by the vision system are not used to update the reference horizon image.

Visually estimated heading direction was not observed to drift relative to the ground-truth during flight testing (Section 6.4.2). The reference horizon image is not static, however, so it is possible that noise or biases in the matching process, or the motion of the sun, cloud patterns, or other distinctive features in the visual compass may cause the reference image to drift over long periods of time. To counter this unwanted drift, and also to give physical meaning to the relative heading estimates, an angular offset between the estimated relative heading direction and magnetic north is estimated according to

$$\Psi_{\text{bias}}^t = \alpha_b \cdot (\mathbb{M} - \Psi^t) + (1 - \alpha_b) \cdot \Psi_{\text{bias}}^{t-1}, \tag{6.6}$$

where Ψ_{bias} is the angular offset between magnetic north and the visual heading direction, $\mathbb{M}$ is the direction of magnetic north (obtained from an IMU on-board the aircraft), Ψ is the estimated visual heading direction, and α_b is the update rate of the heading bias (here $\alpha_b \sim 2.6 \times 10^{-3}$, which corresponds to $\tau_b = 15\text{s} \,@\, 25\text{Hz}$).

Measuring angular offset between the horizon reference image and magnetic north allows heading direction to be given relative to magnetic north, rather than the arbitrary initial orientation of the aircraft. However, magnetometer measurements are typically of much higher latency and lower precision than the visual approach described here. The visual heading estimate is therefore only loosely coupled with the magnetic estimate.

The corrected heading direction is given by

$$\Psi_{\text{corrected}} = \Psi + \Psi_{\text{bias}}. \tag{6.7}$$

For the results presented in this Section 6.4, however, the uncorrected visual heading estimate, Ψ, is used, so that an unbiased appraisal of the performance of the system can be made.

6.4 Flight test results

Two separate flight tests were performed to analyse the performances of the 2-DOF attitude and the heading direction estimation algorithms. During the flight tests, omnidirectional FOV images were captured at 25Hz using the *iEye* vision system (described in Section 5.3), which was mounted on the nose of a small-scale UAV (pictured in Figure 6.7). The on-board computing hardware and the aircraft (excluding the vision system) were identical to those described for the collision avoidance flight tests in Section 3.3.3.

A ground-truth measure of the aircraft's attitude was obtained manually for each flight test to quantify the accuracy of the visual algorithms. Additionally, an IMU (pictured in Figure 5.3) was mounted rigidly to the vision system and its output was logged during testing to provide a comparative measure of 3-DOF attitude. In this section, procedures for obtaining the ground-truth data sets are explained and results from each of the flight tests are analysed.

Figure 6.7: The aircraft used for the flight tests discussed in Part II with the *iEye* vision system mounted on the nose.

6.4.1 Estimating 2-DOF attitude using the visual horizon

The first flight test was conducted to analyse the performance of the visual horizon detection and 2-DOF attitude estimation scheme, introduced in Section 6.3.1. This flight test is shown in the accompanying video[2]. The analysed flight segment covers a complete flight, from take-off to touchdown, and includes three aerobatic manoeuvres.

At the commencement of this flight test, the classifier was completely untrained. Figure 6.5 indicates that the 2-DOF attitude estimation algorithm converges stably from a wide range of initial attitudes. Simply assuming that the aircraft is initially upright with wings level is therefore sufficient for convergence. An example of the system initialising successfully during flight is shown in the accompanying video, confirming that this assumption is valid for a wide range of initial attitude errors.

A ground-truth measure of the aircraft's 2-DOF attitude during the flight test was obtained by manually tracking the position and orientation of the visual horizon. Nine points, on average, evenly spaced along the visual horizon were selected in every tenth frame of the flight test image sequence, and a least-squares fit of a planar surface was made to the corresponding unit view sphere vectors to obtain the horizon normal. The ground-truth measure used here is described in greater detail in a co-authored publication on an alternative attitude estimation scheme [Thurrowgood et al., 2010].

The aircraft's roll and pitch angles, estimated by the vision system during the first flight test, are plotted in Figure 6.8 alongside the ground-truth measure and

[2]Attachment AV05 (available at http://youtu.be/4lNj0SPgVGg).

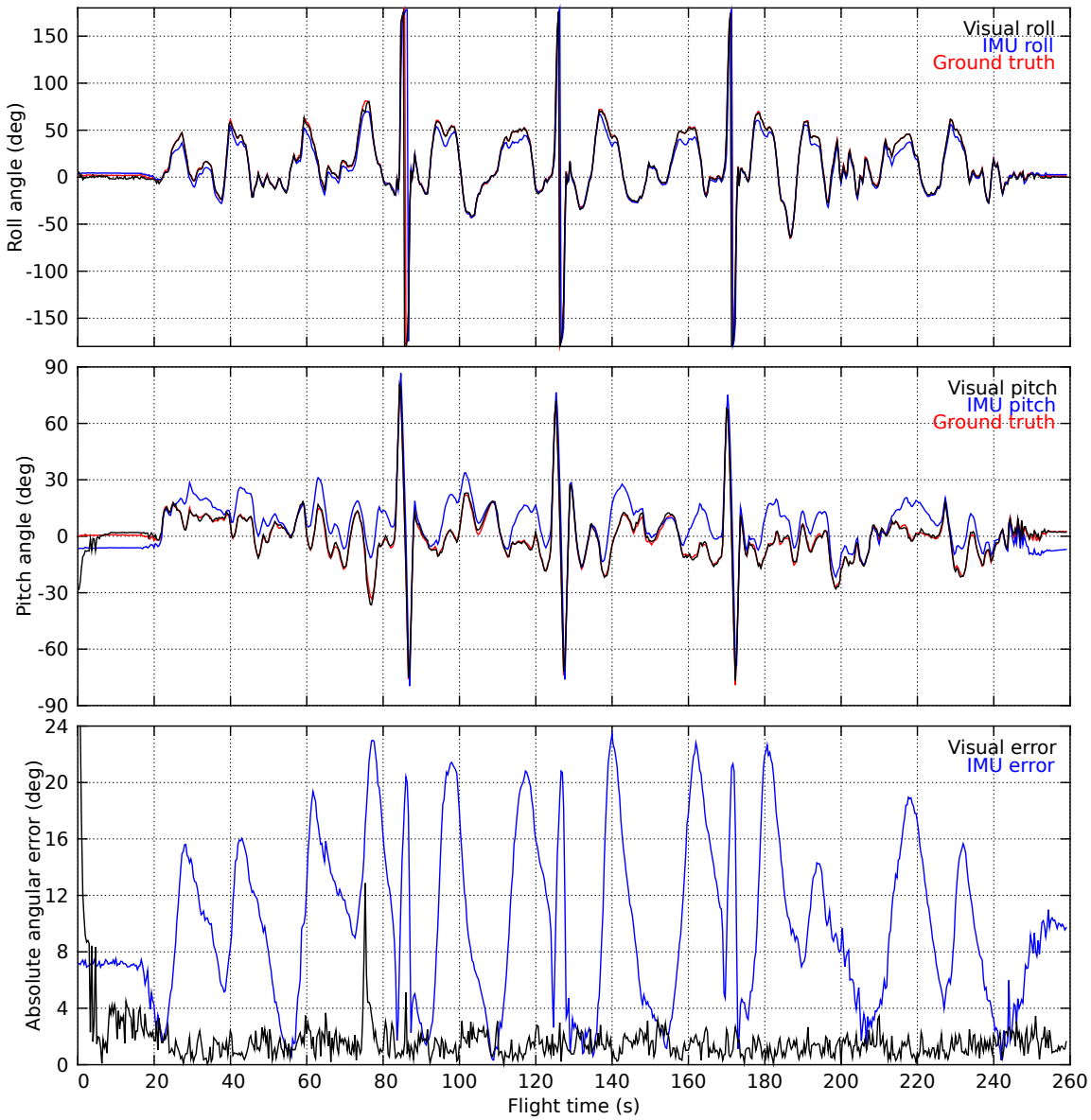


Figure 6.8: Roll angle (top) and pitch angle (centre) of the aircraft during the first flight test as measured automatically from the horizon using the proposed method (black), measured manually from the horizon (red), and reported by the IMU (blue); and the absolute angular errors (bottom) for the 2-DOF attitude measured by the proposed approach (black) and by the IMU (blue). Absolute angular errors were computed by measuring absolute angular separation between the *up* vector computed manually from the horizon and the *up* vectors computed by the vision system and IMU respectively.

2-DOF attitude reported by the IMU on board the aircraft for comparison. It can be seen that error associated with visually estimated 2-DOF attitude quickly approached zero as the classifier learned to separate the spectral and intensity properties of the sky and ground classes. In comparison, error between the attitude reported by the on-board IMU and the ground-truth was significant throughout the entire flight and showed periodic variation that correlated with the turning motions of the aircraft. This strongly suggests that the horizon-based estimates of roll and pitch are much more accurate than those derived from the IMU.

To quantify the accuracy of the proposed automatic horizon detection algorithm, angular 2-DOF attitude errors during the flight test were computed by finding the absolute angular separation between the *up* vector produced by the ground-truth data and the *up* vector computed automatically from the visual horizon for each frame (plotted in Figure 6.8). Using this metric, *average* angular error for the proposed scheme was 1.49°, once the classifier had initially learned the sky and ground classes ($>$ 10s in Figure 6.8). For comparison, average angular error for the IMU was 10.0° during the same period. This comparison may be slightly unfair, since the same attitude reference (the visual horizon) is used for both the ground-truth and the proposed system, while the IMU measures a gravity-referenced attitude. In the area where the flight testing was conducted, however, the visual horizon deviates from a true ground plane by a maximum of $\sim 3°$. It is expected, therefore, that no more than a couple of degrees have been added unfairly to the average angular error computed for the IMU. Furthermore, it can be seen from Figure 6.8 that the maximum angular error for the IMU is significantly greater than that for the automatic horizon detection algorithm.

6.4.2 Estimating heading direction using a visual compass

The second flight test was conducted to analyse the performance of the visual heading direction estimation scheme, introduced in Section 6.3.2. The flight test used for analysis is shown in the accompanying video[3]. The analysed flight segment covers a complete flight, from take-off to touchdown, and comprises a series of banked turns.

A ground-truth measure of the aircraft's heading direction during the flight test was obtained using a novel procedure that involved manually tracking the location of

[3]Attachment AV06 (available at http://youtu.be/zKHh1kndWs4).

the aircraft's shadow in the imagery captured from on board the test aircraft[4]. This is equivalent to tracking the position of the Sun. Using the calibration parameters of the on-board cameras, camera-frame view vectors to the apparent positions of the aircraft's shadow were recorded at regular intervals throughout the flight test. For each frame, the visually estimated roll and pitch angles were used to transpose the view vector corresponding to the shadow into the (rotating) inertial frame (defined in Figure 4.1), and the ground-truth heading estimates were computed according to

$$\Psi^t_{\text{true}} = \arctan(\frac{G^t_y}{G^t_x}),$$

where Ψ^t_{true} is the ground-truth heading direction at time t, and G^t_x and G^t_y are the x and y components of the inertial frame view vector that corresponds to the position of the centre of gravity of the aircraft's shadow at time t.

It was shown in Section 6.4.1, using the visual approach proposed in Section 6.3.1, that the aircraft's 2-DOF attitude is able to be estimated with an average angular error of $\sim 1.49°$, and here the average error for the *manual* tracking of the aircraft's shadow is estimated to be approximately $< 1.5°$ (this value was obtained from the average area covered by the aircraft's shadow in the recorded imagery). It is considered, therefore, that the ground-truth measure described here accurately represents the aircraft's true heading direction during the flight test. Frames in which the aircraft's shadow was not visible are omitted from analysis.

Visually estimated heading direction during the second flight test is shown in Figure 6.9, alongside the ground-truth measure and heading direction reported by the IMU for comparison.

It can be seen from Figure 6.9 that absolute angular error between visually estimated heading direction and ground-truth is consistently $< 10°$ for the entire flight. This indicates that there is very little or no drift of the accumulated reference horizon image, even after a flight time of ~ 250s. Average angular error for visually estimated heading direction was computed to be 2.47°, using the ground-truth

[4]The aircraft's shadow appears in the accompanying video as a dark spot surrounded by a bright halo. This is probably due to a combination of optical phenomena that occur because the shadow of the observer is coincident with the anti-solar point. Mechanisms that may explain the effect include coherent backscattering of the Sun's rays from the crystalline structure of the soil or vegetation; obscuring of shadows that are colinear with the illumination source and the observer – *i.e.* the opposition effect (http://www.atoptics.co.uk/atoptics/oppos2.htm); or the focussing properties of tiny droplets of dew present on the grass – a phenomenon known as *Heiligenschein* (literally "holy shine", http://www.atoptics.co.uk/droplets/heilfrm.htm) [Minnaert, 1954].

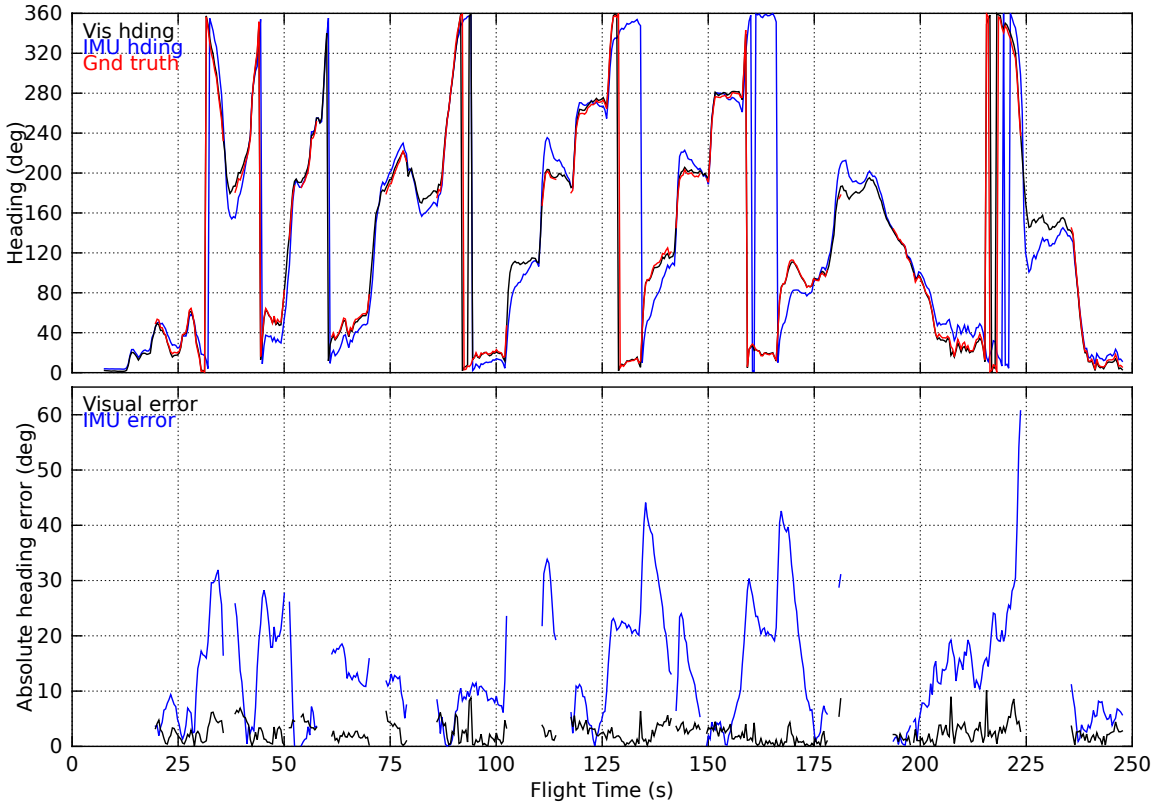


Figure 6.9: Heading direction (top) during the second flight test as measured automatically from the visual compass using the proposed approach (black), measured manually from the position of the aircraft's shadow (red), and reported by the IMU (blue); and absolute angular errors (bottom) between ground-truth and visually estimated heading direction (black) and between ground-truth and heading direction reported by the IMU (blue). Gaps in the ground-truth data represent frames in which the aircraft's shadow was not visible.

measure described above. For comparison, average angular error for IMU heading direction was computed to be 13.7°. Additionally, it can be seen from Figure 6.9 that maximum heading error for the IMU is much greater than that for the visual estimate. These results indicate that the visually estimated heading direction is much more accurate than that measured by the IMU.

Taken together with the results presented in Section 6.4.1, the results presented here demonstrate that the proposed approach is better able to provide accurate estimates of the 3-DOF attitude of an aircraft than a mid-range, off-the-shelf IMU.

6.4.3 Closed-loop 3-DOF attitude control

A closed-loop control system was used to conduct the visual heading estimation analysis discussed above. Data from that analysis, presented in Figure 6.9, includes open-loop segments (in which the aircraft was flown manually under radio control) as well as a closed-loop segment (93s $\sim$ 169s), during which the aircraft was under autonomous control. During the autonomous segment of flight, the aircraft was commanded to perform a series of 90° turns whilst maintaining altitude. The closed-loop flight control system was implemented using PID feedback control loops to enable automatic execution of these manoeuvres. The PID control system is illustrated in Figure 6.10.

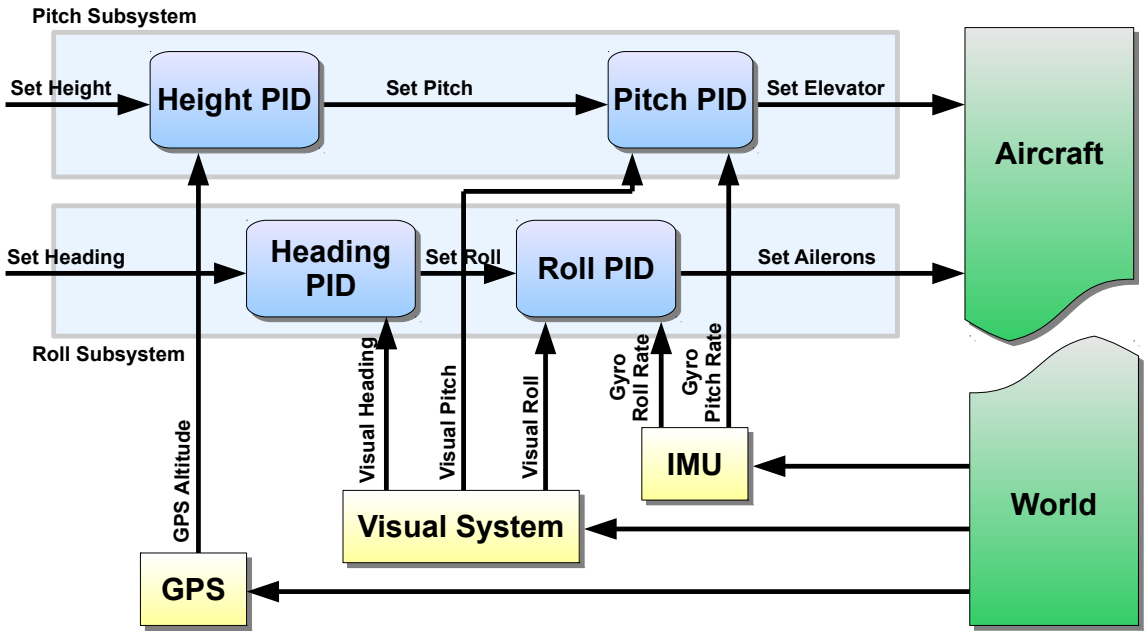


Figure 6.10: Block diagram illustrating the closed-loop control scheme used with the *iEye* vision system for autonomous control of 3-DOF attitude.

The PID control loops gave the guidance system control of both altitude and heading direction, as well as the roll angle and pitch angle of the aircraft. The closed-loop flight test thus examined the ability of the visual guidance system to control the aircraft's 3-DOF attitude. The vision system provided control feedback for heading direction as well as roll and pitch of the aircraft, while the height control feedback was provided by a GPS unit (U-blox LEA-4H[5]), which was mounted on board the

[5]http://www.u-blox.com/images/downloads/Product_Docs/LEA-4x_Data_Sheet%28GPS.G4-MS4-06143%29.pdf

aircraft. An IMU (MicroStrain 3DM-GX2) was rigidly affixed to the vision system and provided roll rate and pitch rate feedback for the derivative components of the roll angle and pitch angle PID controllers. The roll angle and pitch angle PIDs generated roll rate and pitch rate set-points, which were used directly as aileron and elevator commands respectively. This implicitly assumes a particular control surface deflection generates a constant rotation rate about the corresponding aircraft axis. Flight tests conducted with the terrain following system, described in Section 4.3.2, indicated that this allowed a closer approximation of the true aircraft dynamics than treating the control surfaces as accelerating inputs. Efficacies of the control surfaces – the constants of proportionality relating deflections of the control surfaces to rotation rates – depend strictly on the aircraft's airspeed. A method for estimating control surface efficacy online is described in Section 8.4. For the results presented here, however, the aircraft was flown at an approximately constant speed by maintaining a constant throttle setting.

3-DOF attitude of the aircraft during the autonomous section of flight is displayed in Figure 6.11, alongside respective set-points for roll angle, pitch angle, and heading direction. The close correspondence between the set-points and measured roll, pitch, and heading angles shows that the guidance system was able to control the aircraft's attitude precisely. This facilited effective control of both the altitude (not shown) and heading direction of the aircraft. It is noted that the precision of the steady-state response of the aircraft to heading direction commands could have been improved by more finely tuning the heading direction PID – specifically, increasing the gain on the integrated heading error. A 'smooth' heading rate signal was measured by transforming the body-frame yaw-rate (measured by the IMU) into the inertial frame using the visual 2-DOF attitude estimate. Flight data indicated that this heading rate signal was good enough to have permitted a larger derivative error gain for the heading direction PID. In combination with an increased proportional error gain, more aggressive control of heading direction could have been realised – the control system depicted in Figure 6.10, however, was found to be sufficient for the experiments described here.

A modified system for controlling aircraft altitude and attitude is described in Section 6.6.2 which does not incorporate a GPS or IMU as depicted in Figure 6.10. In the modified control system, the signals provided by these devices are obtained visually – resulting in a purely vision-based control system, which is used for terrain following and landing a small-scale UAV automatically in an uncontrolled

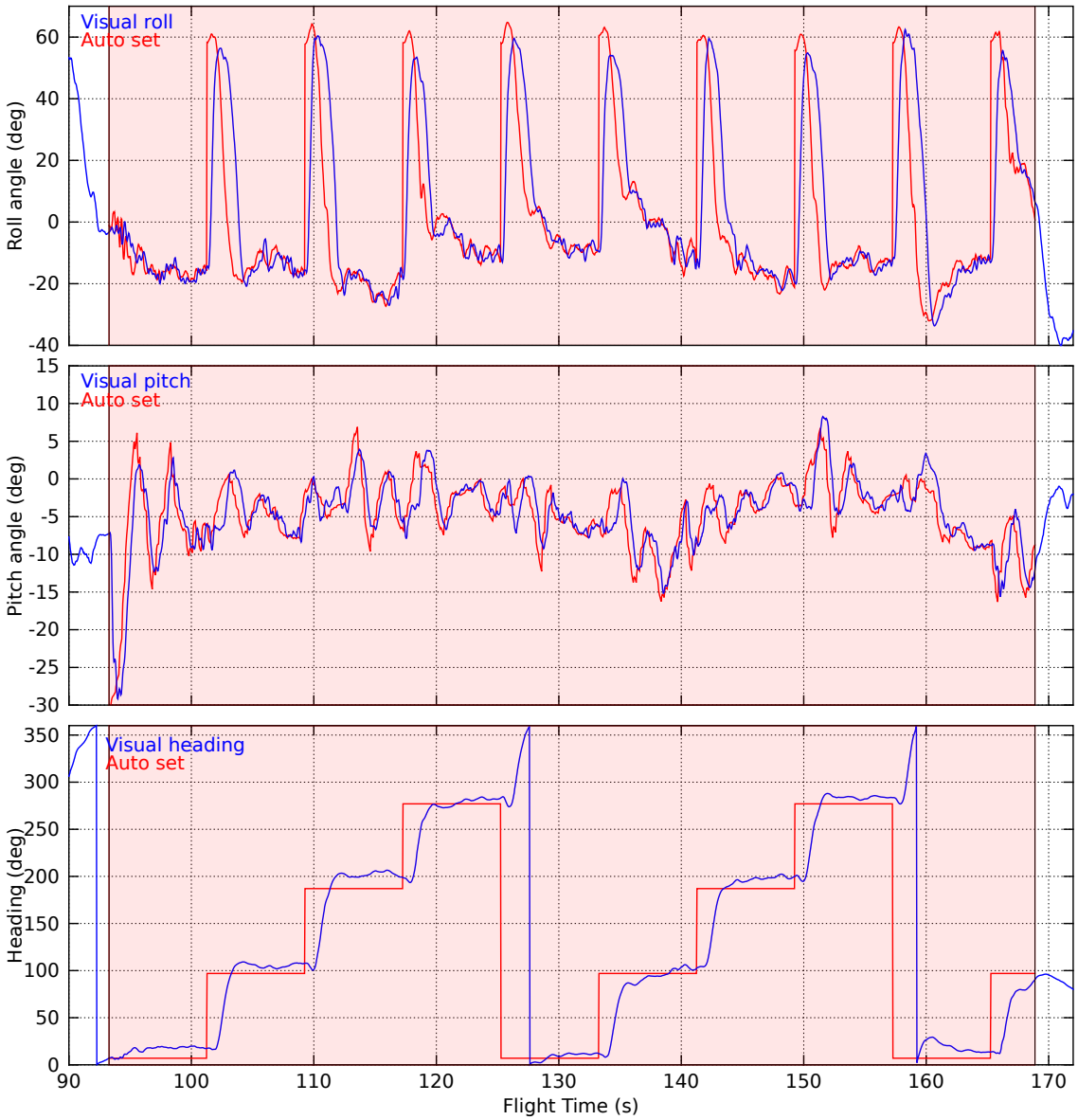


Figure 6.11: Roll angle (top), pitch angle (centre), and heading direction (bottom) of the aircraft during the autonomous segment (shaded red) of the second flight test as estimated by the vision system (blue). Also plotted are roll, pitch, and heading set-points (red) during the period of closed-loop control.

environment. This is the first time a purely vision-based guidance system has been implemented on board a small-scale UAV and used for such manoeuvres. Vision is not typically used for low-level flight control due to the perceived lag of visual rotation rate measurements with respect to their inertial counterparts (due to relatively large transmission and processing times). Interestingly, however, it was found during this study (data not shown) that roll and pitch rates derived visually from optic flow measurements were no more delayed than the corresponding inertial measurements, due to the data filtering performed on the inertial measurements on board the IMU.

6.5 Discussion

The vision-based guidance scheme described in this chapter enables an aircraft's 3-DOF attitude to be estimated and controlled more precisely than by using an IMU. Performance of this approach is compared below with other previously published visual approaches.

6.5.1 Performance

Obtaining a quantitative comparison of the performance of the proposed visual 3-DOF attitude estimation scheme with other previously published approaches is not always possible due to the difficulty of obtaining proper ground-truth measurements of aircraft attitude during flight. The visual horizon detection schemes described in Thurrowgood et al. [2009, 2010] and in Ettinger et al. [2002b][6] have been implemented here, however, to obtain a direct comparison with the proposed approach for estimating 2-DOF attitude. Those algorithms were executed on the same flight imagery used in Section 6.4.1 to analyse the proposed method. All three approaches gave similar results, with Thurrowgood et al. [2009, 2010] and the implementation of Ettinger et al. [2002b] producing average absolute angular errors of 1.45° and 3.92° respectively, compared to 1.49° for the proposed approach.

Mondragón et al. [2010a] compare the performance of their visual 3-DOF attitude estimation scheme to the output from an IMU during four short flight segments (totalling $\sim$ 180s). The average root mean square errors (RMSEs) from their quoted values for the flight segments are 1.68°, 3.30°, and 5.77° for roll angle, pitch angle, and

[6]The implementation here uses the J2 optimisation criterion (see Ettinger et al. [2002b]) and an input image resolution of 80px $\times$ 40px. An angular resolution of 5° is used and an equiangular fit [Shimizu and Okutomi, 2003] is applied to estimate the true attitude.

heading direction respectively. It was shown in Section 6.4, however, that an IMU does not necessarily provide a good comparative measure of attitude. Here, error was computed with respect to ground-truth, giving average absolute angular errors of 1.49° for 2-DOF attitude (roll and pitch angles) and 2.47° for heading direction.

It is expected that methods such as that described in Ettinger et al. [2002b] would outperform the proposed method for data sets containing an extreme amount of image noise, such as might be experienced when transmitting video signals over an analogue interface. As the proposed system is implemented on board the aircraft, however, there is no such extreme image noise and hence no such performance cost. The approaches described by Thurrowgood et al. [2009] and Mondragón et al. [2010a] for computing the 2-DOF and 3-DOF attitude respectively from visual input are similar in purpose to the proposed approach. It is expected, however, that the use of an adaptive classifier to segment the sky and ground classes and an adaptive visual compass to estimate heading direction would permit the proposed approach to function automatically (without requiring retraining) for a wider range of environments and vision systems than the methods described in Mondragón et al. [2010a], Thurrowgood et al. [2009]. Furthermore, the present work stabilises the visual compass using estimated 2-DOF attitude. It is thus expected that the proposed approach for visually estimating heading direction would function more accurately over a wider range of aircraft attitudes than the approach described in Mondragón et al. [2010a]. In fact, the aircraft reached roll and pitch angles of ±80° during the second attitude flight test (analysed in Section 6.4.2), compared with ±15° for Mondragón et al. [2010a] during their flight tests. Due to their unstabilised visual compass, Mondragón et al. [2010a] were only able to compute the change in body-frame yaw angle between successive images. To extract absolute heading direction, then, they must transpose measured yaw deltas into the inertial frame and integrate over time. The integration of noisy yaw deltas will likely result in a heading direction estimate that drifts from ground-truth more quickly than that for the approach proposed here.

Finally, in addition to providing accurate visual estimates of 3-DOF attitude, another advantage of the proposed approach is its rapid execution. Just 1.3ms @ 1.5GHz are required to compute 2-DOF attitude from the visual horizon using input images with a resolution of 80px × 40px and a FOV of 360° × 180°; and 1.6ms @ 1.5GHz to compute heading direction from the visual compass using input images with a resolution of 80px×20px and a FOV of 360°×90°. This gives a total of 2.9ms @ 1.5GHz

	2-DOF attitude		Heading direction	
	mean error (°)	execution (ms)	mean error (°)	execution (ms)
Thurrowgood et al. [2009]	1.45	1.6	-	-
Ettinger et al. [2002b]	3.92	20	-	-
Mondragón et al. [2010a]	2.94	~ 16	4.59	~ 16
IMU	10.0	-	13.7	-
Infrared horizon	< 5	-	-	-
Proposed	**1.49**	**1.3**	**2.47**	**1.6**

Table 6.1: Comparison of errors and execution times computed for the various approaches discussed in Section 6.5.1 with errors and execution times measured for an IMU and the proposed approach during the flight tests discussed in Section 6.4. Reported RMSE values have been converted to estimates of average unsigned angular errors by multiplying by $\sqrt{\frac{2}{\pi}} \approx 0.798$, which assumes normally distributed data, and execution times are stated as equivalent execution times for a 1.5GHz processor.

to compute full 3-DOF attitude. No significant increase in accuracy is observed when higher resolution input images are used, although the execution time of the algorithm is increased significantly. For comparison, to estimate 2-DOF attitude, Ettinger et al. [2002b] and Thurrowgood et al. [2009] quote execution times equivalent to 20ms @ 1.5GHz and 1.6ms @ 1.5GHz respectively. Mondragón et al. [2010a] do not quote an execution time for visually computing full 3-DOF attitude but instead state that their algorithm executes in "near real time". Given that they capture images at 30Hz and utilise a 1.5GHz processor on board their aircraft, it can be assumed that their computational cost is equivalent to $\sim$ 33ms @ 1.5GHz. The proposed approach, therefore, is more computationally efficient than other state-of-the-art visual methods.

Performance attributes of the different visual approaches discussed above are summarised in Table 6.1. It can be seen that the proposed approach compares favourably with approaches previously published by Ettinger et al. [2002b], Mondragón et al. [2010a], Thurrowgood et al. [2009, 2010]. Additionally, the proposed approach is approximately an order of magnitude more accurate than an off-the-shelf, mid-range IMU, and also outperforms other inexpensive autopilot solutions such as infrared horizon detectors [Egan and Taylor, 2007, Rogers et al., 2011]. This suggests that the vision-based approach presented here would be ideally suited to providing fast and accurate estimates of an aircraft's 3-DOF attitude during autonomous flight, and this is confirmed by the closed-loop flight testing results presented in Section 6.4.3.

6.5.2 Limitations

The visual horizon detection algorithm was designed to function with very wide FOV vision systems (*e.g.* the *iEye* vision system described in Section 5.3). A weakness of the proposed approach is that for vision systems with a restricted FOV (much lower than 180°), the horizon may not always be visible. In these cases, the 2-DOF attitude cannot be determined uniquely, which may lead to erroneous estimates of the roll angle or pitch angle of the aircraft. Furthermore, this may corrupt the sky/ground mask, which is generated from the attitude estimate, and for prolonged sequences where the horizon is not visible, this may cause the classifier to learn an incorrect sky/ground association for some YUV combinations. This problem could be mitigated by biasing the generated sky/ground mask towards an upright and wings-level attitude (*i.e.* biasing $\hat{\mathbf{n}}$ in Equation 6.3 towards a level attitude, rather than the estimated direction of *up* in the camera frame) for vision systems with a restricted FOV. This procedure would rely on the assumption that the aircraft is statistically most likely to be in this orientation, which would have little effect on the precision of the system under normal operation but would improve the ability of the system to recover quickly from situations where the horizon was not visible.

The main failure mode for the visual heading estimation algorithm is likely to be when there is insufficient texture or features for the visual compass to uniquely describe the aircraft's heading direction. This might occur on days that are completely overcast or on blue-sky days when the only significant feature is the Sun, which can be obscured by the body of the aircraft. However, by not explicitly extracting prominent features to track, and by matching the stabilised horizon images using an appearance-based algorithm instead, the visual compass is able to make use of other heading cues present in the environment – such as intensity gradients in the sky, which are present even when there are no prominent features. For example, Figure 6.6 shows a single frame from each of two data sets, one of which was collected on a completely overcast day. It can be seen that the visual compass is able to match the stabilised horizon image correctly despite a lack of distinctive features and large areas of both horizon and sky being obscured by the body of the aircraft. Accompanying video of the heading direction flight test (analysed in Section 6.4.2) shows that the visual compass is able to make use of the intensity gradient over the whole sky to match the stabilised horizon image correctly when the only distinctive feature (the Sun) is obscured by the body of the aircraft.

To achieve stable estimates of heading direction over the course of hours or even days, an angular offset between the visually estimated heading direction and the direction of magnetic north is estimated according to Equation 6.6. Visually estimated heading direction is then corrected according to Equation 6.7 to obtain an estimate of true heading direction.

Nearby objects that protrude through the horizon plane will affect the accuracy of the visual compass if they subtend a large enough angular size and move significantly through the aircraft's FOV as it translates (*i.e.* if the aircraft flies closely past a tall city building). Here, the visual compass is represented by a greyscale image with dimensions 80px × 20px. Given the small resolution and large FOV of the visual compass, the aircraft must therefore be very close to a very large object before it affects the visual heading estimate significantly. Furthermore, because the reference horizon image is accumulated slowly over time, nearby objects will not persist for long in the FOV of the aircraft as it translates, and hence their impact on the visual heading estimate will be mitigated. The precision of the system would be affected, however, if the aircraft were to be surrounded by such objects – such as in built-up urban environments or under the canopy level in a forest.

6.6 Applications

Section 6.4 demonstrated that the visual approach described in this chapter enables an aircraft to estimate and control its 3-DOF attitude precisely. This ability is crucial, not just for stabilisation, but also for a wide variety of tasks related to autonomous guidance of UAVs. For instance, ability to directly measure absolute heading of a vehicle significantly reduces the rate at which 2D position errors are accumulated through odometry, as it avoids the need to integrate (noisy) heading rate measurements [Cheung et al., 2008]. Similarly, ability to directly measure 3-DOF attitude of an aircraft significantly reduces drift of the vehicle's 3D position estimated from visual odometry (for a demonstration, see the accompanying video[7]).

[7]Attachment AV07 (available at http://youtu.be/R5t3zHNic3o). Note that in this video the faint hemispherical dome drawn in the sky represents 3-DOF attitude, estimated visually from the horizon, and the grid drawn on the ground represents 3D position of the aircraft with respect to the point of take-off, integrated from visually estimated egomotion; direction and magnitude of estimated 3D translation are represented by a yellow circle ahead of the aircraft. A 2D projection of the aircraft's path is also drawn on the ground in yellow, with circles spaced at 5m intervals. This video was produced by Mr. Saul Thurrowgood.

Some real-world applications have been developed recently within the biorobotics laboratory at the Queensland Brain Institute[8] that are dependent on the vision-based attitude estimation scheme described in this chapter. These applications are discussed below briefly and demonstrate potential of the system developed within this thesis to be usefully applied to real-world scenarios.

6.6.1 Vision-only estimation of wind field strength and direction

A new approach is described here for estimating strength and direction of the wind field from a UAV using visual input only. This section (Section 6.6.1) largely reproduces research that was originally published by Moore et al. [2012]. The intended purpose of this research was to enable a UAV to select the safest direction from which to approach a landing site. Knowledge of local wind field dynamics is also useful for many other aerial applications, such as trajectory planning, plume tracking and gas leak detection, meteorological observation, and other forms of environmental monitoring such as fire-front tracking. The effect of the local wind field is of particular concern to small-scale UAVs, where the wind speed may be of the same order – or greater – than the aircraft's airspeed. The vertical component of the wind field is often negligible for powered aircraft. Aircraft and wind velocities are therefore considered in the horizontal plane only, unless specified otherwise. Geometry of the problem is captured simply by the so-called wind triangle, shown in Figure 6.12.

Thrust exerted by the aircraft and aerodynamic forces exerted by its control surfaces all act on the local air mass. Velocity of the aircraft in the inertial frame, $\mathbf{w}$, therefore results from its velocity in the flight frame, $\mathbf{v}$, and velocity of the local air mass, $\mathbf{u}$, as illustrated in Figure 6.12. Two of the three vectors or four of the six parameters (u, v, w, α, β, γ) of the wind triangle are needed to derive remaining parameters. In many cases, the aircraft's ground velocity, $\mathbf{w}$, can be measured with GPS/IMU or similar instruments. Measuring flight velocity, $\mathbf{v}$, is not so straightforward for small-scale aircraft, however, because air flow angles (*i.e.* angle of attack and angle of sideslip) should be measured in addition to the aircraft's airspeed and attitude.

The M^{2}AV small-scale UAV described by van den Kroonenberg et al. [2010] was

[8]The author is affiliated with this laboratory and contributed to all research presented in this section; contributions by others are stated explicitly where applicable.

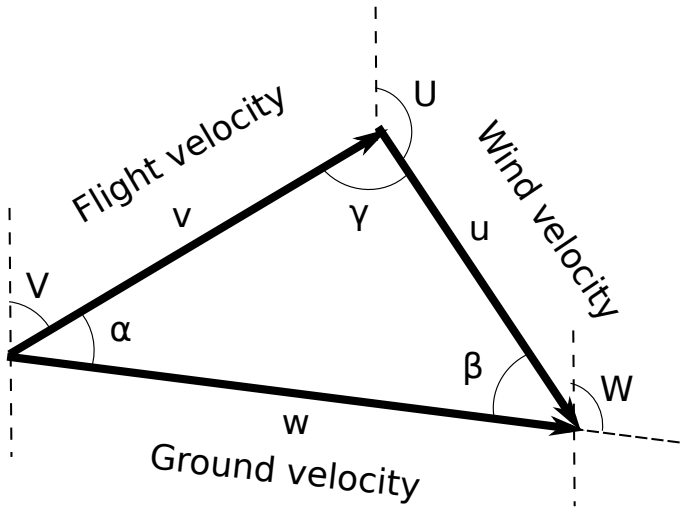


Figure 6.12: The wind triangle is defined by wind velocity, **u**, flight velocity, **v**, and resultant ground velocity, **w**. Magnitudes of the velocity vectors are given by u, v, and w respectively, and their directions with respect to the inertial frame are given by U, V, and W respectively.

designed for making meteorological recordings of wind fields in three dimensions. Position, velocity, and attitude of the M^{2}AV is estimated by fusing measurements from a GPS and an IMU via a Kalman filter [Kalman, 1960], and flight speed and air flow angles are measured with a five-hole-probe[9], which was calibrated for the M^{2}AV in a wind tunnel. For GPS-denied environments, Zachariah and Jansson [2011] describe a dead-reckoning approach, which utilises a Sigma-point Kalman filter [Julier and Uhlmann, 1997] to fuse feature points extracted from a forward looking camera, with measurements from an IMU and a 3-axis anemometer to estimate flight velocity and wind velocity. Often, for fixed-wing aircraft, air flow angles are assumed to be negligible. Under this assumption, flight velocity is parallel with the aircraft's fuselage and can be measured with a single anemometer and knowledge of the aircraft's 3-DOF attitude (as in Cho et al. [2006], Langelaan et al. [2011], Palanthandalam-Madapusi et al. [2011]). Barber et al. [2006] describe an approach that does not require knowledge of the aircraft's heading direction. They use a non-linear optimisation to solve for unknown components of the wind triangle from multiple measurements of the aircraft's ground speed and track (GPS) and airspeed (anemometer) as the aircraft manoeuvres.

In all of these approaches, aircraft flight speed is derived from approximations of

[9]An Pitot tube-like instrument for determining flow speed and angularity. See, for example, http://www.grc.nasa.gov/WWW/k-12/airplane/tunp5h.html.

Bernoulli's principle for incompressible fluids and measurements of dynamic pressure, taken using a Pitot tube or similar instrument. These instruments must be calibrated accurately for each aircraft over the range of flight conditions it is expected to experience and are less applicable to small-scale aircraft or rotor-craft because of low flight speeds and inconsistent flight direction (in the case of rotor-craft). Several approaches have been described previously for estimating wind field velocity that do not rely on an explicit measurement of aircraft airspeed. Neumann et al. [2012] estimate direction and speed of the wind field from a quadrotor platform using a GPS and an IMU. Ground velocity, $\mathbf{w}$, of the quadrotor is measured by GPS and flight velocity, $\mathbf{v}$, is approximated using the aircraft's attitude, measured by IMU, and a system identification performed offline. Wind velocity, $\mathbf{u}$, is then computed using the law of cosines and the wind triangle. Alternatively, Premerlani [2009] describes a method for estimating the airspeed of a manoeuvring fixed-wing aircraft using a GPS and an IMU only. Velocity of the wind field can then be computed simply from ground velocity (from GPS), as in Figure 6.12.

It is shown here that knowledge of an aircraft's ground track, W, and heading direction, V, is sufficient to estimate strength, u, and direction, U, of the wind field, if it is assumed that dynamics of the air mass remain relatively constant over short periods of time and are homogeneous over the area in which the aircraft flies. Ground track and heading direction of an aircraft can both be estimated accurately using visual information only. This approach therefore minimises the number of sensors that need to be installed on board an aircraft and calibrated – or, alternatively, increases redundancy by providing a completely stand-alone means of estimating wind field strength and direction. Vision-based systems can provide lower latency estimates of ground track and more precise estimates of heading direction than is obtainable from a GPS or an IMU respectively.

Estimating wind field strength and direction

In the presence of a slowly changing wind field, the drift angle, α, between ground track, W, and heading direction, V, of a small-scale fixed-wing aircraft is a function of heading direction, V. When the aircraft is headed directly against (or with) the wind, for example, $\alpha = 0$; and when the aircraft is headed directly perpendicular to the direction of the wind, α will be a maximum to the leeward side of the heading direction. By modelling how strength, u, and direction, U, of the wind field affects

ground track, W, of the aircraft as a function of heading direction, V, the model can be fitted to observations of V and W during flight to recover u and U.

Ground track, W, can be given as

$$\tan(W) = \frac{w_x}{w_y},$$

where w_x and w_y are components of ground velocity, $\mathbf{w}$, in easterly and northerly directions respectively. Now, from Figure 6.12,

$$\begin{aligned} w_x &= u_x + v_x = u \cdot \sin(U) + v \cdot \sin(V), \text{ and} \\ w_y &= u_y + v_y = u \cdot \cos(U) + v \cdot \cos(V). \end{aligned}$$

The relationship between ground track, W, and heading direction, V, can therefore be given as

$$\tan(W) = \frac{u \cdot \sin(U) + v \cdot \sin(V)}{u \cdot \cos(U) + v \cdot \cos(V)}. \tag{6.8}$$

It is assumed that wind velocity, $\mathbf{u}$, varies much more slowly than the aircraft's heading direction, V, or ground track, W, thus u and U are approximately constant for consecutive measurements of V and W. If it also assumed that the aircraft's airspeed, v, is relatively constant over multiple consecutive measurements, then

$$u = s_u \cdot v, \tag{6.9}$$

where wind strength, s_u, is simply a scalar parameter that defines wind speed, u, in terms of airspeed, v. Now, substituting Equation 6.9 into Equation 6.8,

$$\begin{aligned} \tan(W) &= \frac{s_u \cdot v \cdot \sin(U) + v \cdot \sin(V)}{s_u \cdot v \cdot \cos(U) + v \cdot \cos(V)} \\ &= \frac{s_u \cdot \sin(U) + \sin(V)}{s_u \cdot \cos(U) + \cos(V)}, \end{aligned} \tag{6.10}$$

where s_u and U are static with respect to V and W. Therefore, if measurements of ground track, W_i, are taken at n heading directions[10], V_i, then the unknown

[10]In principle, only $n = 2$ such measurements need to be taken.

parameters s_u and U can be computed by minimising the objective function

$$\sum_{i=0}^{n} \big[\arctan\big(\frac{s_u \cdot \sin(U) + \sin(V_i)}{s_u \cdot \cos(U) + \cos(V_i)}\big) - W_i \big]^2. \tag{6.11}$$

A non-linear optimisation routine [Johnson, 2009, Powell, 2009] was used to minimise Equation 6.11 and recover wind direction and its strength relative to airspeed of the aircraft.

Assumptions required for this determination are:

- The aircraft manoeuvres such that measurements of ground track, W_i, can be made at multiple heading directions, V_i.
- Wind speed, u, and direction, U, are relatively constant over the time period in which measurements of V and W are taken. It will be seen that a few tens of measurements taken over a period of a couple of seconds are sufficient for this algorithm to converge accurately on real-world flight data.
- Airspeed, v, of the aircraft is also relatively constant over the same time period. This assumption is not too restrictive because a constant throttle setting will produce a largely constant airspeed for an aircraft manoeuvring in the horizontal plane, even if airspeed is not explicitly measured or controlled.
- Air flow angles are negligible. Hence direction, V, of the flight vector, $\mathbf{v}$, is well approximated by the aircraft's heading direction.
- The vertical component of the wind field is neglected.

Simulation results

Due to the difficulty in obtaining ground-truth measurements of wind field strength and direction at flight altitudes (10m $\sim$ 100m), performance of the proposed approach was quantitatively analysed in simulation before being implemented on board a real aircraft. The primary goals of the simulated flight tests were to quantify accuracy of the proposed approach under various conditions, and to analyse convergence rate of the algorithm. The simulator that was used was based on the open-source *zSim* flight simulator [Hubert, 2008]. 168 simulated flights were performed in which an aircraft was commanded to hold an altitude of 100m AGL and a heading rate of approximately 4°/s, meaning that the aircraft yawed through 360° in approximately

90s, at which point each flight was terminated. Airspeed, v, of the aircraft was not controlled, although throttle was held constant so that the aircraft held an approximately constant airspeed of $34 \pm 1.5\text{ms}^{-1}$, where the error bound represents $\pm 2\sigma$.

Measurements of heading direction, V, and ground track, W, were computed from the exact position and orientation of the simulated aircraft at each time step. A wind field was generated for each flight that was constant throughout the flight and globally homogeneous. Wind fields were generated with an evenly distributed random speed, $2\text{ms}^{-1} \leq u \leq 60\text{ms}^{-1}$, and direction, $0° \leq U < 360°$. Note that the proposed approach is capable of estimating wind speeds of any magnitude (*i.e.* $0 \leq s_u < \infty$). For the simulated flight tests, however, wind speeds were not generated below 2ms^{-1} to avoid wind direction becoming undefined. For presentation and analysis, estimated wind strength, s_u, was converted to wind speed, u, using average aircraft airspeed (34ms^{-1}) during the simulations and Equation 6.9. A typical plot of the estimated wind direction and speed during one of the simulated flights is shown in Figure 6.13. It can be seen that estimated wind direction and strength stabilised accurately within approximately $\frac{1}{16}^{\text{th}}$ of a complete turn.

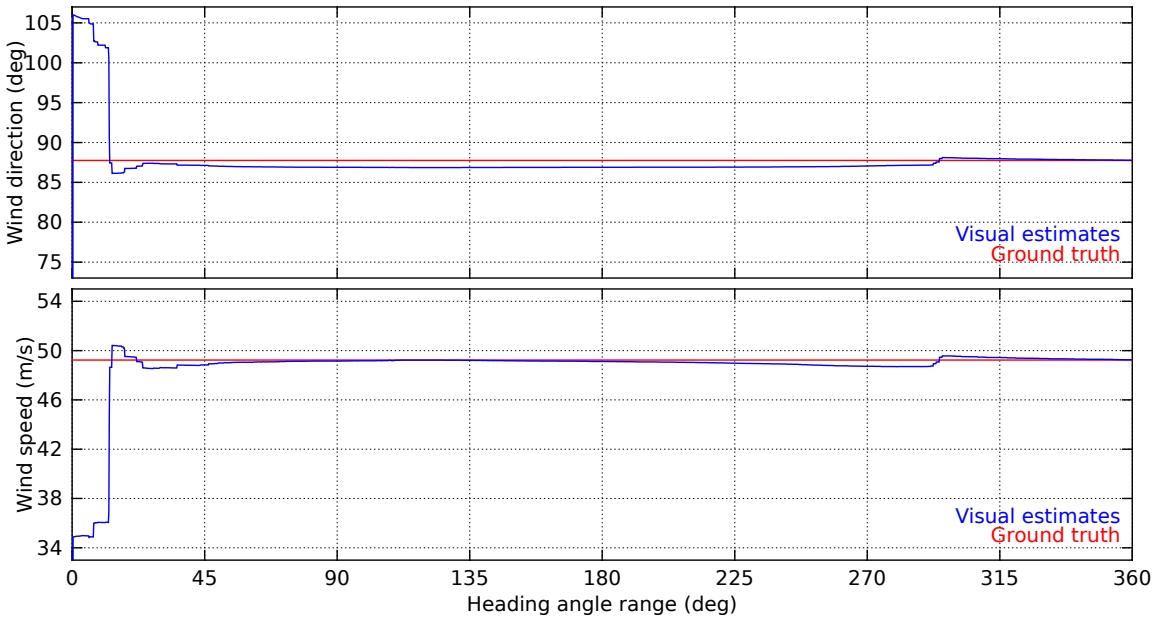
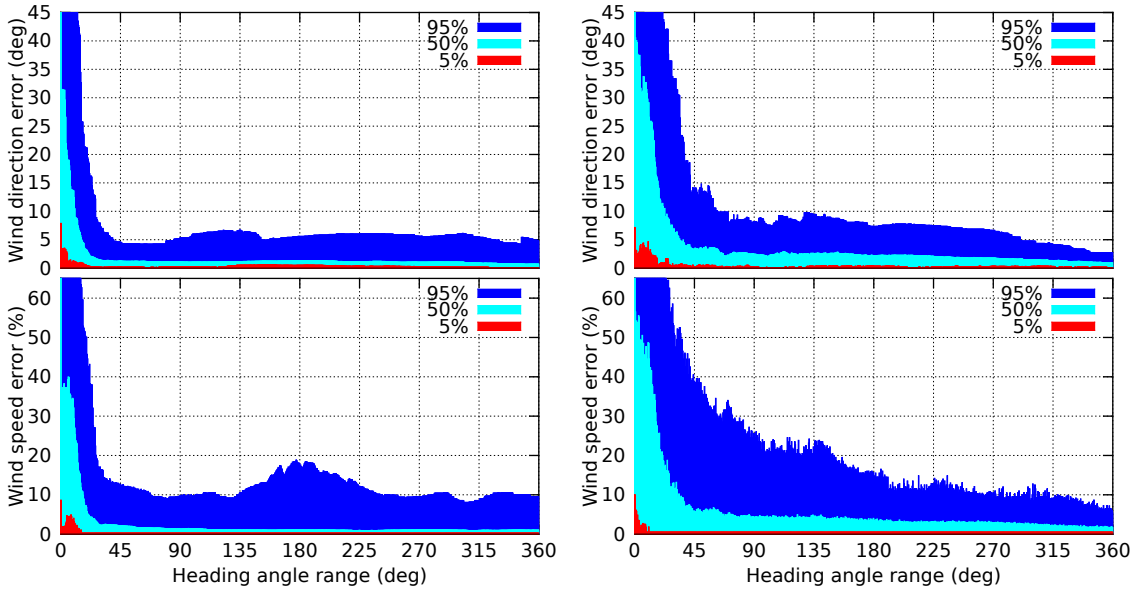


Figure 6.13: Evolution of estimated wind direction (top) and wind speed (bottom) during a simulated flight test. The simulated aircraft flies a curved path at a constant turning rate, hence elapsed flight time is equivalent to heading range through which the aircraft has turned, used here for display (x axis).

Time-plots of errors for estimated wind speed, u, and direction, U, from the 168 simulated flight tests were collated to obtain statistics on accuracy and convergence of the proposed algorithm. Results of this analysis are plotted in Figure 6.14a. Data shows that estimated wind direction approached ground truth quickly for the majority of simulated flights. Median flight time taken for estimated wind direction to converge to within 10° of the true value was approximately 2.7s. The important parameter here, however, is the angular range of heading directions over which measurements of V and W were made. Since heading rate was approximately 4°/s during the simulated flight tests, a convergence time of 2.7s corresponds to a heading angular range of 11°. This value represents the range of heading directions through which the aircraft had to rotate for estimated wind direction to converge to within 10° of the true value for 50% of the simulated flights. After the aircraft had rotated through an angular range of approximately 30°, estimated wind direction had converged to within 10° of the true value for 95% of the simulated flights. Similarly, estimated wind speed had converged to within 15% of the true value after the aircraft had rotated through approximately 15° of heading for 50% of the simulated flights, and approximately 32° for 95% of the simulated flights. Median wind direction error at the conclusion of each simulated flight (after the aircraft had yawed through a complete 360° turn) was computed to be 7.2×10^{-1}°, and median steady-state error of estimated wind speed was computed to be 1% of the true wind speed in each case, or $\sim 3.4 \times 10^{-1}\text{ms}^{-1}$ for a wind speed equal to average aircraft airspeed.

To analyse the proposed algorithm's robustness to measurement noise and fluctuations of the wind field, an additional 80 simulated flight tests were conducted. The second batch of simulated flight tests were constructed similarly to the first, except normally distributed noise ($\sigma = 2°$) was added to measurements of ground track, W, and heading direction, V, as well as generated wind direction, U. Results collated from the second batch of simulated flight tests are presented in Figure 6.14b. It can be seen that addition of Gaussian noise decreased the rate of convergence of the algorithm, however steady-state accuracy (residual error once the aircraft had rotated through 360°) was not significantly affected and both metrics still indicated that the algorithm converges stably. Statistics for all simulated flight tests are summarised in Tables 6.2 & 6.3. Data shows that 95% of the simulated noisy trials converged on estimates for wind direction and wind speed that differed from the true direction and speed by less than 2.7° and 6.0% respectively.



(a) Simulation with no noise.

(b) Simulation with added noise.

Figure 6.14: Convergence statistics for estimated wind direction (top) and wind speed (bottom), collated from (a) 168 simulated flight tests and (b) 80 simulated flight tests with added normally distributed measurement noise. Larger heading ranges correspond to more data points for minimisation of Equation 6.11, and hence generally more accurate estimates of the wind field's properties. The 5, 50, and 95 percentiles are shown in red, cyan, and blue respectively (*i.e.* 95% of the simulated trials produced wind speed or direction estimates with an error less than the (blue) value indicated for each range of heading angles). Wind speed errors are represented as a percentage of true wind speed.

Flight test results

Performance of the proposed approach for estimating wind direction and strength in an uncontrolled outdoor environment was analysed using vision captured during a flight test in which the aircraft was commanded to fly a circular pattern. The flight path allowed drift due to wind of the aircraft's position (Figure 6.15a) to be determined over a short period of time as well as variation of the aircraft's GPS-derived ground speed with heading direction (Figure 6.15b) to be observed, both of which provided corroborating measurements of average wind speed and direction during the flight test. The aircraft, vision system, and flight hardware were identical to those used for attitude flight testing, and are described in Section 6.4. Aircraft attitude (including heading direction) was computed using the approach described in Section 6.3. Wide-field optic flow was computed from omnidirectional imagery captured by the *iEye* vision system using a hierarchical block matching

Convergence criteria	Simulation		Sim + noise	
	50%	95%	50%	95%
Wind heading error $< 10°$	11°	30°	23°	61°
Wind speed error $< 15\%$	15°	32°	26°	187°

Table 6.2: Heading ranges required to satisfy convergence criteria for the simulated flight tests. Statistics correspond to data presented in Figure 6.14.

Estimated parameter	Simulation		Sim + noise	
	50%	95%	50%	95%
Wind heading	0.72°	7.1°	0.95°	2.7°
Wind speed	1.0%	7.9%	1.8%	6.0%

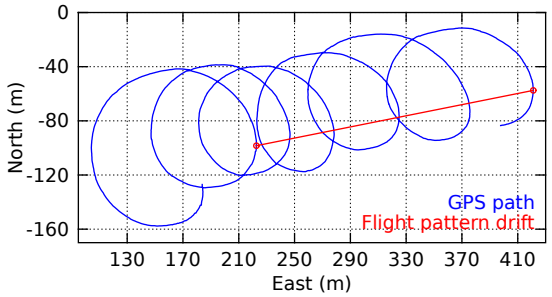
Table 6.3: Steady-state errors for the simulated flight tests. Statistics correspond to data presented in Figure 6.14.

approach in which input intensity images were band-pass filtered to remove high- and low-frequency components at a range of image scales and motion correspondences were found for an approximately regular grid of points following a coarse-to-fine procedure. The aircraft's direction of translation was then estimated from the computed optic flow field using an iterative 'hypothesise and test' approach. The algorithms for computing optic flow and translation direction were developed by Mr. Saul Thurrowgood and no further details are provided here, although there are various well known approaches for extracting an observer's direction of translation from monocular imagery based on epipolar constraints (*e.g.* the eight-point algorithm [Longuet-Higgins, 1981] or five-point algorithm [Nister, 2003]). To compute the aircraft's ground track, the estimated body-frame translation direction was first rotated into the inertial frame of reference using the estimated 3-DOF attitude of the aircraft, and then projected onto the horizontal plane.

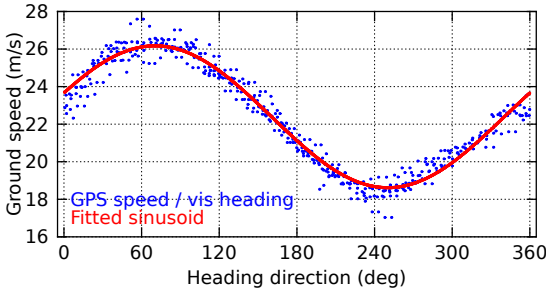
A segment of the wind estimation flight test is analysed in Figure 6.15 and is also shown in the accompanying video[11].

In the absence of wind, an aircraft's 2D translation direction (ground track) and heading direction would be expected to be identical. Visual estimates of the aircraft's

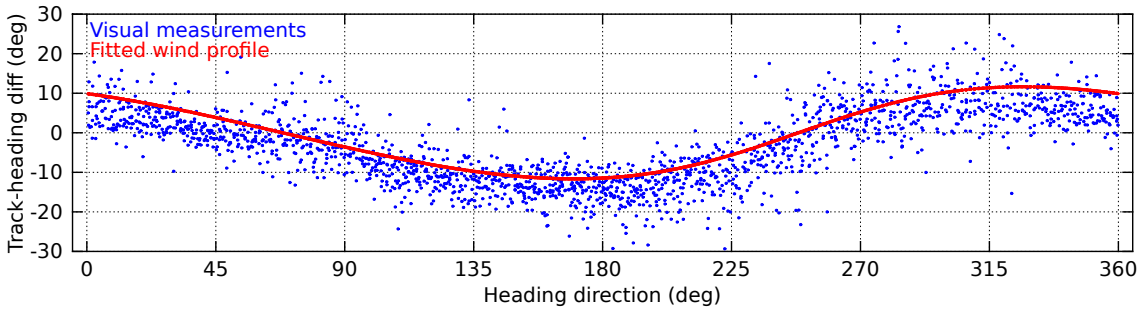
[11] Attachment AV08 (available at http://youtu.be/yHNib60qRYI). Note that this video shows a plot of estimated ground track, W, for each estimated heading direction, V, during the flight test. Time evolution of best fit of Equation 6.11 is also shown, along with corresponding estimates of wind strength, u, and direction, U.



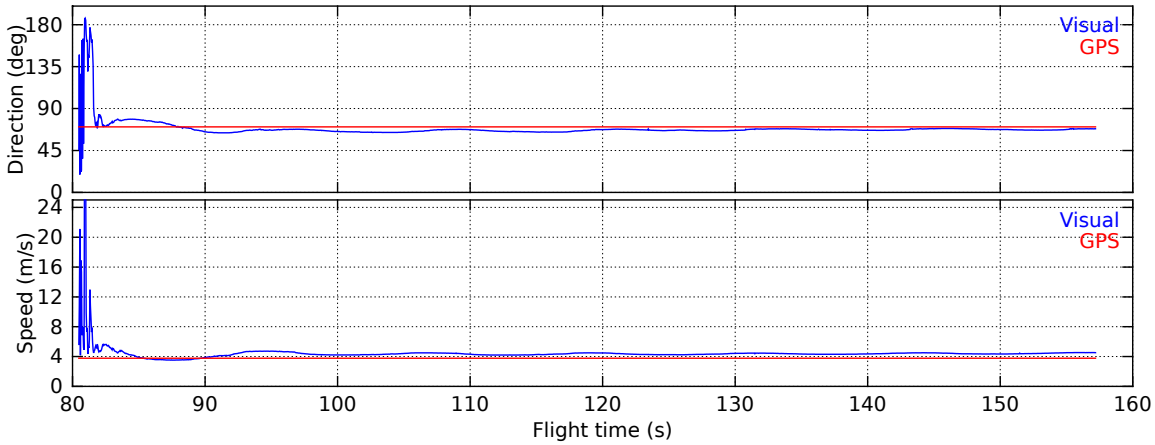
(a) GPS flight path showing drift.



(b) Variation of GPS ground speed with heading.



(c) Variation of visual track with heading direction due to wind, showing fitted wind profile.



(d) Evolution of visual wind speed and direction estimates with GPS comparisons.

Figure 6.15: Wind estimation flight test data: (a) GPS flight path (blue) during test, 2D drift (red) indicates wind speed 3.31ms^{-1} at compass bearing 78.4°; (b) variation of GPS ground speed with respect to visual heading direction (blue) during test, fitted sinusoidal (red) mean, amplitude, and phase indicates average ground speed, wind speed and wind direction of 22.4ms^{-1}, 3.78ms^{-1}, and 70.3° respectively; (c) difference between visual ground track and heading, plotted against heading direction (blue) during test, fitted wind profile (red) indicates average wind speed and direction of 4.52ms^{-1} and 68.3° respectively; and (d) evolution of visually estimated wind speed and direction (blue) during test shown with GPS-based measurements (red) from (b) for comparison. Visually estimated wind strength, s_u, was converted to metric wind speed, u, using the aircraft's assumed average airspeed during the test of 22.4ms^{-1}.

heading and ground track during the flight segment are plotted in Figure 6.15c (difference between measured ground track and heading direction has been plotted instead of raw ground track measurements for clarity). The regular variation between ground track and heading direction with respect to the orientation of the aircraft indicates that a wind was present in this case. Wind-induced variation between ground track and heading direction is predicted by Equation 6.10. Minimising the sum of squared errors (SSE) between ground track predicted by Equation 6.10 and ground track measured at each heading direction (using Equation 6.11) gave estimates of average wind strength and direction during the flight segment of $s_u = 2.02 \times 10^{-1}$ and $U = 68.3°$ respectively. To aid analysis, estimated wind strength, s_u, was converted to metric wind speed, u, using the assumed average airspeed of the aircraft during the flight segment of $v = 22.4\text{ms}^{-1}$ (computed from the average ground speed) and Equation 6.9. Visually estimated average wind speed can therefore be given as $u = 4.52\text{ms}^{-1}$.

GPS-derived position and speed of the aircraft during the flight test were used to obtain corroborating measurements of average wind speed and direction. It can be seen from Figure 6.15a that the aircraft's flight pattern drifted during the test. Magnitude and direction of drift (shown in Figure 6.15a) indicated an average wind speed of 3.31ms^{-1} at a compass bearing of 78.4°. A second, and probably more precise, measurement of wind during the test was obtained by observing variation in the GPS-derived ground speed of the aircraft with respect to heading direction (measured visually). Ground speed would be expected to vary sinusoidally with heading direction for an aircraft flying at a constant airspeed in the presence of wind. Figure 6.15b shows that a sinusoidal variation was indeed observed during the test, and a sine curve fitted to measured data gave an average wind speed of 3.78ms^{-1} at an average compass bearing of 70.3°.

Visually estimated wind speed and direction correlated closely with GPS-derived values used for comparison – the visual approach described here gave estimates of 4.52ms^{-1} and 68.3° for wind speed and direction respectively, compared to wind speed measurements of 3.31ms^{-1} and 3.78ms^{-1} as well as wind direction measurements of 78.4° and 70.3° for the two GPS-based approaches. The second GPS-based method was assumed to provide a better approximation of ground truth because it incorporated measurements made throughout the test. Average errors for the visual approach described here can therefore be given as approximately $7.4 \times 10^{-1}\text{ms}^{-1}$ (or $\sim 20\%$) for wind speed and 2.0° for wind direction, from data accumulated during

77.6s of circular flight at an approximately constant heading rate of 29.4°/s. In the context of landing automatically in windy conditions, estimating wind direction is of greater importance than estimating wind speed accurately. The precision metrics quoted here indicate, therefore, that the proposed system would be more than capable of enabling a small-scale UAV to independently determine the safest direction from which to approach a landing site. Landing into the wind maximises airspeed and minimises ground speed, leading to greater control and safety during landing, and hence tends to be the approach taken by human pilots as well as winged insects, such as honeybees [Liebsch and Srinivasan, 2010].

Evolution of visual wind speed and direction estimates computed in real time during the flight tests are plotted in Figure 6.15d. It can be seen that the wind estimation algorithm presented here converged rapidly onto stable estimates of wind speed and direction. Visually estimated wind speed converged to within 1ms^{-1} of the steady-state value in $\sim$ 2.1s and visually estimated wind direction converged to within 10° of the steady-state value within $\sim$ 1.8s. Given average heading rate during the flight test (29.4°/s), these convergence metrics correspond to heading ranges of approximately 63° and 51° required respectively for convergence of wind speed and wind direction estimates using real-world flight data. Convergence rates of the proposed algorithm indicate that properties of a varying wind field will be estimated accurately if wind dynamics vary only over time periods longer than a couple of seconds; wind estimates in gusty or wildly changing conditions will in effect be averaged over a time period equivalent to convergence time of the algorithm.

Limitations and conclusion

Simulated and real-world flight testing has demonstrated that the proposed algorithm is able to estimate direction and strength of a wind field accurately, using measurements of heading direction and ground track only. Aircraft heading and translation direction can both be estimated accurately and robustly using the *iEye* vision system. This approach therefore constitutes a purely visual method for estimating the properties of the local wind field. Dynamics of the wind field cannot be determined instantaneously using this approach, because Equation 6.10 is underdetermined. An iterative optimisation must be performed instead using multiple measurements of V and W (typically a few tens of measurements), which is a common characteristic of approaches where too few elements of the wind triangle are measured to enable an

instananeous estimation of wind velocity (*e.g.* Barber et al. [2006], Premerlani [2009]). This is a drawback of this method compared with approaches that explicitly measure the velocity (including air flow angles) of the wind, such as those described by van den Kroonenberg et al. [2010], Zachariah and Jansson [2011]. However, simplicity of the proposed approach and its applicability to any fixed-wing platform equipped with a camera make it a viable alternative for small-scale, low-cost aircraft.

Finally, it is noted that although the proposed algorithm estimates strength of the wind field with respect to the (unknown) airspeed of the aircraft, if an estimate of the aircraft's ground speed is also available then both airspeed and wind speed can be resolved explicitly.

6.6.2 Automatic landing using vision only

Honeybees are able to perform grazing landings on flat surfaces safely by maintaining the magnitude of the apparent optic flow generated by the surface at a constant value as they approach the surface [Srinivasan et al., 1996, 2000]. This ensures that their forward speed is in proportion to their height above the surface, and reaches zero at touchdown. Such a scheme enables the honeybee to land safely without explicitly measuring flight speed or distance to the surface. This strategy is obviously of interest for using vision to land a UAV automatically. It is not directly applicable to fixed-wing aircraft, however, as they require a non-zero forward speed at touchdown to avoid stalling. Regardless, modified versions of this strategy have been applied previously to demonstrate safe and autonomous landing of fixed-wing UAVs in indoor and outdoor environments [Beyeler et al., 2009a, Chahl et al., 2004, Green et al., 2003, 2004, Oh et al., 2004].

A strategy for landing a small-scale, fixed-wing UAV in an uncontrolled environment using vision only is described here in brief. This strategy uses the aircraft's 2-DOF attitude, measured via the approach described in this chapter; measurements of the apparent translation of the aircraft with respect to the ground (*i.e.* the translational component of the optic flow generated by the ground); and, at the final stages of touchdown, stereo vision-based measurements of distance to the ground. All measurements were performed with the *iEye* vision system. The algorithms for computing the raw optic flow field and extracting the translational component, as well as computing stereo disparity were developed primarily by Mr. Saul Thurrowgood; the algorithm for computing attitude (Section 6.3) and the closed-loop control system

were developed by the author; and overall landing strategy and guidance system were developed jointly.

As with the approaches described previously for landing a UAV automatically using optic flow, the aircraft's thrust is reduced in an open-loop fashion. As forward speed of the aircraft decreases, it must descend to maintain airspeed. Descent is controlled by moderating the aircraft's pitch angle such that the magnitude of the translational component of optic flow follows a predefined trajectory. The result is a strategy that is able to bring a fixed-wing UAV from a wide envelope of flight conditions to a repeatable airspeed and height immediately prior to touchdown. Touchdown is controlled via stereo vision-based height estimates. Throughout the manoeuvre, roll angle and pitch angle of the aircraft are controlled via the visual horizon estimation algorithm described in Section 6.3.1.

Automatic landing using visual input only was demonstrated (shown in the accompanying video[12].) using the procedure described above by coupling the *iEye* vision system with a PID-based control system, which was similar to the scheme described for the control of 3-DOF attitude in Section 6.4.3. In the modified control system used for the landing flight test, heading direction was not controlled; 'height' was controlled using the translational component of optic flow measured from the ground plane and – in the final stages of landing – stereo-based height estimates; and aircraft roll and pitch rates (for derivative control of the aircraft's roll and pitch angles) were obtained from the rotational component of optic flow measured from across the view sphere. In this way, the IMU- and GPS-derived signals depicted in Figure 6.10 were replaced by visual measurements, which allowed completely vision-based autonomous flight during terrain following and landing.

The aircraft's attitude and height during a landing manoeuvre are plotted in Figure 6.16. The period of autonomous control begins at ~ 169s and continues until the aircraft is at rest on the ground. The aircraft follows a steep linear descent during the first phase of landing, followed by a shallower descent controlled by the stereo vision-based height estimates until touchdown. Optic flow-based pseudo-height estimates diverge from stereo vision-based height estimates as the aircraft's airspeed decreases prior to touchdown. Loss of airspeed also reduces efficacy of the control surfaces. This is not explicitly accounted for here by the PID-based control system,

[12]Attachment AV09 (available at http://youtu.be/U9iy1B5QG-0). Note that in this video visually estimated attitude and the translational component of the aircraft's egomotion have been drawn on the on-board imagery.

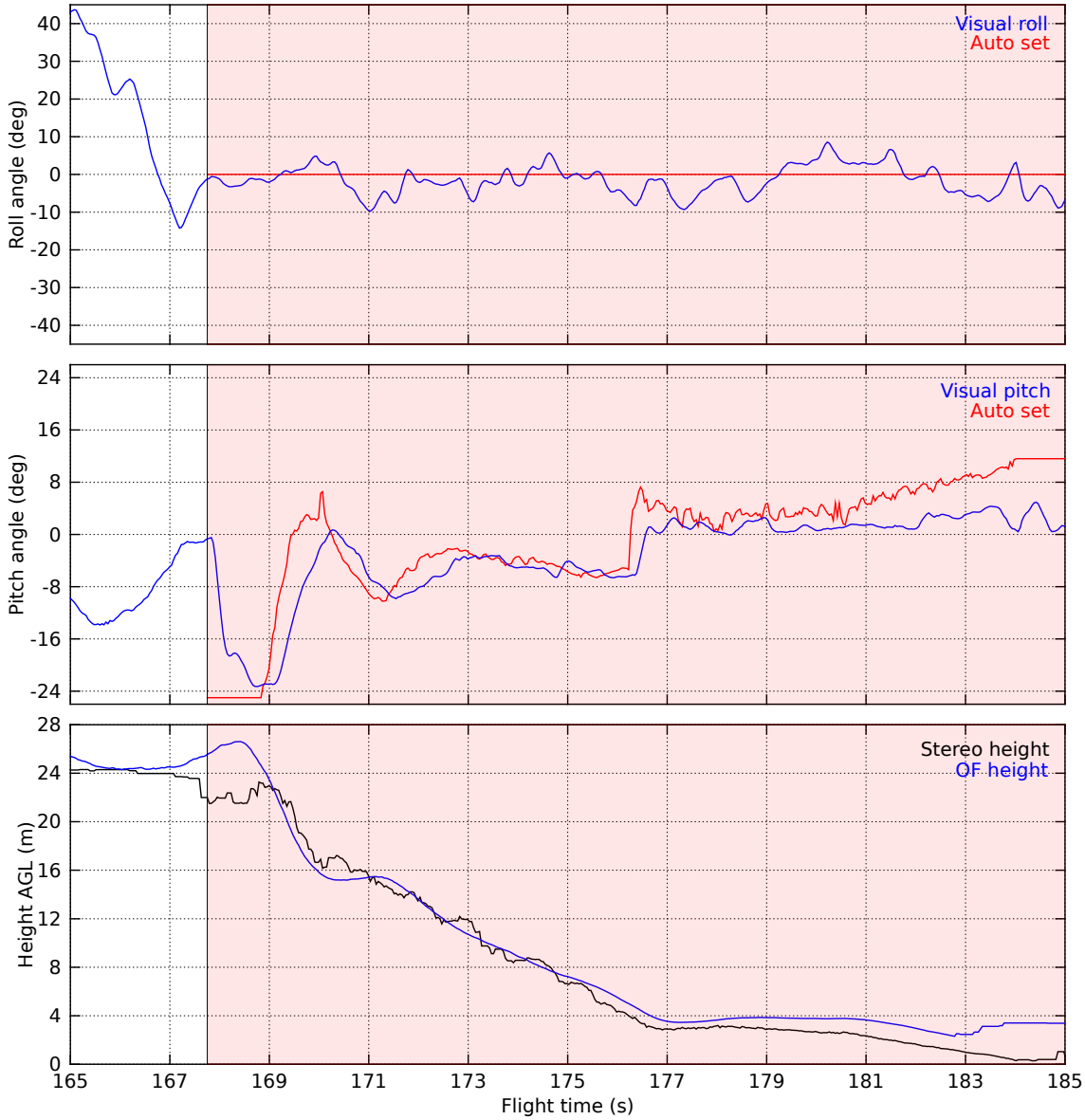


Figure 6.16: Roll angle (top), pitch angle (centre), and height (bottom) of the aircraft during an automatic landing. Attitude of the aircraft is estimated via the visual approach described in this chapter and controlled according to the landing procedure described in Section 6.6.2 during autonomous control (shaded red). Height of the aircraft is estimated from the translational component of measured optic flow (blue) using an assumed flight speed, and from stereo vision-based measurements of the range to the ground plane (black). Set-points for roll angle and pitch angle during autonomous control are also shown (red).

hence the guidance system is not able to achieve the desired pitch angle during touchdown, as can be seen from the discrepancy between commanded and measured pitch angle in Figure 6.16. A method for automatically estimating trim positions and control surface efficacies during flight is described in Section 8.4. Here, however, the mis-match between commanded and actual pitch angle is not too undesirable, as it avoids the aircraft entering a stall whilst maintaining a flaring manoeuvre.

This automatic landing strategy was (and continues to be) used during all following flight tests to recover the aircraft. Observations from many landings indicate that this scheme allows a fixed-wing UAV to land automatically with greater repeatability and safety than a ground-based human controller, even in gusty conditions.

6.6.3 Autonomous aerobatics using visual attitude

The first attitude flight test (Figure 6.8) demonstrated that the visual attitude estimation scheme was able to track the aircraft's orientation robustly as the aircraft flew vertical 360° loops. As a further test of the robustness of the visual guidance system, a series of flight tests were conducted to examine whether aerobatic flight manoeuvres could be executed autonomously. In one such flight test (shown in the accompanying video[13]) the aircraft was commanded to perform Immelmann manoeuvres[14] autonomously. This research was performed by the author based on prior work described by Thurrowgood et al. [2010].

To perform each Immelmann manoeuvre, the aircraft was commanded to maintain a level attitude at an altitude of approximately 30m AGL for a short period; an open-loop elevator command was then applied to commence the half-loop, whilst the roll angle of the aircraft was maintained such that the wings remained perpendicular to the plane of the loop; once the aircraft has passed through the top of the loop and was descending again, but upside down, roll control was reverted to normal, which caused the aircraft to roll 180°; finally, once upright again, the aircraft was commanded to descend to approximately 30m AGL and maintain wings-level to complete the manoeuvre. The control system was similar to the PID-based scheme described for control of 3-DOF attitude in Section 6.4.3 except for the roll angle and elevator commands issued during the half-loop phase. Heading direction was not controlled

[13] Attachment AV10 (available at http://youtu.be/yGhDWrRr6QY).

[14] An aerobatic turn in which the aircraft performs a half-loop followed by a half-roll, after which the aircraft is facing 180° to the initial heading direction.

in this flight test.

The 3-DOF attitude of the aircraft during a segment of the flight test in which the aircraft performed two Immelmann turns is plotted in Figure 6.17. It can be seen that the aircraft performed the manoeuvres effectively, and heading direction changed by approximately 180° during each manoeuvre. Ideally, roll angle should jump instantaneously from 0° to ±180° as the aircraft passes through vertical. However, the aircraft showed a tendency to roll to the right slightly when it 'pulled up' (possibly as a result of being put through numerous aerobatic manoeuvres, for which the airframe was not designed). Roll angle thus varied continuously between upright and almost exactly upside down during the manoeuvres. It can be seen from Figure 6.17 that 3-DOF attitude, in particular pitch angle and heading direction, reported by the IMU on board the aircraft varied significantly from the visual estimates during the flight test. The findings presented in Section 6.4 indicate that the visual estimates are likely to more closely represent true attitude of the aircraft.

6.7 Conclusion

This chapter has described a new vision-based method for automatically obtaining 3-DOF attitude of an aircraft. A very wide-angle view of the environment, including the horizon, is captured and a classifier is used to segment the scene into fuzzy sky and ground regions, using the spectral and intensity properties of the input pixels. The classifier is updated continuously using an online reinforcement strategy. It is therefore able to adapt to changing environmental conditions and does not require training offline prior to use. The aircraft's 2-DOF attitude is estimated by comparing the classified images with a database of reference image classifications. A novel matching technique is described, which reduces both the complexity of the matching process and the data required to store the database.

Using estimated 2-DOF attitude, a stabilised panoramic image of the horizon profile and sky is generated. This image is then used in a 'visual compass' to determine the aircraft's heading direction, relative to a reference horizon image. An appearance-based matching procedure allows heading direction to be estimated visually, even when there are no prominent features in the input image. Furthermore, by coupling the visual heading direction estimation with the visual horizon attitude estimation, full 3-DOF attitude is able to be determined robustly, irrespective of the orientation of the aircraft. The proposed approach can accurately estimate the 3-DOF attitude of

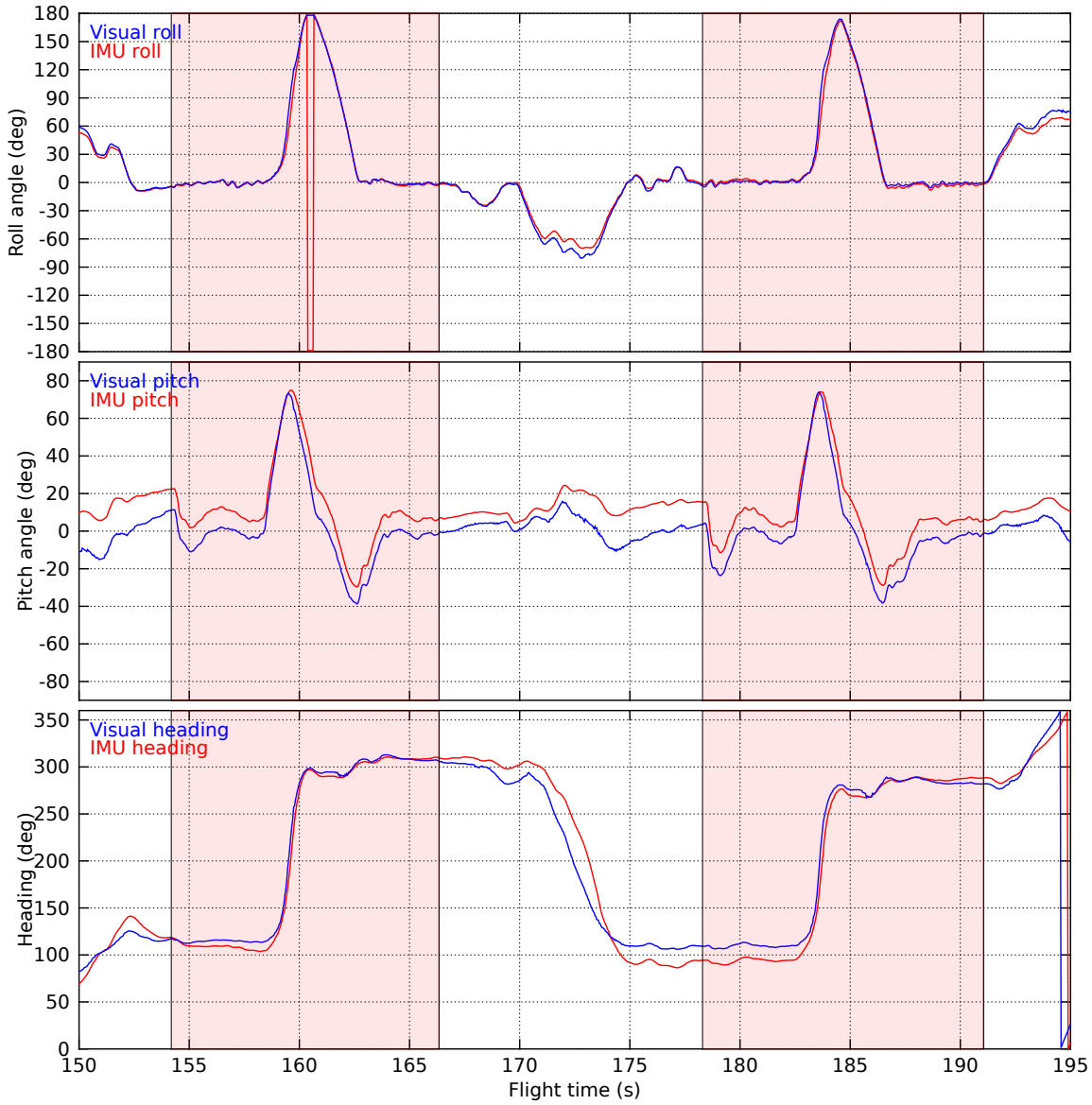


Figure 6.17: Roll angle (top), pitch angle (centre), and heading direction (bottom) of the aircraft as estimated by the vision system (blue) and reported by an IMU on board the aircraft (red) during a segment of an autonomous, aerobatic flight test. During this segment of flight the aircraft performed two Immelmann manoeuvres autonomously (shaded red) using the visual attitude estimate. Details in text.

the aircraft despite arbitrarily large changes in attitude between frames, once trained.

Results from several flight tests have demonstrated the ability of this vision-based approach to outperform the attitude estimates obtained from the IMU used in Section 6.4. Furthermore, it was shown in Section 6.5 that the accuracy of this approach is equal to, or better than, that of other previously published vision-based approaches, whilst being computationaly efficient and faster to execute. This system's robustness to varying environmental conditions and aircraft motions, and ability to be integrated with any calibrated wide-FOV colour vision system therefore makes it suitable for any UAV application where the aircraft's attitude must be known precisely.

This work has demonstrated that the robustness afforded by an omnidirectional FOV and the computational efficiency gained through the design of simple sensing algorithms can result in a guidance system that is superior to the current state of the art. An additional advantage of this approach is that it can be mated easily with other visual sensing and guidance strategies in order to demonstrate complex behaviours. The applicability of this visual guidance scheme to various real-world tasks was demonstrated with several brief investigations that combined this approach with measurements of optic flow: techniques for landing a fixed-wing UAV in an uncontrolled environment and for estimating the strength and direction of the local wind field were described and demonstrated using visual input only.

7

Visual Target Detection and Tracking

7.1 Summary

Autonomous vision-based interception of a moving ground-based vehicle using a small-scale UAV is a complex task that requires multiple sensing and guidance algorithms to operate simultaneously and in real time on board the UAV. Visual detection and tracking of independently moving objects are prerequisites for interception. The tracking and guidance systems described in this chapter rely on wide FOV images captured by the *iEye* vision system (described in Chapter 5), and the aircraft's 3-DOF attitude, estimated via the approach described in Chapter 6.

To track a target vehicle or object, an observer must be able to identify or recognise the target from other objects within a scene. This can be accomplished either by apparent motion of the target with respect to the expected motion of objects within the scene, or by its appearance. These approaches are discussed further in Section 7.2. A computationally lightweight appearance-based tracker is described in Section 7.3 that forms the basis for the interception scheme described in Chapter 8. The tracker operates on RGB properties of the target and its background. Flight test results demonstrate that the tracking algorithm is effective at localising natural and man-made targets within the environment despite motion of the aircraft or target, changing light conditions, and occlusions.

7.2 Detecting targets visually

Visual tracking can be initiated either by detecting independent motion within the scene or by recognising the shape or appearance of a target directly.

7.2.1 Detecting independent motion

Recognising the presence of moving objects in a scene is an important function of visual guidance systems, not only for the purpose of interacting with those objects, but also for discounting those parts of the scene from visual estimation of egomotion. As such, it is an area that has received much attention within the research field of computer vision.

Structure from motion

If the observing camera is stationary and the illumination is constant, the problem of detecting motion can be reduced to simply comparing successive frames for differences in image intensity or colour [Thompson and Pong, 1990]. For a camera undergoing self-motion, however, the problem is considerably more complex, because even the image of a stationary environment will be in motion. In a stationary environment, translational motion by the camera will induce deformations in the image that depend on ranges to objects in the environment. This deformation is predictable, however, as it is the same deformation that is measured by Equation 2.1. If 3D structure of the environment, rotational and translational motion of the observer, and the imaging properties of the camera are known, then the new projection of the environment onto the image plane can be estimated. Any gross differences between the captured and predicted images (either intensity images or range images) can be attributed to locomoting objects (as in Agrawal et al. [2005]).

In practice, typically, neither 3D structure of the environment nor motion of the camera is known *a priori*, so they must be estimated simultaneously. The classical approach, as discussed in Section 2.2.3, is to solve the SFM problem. This problem is complex but can be simplifed somewhat by incorporating an independent measure of the 3D structure of the scene, thus reducing the problem to obtaining *motion from structure*. Stereo cameras have been used for this purpose [Agrawal et al., 2005, Badino et al., 2006, Demirdjian and Horaud, 2000, Franke et al., 2005, 2008, Maimone et al., 2007, Mallet et al., 2000, Mandelbaum et al., 1999, Nister et al., 2004, Olson et al., 2003, Rabe et al., 2007, Talukder and Mathies, 2004]. Conversely, an independent measure of the observer's motion can reduce the problem to obtaining *structure from motion*. Once both the camera's egomotion and the 3D structure of the environment are measured or computed, the new projection of the environment onto the image plane can be estimated, as described above. A conundrum arises, however,

when dealing with scenes containing independently moving objects. Objects in the scene undergoing self-motion must be discounted when estimating the egomotion of the camera, otherwise their motion will contaminate the egomotion estimate. But, to identify those objects that are moving, the egomotion of the camera must first be known!

This conundrum is typically avoided by estimating egomotion of the camera robustly and removing outliers from the motion model. However, outliers can be due to measurement errors, or they can be due to the presence of locomoting objects. To detect self-moving objects reliably in ground-based urban environments, Badino et al. [2006], Demirdjian and Horaud [2000], Franke et al. [2005], Rabe et al. [2007], Talukder and Mathies [2004] cluster SFM outliers. This strategy is sensitive to errors in the estimates of scene structure and camera egomotion and can be unreliable for scenes where depth variations are not significant [Adiv, 1989, Irani and Anandan, 1998]. Furthermore, implementations are typically computationally intensive and hence have not been demonstrated in real time on board aerial platforms.

Epipolar constraints

Apparent motion of a scene – that is due to purely translational motion of the camera – radiates from the FOE, or epipole, which corresponds to the line of sight coincident with translation direction. When projected on the surface of the view sphere, therefore, the direction of motion measured from each optic flow vector must align with the epipolar lines, which travel from the FOE to the FOC. This is known as the motion epipolar constraint. Optic flow vectors that violate this contraint must be due to independently moving objects [Jain, 1984].

Strategies for detecting independently moving objects based on the epipolar constraint depend critically on the ability to estimate location of the epipole, or equivalently direction of translation of the observer, accurately. Frazier and Nevatia [1992], Jain [1984] detect locomotion but assume that the camera translates with a known or purely forward self-motion, *i.e.* location of the epipole is known *a priori*. Salgian et al. [2006] determine translational motion of the camera using a stereo SFM approach. Methods for dynamically estimating location of the epipole in image space have also been described [Burger and Bhanu, 1989, Heeger and Jepson, 1992, Verri and Trucco, 1999], but are susceptible to failure in the presence of independent or rotational motion or small translations due to the restricted FOV

of the systems described. Furthermore, it is true that optic flow vectors violating the motion epipolar constraint must be due to independent motion. The converse, however, is not necessarily true. It is possible that independently moving objects can coincidentally generate image motion compatible with this constraint – *e.g.* if motion of the independently moving object is parallel or anti-parallel to direction of translation of the camera.

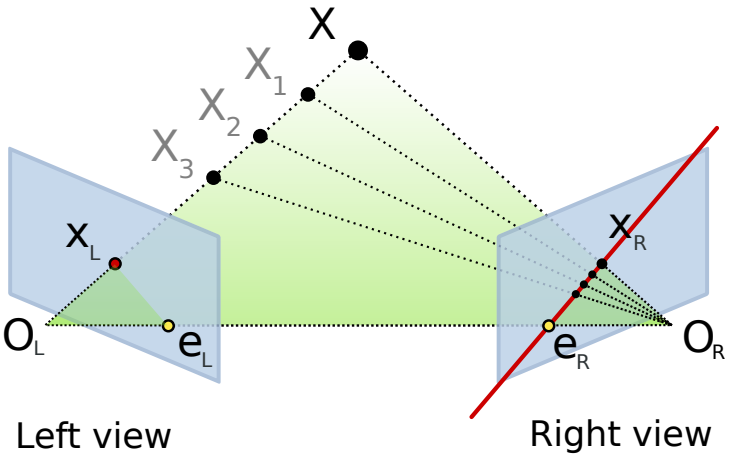


Figure 7.1: Projection, X_{L}, of a feature, X, in the image plane of the first camera constrains the projection, X_{R}, of the feature in the image plane of the second camera to lie on an epipolar line, $e_{\mathrm{R}}X_{\mathrm{R}}$. Image credit: http://commons.wikimedia.org/wiki/User:Norro.

The trifocal constraint is an extension of the epipolar constraint to three distinct views, or cameras. When using two views (as in the epipolar constraint), the projection of a feature in the image plane of the first camera constrains the projection of the feature in the image plane of the second camera to lie on an epipolar line (refer to Figure 7.1), whereas by using three views, the projection of the feature in the image plane of the third camera is constrained to lie at the intersection of two epipolar lines – *i.e.* a point. Algorithms utilising the trifocal constraint [Klappstein et al., 2009, Torr et al., 1995] are therefore able to resolve a greater range of independent motions than algorithms utilising the epipolar constraint, at the cost of increased susceptibility to degeneracy and instability, and the delayed identification of independently moving objects. Similar multiframe approaches are utilised by Irani and Anandan [1998], Kang et al. [2005] for the segmentation of independent motion in scenes containing sparse 3D structure.

Image registration

For scenes that do not contain significant depth variations, an effective approach for accounting for camera-induced image motion is to model image motion in terms of a global 2D parametric transform, or homography. Frietsch et al. [2007], Salgian et al. [2006] track features in sequences of aerial imagery to compute the dominant affine homography between successive frames. Warping the previous image into the reference frame of the current image allows them to detect locomoting objects as non-matching image regions. Similarly, Jung and Sukhatme [2004] use a bilinear projective transformation model to compensate for camera egomotion, then use frame differencing to detect locomotion from ground-based and aerial platforms.

Image-based approaches have the benefits that they require just a single camera, are fast to compute, and are theoretically able to detect image regions undergoing independent motion as small as one pixel, which allows for the detection of locomoting objects that are small or far away. Image-based approaches are thus well suited for implementation on board aerial platforms. Egomotion of the camera, however, may only be approximated by a global 2D parametric transformation if camera motion is pure rotation, if the scene is in fact predominantly planar, or if it can be approximated as being predominantly planar (as is the case for high altitude aerial imagery). If the above conditions cannot be met, 3D structure of the scene will produce significant parallax effects that will obfuscate the detection of independently moving objects.

There exist, then, two disparate and incompatible approaches for detecting locomoting objects. This presents a problem, for instance, in the case where an aircraft detects a moving target at range but must manoeuvre closer to intercept or dock with the target. To address this issue, Irani and Anandan [1998] propose a stratification of the problem of detecting independently moving objects into scenarios with increasing complexity. They compensate for camera egomotion by first registering successive frames using the dominant 2D homography. Residual image motion can be from two sources: 3D parallax, produced by purely translational motion of the camera and deviations in scene structure from a planar surface, or independently moving objects. Further constraints must be applied to distinguish between the two sources of residual motion.

Biological systems

Many flying insects have seemingly mastered the ability to detect and pursue or intercept their prey or other conspecifics, despite their own motion and possessing simple visual systems with limited processing power. Land and Collett [1974] report that chase sequences in some species of fly can be initiated by throwing small stones or painted peas near patrolling flights of males. Furthermore, flies are less inclined to pursue larger targets [Boeddeker et al., 2003] and many insects have been found to possess specialised zones within their eyes that are acutely sensitive to motion of small objects [Collett and King, 1975, Gilbert and Strausfeld, 1991, Olberg, 1981, Olberg et al., 2000]. For insects, therefore, detection of independently moving prey or a likely mate is probably achieved using a combination of motion cues and *a priori* information on approximate size, shape, and appearance of the target. Once a moving target has been detected initially, it can be tracked based on its appearance alone – assuming that the target is not camouflaged against its background. In fact, biological systems probably employ two separate systems: one that detects a potential target and one that drives the insect towards the target using a retinal position feedback control loop.

7.2.2 Appearance-based target detection

Detecting independent motion robustly is a difficult task and many of the implementations discussed in Section 7.2.1 are valid under certain conditions only. Perhaps a more robust approach would be to seed a tracking algorithm with the output of a motion detection algorithm, such that the target is detected initially from its independent motion, but then subsequently tracked primarily via other methods. Alternative methods for target tracking that rely on visual appearance of the target include (but are not limited to) algorithms based on supervised classifiers (*e.g.* Freund and Schapire [1995]); image-space kernels or templates (*e.g.* Kalal et al. [2009]); *a priori* models of the target's 3D structure (*e.g.* Song and Nevatia [2005]); image points or more complex features (*e.g.* Lowe [2004]); or simply segmentation of foreground objects from their background via colours (*e.g.* Cheng [1995]), contours (*e.g.* Caselles et al. [1997]), or background modelling (*e.g.* Mahadevan and Vasconcelos [2008]). A thorough review of the multitude of appearance-based techniques for target tracking is given in Yilmaz et al. [2006]. A brief overview of some relevant implementations is as follows:

Colour, in particular, is one of the most widely used features for appearance-based tracking. Many implementations are based on colour histograms of the target and utilise the 'mean shift algorithm' [Cheng, 1995] or similar adaptions [Bradski, 1998] to localise the tracker. A primary advantage of appearance-based tracking schemes is that they typically support implementations that are computationally efficient. Comaniciu et al. [2000, 2003] use the mean shift algorithm to demonstrate real-time tracking of non-rigid objects from a moving camera. A simliar approach is also implemented by Teuliere et al. [2011] to demonstrate autonomous tracking and chasing of a moving ground-based target by a quadrotor. In their implementation, however, image processing and tracking is performed on a ground-based workstation.

Kalal et al. [2009, 2010] describe a robust, kernel-based tracker that constructs a reliable object detector online, even in highly dynamic environments. Their implementation executes in real time and highlights the benefit of maintaining positive and negative target exemplars for robust tracking.

Other implementations of appearance-based trackers that have been applied to the problem of tracking a target from a small-scale UAV include Proctor et al. [2006], who describe and demonstrate a classifier-based approach that enables a small glider to track and fly through a stationary window; McGee et al. [2005], who describe a vision-based obstacle detection algorithm that identifies and tracks obstacles as non-sky regions lying above the horizon, where sky and non-sky regions are classified using a SVM that operates on the YUV colour space of input pixels; and Johnson et al. [2007], who use a contour-based approach to segment a lead aircraft using vision from on board a chase aircraft, and compute range and relative motion of the leader with respect to the chaser. In each case, however, image processing and control was either computed on a ground-based workstation, or was performed on board the aircraft but was not fast enough to permit closed-loop flight.

7.3 A new appearance-based tracker

A system for recognising and/or detecting a ground-based target is required to accomplish the goal for Part II – to demonstrate the visual tracking and interception of a ground-based vehicle using a small-scale UAV. Tracking algorithms developed previously by others have typically been optimised for performance, with speed of execution a secondary consideration. Here, minimal execution time is crucial for this tracker to be integrated with the other components of the vision system on

board the aircraft. The tracker should also be robust to uncontrollable and changing environmental conditions, but may leverage properties of the rural environment in which the tracker will operate. To satisfy these criteria, a new but simple tracking algorithm was developed. It is obvious from the discussion above that tracking schemes based on appearance of the target are more suited to aerial applications than schemes based on motion of the target due to the reduced computational and sensory requirements of appearance-based schemes. Thus the proposed algorithm is a computationally lightweight appearance-based tracker. The proposed algorithm shares some similarities with the 'CAMSHIFT' algorithm of Bradski [1998], in that colour is used to distinguish the target from the background and the target is localised and oriented using the estimated moments of segmented distribution. The proposed strategy, however, makes several simplifications and is tightly coupled with the visual attitude estimation process to enable rudimentary but efficient tracking of targets within a natural setting.

7.3.1 A colour-based blob tracker

Fundamentally, the tracking algorithm developed here attempts to localise a small contiguous region, or blob, whose colour is approximately uniform and distinguishable from the background. Inputs to the tracking algorithm are two raw fisheye images captured simultaneously by the *iEye* vision system, described in Chapter 5, and absolute 3-DOF attitude of the aircraft, which is estimated visually according to the procedure described in Chapter 6. Properties used by the algorithm to track the blob are:

- A unit view vector (target vector), $\hat{\mathbf{v}}_t$, which describes direction from the aircraft to the centre of the blob in the inertial coordinate frame.
- A zoom parameter (target zoom) that describes FOV of the tracking window.
- Two three-element vectors that describe average RGB values of blob (target) and non-blob (background) pixels in the tracking window.

Operation of the tracking algorithm is depicted in Figure 7.2. A region of interest (ROI), or tracking window, centred on the target vector is extracted from raw input images in order to localise the blob. At each frame, the target vector is updated according to the tracked position of the centre of the blob and the target zoom is

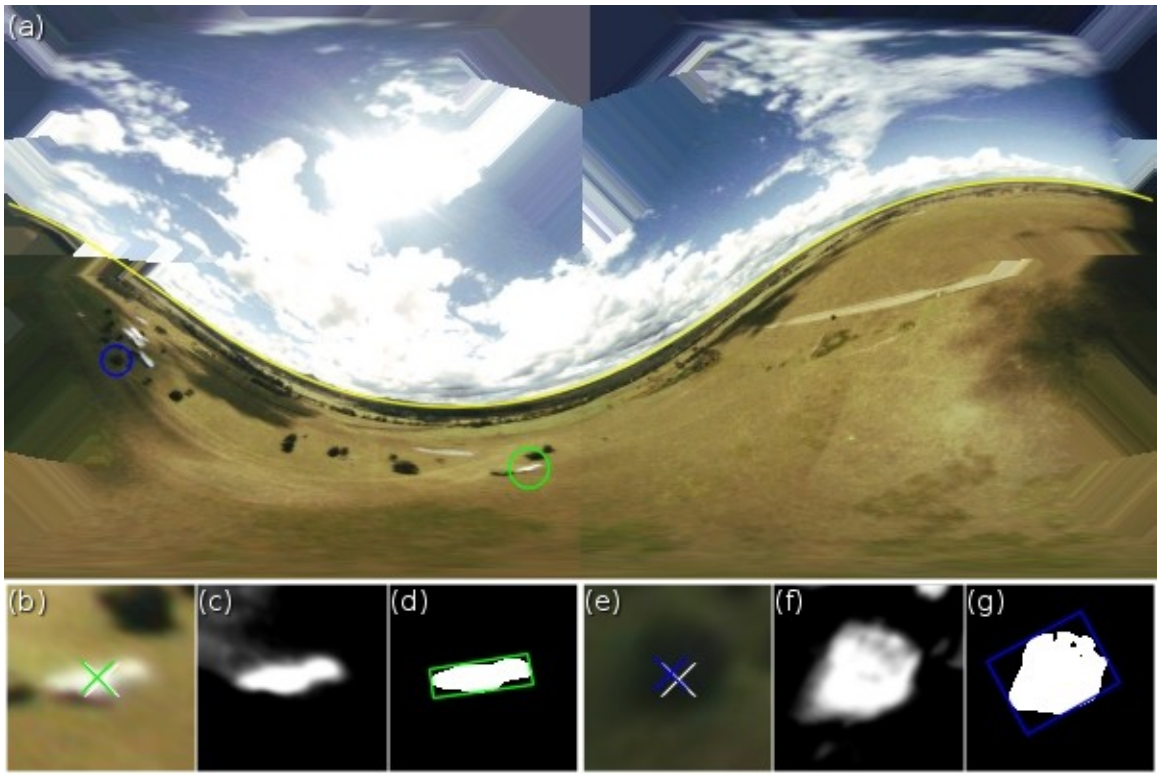


Figure 7.2: Operation of the appearance-based tracker showing a stitched image from the *iEye* vision system (a) overlaid with the estimated horizon (yellow line) and two tracked targets (coloured circles); and intermediate tracking steps for each target – tracking window (b,e) and its centre (white cross) with the computed target centre (coloured cross), greyscale transformed tracking window (c,f), and segmented target blob (d,g) with computed bounding box.

adjusted such that the blob fits entirely within the tracking window. RGB properties of both the tracked blob and the background pixels within the tracking window are also updated at each frame so that the blob can be re-localised within the subsequent frame. Complete operation of the algorithm is described by five steps performed in turn on each input frame:

1. *Propagate/Initialise* – A camera view vector, $\hat{\mathbf{v}}_t^{\text{cam},k+1}$, pointing to the predicted centre of the target blob in the current, $k^{\text{th}} + 1$, frame (white cross in Figure 7.2b,e), is obtained by transforming the previous localised inertial target vector, $\hat{\mathbf{v}}_t^{k+}$, into the current camera coordinate frame, using the aircraft's current 3-DOF attitude. The target zoom and average RGB of the blob and background pixels surrounding the blob are given by the sames values measured in the previous frame. Alternatively, the target vector, target zoom,

and target/background RGBs can be initialised in this step either manually or automatically according to some heuristic for re-initialising the blob.

2. *Transform* – A square ROI centred on the current target vector is extracted from raw input fisheye images using the current target zoom. This is the tracking window (Figure 7.2b,e). A transform is applied to pixels within the tracking window to enhance appearance of the target blob (Figure 7.2c,f). The transform maps the RGB values of all pixels within the tracking window to greyscale values in the range $[0, 255]$. The transform operates on pseudo-distance in RGB-space between RGB of the pixels within the tracking window and RGB averages for the target blob and background pixels. The transform can be visualised as mapping the RGB of the pixels within the tracking window to a 3D line through RGB-space that runs from the background RGB point to the target RGB point, where the value output by the transform, $[0, 255]$, corresponds to the distance along the line to which a particular pixel maps.

3. *Segment* – A blurring operation is performed on the greyscale tracking image, after which highest scoring pixel/s are labelled as *target* and remaining pixels are labelled as *background*. The target blob is segmented via an iterative region-growing operation as follows:

   - Mean greyscale values for *target* and *background* pixels are computed for the current iteration.
   - For each *target* pixel, neighbouring pixels are also labelled as *target* if their greyscale value is at least three times closer to the *target* mean than the *background* mean – *i.e.* $3 \times |g_i - \bar{g}_t| \leq |g_i - \bar{g}_b|$, where g_i is the greyscale value of the i^{th} pixel in the transformed tracking window, and $\bar{g}_t$ and $\bar{g}_b$ are mean greyscale values for *target* and *background* regions respectively. This ratio was empirically determined to provide a good compromise between segmentation selectivity and robustness to changing appearance of the target.

   The region-growing steps are iterated until no new pixels are added to the *target* class. The output from this step is a contiguous region within the tracking window that is labelled as *target* (Figure 7.2d,g).

4. *Localise* – The target blob is localised within the tracking window by fitting

an oriented bounding box to pixels labelled as *target* (coloured box in Figure 7.2d,g). The bounding box is fitted using the Geometric Tools [2010] mathematics library.

5. *Update* – The current localised camera view vector to the target, $\hat{\mathbf{v}}_t^{\text{cam},k+1+}$, is extracted from the computed centre of the *target* bounding box (coloured cross in Figure 7.2b,e) using the calibration model for the *iEye* vision system. Then, the localised camera target view vector is transformed to the inertial coordinate frame, $\hat{\mathbf{v}}_t^{k+1+}$, using the aircraft's current 3-DOF attitude. Target and background RGBs are updated by finding the mean RGB values of the pixels lying inside and outside the bounding box respectively. Finally, the target zoom is updated via a feedback process that maintains the width of the (square) tracking window to be twice the longest dimension of the bounding box. The exact value of the feedback gain for the target zoom process is not critical and it permits a trade-off between zoom dynamicity and stability of the tracking window.

The result of the tracking stage is a blob tracker capable of tracking a contiguous region in image space distinguishable from its background based on colour or intensity. The tracked target is represented by a unit view vector, which is given in the inertial frame so that tracking is independent of the orientation of the aircraft. The target can be tracked through multiple scales via an active zooming mechanism, which is driven by apparent size of the target on the image plane.

For clarity, the various superscripts used to describe the inertial target vector, $\hat{\mathbf{v}}_t$, are: $k+$, for the localised vector in the previous (time) frame, which is equivalent to the predicted vector in the current frame, $k+1$; and $k+1+$, for the localised vector in the current frame. The superscript *cam* is included to indicate when the view vector in given in the coordinate frame of the vision system (see Figure 4.1 for an illustration of the coordinate frames used here).

7.3.2 Tracking a ground-based target

The purpose of the tracking algorithm described in Section 7.3.1 is to localise a ground-based target, which will provide information to be used by an interception guidance system (this is the topic of Chapter 8). To provide a qualitative measure of performance of the tracker, a static, ground-based target was tracked throughout

a segment of open-loop flight. During the flight test, omnidirectional images were captured at 25Hz using the *iEye* vision system (described in Section 5.3), which was mounted on the nose of a small-scale UAV (pictured in Figure 6.7). The tracking algorithm was analysed offline by executing it on recorded flight data.

A single frame from the *iEye* vision system during the flight test is displayed in Figure 7.3a. The target is a white concrete rectangle with dimensions 8m × 4m and which is level with the ground. A segment of the flight test, during which the target is tracked continuously, is shown in the accompanying video[1]. Cropped images from the stabilised tracking window (Figure 7.3e) during the same period of flight are shown in Figure 7.4. It can be seen that the target was localised successfully in every frame of the ∼ 150s flight segment, except for frames in which the target was obscured by the body of the aircraft or not imaged by the vision system. Tracking was initialised manually prior to the analysed segment of flight.

Frames in which predicted viewing direction to the target was obscured were detected and tracking was suspended automatically until the target reappeared. Note that tracking was *not* reinitialised by hand at any stage during the analysed flight segment. Estimated translational motion of the aircraft with respect to the ground plane was incorporated into the *Propagate* stage of the tracking algorithm (described in Section 7.3.1) to improve the ability of the tracking system to reinitialise automatically on the target following an obstruction. Prior to transforming the previous localised inertial target view vector, $\hat{\mathbf{v}}_t^{k+}$, into the current camera frame, $\hat{\mathbf{v}}_t^{\text{cam},k+1}$, the target vector was repositioned according to apparent translational motion of the ground at the point where the target vector intersected the ground plane. The aircraft's 3-DOF translational motion with respect to the ground plane was computed from wide-field optic flow measured with the *iEye* vision system using an algorithm developed by Mr. Saul Thurrowgood. To compute 3-DOF translation between two frames, the aircraft was assumed to be a unit distance from the ground plane in the first frame. Hence, the new transformation of the target vector, $\mathbf{v}_t^{k+1}$, during the *Propagate* stage is given by

$$\mathbf{v}_t^{k+1} = \frac{\hat{\mathbf{v}}_t^{k+}}{v_{tz}^{k+}} + \mathbf{T}_a^{k,k+1}, \tag{7.1}$$

where $\hat{\mathbf{v}}_t^{k+}$ is the unit inertial view vector to the target after being localised in the

[1] Attachment AV11. Note that data and imagery shown in this video are explained in Figure 7.3.

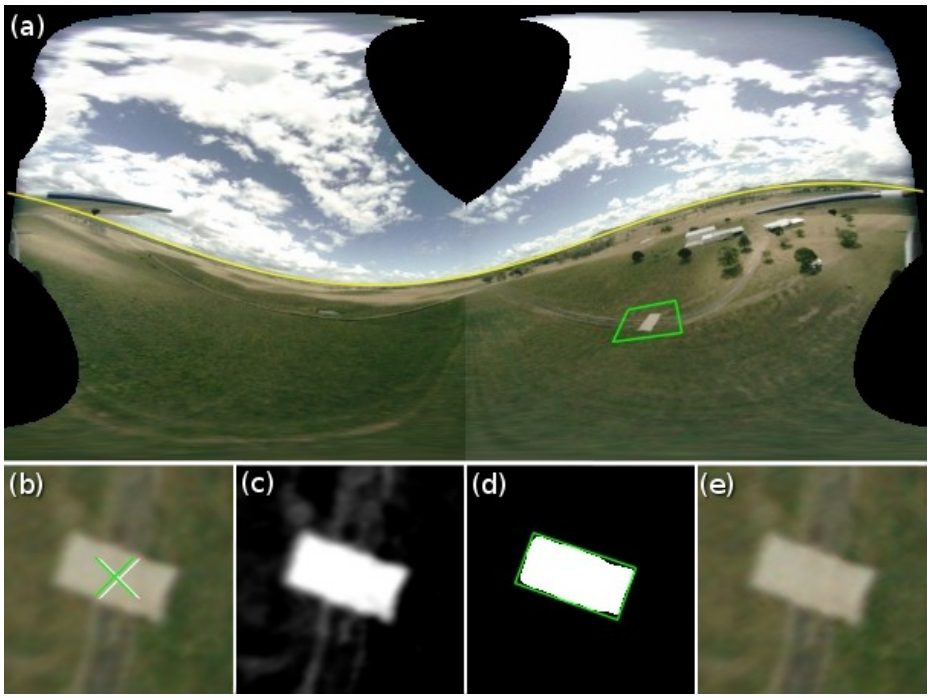


Figure 7.3: Single frame from the video of the tracking flight test during which a flat concrete target was tracked. The stitched image from the *iEye* vision system is shown in (a), along with the estimated horizon (yellow line) and approximate projection of the limits of the tracking window (b) onto the stitched image (green quadrangle). The centre of the tracking window (*i.e.* the predicted position of the target in this frame) is shown as a white cross in (b), along with the localised target centre (green cross). The transformed greyscale tracking window is shown in (c) and the segmented target is shown in (d), along with the computed bounding box (green rectangle). The computed centre of the bounding box (*i.e.* the localised target vector) is centred to produce (e), the stabilised tracking window. Details in text.

previous frame and v_{tz}^{k+} is the z (vertical) component of the same vector; and $\mathbf{T}_a^{k,k+1}$ is the inertial 3-DOF translation of the aircraft between the previous and current frames. The previous target vector, $\hat{\mathbf{v}}_t^{k+}$, is normalised with respect to its z component because the computation of the aircraft's translation with respect to the ground plane assumes an initial unit height. The predicted unit inertial view vector to the target in the current frame is then given by $\hat{\mathbf{v}}_t^{k+1} = \frac{\mathbf{v}_t^{k+1}}{|\mathbf{v}_t^{k+1}|}$, which is transformed into the coordinate frame of the camera using current 3-DOF attitude of the aircraft. The relevant coordinate frames and view vectors are illustrated in Figure 7.5. This procedure assumes that the target is stationary on the ground and hence apparent motion of the target in the image is due only to the motion of the aircraft with respect to the ground plane. In addition to propagating the estimated target location, the

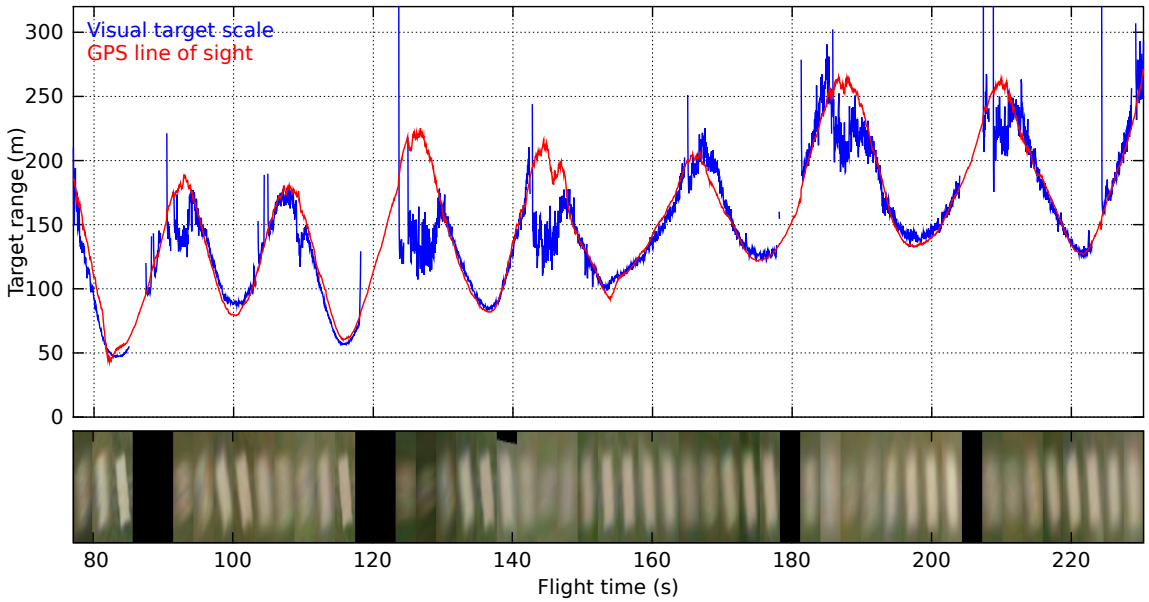


Figure 7.4: Estimated LOS distance to the concrete target (top) during the open-loop tracking flight test as estimated from scale of the tracked target (blue) and from the intersection between the target view vector and the ground plane (red), computed with the aid of the GPS height of the aircraft above the target. Visual range estimates have been scaled to aid comparison. Cropped images from the tracking window at regular intervals throughout the flight test are also shown. Frames in which the tracking window is (partially) missing, or for which there is no visual estimate for the range to the target, correspond to frames in which the target was obscured by the body of the aircraft or otherwise not visible.

tracking zoom parameter was also halved upon occlusion to improve the robustness of the tracking system to errors in estimated translational motion of the target with respect to the aircraft.

To analyse the capabilities of the tracker, an estimate of range to the target was obtained from the apparent scale of the target within the tracking window and compared with LOS range computed from the target view vector, $\hat{\mathbf{v}}_t$, and altitude of the aircraft above the target (obtained from a GPS unit on board the aircraft)[2]. Range estimates computed from the visual size of the target, and those derived from the intersection between $\hat{\mathbf{v}}_t$ and the ground plane during the flight test, are plotted in Figure 7.4. To aid comparison between the two methods, the range estimates computed from the visual size of the target have been scaled by a constant factor, which was determined empirically.

[2]This metric was chosen over other more common tracking metrics for reasons which will become apparent in Chapter 8 – target size is important for applications such as visual height estimation and time to contact (TTC) estimation.

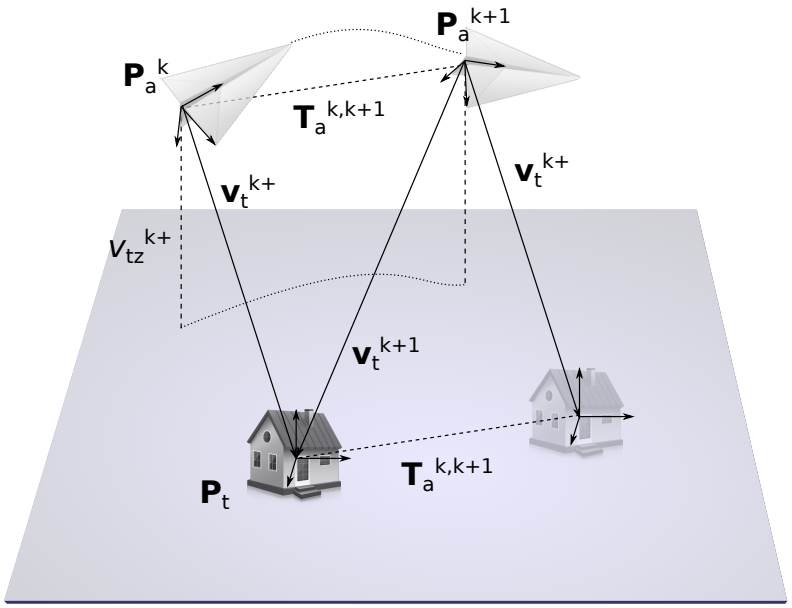


Figure 7.5: Coordinate frames and view vectors for the *Propagate* stage of the tracking algorithm for a stationary ground-based target. At frame k the aircraft and target are at positions $\mathbf{P}_a^k$ and $\mathbf{P}_t$ respectively and the localised target view vector is given by $\mathbf{v}_t^{k+}$ in the inertial coordinate frame. At some later frame, $k+1$, the aircraft has moved to $\mathbf{P}_a^{k+1}$ and the view vector to the target in the inertial frame, $\mathbf{v}_t^{k+}$, no longer points to the true position of the target. The estimated translation of the aircraft between frames k and $k+1$ is given by $\mathbf{T}_a^{k,k+1}$ and is computed assuming that the aircraft is initially at unit height at frame k. The true current position of the target at frame $k+1$ can be predicted, therefore, by accounting for the translational motion of the aircraft: the target view vector in the previous frame, $\mathbf{v}_t^{k+}$, is scaled such that the z (vertical) component, v_{tz}^{k+}, is normalised; the normalised target view vector is then added to $\mathbf{T}_a^{k,k+1}$ according to Equation 7.1 to give the predicted viewing direction to the target in the current frame, $\mathbf{v}_t^{k+1}$. House and paper plane image credit: http://www.graphicsfuel.com.

Obviously apparent scale of a 2D target within the tracking window depends not only on LOS distance between observer and target, but also on visual distortion experienced when observing the 2D target from anywhere but directly overhead. To correct for this distortion, an orthographic projection of the target onto the ground plane is generated each frame using the predicted target vector, $\hat{\mathbf{v}}_t$, and current 3-DOF attitude of the aircraft. This projection ideally removes any distortion introduced by the angle at which the aircraft views the target.

To generate the orthographic projection, a regular grid of points is placed on the ground plane surrounding the intersection between the inertial target vector, $\hat{\mathbf{v}}_t$, and the ground plane, using an assumed height for the observer. The grid is positioned

such that n points are evenly (metrically) distributed over the range $[-0.5\text{m}, +0.5\text{m}]$, where $n = r \times c$ is number of pixels within the tracking window and r and c are number of rows and columns respectively within the tracking window ($r = c = 151$ for data presented in this section). Axes of the projected grid are aligned with the x, y axes of the inertial coordinate frame (*i.e.* north and east). To produce the orthographic tracking window, pixels in the raw fisheye images captured by the *iEye* vision system, corresponding to the projected grid of points, are found by transposing the inertial viewing directions to the points into the current coordinate frame of the vision system, using current 3-DOF attitude of the aircraft. The resulting orthographic projection of the target can be seen in Figure 7.3b and the corresponding region in the stitched omnidirectional image is shown in Figure 7.3a.

Put simply, a vector pointing to the target in the coordinate frame of the vision system is obtained from the tracking algorithm described in Section 7.3.1 and that vector is transformed into the inertial coordinate frame using visually estimated 3-DOF attitude of the aircraft. Using an assumed height for the aircraft, the inertial target vector can be scaled to represent 3D position of the target with respect to the aircraft in the inertial frame. A regular grid of 3D points is placed on the ground surrounding the estimated 3D position of the target such that width and height of the grid are equal to one unit of assumed height for the aircraft. If *assumed* height of the aircraft (the observer) is equal to *true* height of the aircraft (true height above the ground at which the images were captured), then the orthographic projection covers a square region with bounds ± 0.5m from the centre of the target. If *true* height is actually twice *assumed* height, however, then the region of the projected grid is actually ± 1.0m from the centre of the target (since one unit of *assumed* height would be equal to two units of *true* height, or 2m). Therefore, true width (and height) of the orthographic projection corresponds to a scaling factor relating assumed height of the observer to true height of the aircraft. If an object of known size is imaged within the tracking window, then true height of the aircraft can be computed simply from the assumed height at which the projection was generated.

In Section 7.3.1, the FOV (or zoom) of the tracking window was automatically and dynamically scaled to maintain the apparent size of the target. For the orthographic projection discussed in this section, the 'zoom' is in fact assumed height of the observer. Dynamically adjusting the assumed height of the observer over time to maintain apparent size of the target therefore permits true height of the aircraft to be measured (correct to a scale factor) whilst the target is tracked. True 3D position

(and hence LOS range) of the target, relative to the aircraft, can then be computed simply by adjusting the length of the inertial target vector, $\hat{\mathbf{v}}_t$, such that the vertical component is equal to true height of the aircraft at each frame.

A scale factor of 15 was used to plot the visual range data presented in Figure 7.4. This means the true dimensions of the orthographic projection in this case are 15m $\times$ 15m and thus, since the zoom is adjusted dynamically to maintain the largest dimension of the target bounding box to be equal to half the tracking window size, the estimated dimensions of the target are approximately 7.5m$\times$3.8m. This is close to the measured dimensions of 8m $\times$ 4m. The scale factor was determined empirically from correspondence between the GPS-derived target LOS range and visually estimated range (plotted in Figure 7.4). Error in the estimated target dimensions is due entirely to error in the estimated scale factor and could therefore be due to bias in the altitude reported by the GPS unit on board the aircraft, or bias introduced by the method for segmenting the target from the tracking window (described in Section 7.3.1).

7.3.3 Tracking a moving target

Section 7.3.2 described an implementation of the appearance-based tracker designed specifically for robustly tracking stationary ground-based targets. In this section it is shown that this same implementation can be applied to *moving* ground-based targets, without modification. The tracking algorithm relies on appearance of the target. Hence, if the viewing direction to the target does not change so much between frames that the target leaves the tracking window entirely, the target will be localised successfully. A caveat to this approach is that, if the target is obscured for a period of time (*e.g.* when blocked by the body of the aircraft), then the predicted target vector, $\hat{\mathbf{v}}_t$, which is propagated between frames during the obstruction, will drift away from the true viewing direction in the inertial coordinate frame. This is because independent motion of the target differs from expected motion of the ground plane. Therefore, the tracker will not necessarily reacquire the target successfully, following occlusion.

To compensate for this deficiency, 2D motion of the target along the ground plane can be estimated from rate-of-change of the viewing direction to the target in the inertial coordinate frame and visually estimated 3-DOF translation of the aircraft with respect to the ground plane. Using this approach, the transformation applied to

the target vector during the *Propagate* stage (Equation 7.1), can be modified to give

$$\mathbf{v}_t^{k+1} = \frac{\hat{\mathbf{v}}_t^{k+}}{v_{tz}^{k+}} + \mathbf{T}_a^{k,k+1} + \mathbf{T}_t,$$

where $\mathbf{T}_t$ specifies predicted 2D motion of the target in the inertial frame, and the other parameters are as defined for Equation 7.1. The relevant coordinate frames and view vectors for tracking a moving ground-based target are illustrated in Figure 7.6. Only 2D motion of the target is used because it is assumed that the target does not leave the ground plane.

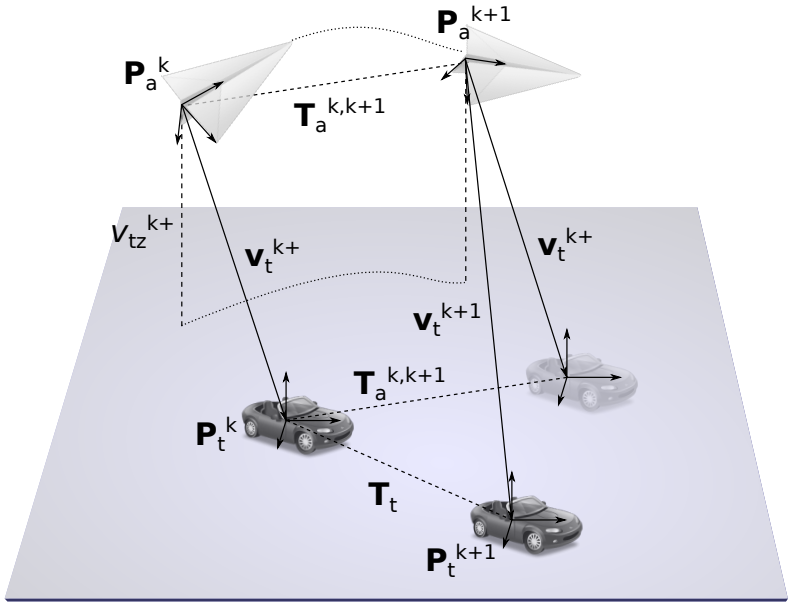


Figure 7.6: Coordinate frames and view vectors for the *Propagate* stage of the tracking algorithm for a ground-based target moving with constant velocity. The predicted location of the target in the current frame, $\mathbf{v}_t^{k+1}$, is computed as for the stationary target case (*e.g.* Figure 7.5), except here an additional term accounting for the observed velocity of the target, $\mathbf{T}_t$, is required. This term can be measured each frame from the difference between the predicted viewing direction to the target (without accounting for its motion) and the localised viewing direction to the target (output from the tracking stage). The computed vector, $\mathbf{T}_t$, will have units of the height of the aircraft at frame k. To allow multiple estimates of $\mathbf{T}_t$ to be combined across several frames, or to predict the viewing direction to the target through extended periods of occlusion, the change in height of the aircraft between frames must be incorporated in the propagation of $\mathbf{T}_t$, as per Equation 7.2. Image credits for paper plane and car: http://www.graphicsfuel.com and http://www.fordesigner.com.

The aircraft's 3-DOF translation with respect to the ground plane is estimated for each pair of successive frames and used to propagate the predicted viewing direction

to the target each frame (Equation 7.1). For a static target on the ground, therefore, the localised target vector, $\hat{\mathbf{v}}_t^{k+}$, ideally coincides exactly with the predicted target vector, $\hat{\mathbf{v}}_t^k$, each frame. The 2D vector difference between the two intersection points of the predicted and localised target vectors, $\hat{\mathbf{v}}_t^k$ and $\hat{\mathbf{v}}_t^{k+}$, with the ground plane, thus corresponds to 2D translation of the target along the ground. The estimated translation of the target has units equal to the height of the aircraft AGL. If height is known, metric velocity of the target can thus be estimated. In this case, height of the aircraft is unknown, but *change in height* of the aircraft over a short sequence of frames can be estimated by integrating the z (vertical) component of the aircraft's estimated 3-DOF translation vector, $\mathbf{T}_a$, with respect to the ground plane. The compounded change in height of the aircraft can be used to propagate the target velocity estimate, $\mathbf{T}_t$, according to

$$\mathbf{T}_t^{k+1} = \frac{\mathbf{T}_t^k}{1 - T_{az}^{k,k+1}}, \tag{7.2}$$

where $\mathbf{T}_t^k$ and $\mathbf{T}_t^{k+1}$ are the velocity of the target in the previous frame and predicted velocity of the target in the current frame respectively; and $T_{az}^{k,k+1}$ is the z (vertical) component of $\mathbf{T}_a^{k,k+1}$, the aircraft's estimated translation between frames k and $k+1$. Because $\mathbf{T}_a^{k,k+1}$ is specified in the NED inertial frame and is computed assuming an initial unit height at frame k, $-\infty < T_{az}^{k,k+1} < 1$ and Equation 7.2 is well defined. Thus, for a target travelling at constant velocity, the apparent displacement of the target with respect to its predicted position each frame will be consistent and the apparent motion of the target in 2D, $\mathbf{T}_t$, can be estimated with respect to the ground plane, even though metric velocity of the target will be unknown.

In practice, of course, there is some noise associated with visual estimation of the aircraft's translation, $\mathbf{T}_a$, so the target must be localised every frame. Apparent motion, $\mathbf{T}_t$, of a constant-velocity, ground-based target can therefore be estimated by computing the average displacement between predicted and localised positions of the target on the ground plane over a number of frames. If an estimate of the target's translation, $\mathbf{T}_t$, is built up over a number of frames in which the target is tracked, then its predicted position can be propagated when it is obscured, if relative height of the aircraft is computed throughout the tracking sequence.

This approach was implemented to track a vehicle moving along a straight section of road during an open-loop flight test. A segment of the flight test in which the vehicle is tracked, then obscured by the body of the aircraft, and

then reacquired automatically is shown in the accompanying video[3]. This result demonstrates successful tracking of a constant-velocity, ground-based target under real-world conditions. Tracking was initiated manually at the beginning of the flight segment.

It is important to reiterate that scale of the aircraft's estimated 3-DOF translation, $\mathbf{T}_a$, is unknown, hence it cannot be specified metrically without additional measurements of the aircraft's ground speed or height AGL. $\mathbf{T}_a$ does, however, exactly specify *apparent* translation of the ground plane with respect to the aircraft and *vice versa*. Thus it can be used to describe, or predict, the observed motion of objects on the ground, as described above. The same principle is true of estimated 2D velocity of the target. Because these motions cannot be described metrically, using only the visual information at hand here, they cannot be used to compute a 3D trajectory for the aircraft to intercept the target.

7.4 Discussion

Results presented in Figure 7.4 demonstrate that the approach for tracking a ground-based target described in this chapter leads to an accurate representation of the position of the target and its size. The tracking algorithm is robust to temporary occlusions, evident from the tracking segments analysed in Section 7.3.2, and robust to changing appearance of the target induced by relative position of the aircraft and changing environmental conditions. If a target of known size is tracked, this approach permits the aircraft's height above the target, or equivalently range to the target, to be estimated.

Visually estimated range to the stationary target (Figure 7.4) is reasonably accurate throughout the flight test. Several periods are visible, however, where range estimates become noisy and inaccurate. Cropped images of the tracking window during tracking, shown in Figure 7.4, indicate that these periods occur when the effective resolution of the tracking window is low and the shape of the target is poorly defined – in fact for many of these instances it is difficult to distinguish the target from the background by eye. It can be seen from the accompanying video of the static target flight test that these periods correspond to flight segments where the aircraft is far from the target and low to the ground. At these times the aircraft views

[3]Attachment AV12. Note that data and imagery shown in this video are explained in Figure 7.3. It was found to not be necessary to halve tracking zoom during occlusion of the moving target.

the target at a very shallow grazing angle and, although the orthographic projection accounts for the distortion, there are very few pixels in the raw fisheye images that correspond to the target. Hence, the effective resolution of the tracking window is very low, which leads to less precise estimates of target dimension and range.

The new tracking algorithm described above takes 35ms @ 1.5GHz to execute using a tracking window resolution of 151px×151px. In practice, it was found that a window size of 15px × 15px could be used without significant loss of tracking precision or robustness, which reduced execution time to 0.5ms @ 1.5GHz. This lower resolution was used for closed-loop target tracking and interception in Chapter 8. A significant portion of tracking time is consumed during generation of the orthographic projection – which has to be performed each frame. It was found that this process could be simplified significantly by using an affine transform to approximate and remove distortions produced by viewing angle between the aircraft and the target. This approach further reduced execution time of the tracking algorithm to 0.2ms @ 1.5GHz, again without significant loss of tracking precision or robustness. The new algorithm proposed in this chapter was therefore found to permit robust and efficient tracking of static or moving ground-based targets. Furthermore, the computational efficiency of the algorithm allowed multiple tracking instances to be executed in parallel, permitting tracking of multiple targets.

7.5 Conclusion

A simple, yet effective, appearance-based tracker has been described here that uses RGB properties of an image patch to segment a blob from the background. Most objects in a rural environment, such as trees, buildings, vehicles, patches of vegetation, or people, appear largely uniform in colour from above and are distinguishable from the grass, earth, or sky, which surrounds them. Flight test results demonstrate that the tracker presented here is suitable for localising such objects from omnidirectional imagery provided by the *iEye* vision system over extended sequences. The predicted position of the tracked object within the view sphere is propagated during occlusions so that the target can be reacquired successfully when it reappears. Using this approach, static or moving ground-based targets can be tracked throughout an entire flight.

The simplicity of the proposed tracker means that it is very computationally efficient, and hence multiple instances can be executed in parallel to track many

objects simultaneously. This system may therefore be used to improve precision of visual odometry by providing robust image correspondences over extended sequences. By describing each tracked object's colour, scale, or dimensions (from the bounding box), individual targets or classes of objects (*i.e.* trees, buildings, vehicles, *etc.*) may be recognised and reacquired after long periods of occlusion – *e.g.* when the aircraft flies away from a particular area. This system could therefore be used within a real-time visual SLAM architecture to improve long-term navigation.

A guidance scheme that allows an aircraft to intercept a moving object using only apparent viewing direction to the target in the inertial coordinate frame can now be proposed. The target vector, $\hat{\mathbf{v}}_t$, computed via the tracking algorithm described in Section 7.3, provides exactly the information required for interception. This tracker therefore forms the basis for visual interception of a moving, ground-based vehicle. A task as complex as this can be demonstrated in real time on board a small-scale UAV by combining several simple and bioinspired vision-based sensing and guidance algorithms. The efficiency and robustness of these algorithms stem from the insect-like omnidirectional FOV of the *iEye* vision system.

8

Vision-based Interception of Moving Objects

8.1 Summary

A requirement for the next generation of small-scale UAVs is the capability to interact with other moving vehicles and objects. A novel and bioinspired guidance strategy is introduced here that allows an aircraft to intercept an independently moving target using vision only. The strategy relies on visual input from the *iEye* vision system (Chapter 5) in combination with the aircraft's 3-DOF attitude (Chapter 6) and the output from an appearance-based tracker (Chapter 7). All components of the guidance system are implemented on board a small-scale UAV to demonstrate autonomous interception of an independently moving object.

Tracking other moving vehicles or objects and computing trajectories for interception are important capabilities for UAVs because they facilitate landing or docking manoeuvres (*i.e.* for return-to-ship or refuelling) and can also provide forewarning of independently moving objects on a collision course. The application of various visual tracking schemes to the problem of interception is discussed in Section 8.2 and a distinction between *interception* and simple *pursuit* is made using examples from biology. Uncommonly for insects, dragonflies intercept their prey. A novel control scheme based on the dragonfly's strategy is described in Section 8.3, which allows a fixed-wing UAV to intercept a moving target using visual input only. Autonomous interception of a moving, ground-based vehicle is demonstrated with simulated and real-world flight tests in Section 8.4, proving the advantages of this bioinspired, visual approach. Limitations are discussed in Section 8.5 and applications of this method to

other aspects of UAV guidance, such as landing in windy conditions, are investigated in Section 8.6.

8.2 Intercepting a moving target

Planning a course for interception intuitively requires prediction of the target's trajectory and, therefore, knowledge of the target's 3D position and velocity relative to the observer. Computer vision literature describes several approaches for obtaining such information visually. Examples from biology demonstrate, however, that this information is not necessary to intercept independently moving objects, leading to an efficient, vision-based strategy for interception.

8.2.1 Tracking and interception in 3D

Tracking independently moving objects from a moving platform is a problem that has received a great deal of attention from the computer vision community in the past (for reviews see Jia et al. [2008], Yilmaz et al. [2006]). For all vision-based approaches, the target is intially localised in image space (*e.g.* the appearance-based tracker described in Chapter 7). Reconstructing the target's trajectory, or planning trajectories for the pursuer to intercept or dock with the target, both intuitively require knowledge of 3D position and velocity of the target with respect to the pursuer.

For the case where no additional measurements of the target's motion are available, other than viewing direction to the target, 3D trajectory of the target relative to the observer can only be recovered by iteratively estimating a kinematic model for the target according to available measurements. This technique is known as bearings-only range estimation, or bearings-only target motion analysis [Nardone et al., 1984]. The measurement model in this case is highly non-linear. Extended Kalman filters (EKFs), unscented Kalman filters (UKFs), or particle filters are therefore typically used to propagate estimates of the target's kinematics. Range is unobservable except during certain manoeuvres, however, and target accelerations can cause the filters to diverge. Implementations based on this approach are described by Aidala and Hammel [1983], Allen and Blackman [1991], Johnson et al. [2007] for the purpose of tracking trajectories of target vehicles from airborne platforms.

Alternatively, the position of the target relative to the observer can be estimated in 3D space from location and apparent size of the target on the image plane (*e.g.*

Section 7.3.2). This technique is used by Rosales and Sclaroff [1999] to track the trajectories of people moving at ground level and by Johnson et al. [2007] to compute the trajectory of a lead aircraft from a pursuing aircraft. This approach can only provide metric estimates of position or velocity of targets, however, if their dimensions are known *a priori*.

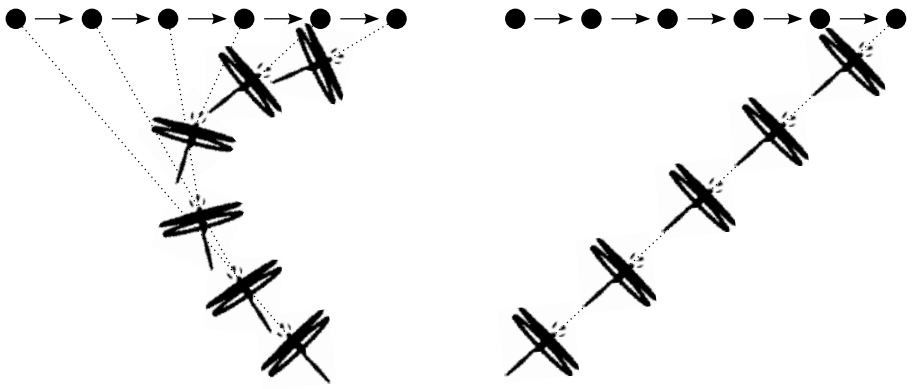
If the target is at ground level and the observer is airborne, LOS range to the target can be estimated by extending the vector describing the viewing direction to the target so that it intersects the ground plane, using the height of the observer above ground. Frietsch et al. [2007] obtain aircraft height from GPS and use this approach to reconstruct trajectories of moving ground-based vehicles from a hovering MAV (this is very similar to the technique used in Section 7.3.2 to provide comparative estimates of LOS range to the target during flight testing). Saripalli and Sukhatme [2007], Saripalli et al. [2003] also obtain the aircraft's height from GPS and track a moving, ground-based platform using vision, enabling a small-scale UAV to dock with it. Han and DeSouza [2009] use a similar approach, but measure the aircraft's height above the target using multi-frame triangulation instead of directly via GPS. This technique obviates the need to assume a locally flat ground plane and may provide more accurate range estimates for large ranges, however it requires that the observer's 3D translation be known or measured accurately to compute target range.

Triangulation through stereo vision has also been used to track moving targets in 3D space from stationary platforms [Beymer and Konolige, 1999] and from moving ground-based vehicles [Badino et al., 2006, Dang et al., 2002, Leibe et al., 2007, Rabe et al., 2007]. A very wide stereo baseline is typically required, however, to triangulate a target precisely from an airborne platform. Stereo-based approaches are therefore not ideally suited to implementation on board small-scale UAVs.

8.2.2 Biological systems

In many instances in nature, the pursuer apparently has no accurate information on range to its quarry, yet it is still able to catch it. Some species of fly, for example, chase other flies simply by continually aiming at the perceived location of the target on the retina [Land and Collett, 1974]. This is an example of *pursuit* (Figure 8.1). Individuals following a pure pursuit trajectory typically end up in a tail-chase situation. The actions of the pursuer are governed by a completely reactive control law, although Land and Collett [1974] show that pursuing flies anticipate the

heading of leading flies to some extent by also controlling on the apparent angular velocity of their target. If the pursuer flies faster than the target, this strategy leads to a spiralling flight path ending in capture. Most insects that have been studied use this simple strategy to capture moving objects.



(a) Pursuit. (b) Interception.

Figure 8.1: Two alternative strategies for approaching a moving object – (a) pursuit and (b) interception. Modified from Olberg et al. [2000].

A more efficient strategy, perhaps, is for the pursuer to travel in a relatively straight line intersecting the projected flight path of the target. This is an example of *interception* (Figure 8.1). Intuitively, this strategy requires knowledge of the target's 3D position and 3D velocity relative to the interceptor – quantities that are typically not able to be estimated accurately. Male hoverflies follow intercepting trajectories in their pursuit of female hoverflies [Collett and Land, 1978]. The algorithm used by the hoverfly, however, requires targets of known size and flight speed.

Dragonflies have also been shown to exhibit intercepting flight paths when pursuing prey [Olberg et al., 2000]. Unlike hoverflies, however, dragonflies can intercept prey of various sizes and flight speeds. By studying the trajectories of dragonflies and their prey, Olberg et al. [2000] showed that dragonflies maintain the absolute angle between the apparent LOS to their target and the horizon constant during interception. By maintaining a constant absolute target direction (CATD), the dragonflies were able to follow a nearly time-optimal trajectory to interception, without explicitly requiring knowlege of their target's relative position or velocity. It has since been shown that bats [Ghose et al., 2006] and humans [Fajen and Warren, 2007] also use this strategy to capture moving targets.

Interestingly, if the pursuer holds the absolute bearing to the target constant,

the pursuer appears stationary against a distant background (or, effectively, appears to be infinitely distant) from the viewpoint of the target. Trajectories of the flights of hoverflies [Srinivasan and Davey, 1995] and dragonflies [Mizutani et al., 2003], in which the pursuer follows a CATD strategy, have thus been interpreted as camouflaging behaviour on the part of the pursuer.

8.2.3 Parallel navigation

The practical advantages to using strategies such as CATD to intercept moving targets have not gone unnoticed. In the field of missile guidance, the CATD strategy is referred to as parallel navigation [Yanushevsky, 2007]. The basic philosophy behind parallel navigation is based on recognition that if two bodies are closing on each other, they will eventually intercept if the absolute bearing between the two does not rotate relative to the inertial frame. In an active sense, parallel navigation requires the pursuer to nullify rate of change of the inertial bearing to the target.

Guidance laws based on parallel navigation have been derived for interception of moving targets by ground-based vehicles [Belkhouche et al., 2006] and for motion camouflage (which is in practice equivalent to a CATD guidance strategy) [Carey et al., 2004, Justh and Krishnaprasad, 2006, Reddy et al., 2006]. Interception of moving objects by a ground-based vehicle is demonstrated by Manchester and Savkin [2002], Manchester et al. [2007] using a visual guidance strategy, but they require estimates of the target's velocity.

8.3 A control strategy for interception

Specific control laws for achieving parallel navigation have been developed since the 1940s [Yuan, 1948]. Perhaps the most widely known and used guidance law is proportional navigation (PN), due to its inherent simplicity and ease of implementation [Siouris, 2004, Yanushevsky, 2007]. In this section, a control strategy based on proportional navigation (PN) is described that is designed specifically for guidance of a small-scale, fixed-wing UAV.

8.3.1 Proportional navigation

The PN guidance law seeks to nullify rate of change of inertial bearing to the target by controlling acceleration of the interceptor such that

$$a_{\perp} = NV_r \left(\frac{d\lambda_t}{dt} \right), \tag{8.1}$$

where N is a positive real number known as navigation gain; V_r is relative closing speed between interceptor and target; λ_t is instantaneous bearing to the target with respect to the inertial frame ($d\lambda_t/dt \equiv \dot{\lambda}_t$ is its time derivative); and $a_{\perp}$ is commanded acceleration of the interceptor, which acts perpendicular to λ_t and in the direction to reduce change in λ_t [Siouris, 2004]. Under this guidance law, therefore, the interceptor's forward speed remains relatively constant and commanded accelerations act to keep the target on a constant inertial bearing (for the definition of the inertial frame of reference, see Figure 4.1).

Navigation gain, N, is a function of the interceptor's acceleration capabilities and expected manoeuvres of the target. For $N \to \infty$, the interceptor's trajectory will be straighter. If the interceptor can accelerate (*i.e.* change course) rapidly with little delay, N can be large. A large navigation gain allows the control system to nullify any change in inertial bearing to the target promptly, and thus the interceptor's trajectory will be straight for a constant-velocity target. Setting N too large, however, results in an oscillatory trajectory.

For many aerial vehicles, applied turning accelerations act perpendicular to the instantaneous velocity vector of the vehicle. In practice, therefore, $a_{\perp}$ is often approximated by a_l, a lateral acceleration that is applied normal to instantaneous velocity [Siouris, 2004]. Therefore, by making the substitution

$$a_{\perp} \approx a_l,$$

and by noting that lateral acceleration, a_l, can be represented in terms of forward velocity, V_m, and turn rate, $\dot{\gamma}_m$, of the interceptor

$$a_l = V_m \left(\frac{d\gamma_m}{dt} \right),$$

Equation 8.1 can be rewritten as

$$V_m \left(\frac{d\gamma_m}{dt} \right) = N V_r \left(\frac{d\lambda_t}{dt} \right).$$

If it is assumed then, that the interceptor's speed is greater than the target's speed ($V_m \gg V_t$), then closing speed between the interceptor and target, $V_r = \|\mathbf{V}_m - \mathbf{V}_t\|$, will be approximately equal to the interceptor's speed, $V_r \approx V_m$, and thus the modified PN guidance law can be written

$$\frac{d\gamma_m}{dt} = N \left(\frac{d\lambda_t}{dt} \right). \tag{8.2}$$

Applying a turning rate proportional to the measured rate of change of the target's inertial bearing, therefore, will satisfy the CATD rule and the pursuer will intercept a constant-velocity target along a time-optimal (*i.e.* straight-line) trajectory. For the case of a manoeuvring target, Ghose et al. [2006] show that the interception trajectory will still be close to time-optimal if the target's trajectory can be represented by a piecewise constant-velocity trajectory.

8.3.2 An interception strategy for fixed-wing aircraft

The PN-based guidance law written in Equation 8.2 defines the 2D turning rate commands necessary for interception in a planar engagement scenario. Commanded turning rate, $\dot{\gamma}_m$, and observed target bearing rate, $\dot{\lambda}_t$, are both given in the inertial coordinate frame. For the 3D case, therefore, it is trivial to define two guidance equations,

$$\dot{\gamma}_m^h = N^h \dot{\lambda}_t^h, \tag{8.3a}$$

$$\dot{\gamma}_m^v = N^v \dot{\lambda}_t^v, \tag{8.3b}$$

where $\dot{\gamma}_m^h$ and $\dot{\gamma}_m^v$ define interceptor turning rates required in the horizontal and vertical inertial planes respectively; and $\dot{\lambda}_t^h$ and $\dot{\lambda}_t^v$ are observed target bearing rates in the corresponding planes. Now, $\dot{\gamma}_m^v$ and $\dot{\gamma}_m^h$ are exactly the required inertial-frame pitch rate, Q_m^*, and heading rate, R_m^*, for an aircraft following an interception trajectory. To implement this guidance law, a control strategy must therefore be derived that permits the aircraft to obtain Q_m^* and R_m^* simultaneously.

Inertial rotation rates, P, Q, and R, around the inertial coordinate frame x, y, and z axes respectively, are completely specified by body rotation rates (rotation rates measured in the body frame of reference), p, q, and r, and the aircraft's roll angle and pitch angle, ϕ and θ, according to

$$\begin{bmatrix} P \\ Q \\ R \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\phi \cdot \sin\theta & -\cos\phi \cdot \sin\theta \\ 0 & \cos\phi & \sin\phi \\ \sin\theta & -\cos\theta \cdot \sin\phi & \cos\phi \cdot \cos\theta \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix}.$$

Thus, inertial pitch rate, Q, and heading rate, R, can be defined with respect to body rotation rates and aircraft attitude,

$$Q = \cos\phi \cdot q - \sin\phi \cdot r, \tag{8.4a}$$

$$R = -\sin\theta \cdot p + \sin\phi \cdot \cos\theta \cdot q + \cos\phi \cdot \cos\theta \cdot r. \tag{8.4b}$$

It is proposed that there exists a certain combination of roll angle, ϕ, and body pitch rate, q, which will produce the desired inertial pitch rate, Q, and inertial heading rate, R, simultaneously, over the space of achievable combinations of Q and R. This is intuitive for a bank-to-turn aircraft such as the vehicle under consideration (Figure 6.7). It is thus desired to express ϕ and q in terms of Q and R.

To begin with, a novel approximation is made to simplify Equation 8.4,

$$r \approx \alpha \cdot \sin\phi. \tag{8.5}$$

Equation 8.5 was derived empirically and states that body yaw rate is approximately proportional to a sinusoidal function of roll angle, where α is a constant of proportionality. The justification is as follows: as the aircraft banks to enter a turn, and if no rudder is applied, the aircraft will tend to slip towards the lower wing, but asymmetrical airflow over the body of the aircraft will produce a yawing torque, which will act to realign airflow over the nose of the aircraft. Conversely, during a coordinated turn, the rudder is actuated to maintain airflow over the nose of the aircraft throughout the manoeuvre. In both cases, the effective body yaw rate of the aircraft is a function of bank (*i.e.* roll) angle.

The relationship between roll angle and body yaw rate, for moderate deviations from a wings-level, equilibrium flight condition, is plotted in Figure 8.2 using empirical data obtained from simulation and real-world flight testing. For reasonable roll angles,

it can be seen that body yaw rate, r, is well modelled by a sinusoidal function of roll angle, confirming the validity of the approximation written in Equation 8.5.

Inertial pitch rate, Q, defined in Equation 8.4a, can be restated by substituting body yaw rate, r, for Equation 8.5 right-hand side (RHS),

$$Q = \cos\phi \cdot q - \sin^2\phi \cdot \alpha,$$

and rearranged to give

$$q = \frac{Q + \sin^2\phi \cdot \alpha}{\cos\phi}. \tag{8.6}$$

Similarly, inertial heading rate, R, defined in Equation 8.4b, can be restated as

$$R = -\sin\theta \cdot p + \sin\phi \cdot \cos\theta \cdot q + \cos\phi \cdot \cos\theta \cdot \sin\phi \cdot \alpha.$$

Then, substituting body pitch rate, q, for Equation 8.6 RHS,

$$\begin{aligned} R &= -\sin\theta \cdot p + \sin\phi \cdot \cos\theta \left(\frac{Q + \sin^2\phi \cdot \alpha}{\cos\phi}\right) + \cos\phi \cdot \cos\theta \cdot \sin\phi \cdot \alpha \\ &= -\sin\theta \cdot p + \cos\theta \left(\tan\phi \left[Q + \sin^2\phi \cdot \alpha\right] + \cos\phi \cdot \sin\phi \cdot \alpha\right) \\ &= -\sin\theta \cdot p + \cos\theta \left(\tan\phi \left[Q + \sin^2\phi \cdot \alpha + \cos^2\phi \cdot \alpha\right]\right) \\ &= -\sin\theta \cdot p + \cos\theta \cdot \tan\phi \left[Q + \alpha\right]. \end{aligned}$$

Rearranging to solve for ϕ gives

$$\phi = \arctan\left(\frac{R + \sin\theta \cdot p}{\cos\theta \left[Q + \alpha\right]}\right),$$

which can be simplified further, because close to equilibrium flight conditions $\sin\theta$ is small and body roll rate is negligible, $p \approx 0$. Thus, $\sin\theta \cdot p$ vanishes and

$$\phi = \arctan\left(\frac{R}{\cos\theta \left[Q + \alpha\right]}\right). \tag{8.7}$$

Equations 8.6 & 8.7 comprise a system that defines a particular combination of roll angle, ϕ^*, and body pitch rate, q^*, that will simultaneously achieve the inertial pitch rate, Q^*, and heading rate, R^*, demanded by the *interception guidance law*

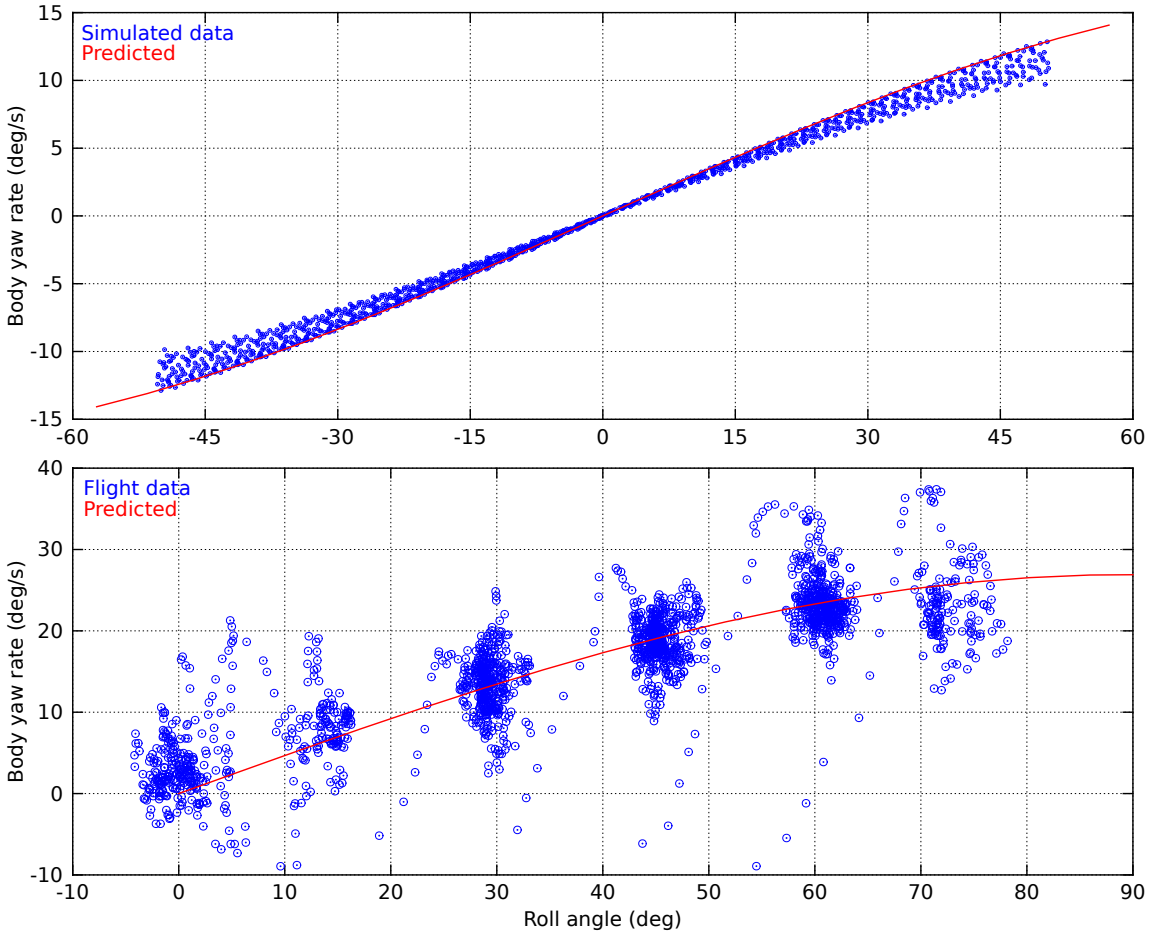


Figure 8.2: Relationship between roll angle and body yaw rate. Plots show recorded flight data (blue) obtained from simulation (top) and from an IMU on board the test aircraft during a flight test (bottom). Predicted relationship between roll angle and body yaw rate (Equation 8.5) is also plotted (red). For simulated data $\alpha = 0.29$ and for flight test data $\alpha = 0.47$. Simulations were based on the open-source *zSim* flight simulator [Hubert, 2008] and used a trainer aircraft model similar, but not identical, to the flight test aircraft. For each *simulated* flight test, the aircraft was commanded to hold pitch angle constant whilst roll angle was varied linearly over the range $[-50°, 50°]$. A number of simulated flight tests were conducted at pitch angles within the range $[0°, 30°]$. The predicted curve for simulated data is fitted at 0° pitch. The spread of the simulated flight data indicates variation for increasing pitch angle. Simulated data for negative pitch angles is not shown but is similar to the data for positive pitch angles (*i.e.* $\alpha \to 0$ as $\theta \to \pm 90°$). For the *real-world* flight test the aircraft was commanded to hold 0° pitch and simultaneously roll angles of 0°, 15°, 30°, 45°, 60°, and 75° for periods of 5s $\sim$ 20s.

(Equation 8.3). The *interception control strategy* can thus be written:

$$\phi^* = \arctan\left(\frac{R^*}{\cos\theta\,[Q^* + \alpha]}\right), \tag{8.8a}$$

$$q^* = \frac{Q^* + \sin^2\phi^* \cdot \alpha}{\cos\phi^*}. \tag{8.8b}$$

The system of interception control equations (Equation 8.8) can be solved sequentially (first ϕ^* is evaluated and then used to compute q^*) due to the simplification brought about by the empirically defined constant α (Equation 8.5), which allows the aircraft's body yaw rate to be predicted for stable flight at a particular roll angle. In fact, the two control Equations 8.8a & 8.8b can be unlinked and solved separately, because the aircraft's roll angle dynamics are much slower than its body pitch rate dynamics. Body pitch rate commands, q^*, act (almost) directly on the elevators whilst roll angle commands, ϕ^*, are a degree of separation further from the ailerons. Thus q^* will typically be realised much more quickly than ϕ^*. To prevent body pitch rate from 'pre-empting' roll angle during abrupt control changes, it is more appropriate to solve Equation 8.8b for q^* using the current roll angle, ϕ, rather than the targeted roll angle, ϕ^*. This approach was used to implement the control system for simulated and real-world flight testing (Section 8.4).

If the desired inertial pitch rate is less than the body yaw rate constant (*i.e.* $Q^* < \alpha$) then Equation 8.8a denominator will be negative, and hence the aircraft will roll in the opposite direction to that expected and apply a negative body pitch rate to achieve the desired inertial pitch rate and heading rate. This might seem somewhat unintuitive but the only other option to achieve, for example, a large negative Q^* and a large positive R^*, would be for the aircraft to roll upside down ($\phi > 90°$) and apply a positive body pitch rate. This is obviously undesirable behaviour because dynamics of the aircraft in this condition would differ greatly from those in typical operating conditions and, furthermore, switching between positive and negative pitch rate set-points, Q^*, would require the aircraft to roll back and forth between upside down and upright configurations. To prevent the aircraft from flipping between symmetric solutions to Equation 8.8 (*i.e.* $+q^*/+\phi^*$ and $-q^*/-\phi^*$), the commanded inertial pitch rate is clamped such that $Q^* \geq -\alpha$. Given that $\alpha = 0.47$ for the real aircraft, $Q^* \geq -0.47\text{rad/s}$ ($\approx -27°/\text{s}$).

8.4 Flight test results

The guidance scheme proposed for autonomous interception can be summarised as follows:

- Omnidirectional imagery is captured by the *iEye* vision system (Section 5.3) and the aircraft's 3-DOF attitude is estimated via the visual method described in Section 6.3.
- The target is localised within captured imagery using the appearance-based tracker described in Section 7.3. The output of the tracker is a vector, $\hat{\mathbf{v}}_t$, which describes direction from the aircraft to the target in the inertial coordinate frame – *i.e.* the target LOS vector.
- The *interception guidance law*, Equation 8.3, relates rate of change of $\hat{\mathbf{v}}_t$ in the vertical and horizontal inertial planes, λ_t^v and λ_t^h respectively, to the inertial pitch rate, Q^*, and heading rate, R^*, required by the aircraft to intercept the target.
- The *interception control strategy*, Equation 8.8, defines a combination of roll angle, ϕ^*, and body pitch rate, q^*, which will produce turning rates specified by the *interception guidance law*, Q^* and R^*, for a small-scale, fixed-wing UAV.
- A PID-based control system 'closes the loop' by generating control surface commands for the aircraft to obtain ϕ^* and q^*, thereby realising the interception guidance scheme.

The PID-based control system is illustrated in Figure 8.3. Roll angle and body pitch rate set-points are generated by the *interception control strategy* using inertial pitch and heading rates required for interception and the aircraft's current pitch angle, θ. A PID controller compares the aircraft's current roll angle with the set-point to generate a roll rate set-point using the aircraft's current roll rate as derivative feedback. Body roll rate and pitch rate set-points, p^* and q^* respectively, are translated to aileron and elevator commands, a^* and e^* respectively, according to a linear model:

$$a^* = m_a \cdot p^* + c_a,$$
$$e^* = m_e \cdot q^* + c_e,$$

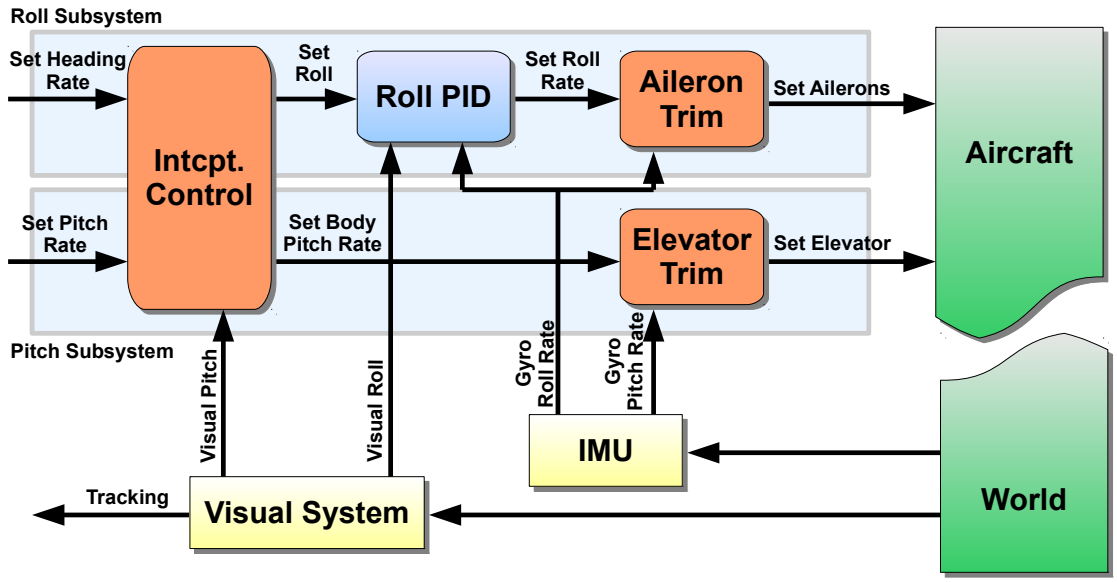


Figure 8.3: Block diagram illustrating the closed-loop control scheme used for interception guidance in simulated and real flight tests. Details in text.

where m_a and m_e are the efficacies of the ailerons and elevators respectively, and c_a and c_e are the corresponding trim mid-points. Efficacies and mid-points were estimated online continuously by comparing commanded aileron and elevator positions with measured body roll and pitch rates. Control lag present in the system was accounted for by choosing a previous control surface command to match current measured rotation rate (a lag of $\sim$ 80ms was used for the test aircraft). A scrolling window of the previous 100 matching command and measurement pairs was used to apply a separate linear least-squares fit for each aircraft axis to obtain m_a and c_a for ailerons, and m_e and c_e for elevators. Regression was performed using the Geometric Tools [2010] mathematics library.

Aileron and elevator efficacies and mid-points estimated automatically during a flight test are plotted in Figure 8.4. The estimated trim parameters (m_a, m_e, c_a, and c_e) evolved from their initial (guessed) values and then remained approximately constant whilst the aircraft flew at approximately constant speed (80s $\sim$ 460s), demonstrating that consistent measurements were made by the system. Figure 8.4 shows that significant negative elevator trim (a quarter of the total available negative deflection!) and less significant negative aileron trim were required for level flight during this flight test. These values roughly correspond to trim positions set by the human pilot for manual control during this flight test. Averaging efficacies estimated

during the flight indicates that maximum achievable rotation rates about the body roll and pitch axes were both approximately $100°s^{-1}$ at this flight speed.

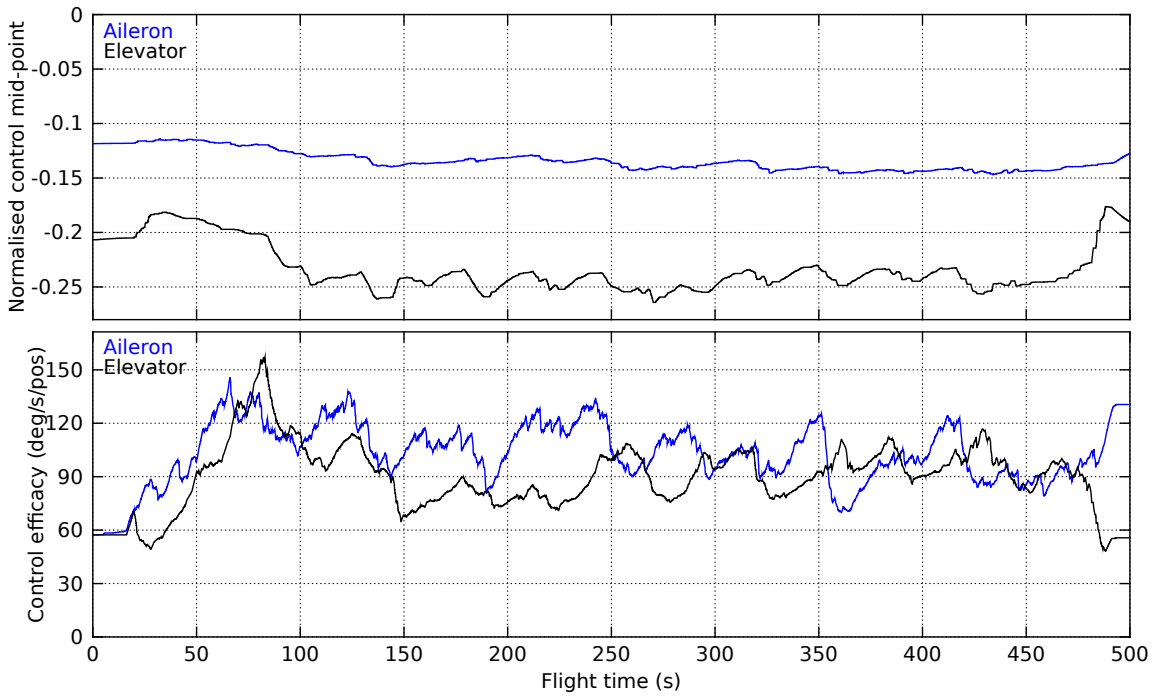


Figure 8.4: Aileron (blue) and elevator (black) trim mid-points (top) and efficacies (bottom) estimated automatically during a flight test. Control surface deflections are normalised so full range is $[-1, 1]$. Efficacies are given as rotation rate (°/s) per unit deflection, *pos* (*i.e.* rotation rate that would be experienced at maximum deflection). A complete flight is shown. Take-off occurs at $\sim$ 80s and touchdown at $\sim$ 460s. The aircraft travelled at approximately constant speed during flight.

The novel approach described above allows control surface efficacies and mid-points to be estimated online during flight, which permits the control system (depicted in Figure 8.3) to adapt to changing flight conditions (*e.g.* airspeed) automatically. Efficacy and mid-point estimates are updated continuously, during manual or automatic control, but were only used to modify control commands during autonomous flight.

8.4.1 Simulated interception

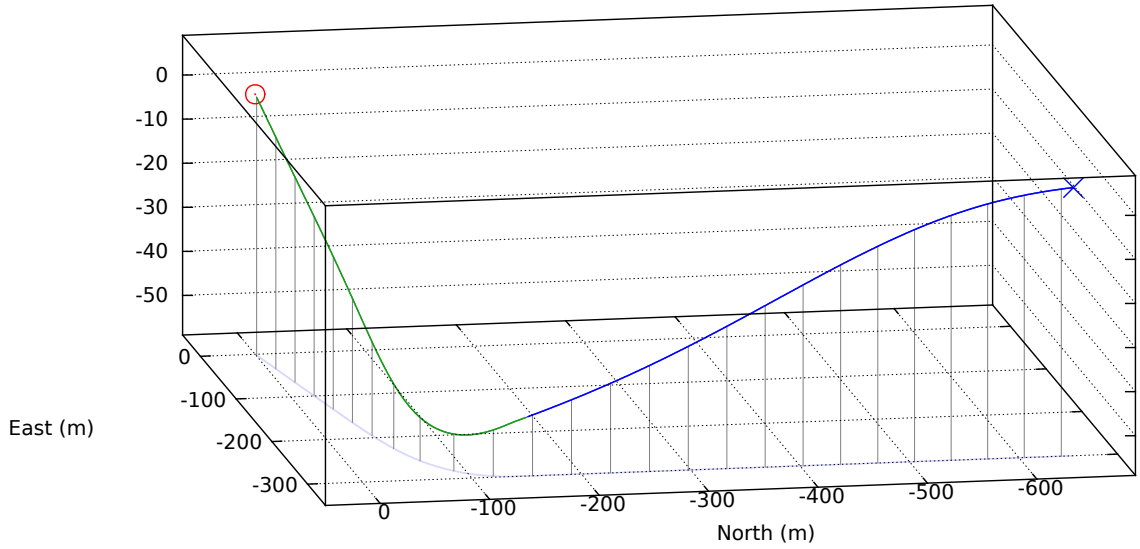
The *interception control strategy* described in Section 8.3 was validated by simulating various interception scenarios. Simulations were based on the open-source *zSim* flight simulator [Hubert, 2008] and used a trainer aircraft model similar, but not identical,

to the flight test aircraft. The closed-loop control system (Figure 8.3) operated at 25Hz using PID parameters tuned for dynamics of the simulated aircraft and feedback (aircraft attitude and rotation rates) provided by the simulator. An easily distinguishable spherical target was used during simulated flight tests to simplify the tracking procedure. Ideal navigation gain, N, was determined empirically to be $N = 30$ for the horizontal and vertical axes (refer to Equation 8.3).

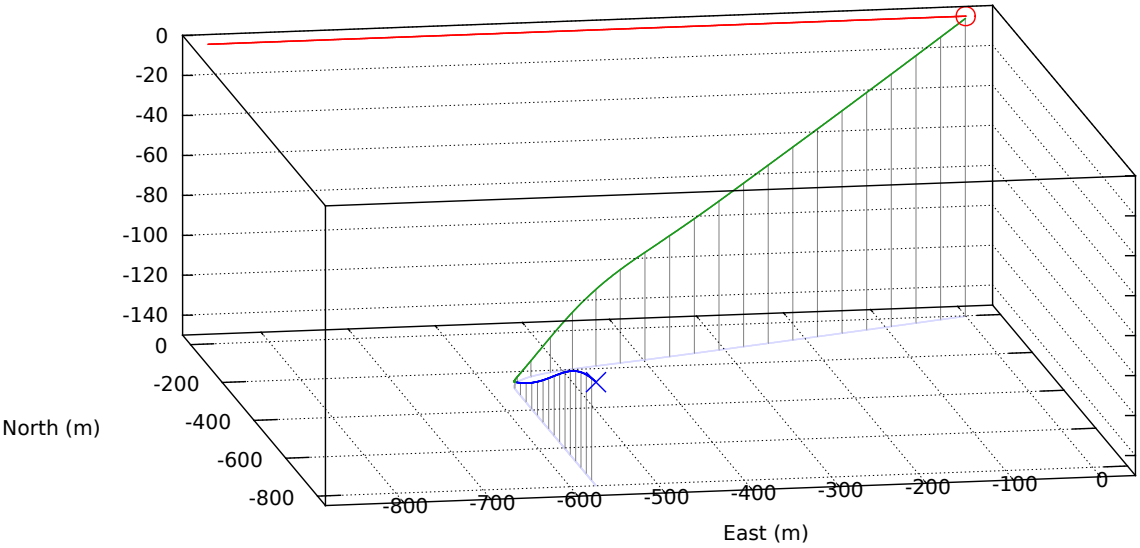
Different scenarios were simulated to analyse capabilities of the interception guidance scheme. Results of two trials are plotted in Figure 8.5, in which the aircraft intercepted a stationary and a constant-velocity target. The static target and constant-velocity target simulated flight tests are shown in accompanying videos[1]. The aircraft initially flew an open-loop path at level attitude, which resulted in a straight-line north-bound course and loss of altitude (altitude was lost at a level pitch angle because the simulated aircraft would have required a positive angle of attack to maintain altitude at this flight speed). After some time, automatic control was enabled and the aircraft turned quickly onto a course for interception of the target. For the static target test, the aircraft was approximately 335m from the target in the horizontal plane and 45m below the target in the vertical plane at the time automatic mode was enabled. For the constant-velocity target, horizontal range to the target at the moment automatic mode was enabled was approximately equal to the static target case. However, the target's motion – east-bound at a constant 100m above the initial altitude of the aircraft – was perpendicular to the original course of the aircraft, hence the aircraft had to travel further to reach the point of interception.

It can be seen from Figure 8.5a that the aircraft followed a straight-line 3D course towards the static target once automatic mode was enabled. This trajectory could have been produced simply by guiding the aircraft towards the apparent relative position of the target, because the target was stationary in this case. For a constant-velocity target, however, simple pursuit would have resulted in a curved trajectory for the pursuer (*e.g.* Figure 8.1a). It can be seen from Figure 8.5b that during the period of autonomous control, the aircraft in fact flew a straight-line 3D course towards a point in front of the instantaneous position of the moving target, thus resulting in a perfect interception of the target via the shortest possible route. This result validates the guidance scheme described in Section 8.3 and demonstrates autonomous

[1]Attachments AV13 & AV14 respectively. Note that position of the aircraft is plotted in these videos showing open-loop flight (black) and closed-loop interception (red); position of the target is also plotted (blue).



(a) Simulated interception of a *static* target.



(b) Simulated interception of a *constant-velocity* target.

Figure 8.5: 3D trajectories of the simulated aircraft and target during interception of (a) a static target and (b) a target moving with constant velocity. In both tests, the aircraft was initially at the position marked by (×) and flew an open-loop path (blue) for 14.5s. Autonomous tracking was then enabled (green) and the aircraft turned quickly onto a straight-line 3D course to intercept the target (red). The point of interception is indicated by (○). The 2D projection of the flight paths is plotted (light blue) and height indicators (grey) are drawn at 1s intervals. Distances are given with respect to the final position of the target. Average aircraft speed was $\sim 34\text{ms}^{-1}$ during autonomous control and the constant-velocity target moved at 25ms^{-1}.

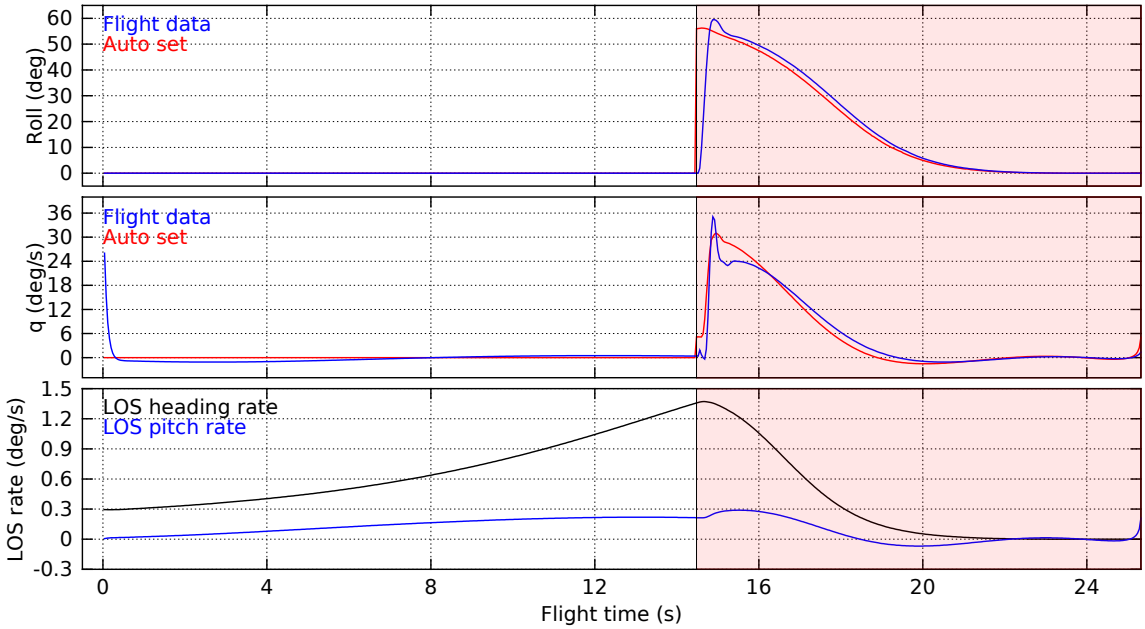
interception of a moving target in 3D space, using only the apparent direction to the target and 3-DOF attitude of the aircraft.

Performance of the *interception control strategy* is analysed in greater detail in Figure 8.6. It can be seen that the rate of change of the viewing direction (LOS) to each target was quickly nullified for both the static and constant-velocity target scenarios, once automatic control was enabled. Roll rate and body pitch rate set-points computed by the *interception control strategy* were followed closely, although not ideally, by the simulated aircraft. Despite small inaccuracies in the closed-loop response of the aircraft, simulated flight test data presented in Figure 8.5 & 8.6 demonstrates that the principles of the *interception guidance law* are valid and that the *interception control strategy* is functioning as expected.

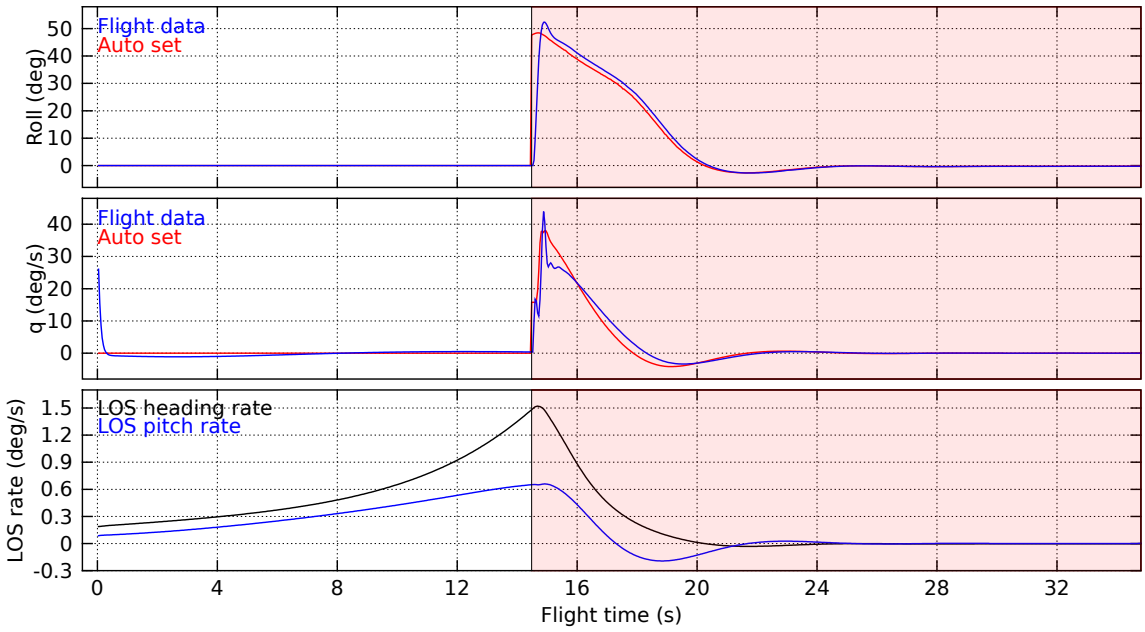
8.4.2 GPS-based interception using a small-scale UAV

The interception guidance scheme was implemented on board a small-scale UAV to analyse performance in an uncontrolled, outdoor environment. Target motion and tracking were simulated via GPS so that the guidance strategy could be analysed independently of the target tracking front-end. Targets were represented by either stationary or moving GPS coordinates, and a GPS unit (U-blox LEA-4H) was mounted on board the aircraft to provide the guidance system with real-time measurements of the aircraft's inertial position. GPS coordinates of the target and aircraft were used to compute a 3D vector that defined metric position of the target relative to the aircraft. This vector exactly specified instantaneous viewing direction, or LOS, to the target from the observing aircraft in the inertial coordinate frame, which would have been the output of the tracking stage, $\hat{\mathbf{v}}_t$. Computing rate of change of $\hat{\mathbf{v}}_t$ in the horizontal and vertical inertial planes gives λ_t^v and λ_t^h, which are the inputs for the *interception guidance law* (Equation 8.3). Visual tracking was thus simulated using GPS to test performance of the interception guidance scheme with hardware in the loop. Note that GPS-defined positions of the aircraft and target were used only to compute viewing direction to the simulated target, which would otherwise have been measured by the appearance-based tracker described in Section 7.3. Neither the 3D position of the aircraft nor that of the target was used (or is required) for guidance.

Ideal navigation gain, N, was identified for conditions under which the flight testing was conducted by flying a number of trials in which the aircraft intercepted



(a) Simulated 3D interception of a *static* target.



(b) Simulated 3D interception of a *moving* target.

Figure 8.6: Flight data from simulated 3D interception of (a) a static target and (b) a target moving with constant velocity. Each subfigure shows roll angle (top) and body pitch rate (centre) commanded by the control system (red) and reported by the simulator (blue). Angular rates of change of LOS to the target (bottom) in the horizontal (black) and vertical (blue) planes are also plotted. The autonomous segment of each flight is highlighted (shaded red). Data corresponds to trajectories plotted in Figure 8.5.

a static target and the value of N was increased for each trial. Figure 8.7 shows the trajectory of the aircraft during five trials, for $N = 2, 4, 6, 8, 10$. In most cases the aircraft was able to pass through the coordinates of the target. However, low values of N resulted in curved aircraft trajectories to the target and $N = 10$ produced an unstable trajectory that was terminated early. $N = 7$ was selected as the optimal navigation gain for flight testing because it enabled the aircraft to turn quickly onto a course for interception without introducing oscillations to the trajectory.

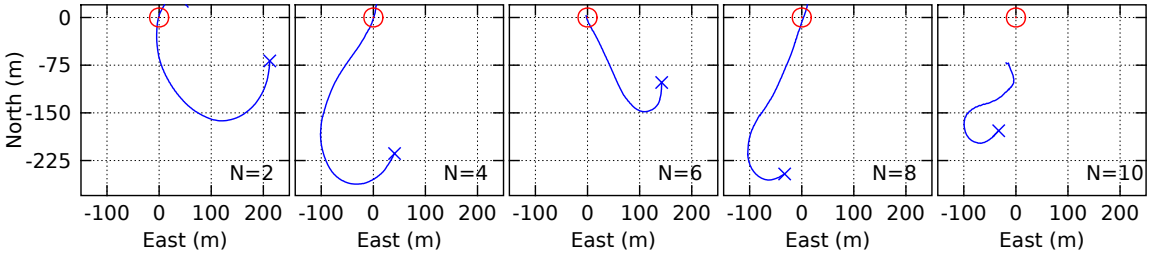


Figure 8.7: Interception of a static GPS target with navigation gains $N = 2, 4, 6, 8, 10$ (left to right). Autonomous flight segments, commencing at ($\times$), are plotted (blue) with the GPS-defined position of the target ($\circ$). Only horizontal 2D components of trajectories are shown for clarity.

Interception of a stationary GPS-defined target was demonstrated (Figure 8.8) using the approach described above with $N = 7$. The aircraft turned quickly onto a straight-line 3D course to intercept the target, confirming that the *interception control strategy* is capable of providing guidance for interception in a real-world environment. Performance of the control strategy is analysed in greater detail in Figure 8.9; flight data corresponds to the 3D trajectory plotted in Figure 8.8.

The *interception control strategy* ideally generates roll angle and body pitch rate commands that nullify the observed rate of change of the direction of the target LOS vector (as for the simulated trials, Figure 8.6). It can be seen from the horizontal and vertical components of the target LOS rate plotted in Figure 8.9 that the control strategy was able to achieve this for a real-world interception scenario. As the aircraft neared the target, however, small deviations in the aircraft's position produced increasingly larger changes to the target LOS vector, which were translated to increasingly larger roll angle and body pitch rate commands by the *interception control strategy*. This phenomenon effectively increases navigation gain as the aircraft nears the target. If control lag – due to measurement delays as well as aircraft dynamics – is significant, then an increase in effective navigation gain could introduce oscillations to the aircraft's trajectory in the terminal stages of interception. The

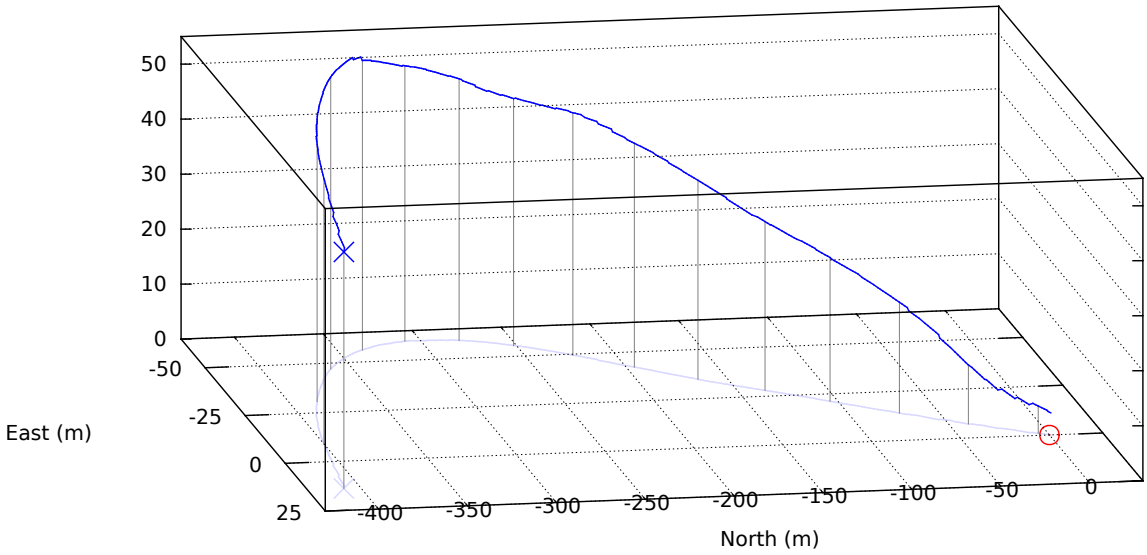


Figure 8.8: 3D interception of a stationary GPS target. A segment of autonomous flight, commencing at (×), is plotted (blue) with the GPS-defined position of the target (○). The 2D projection of the aircraft's trajectory is plotted (light blue) and height indicators (grey) are drawn at 1s intervals. Distances are given with respect to the position of the target.

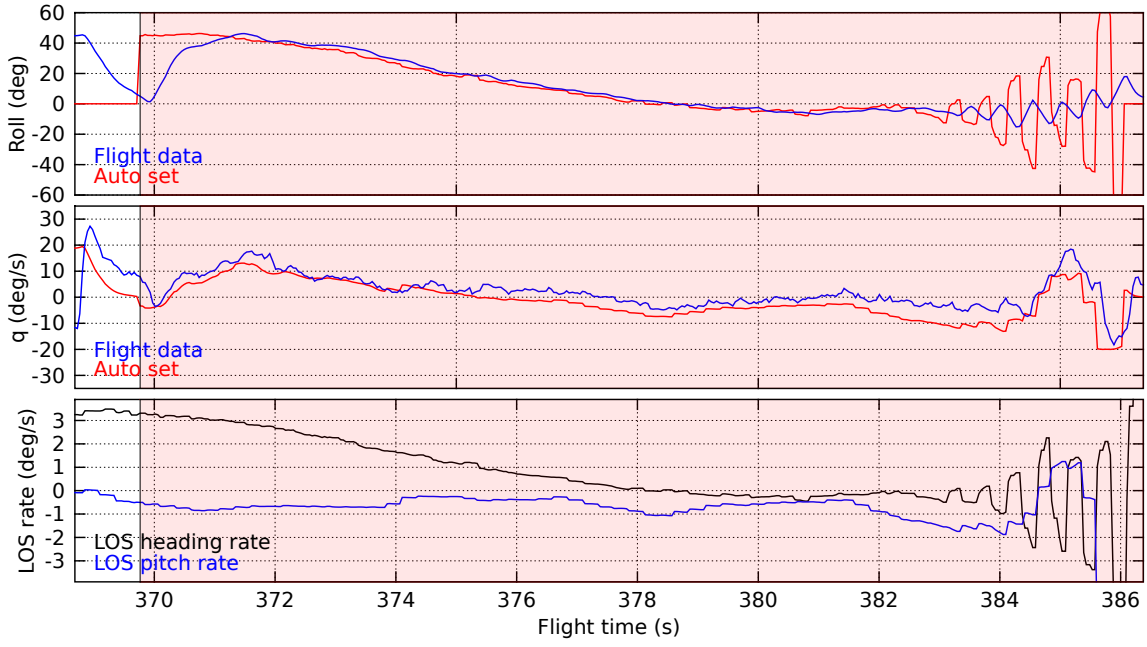


Figure 8.9: Flight data from 3D interception of a stationary GPS target. The plot shows roll angle (top) and body pitch rate (centre) as commanded by the control system (red) and estimated visually (blue). Angular rates of change of LOS to the target (bottom) in the horizontal (black) and vertical (blue) planes are also plotted. The period of autonomous control is highlighted (shaded red). Data corresponds to the trajectory plotted in Figure 8.8.

aircraft's 3D trajectory (Figure 8.8) shows that in this case there were no significant deviations from an ideal straight-line trajectory towards the target. Oscillations in the commanded roll angle and body pitch rate immediately prior to interception, however, may have contributed to the 'miss distance' (the smallest separation between the aircraft and target during the test) of 4.1m.

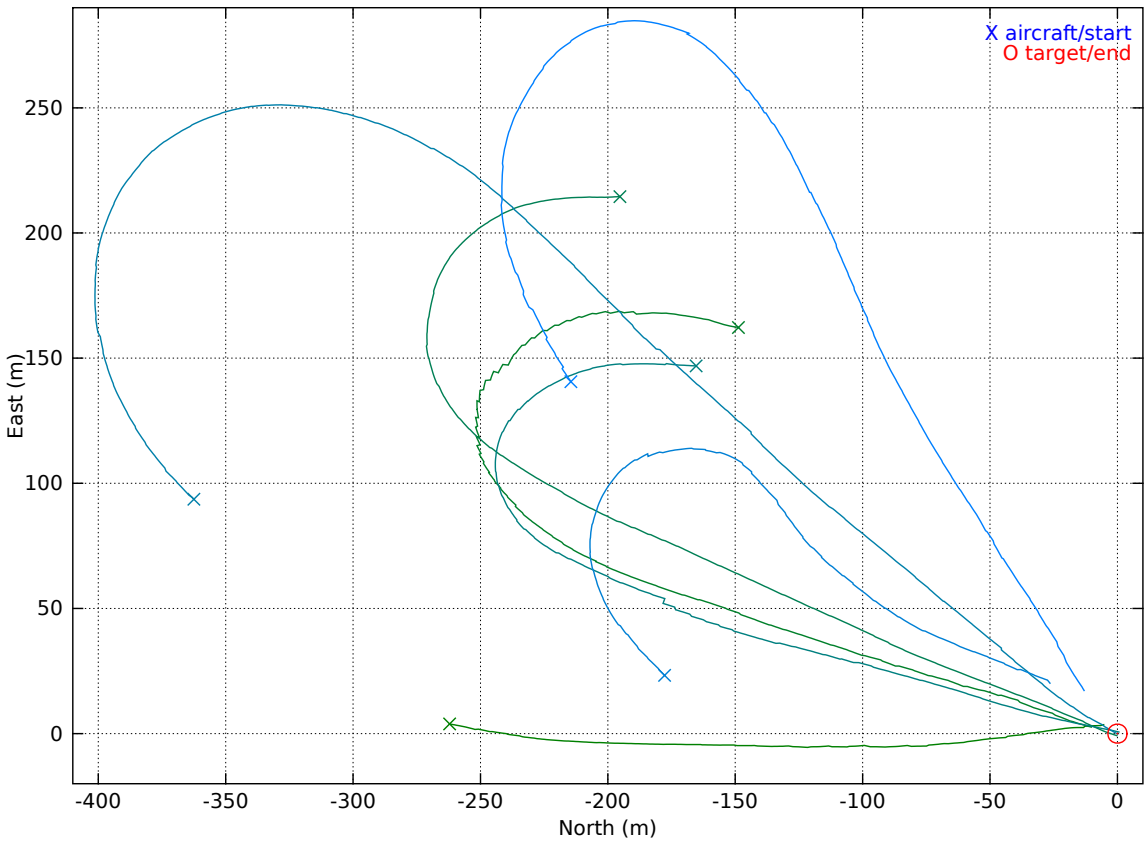
The simulated flight data (Figure 8.6) demonstrates that in the absence of control lag and measurement noise, oscillations in the aircraft's trajectory or dynamics, due to an increase in effective navigation gain, are delayed or prevented from occurring. On board the real aircraft, where such delays are inevitable, one strategy for mitigating impacts of increasing effective navigation gain is to reduce N during terminal stages of interception. Justh and Krishnaprasad [2006], Reddy et al. [2006] describe a control strategy for motion camouflage – certain forms of which produce trajectories identical to interception – that uses range to the target to reduce gain in the terminal phase of flight. In nature, animals whose predatory or other behaviours adhere to a CATD strategy might make use of echolocation (*e.g.* some species of bats [Ghose et al., 2006]) or apparent target size to sense the decreasing range to the target and use it to perform a function equivalent to reducing navigation gain at the final stages of interception. If target size is known *a priori*, then metric range to the target can be estimated here via the method described in Section 7.3. This is not too restrictive, because target range is used only to modify navigation gain and therefore does not need to be precise (*i.e.* it would be sufficient to recognise when the target was close by, rather than estimate relative 3D position).

For the results presented here, navigation gain, N, was modified according to

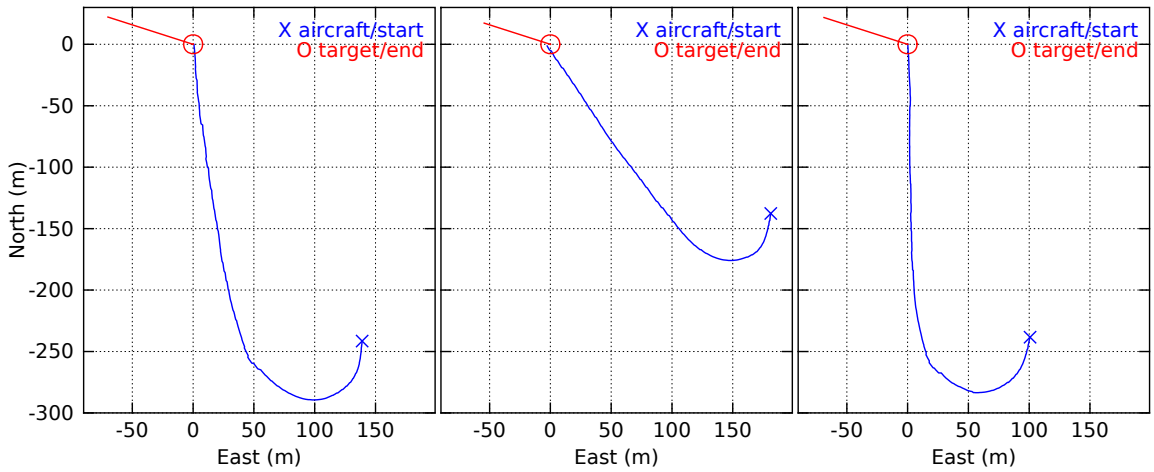
$$N_r = \begin{cases} N \cdot \frac{r_t}{r_0} & \text{if } r_t < r_0, \\ N & \text{otherwise,} \end{cases} \tag{8.9}$$

where N_r is the damped navigation gain; r_t is target range; and r_0 is a scale length for the problem (here $r_0 = 100$m). Navigation gain was reduced in proportion to target range when $r_t < r_0$. During flight testing, r_t was obtained directly from GPS, but it could have been estimated by the appearance-based tracker (Section 7.3). An alternative scheme for damping navigation gain that does not require metric measurements of target range is discussed in Section 8.5.

A flight test comprising a number of static-target and moving-target interception trials was conducted to quantitatively analyse performance of the guidance scheme.



(a) Interception of a *static* GPS target. Colours represent different trials.



(b) Interception of a *constant-velocity* GPS target.

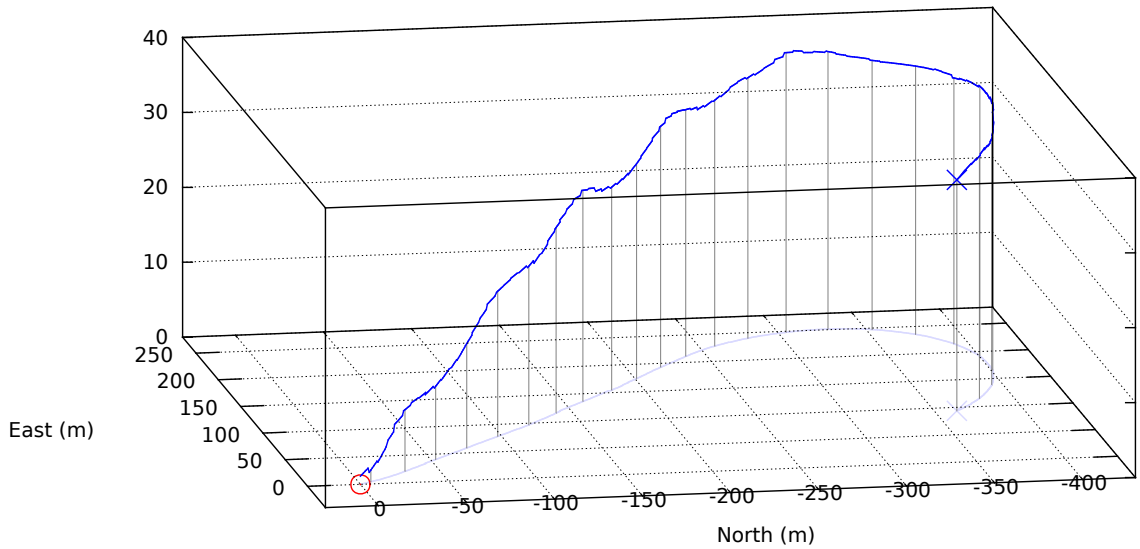
Figure 8.10: Trajectories of the aircraft and target in the horizontal plane during 3D interception of (a) a stationary GPS target and (b) a GPS target moving with constant velocity. The aircraft's trajectory was obtained from GPS. Navigation gain was damped according to Equation 8.9. Autonomous flight segments, commencing at (×), are plotted (green/blue) with the GPS-defined position of the target (red), and the final position of the target at interception (○). Average aircraft ground speed was $\sim 25\text{ms}^{-1}$ during autonomous control and the constant velocity target moved at 5.2ms^{-1}.

Autonomous segments of the aircraft's trajectory during the flight test are plotted in Figure 8.10. Each target was intercepted by the aircraft in the horizontal and vertical planes simultaneously, although for clarity, only 2D (horizontal) positions of the aircraft and target are plotted in Figure 8.10. The 3D positions of the aircraft and target during two of the trials are plotted in Figure 8.11. The aircraft followed a straight-line 3D course towards the point of interception with both stationary and constant-velocity targets, demonstrating that the aircraft was in fact *intercepting* the target, rather than simply pursuing it. This can be seen clearly from horizontal trajectories of the aircraft plotted in Figure 8.10b and the 3D trajectory of the aircraft plotted in Figure 8.11b.

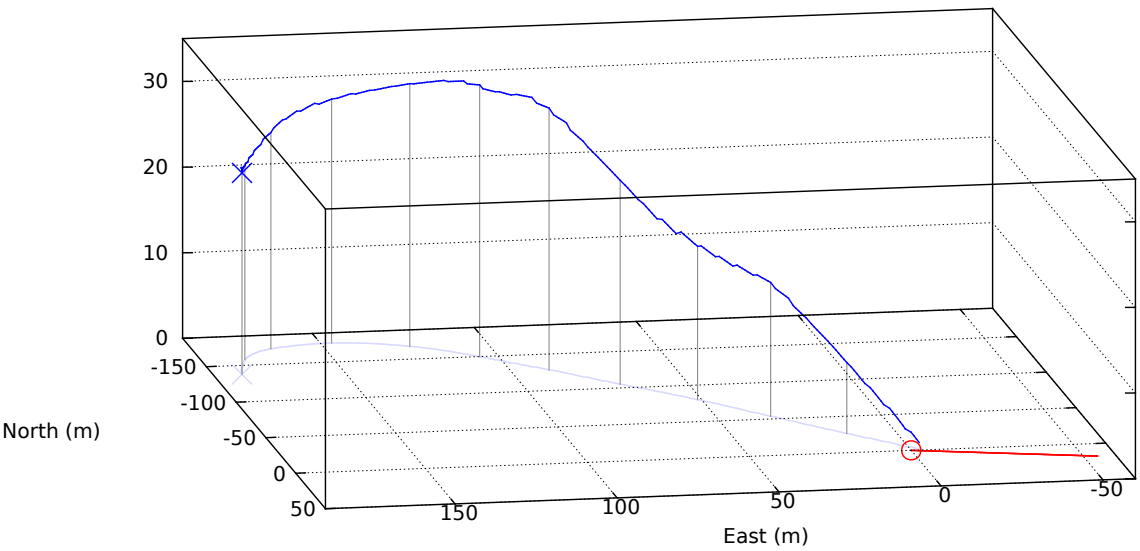
The average closest (3D) distance between the aircraft and targets during the flight test (*i.e.* the average miss distance at interception) was computed to be 1.10m for stationary targets and 1.79m for targets moving at a constant velocity. Target speed during the moving-target trials was 5.2ms^{-1}, or approximately 20% of the aircraft's average ground speed. Hence, the target would have travelled the average miss distance (1.79m) in approximately one quarter of a second. Figure 8.10b indicates that the aircraft was able to select a stable course at least 250m from the eventual point of interception with a moving target. These two observations show that the guidance scheme allowed the aircraft to predict the target's motion so precisely that it was able to fly more than 250m without making any significant course corrections and pass through an arbitrary point in space within $\frac{1}{4}$s of the target.

Guidance commands computed during interception of a moving target are shown in Figure 8.12; flight data corresponds to the 3D trajectory plotted in Figure 8.11b. Comparing the waveforms plotted in Figure 8.12 with those plotted in Figure 8.9 shows that scheduling navigation gain minimised oscillations in the interception guidance commands as the aircraft approached the target. However, the smoothness, straightness, and miss distance pertaining to the 3D interception trajectory resulting from use of a damped navigation gain, N_r (Figure 8.11b), are not very different to those pertaining to the interception trajectory resulting from use of an undamped navigation gain, N (Figure 8.8). This finding supports the claim that accurate target range measurements are not critical for the proposed approach.

The GPS-based flight test results analysed here confirm the findings from the simulated flight tests – that the interception guidance scheme, developed in this research and described in Section 8.3, allows a small-scale UAV to anticipate the motion of an object moving at constant velocity and to follow the shortest path



(a) Interception of a *static* GPS target.



(b) Interception of a *constant-velocity* GPS target.

Figure 8.11: 3D trajectories of the aircraft and target during interception of (a) a stationary GPS target and (b) a GPS target moving with constant velocity. The aircraft's trajectory was obtained from GPS. Navigation gain was damped according to Equation 8.9. Autonomous flight segments, commencing at (×), are plotted (blue) with the GPS-defined position of the target (red), and the final position of the target at interception (○). The 2D projection of the flight paths is plotted (light blue) and height indicators (grey) are drawn at 1s intervals. Distances are given with respect to the final position of the target.

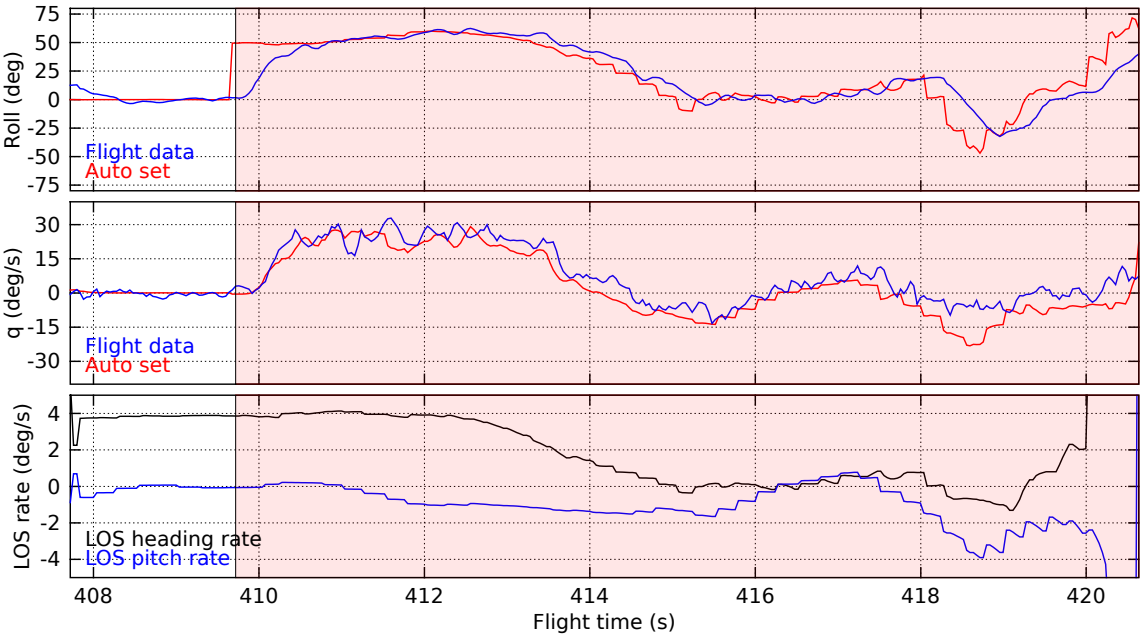


Figure 8.12: Flight data from 3D interception of a GPS target moving at a constant velocity. The plot shows roll angle (top) and body pitch rate (centre) as commanded by the control system (red) and as estimated by the vision system (blue). Angular rates of change of LOS to the target (bottom) in the horizontal (black) and vertical (blue) planes are also plotted. The period of autonomous control is highlighted (shaded red). Data corresponds to the trajectory plotted in Figure 8.11b.

towards a point of interception, using only information provided by the vision system on board the aircraft.

8.4.3 Visual interception using a small-scale UAV

Results presented thus far have demonstrated that a ground-based target can be visually tracked over extended periods of time based on its appearance alone (Section 7.3) and that a small-scale UAV can be guided to intercept a target moving with constant velocity using only observed rate of change of LOS. These two systems were combined to enable a small-scale UAV to track and intercept an independently moving vehicle using only visual information and on-board computing resources for all sensing and guidance. A flight test demonstrating visual interception of a stationary ground-based target is shown in the accompanying video[2]. The ground-

[2]Attachment AV15. Note that data and imagery shown in this video are explained in Figure 7.3; also shown is an indication of when the interception guidance scheme has fully autonomous control of the aircraft, as well as a small crop from the raw omnidirectional imagery that is centred on the

based target was the same target that was used for testing performance of the appearance-based tracker. A tracking window size of 15px × 15px was used and the target was initialised mid-flight at the beginning of each trial using approximate GPS coordinates and colour. This procedure enabled the tracker to reliably select the same target from various initial positions and orientations, after which the target was tracked robustly using the method described in Section 7.3.2. Navigation gain, N_r, was damped according to Equation 8.9 using visual LOS to the target scaled by the height of the aircraft AGL (obtained from GPS). To preserve aircraft, the target was intercepted in the horizontal plane only, but simluated and GPS-based flight tests have demonstrated that this approach is valid for 3D scenarios.

Autonomous segments of the aircraft's trajectory during the flight test are plotted in Figure 8.13 along with position of the stationary target. A visual estimate for position of the aircraft relative to the target at each frame was obtained offline by extending the inertial target view vector, $\hat{\mathbf{v}}_t$, such that it intersected the ground plane (using GPS height); relative horizontal position of the target with respect to the aircraft was then computed simply from the x and y components of $\mathbf{v}_t$. This information was used only for analysis offline and for damping navigation gain during the terminal stages of interception (via Equation 8.9); it was not used directly for guidance. The GPS-derived trajectory of the aircraft during the flight test is shown in Figure 8.13 for comparison. Visual estimates of the aircraft's position relative to the target did not agree well with GPS measurements at the beginning of each trial because at these points the aircraft viewed the ground-based target at a shallow angle and thus small angular localisation errors translated to large position errors. It can be seen from the GPS trajectories that in each case the aircraft turned quickly onto a straight line course to intercept the target. The average (2D) miss distance (computed from visual LOS range) for these trials was 0.80m, indicating that the target was tracked accurately and the guidance system provided stable control commands.

The interception trial corresponding to the trajectory marked by (+) in Figure 8.13 was used to analyse function of the tracking and guidance stages in greater detail. Figure 8.14 displays intermediate tracking steps at regularly spaced intervals throughout the trial. Initially the aircraft was approximately 300m horizontally from the target and about 60m AGL. At this range the 8m × 4m target would subtend approximately 1° if viewed orthogonally to its long axis, which was oriented

direction of vertically down in the inertial reference frame, which shows the miss distance as the aircraft passes over the target.

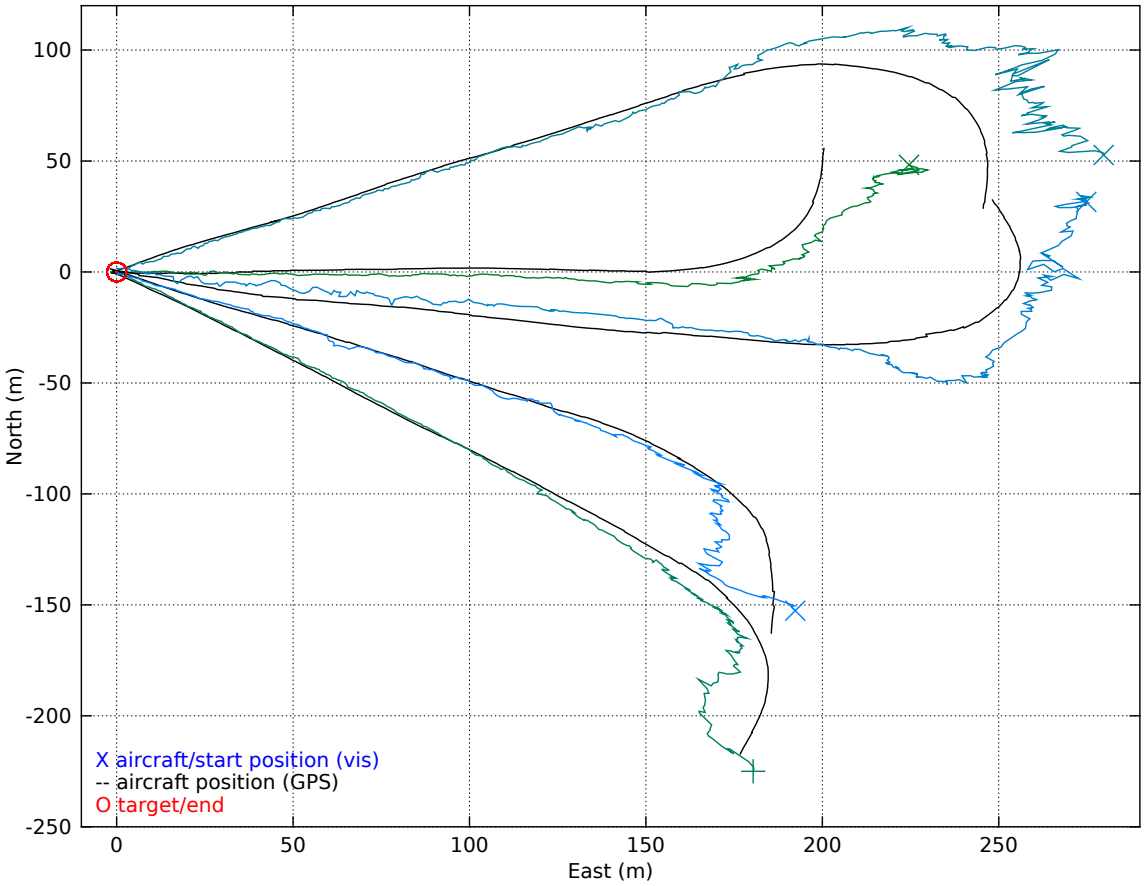


Figure 8.13: Trajectories of the aircraft and position of the target in the horizontal plane during 2D visual interception of a stationary target. Aircraft trajectories (commencing at $\times$) were obtained visually using LOS range and direction to the target (green/blue) and also directly from GPS (black). Distances are given with respect to the position of the target ($\circ$). GPS-derived aircraft trajectories were aligned with visual trajectories using approximate GPS coordinates for the target.

approximately east-west. It can be determined from Figure 8.13 that the initial positions of the aircraft for each of the closed-loop trials would have caused it to perceive a much narrower view of the target. Despite the scarcity of pixels corresponding to the target in the raw omnidirectional images, Figure 8.14 shows that the tracker was able to initialise the target successfully and track it continuously as the aircraft manoeuvred.

Flight data corresponding to the same interception trial is plotted in Figure 8.15. Immediately following activation of autonomous control the *interception control strategy* commanded the aircraft to bank left to cancel the observed rate of change of horizontal LOS to the target. Horizontal LOS rate was quickly nullified, satisfying the

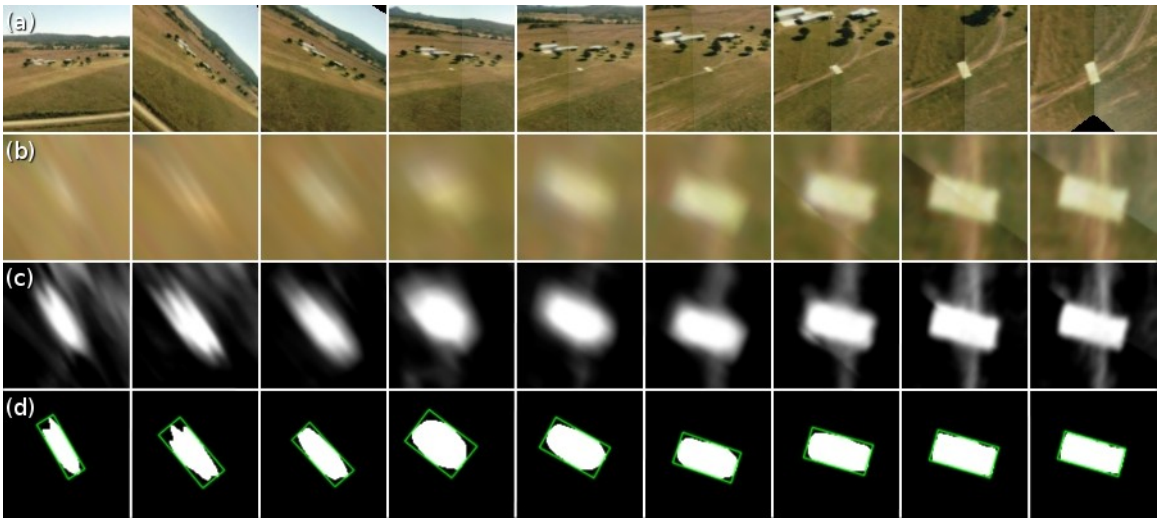


Figure 8.14: Sequence of frames illustrating target tracking during vision-based interception showing: (row a) constant-zoom (40° FOV) crop from raw omnidirectional imagery centred on localised target; (row b) orthographic tracking window; (row c) transformed tracking window; and (row d) segmented target blob overlaid with detected bounding box. See Section 7.3.1 for a description of the intermediate tracking steps. Frames are separated by approximately 2s and cover an entire interception trial – from initialisation at a (horizontal) range of approximately 300m to passing directly over the target.

interception guidance law, and the aircraft followed a stable trajectory to the point of interception. Neither target LOS rate in the vertical plane nor height of the aircraft were controlled in this flight test, hence the guidance system attempted to maintain zero inertial pitch rate, $Q^* = 0$, during interception. Comparing Figure 8.15 with Figure 8.12 indicates that stabler interception trajectories were achieved by tracking the target visually rather than simulating it using GPS, which is probably due in part to the frequency of visual measurements (25Hz) compared to the frequency of GPS measurements (4Hz). This finding is supported by the average miss distances computed for the two methods (0.8m for visual interception and 1.1m for GPS interception), although in this case miss distance is a 2D measurement as opposed to a 3D measurement for GPS-based flight tests.

Visual tracking and interception of a moving ground-based vehicle was not able to be demonstrated during this study due to time constraints. However, accumulated real-world flight test results, from both GPS-based and visual interception trials, demonstrate that the interception guidance scheme derived and presented in this chapter provides a practical and effective means by which a small-scale UAV can interact with an independently moving object or vehicle. Results from vision-only

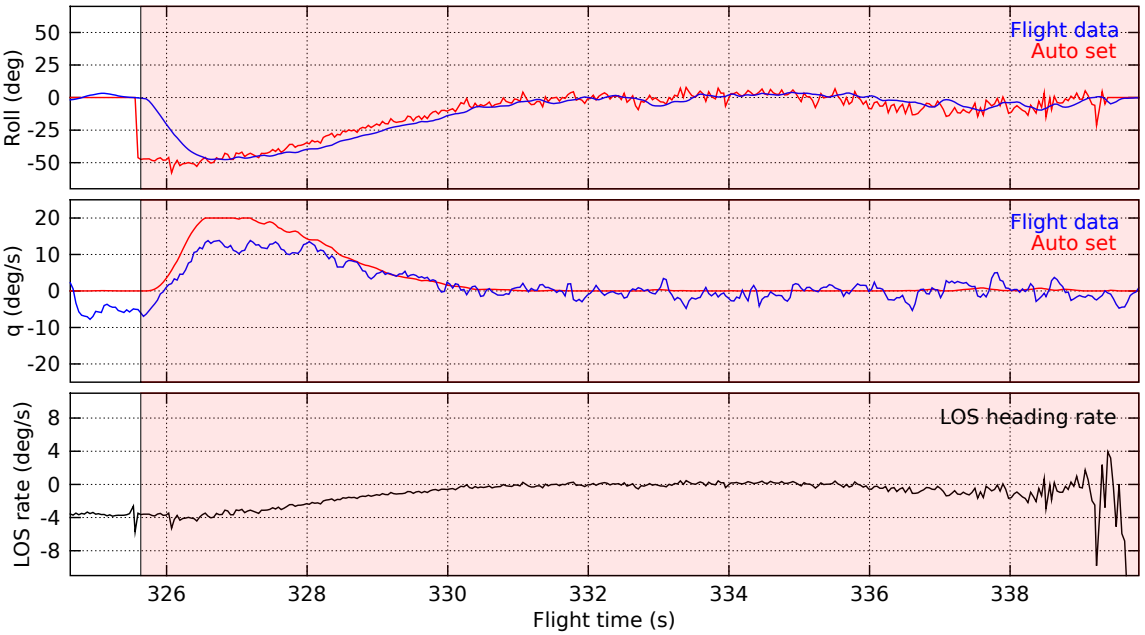


Figure 8.15: Flight data from 2D visual interception of a stationary target. The plot shows (top) roll angle and (centre) body pitch rate as commanded by the control system (red) and estimated visually (blue); and (bottom) angular rate of change of horizontal LOS to the target. The period of autonomous control is highlighted (shaded red). Data corresponds to the trajectory plotted in Figure 8.13 commencing at (+).

trials further demonstrate that the guidance system operates solely and reliably on information provided by an appearance-based tracker, enabling real-time tracking and interception of moving objects using vision for all levels of flight control, sensing, and guidance. It is planned to demonstrate docking between a small-scale UAV and a moving ground-based vehicle in continuing work with this system.

8.5 Discussion

Flight test results have demonstrated that the guidance strategy introduced in this chapter provides an effective means for a small-scale UAV to predict the motion of an independently moving vehicle and to follow a direct path to a point of interception. The guidance strategy comprises a set of simple control equations which rely only on target viewing direction, obtained from an appearance-based tracker. Altogether, just 0.5ms @ 1.5GHz is required to visually track a single target and compute guidance for interception. The algorithm described in Chapter 6 for computing 3-DOF attitude executes in 2.9ms @ 1.5GHz. Total time required to perform visual tracking and to

estimate attitude, as well as to compute interception guidance and flight control, is approximately 3.5ms @ 1.5GHz. An additional 15.5ms @ 1.5GHz was taken to measure optic flow and compute translation direction, which were used by the tracking algorithm to predict viewing directions to ground-based targets during occlusions. All sensing, guidance, and flight control algorithms therefore have a combined execution time $<$ 20ms @ 1.5GHz and, furthermore, they rely solely on visual information provided by the *iEye* vision system. Both of these factors make this system ideal for implementation on board small-scale UAVs.

8.5.1 Damping navigation gain

Navigation gain, N, is a function of the interceptor's acceleration capabilities and expected manoeuvres of the target. For a constant-velocity target and a constant aircraft flight speed, N is approximately constant. As the aircraft approaches the target, however, the effect of navigation gain on inertial pitch rate and heading rate set-points (Equation 8.3) is enhanced because deviations in the position of the target will produce larger Q^* and R^* changes when the aircraft is close to the target compared to when it is further away. This phenomenon is desirable because the aircraft needs to respond more quickly to target motions when it is nearby. However, during the terminal stages of interception, small errors in the estimated target LOS vector or the trajectory of the aircraft can cause Q^* and R^* to change rapidly, leading to oscillations in the *interception control strategy*. To overcome this issue during real flight tests (Sections 8.4.2 & 8.4.3), navigation gain, N_r, was decreased proportional to target range (Equation 8.9).

Flight speed also affects response of the aircraft to perturbations of observed target LOS rate: the original PN guidance law (Equation 8.1) states that a pursuer must apply an acceleration to cancel observed target LOS rate, and faster flight speeds will produce greater accelerations for equal turning rates in the case of an aircraft. As with dependence on target range, this phenomenon is desirable because the faster the aircraft flies, the quicker it needs to correct its position in order to intercept a target. Aircraft speed was cancelled, therefore, during derivation of the *interception guidance law* (Equation 8.2). However, during the terminal stages of interception, small errors in observed target LOS rate will cause larger responses from the aircraft for faster flight speeds, possibly leading to control oscillations. In this study the simulated and real world aircraft's flight speeds were held approximately constant by maintaining

a constant throttle setting. In fact, aircraft airspeed would have varied by a small amount during flight testing due to the motions of the aircraft (*i.e.* during gains or losses of altitude and sharp turns). For large variations of flight speed, it may be necessary to measure airspeed so that navigation gain can be damped appropriately.

An alternative scheme for damping navigation gain during the terminal stages of interception, one that would not require metric measurements of target range and aircraft flight speed, relies on estimating the TTC with the target. TTC is a measure of the time that would elapse before an observer reaches an oncoming surface if current relative motion between the observer and the surface were to continue without change. TTC is essentially the ratio of distance to velocity:

$$TTC = \frac{-d_{\text{LOS}}}{\dot{d}_{\text{LOS}}}, \tag{8.10}$$

where d_{LOS} is LOS range to the target and $\dot{d}_{\text{LOS}}$ is its time derivative. For visual systems, TTC can also be defined in terms of fractional expansion of an object as the observer approaches it [Horn et al., 2007]. This representation is unnecessary for this study, however, because range to the target could be provided by the appearance-based tracker (Section 7.3). Range estimates would only be correct to a scale factor (unless metric scale of the target is known *a priori*), but this does not matter for computation of TTC because the scale factor cancels from both the numerator and denominator in Equation 8.10. TTC could therefore be obtained for any tracked object regardless of scale and would provide a measure of 'temporal' proximity of the aircraft to the target, which could be used to dampen navigation gain without requiring metric measurements of the aircraft's flight speed or range to the target.

Flight data recorded during the target tracking flight test (Section 7.3.2) was used to conduct a brief investigation of the practicality of measuring TTC from target range estimates obtained from the appearance-based tracker. A flight segment during which the aircraft approached and flew over a ground-based target was analysed. Visual estimates of both target LOS range and TTC during the pass are plotted in Figure 8.16. GPS-derived altitude was used to provide a comparative measure of LOS range to the target by scaling the inertial target view vector, $\hat{\mathbf{v}}_t$, such that the vertical (z) component was equal to the aircraft's height above the target. GPS-derived target range measurements were then used to compute a comparative measure of TTC.

Figure 8.16 shows that target range estimates obtained from visual scale of the target correspond well with GPS-derived target range estimates. Visual target range

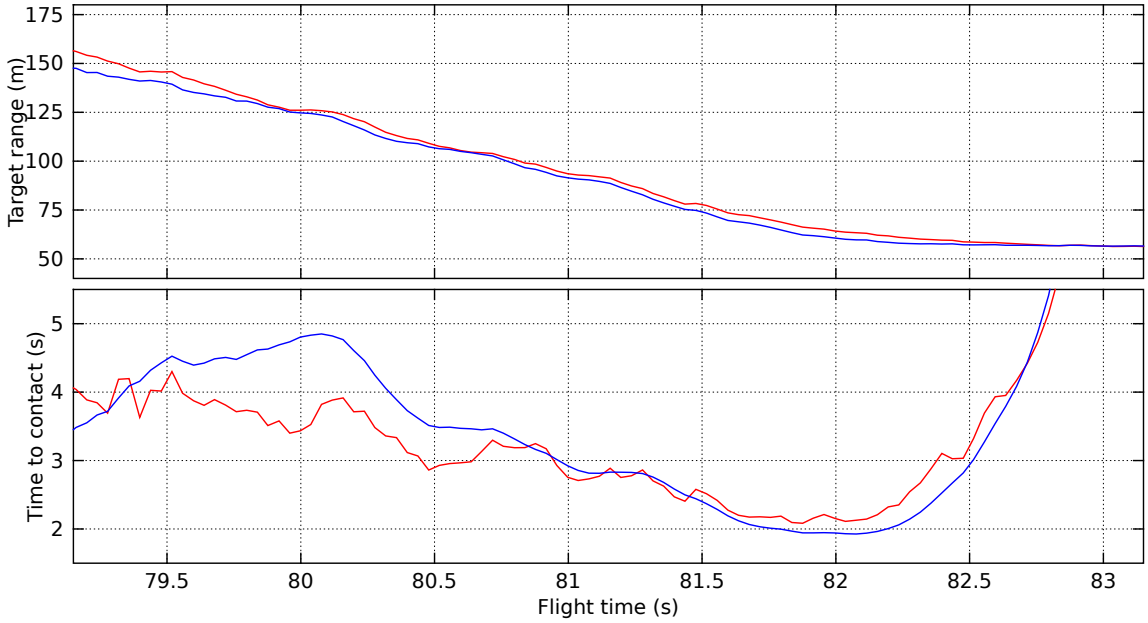


Figure 8.16: An open-loop fly-over of a ground-based target showing (top) range to the target computed from visual target scale (blue) via the approach described in Section 7.3.2, and GPS-based LOS distance to the target (red); as well as (bottom) TTC with the target computed from Equation 8.10 using visual target scale (blue) and GPS-based LOS range (red). Data corresponds to the aircraft's first pass of the target in Figure 7.4.

plotted in Figure 8.16 has been scaled arbitrarily to illustrate its correspondence with GPS-derived LOS range estimates. The scale factor is irrelevant, however, because it cancels during computation of TTC (Equation 8.10). Both visual and GPS-derived range measurements were used to compute separate estimates of TTC during the pass of the target (Figure 8.16). Noise introduced through numerical computation of target range rate, $\dot{d}_{\text{LOS}}$, was reduced by computing average target range rate between target range measurements 1s apart (*i.e.* each target range measurement was compared with target range 1s beforehand to obtain an instantaneous target range *rate* which was then combined with instantaneous target range via Equation 8.10 to obtain TTC).

As range to the target decreased during the first phase of the manoeuvre, TTC decreased approximately proportionally. For an observer travelling at constant velocity towards a stationary target this trend would be expected to continue until collision at $TTC = 0$. In this open-loop flight test, however, the aircraft passed over the target, so as target range plateaued and began to increase, TTC can be seen to increase quickly (*i.e.* as $\dot{d}_{\text{LOS}} \to 0$, $TTC \to \infty$). This observation highlights the fact that the observer must maintain a constant trajectory towards the target for TTC

estimates to be valid. This condition is guaranteed, however, by the interception guidance scheme described here. TTC would thus provide a suitable means for measuring the proximity of a target. Navigation gain, N_r, was damped for target ranges $r_t < 100$m during flight testing. Figure 8.16 demonstrates that TTC estimates obtained visually are reliable over this range. Visual estimates of relative target range provided by the appearance-based tracker could therefore be used to compute TTC, which could be used in turn to dampen navigation gain during the terminal stages of interception without requiring metric measurements of target range or aircraft flight speed.

8.5.2 Intercepting accelerating targets

The PN guidance law (Equation 8.1) guarantees a time-optimal trajectory for intercepting a constant-velocity target, and in both the derivation of the *interception control strategy* as well as the implementation of the guidance scheme, a constant-velocity target was assumed. It is of interest to consider outcomes of this guidance scheme if the target is manoeuvring. Ghose et al. [2006] show that if the trajectory of the target can be represented by a piece-wise constant-velocity trajectory, then the path of the interceptor will still be close to time-optimal – *i.e.* also piece-wise linear. If the target manoeuvres with a constant acceleration, the constant-target-velocity *interception guidance law* will result in a curved trajectory to interception. The path of the interceptor will no longer be the shortest possible, but the target will still be intercepted in most circumstances (*i.e.* if navigation gain can be set high enough to account for the maximum speed and manoeuvrability of the target).

Interception of an accelerating target, using an assumption of constant-velocity, was investigated in a simulated flight test (shown in the accompanying video[3]). Trajectories of the aircraft and target during the flight test are plotted in Figure 8.17. It can be seen that the aircraft successfully intercepted the target, although the path taken by the aircraft was curved and therefore not time-optimal. Acceleration of the target (5.2ms^{-2}) was chosen such that final speed of the target did not exceed that of the aircraft ($\sim 37\text{ms}^{-1}$).

[3]Attachment AV16. Note that position of the aircraft is plotted in this video showing open-loop flight (black) and closed-loop interception (red); position of the target is also plotted (blue).

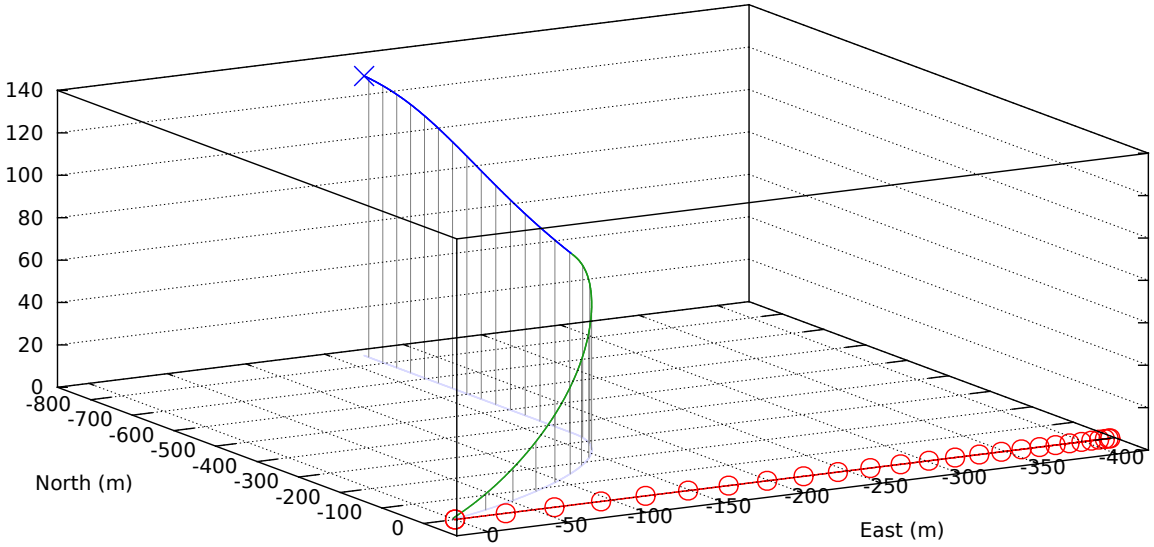


Figure 8.17: 3D trajectories of the simulated aircraft and target during interception of an accelerating target. The aircraft was initially at the position marked by (×) and flew an open-loop path (blue) for 14.5s. Autonomous tracking was then enabled (green) and the aircraft turned to intercept the target (red). The target's position at 1ms^{-1} intervals and the point of interception are marked by (○). The 2D projection of the aircraft's flight path is plotted (light blue) and height indicators (grey) are drawn at 1s intervals. Distances are given with respect to the final position of the target. Average aircraft speed was ~ 37ms^{-1} during autonomous control and the target accelerated at 5.2ms^{-2} from an initial speed of 5.6ms^{-1} (target speed at the beginning of autonomous control was 19ms^{-1}) to a final speed of 32.8ms^{-1}. Note that the aircraft also accelerated slightly (~ 0.7ms^{-2}) as it lost altitude during the period of autonomous control.

8.6 Applications

The interception guidance scheme described in Section 8.3 is useful for various real-world applications where an aircraft must approach an independently moving vehicle – such as for refuelling or return-to-ship manoeuvres – or for detecting other moving vehicles on a collision course. The system is also useful for applications that do not involve a second vehicle, such as landing a fixed-wing aircraft in the presence of a cross-wind. Results of a brief investigation of this application are presented below.

8.6.1 Landing in windy conditions

The *interception guidance law* defines turning rates that should be applied by a pursuing aircraft in order to bring its trajectory onto a straight line 3D path that intercepts the predicted trajectory of a constant-velocity target. From flight test

results presented in Section 8.4 it can be seen that the same guidance strategy also applies to the simpler case of intercepting a stationary target.

The *interception guidance law* operates on rate of change of the inertial target LOS vector. Because the target LOS vector is measured in the inertial coordinate frame, the aircraft is free to rotate about its body axes as necessary during interception. This property would make this guidance scheme particularly useful for landing an aircraft in windy conditions, where the aircraft may need to rotate its body unpredictably in order to maintain a constant descent trajectory towards a landing site. During interception, aircraft body rotations are commanded by the *interception control strategy* to achieve turning rates computed by Equation 8.3. By 'intercepting' a landing site at a shallow descent angle, therefore, the interception guidance scheme described in this chapter could be used to keep an aircraft on a predefined landing trajectory even without knowledge of wind speed or direction.

Application of the interception guidance scheme to landing in windy conditions was investigated in simulation. A flight test in which a simulated aircraft approached a landing strip in the presence of a 10ms^{-1} cross-wind is shown in the accompanying video[4]. The 3D trajectory of the aircraft during the simulated flight test is plotted in Figure 8.18.

During the period of autonomous control, the aircraft was guided to intercept a target positioned on the runway. Control commands were generated by the *interception control strategy* using the setup described for simulated flight testing in Section 8.4. An additional (and optional) input was used here to demonstrate that the aircraft could be guided to approach the landing site at a predefined angle of descent. Target inclination (*i.e.* the vertical component of the inertial target vector, $\hat{\mathbf{v}}_t$) was maintained at a constant angle of $-2.5°$ with respect to the horizontal plane by augmenting the inertial pitch rate set-point (Equation 8.3b) with a term proportional to target inclination error (*i.e* error between the measured target inclination and the set-point of $-2.5°$). Function of the augmented *interception guidance law* during this simulated flight test was therefore equivalent to a proportional-derivative (PD) controller, with $k_P = 3$ and $k_D = N = 30$.

The simulated aircraft initially flew at 34ms^{-1} but its throttle was reduced at the beginning of the flight test to simulate a realistic landing scenario. During

[4]Attachment AV17. Note that position of the aircraft is plotted in this video showing open-loop flight (black) and closed-loop interception (red); position of the target is also plotted (blue). A cross-wind blows from right to left across the runway at 10ms^{-1}.

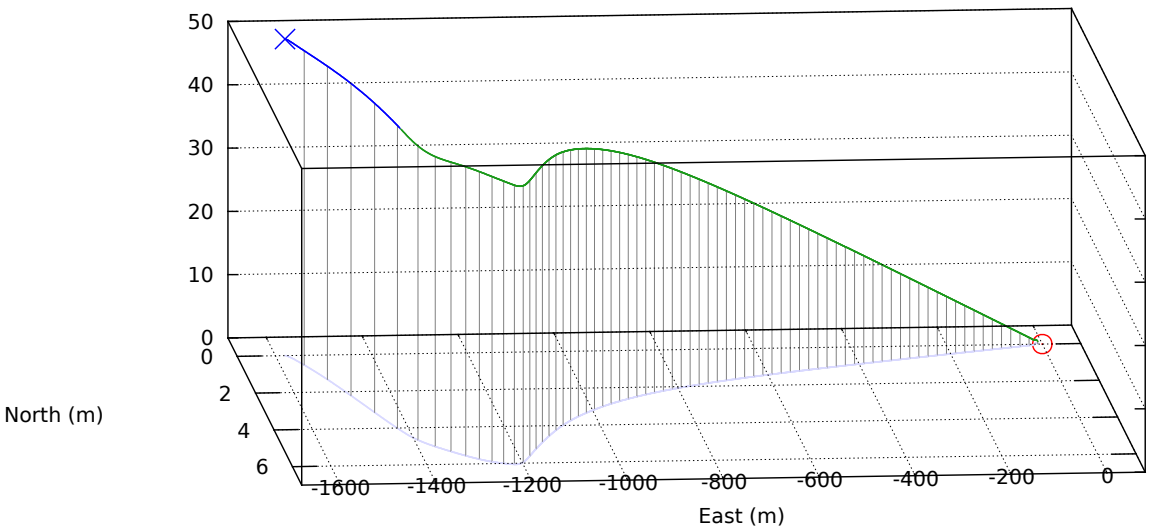


Figure 8.18: 3D trajectory of the simulated aircraft during an autonomous landing in the presence of a 10ms^{-1} cross-wind (simulated wind blew south-north). The aircraft was initially at ($\times$) and flew an open-loop path (blue) until autonomous control was enabled (green) and the aircraft turned to 'intercept' the landing site ($\circ$). The 2D projection of the aircraft's flight path is plotted (light blue) and height indicators (grey) are drawn at 1s intervals. Note the different scales of the north and east axes.

autonomous control the aircraft's flight speed dropped to $\sim$ 20ms^{-1}. The descent angle set-point ($-2.5°$) was chosen to maintain the simulated aircraft's flight speed above a stalling speed. It can be seen from Figure 8.18 that the decelerating flight speed of the aircraft as well as the interplay between the two Q^* set-points (derived from the *interception guidance law* and the descent angle set-point) resulted in a curved trajectory during the initial stages of autonomous flight, but once the flight speed of the aircraft stabilised it was able to follow a straight-line 3D trajectory to the landing site at the desired descent angle of $-2.5°$.

Flight data recorded during the simulated flight test is plotted in Figure 8.19. It can be seen that a non-zero roll angle was commanded by the *interception control strategy* at the commencement of autonomous control to nullify the horizontal component of the rate of change of LOS to the landing strip, which was caused by a sideways drift of the aircraft during its approach due to the 10ms^{-1} cross-wind. Rate of change of the vertical component of target LOS was not nullified as quickly due to the effects of the absolute descent angle set-point.

Figures 8.18 & 8.19 demonstrate that the interception guidance scheme described here is capable of cancelling the effects of an unknown wind field on an aircraft during

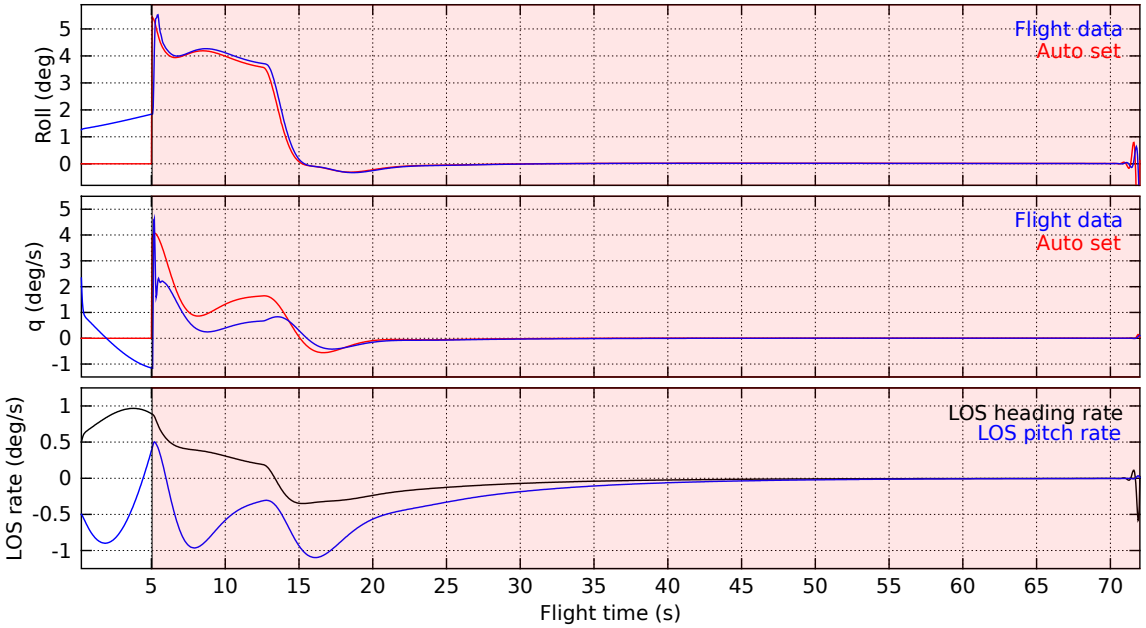


Figure 8.19: Flight data from simulated autonomous landing in windy conditions showing roll angle (top) and body pitch rate (centre) commanded by the control system (red) and reported by the simulator (blue). Angular rates of change of LOS to the target (bottom) in the horizontal (black) and vertical (blue) planes are also plotted. The autonomous segment of each flight is highlighted (shaded red). Data corresponds to landing trajectory plotted in Figure 8.18.

landing approach, allowing the aircraft to follow a predefined trajectory without requiring metric measurements of the aircraft's flight speed or 3D position with respect to the landing site, or measurements of wind speed and direction.

A scheme for landing an aircraft automatically in an uncontrolled environment using visual attitude and optic flow was described in Section 6.6.2. The alternative approach described here, however, enables an aircraft to pin-point an exact landing site and to maintain a constant approach trajectory even in the presence of disturbances. Ideally the two techniques would be combined to enable the aircraft to identify a suitable landing site from the air and to follow an appropriate approach trajectory using the interception guidance scheme; then use the optic flow-based landing scheme to perform a safe touchdown at the desired landing site.

8.7 Conclusion

Ability to intercept a target gives the pursuing animal or vehicle two distinct advantages: the trajectory followed by the pursuer will result in capture of the target in the shortest possible time along the shortest possible path; and the relative motion of the pursuer will cause it to appear to be infinitely far away, and hence stationary against a distant background from the perspective of the target (*i.e.* motion camouflage). Conversely, identifying intercepting trajectories can provide an animal or vehicle with forewarning of others that are on a collision course, or perhaps trying to intercept it. The guidance scheme for interception described here is derived from the PN guidance law, which fundamentally seeks to nullify rate of change of the inertial LOS vector to the target. This strategy, also referred to as CATD, is thought to be used by some animals, such as dragonflies, who keep the angle between their prey and the horizon constant during predatory flights. This *interception guidance law* enables a pursuer to intercept a constant-velocity target along a time-optimal trajectory, without requiring metric measurements of 3D position or velocity of the target.

In this research, 3-DOF attitude of the aircraft was measured with respect to the horizon and appearance of the sky using the approach described in Chapter 6; and LOS to the target was tracked in omnidirectional imagery using the appearance-based method described in Chapter 7. A novel *interception control strategy* was derived in Chapter 8; this enables a fixed-wing aircraft to obtain turning rates demanded by the *interception guidance law*. This bioinspired guidance scheme was used to demonstrate autonomous interception of a moving ground-based vehicle by a small-scale UAV. Such a manoeuvre requires precise flight control and accurate prediction of the target's motion, which would typically be achieved using extensive computational resources and a range of sensors. In the research presented here, vision was used as the primary sensory input for all levels of flight control and all computation was performed on board the aircraft. The efficiency of this approach and its robustness to real-world conditions stem from the simple and bioinspired algorithms used for sensing and guidance as well as the omnidirectional FOV of the *iEye* vision system.

9

General Conclusions and Outlook

9.1 Thesis findings

This thesis presents the methods and findings of an investigation of the use of vision for providing guidance to a small-scale autonomous UAV. Two wide-angle vision systems were developed during the course of this research as well as a number of novel vision-based sensing and guidance algorithms, enabling detection of oncoming obstacles (Chapter 3), estimation of attitude and altitude (Chapters 4 & 6), long-term tracking of features (Chapter 7), and interception of independently moving objects (Chapter 8). Key aspects of all of these algorithms were sole reliance on visual input and extreme computational efficiency, brought about by their simple and bioinspired designs – properties that allowed these sensing and guidance algorithms to be implemented in parallel, using wide-angle vision systems. Complex capabilities such as terrain following at low altitude (Chapters 3 & 4), aerobatics (Chapter 6), landing in an uncontrolled environment (Chapter 6), as well as visual tracking and interception of a moving object (Chapter 8) were all achieved using this approach. This is the first time such complex behaviours have been demonstrated autonomously using only computing resources on board a small-scale UAV and using only vision for all sensing and guidance.

Traditionally, the capability to track an independently moving vehicle and then move to intercept it, or the capability to land a fixed-wing aircraft in an uncontrolled environment would require sensing and processing architectures incompatible with small-scale UAVs. Research conducted over the past few decades has shown that guidance systems inspired from biological vision-based flight control strategies can offer dramatically improved sensing and control efficiencies over more complex

computer vision-based approaches. Bioinspired methods, such as those based on specialised optic flow sensors or electronic *ocelli*, are therefore well suited for stabilising or controlling flight attitude and providing low-level flight control to UAVs flying near obstacles, and are unparalleled for providing guidance information on board vehicles with extremely strict weight or power consumption requirements – such as MAVs. This thesis establishes that *wide-angle* vision systems enable a much broader array of such bioinspired methods, which can be implemented in parallel, in turn enabling behaviours that are more complex than simple sense-and-avoid capabilities demonstrated thus far in bioinspired UAV guidance literature. Furthermore, a large visual field has been shown here to be crucial for robustly estimating attitude and egomotion in real-world conditions. The design methodologies proposed in this thesis therefore represent an important step towards full autonomy for UAVs flying missions in near-Earth environments or in environments where unmanned aircraft must interact with other moving vehicles.

Modern UAVs are typically flown remotely by a human pilot and their autonomous capabilities are restricted to low-level flight control or waypoint following. There is a trend towards more autonomy, however, and over the next decade autonomous UAVs will see increased usage within the civil and commercial sectors for tasks such as infrastructure inspection and maintenance, search and rescue, surveillance, agricultural imaging, transportation of goods, as well as bush fire detection and fighting, to name just a few. Unmanned aircraft are suited to these roles because they can be smaller and lighter than manned aircraft and hence cheaper to operate, as well as being able to perform dull or repetitive tasks with greater precision than human operators, and dangerous tasks with greater safety. Smaller and lighter aircraft also pose less risk to other vehicles and infrastructure should something go wrong (*e.g.* Figure 9.1).

Vision-based guidance systems are ideal for small-scale UAVs because they are compact, lightweight, low-power, and high-bandwidth. A pragmatic finding of this research is that much can be achieved using a single vision sensor, removing the need for multiple sensors on power- or weight-constrained platforms, or alternatively, adding redundancy by offering a completely independent pathway for measuring critical state variables for platforms where safety is paramount or where the vehicle must operate autonomously for extended periods of time – *i.e.* extraterrestrial exploration. Vision will play an important role in the automation of vehicles and aircraft in the future for the reasons outlined above. For equivalent reasons, evolution

Figure 9.1: Failure case for aircraft guidance: detachment of wing from aircraft body. The moment of impact was captured by (a) an external camera and (b) the *iEye* vision system on board the aircraft. Extreme attitude and proximity to the ground in conjunction with the wing (red circle) visibly separated from the aircraft should have alerted the on-board guidance system to an 'abnormal control situation'. Unfortunately the control system was not optimally tuned to cope with such eventualities.

has produced biological vision systems and guidance strategies that are extremely efficient at extracting just the information required by the animal to perform its duties. Biology provides many examples of elegant solutions to problems facing autonomous guidance of small-scale UAVs. Studying biological mechanisms to inform the design of their robotic counterparts has therefore proven to be an effective strategy for producing smaller, lighter, and smarter aerial robots.

9.2 Limitations and future work

Limitations of the various algorithms and approaches developed by this research were discussed in relevant chapters. Some limitations were addressed by additional experiments or analysis, whilst others remained outstanding due to time constraints or the scope of this thesis. These topics shall be addressed by future research. A few important limitations are discussed below.

9.2.1 Collision avoidance

Chapter 3 introduced a novel stereo vision-based guidance system that enabled an aircraft to directly estimate range from its projected flight path to obstacles in the environment. This system was used to demonstrate detection and avoidance of obstacles such as trees, which present a danger to low flying aircraft. Performance of the system was limited by its inability to detect small frontal obstacles coincident with the optic axis. This was due to the coaxial design of the system: the relative positioning of the cameras meant that the fore camera blocked a portion of the rear camera's view; and the geometric mapping (*terrain following* mapping) performed by the vision system stretched image areas close to the optic axis, reducing effective resolution of the vision system directly ahead of the aircraft.

Mirrors could be used to perform the function of the geometric mapping, which might alleviate resolution issues but would introduce other issues, such as optical distortion due to imperfections in the mirrors' surfaces, and the blind zone created by the position of the fore camera (and mirror) would still prevent detection of directly oncoming small obstacles. Positioning the cameras side-by-side (a traditional stereo configuration) would remove the frontal blind zone but also prevent the aircraft from using the *terrain following* mapping to detect obstacles penetrating the free-flight cylinder.

Another approach would be to remove the second camera altogether and to apply the *terrain following* mapping to measured optic flow. A difficulty with this approach would be determining precise rotation and translation of the camera between successive frames, required to estimate range to obstacles in the environment. The *iEye* vision system (Chapter 5) was designed specifically for this purpose, however an inherent limitation of optic flow-based object avoidance strategies is a reduced sensitivity to obstacles directly coincident with the vehicle's translation direction – similar to the limitation faced by the coaxial stereo vision system. Many biological organisms such as fruit flies [Tammero and Dickinson, 2002], blowflies [Kern et al., 2012], and locusts [Simmons and Rind, 1992] are known to perform *saccades*, or sporadic jerking motions, to regularly reorient their direction of motion, thereby revealing oncoming obstacles. Looming cues from expansion of the frontal optic flow field are also known to trigger such behaviours. Similar flight strategies have been demonstrated on board MAVs [Zufferey and Floreano, 2006] and could be employed to overcome the vulnerability of the coaxial stereo system and the *iEye* vision system to

small frontal obstacles. Alternatively, the side-by-side configuration of cameras in the *iEye* vision system enables it to compute stereo range in the region in which optic flow signals are weakest. Due to the small separation of the two cameras, however, reliable ranging would be restricted to distances no greater than approximately $10 - 15$m. Investigating combined stereo and optic flow-based collision avoidance strategies using the *iEye* vision system is therefore a topic for future work.

9.2.2 Moving vehicle interception

The culmination of the research presented in Part II was a guidance system that enabled an aircraft to track and intercept an independently moving vehicle using only visual input for all sensing and guidance. This system demonstrated successful operation and interaction of the various vision-based sensing and guidance strategies that were developed as part of this research and presented in Chapters 6, 7 & 8. Thorough flight testing demonstrated the practicality and effectiveness of the system in the real world. Ongoing research includes applying this method to visually track and intercept a moving ground-based vehicle. If the vehicle maintains a constant velocity and the UAV approaches from the rear, this system should enable the aircraft to descend stably and dock with the vehicle without requiring knowledge of its relative 3D position or velocity and even in the presence of disturbances such as wind. Such behaviour would be useful for retrieving a UAV while in motion. A similar application would be for refuelling a UAV from another aircraft mid-air. As a first step towards these applications it is planned to demonstrate the capability to select a particular landing site and then land autonomously by combining the interception scheme described in Chapter 8 and the optic flow-based landing strategy, described in Section 6.6.2.

9.2.3 Poor weather and flying in low light

A topic not considered within the scope of this thesis is application of visual guidance strategies to poor-visibility environments, such as during inclement weather or low-light conditions. Robustness to varying environmental conditions is important for any guidance system to be practical. It is implicitly assumed that aircraft using the systems developed within this thesis operate at low altitudes (*i.e.* under cloud level), because at higher altitudes precise control of orientation or position is less important. The *iEye* vision system was tested under a variety of conditions (*e.g.* Figure 9.2)

and proved remarkably resilient. The omnidirectional visual field of the *iEye* vision system enabled attitude and egomotion to be computed robustly despite localised deformation of the input image (*e.g.* due to rain drops on the lenses). The algorithm for computing attitude (Chapter 6) relies on intensity and/or spectral separation of ground and sky, hence it would be expected to fail at twilight or under conditions in which there is not enough light for the vision system to capture colour images. In such situations it might be possible to use thermal imaging systems in place of visual-spectrum cameras. Ongoing work involves evaluating performance of 3-DOF attitude and egomotion estimation algorithms on wide-angle thermal images captured in the long wavelength infrared spectrum (8μm–15μm). By modifying these algorithms (if neccessary) it is planned to demonstrate autonomous control of attitude and altitude in low-light conditions and at night.

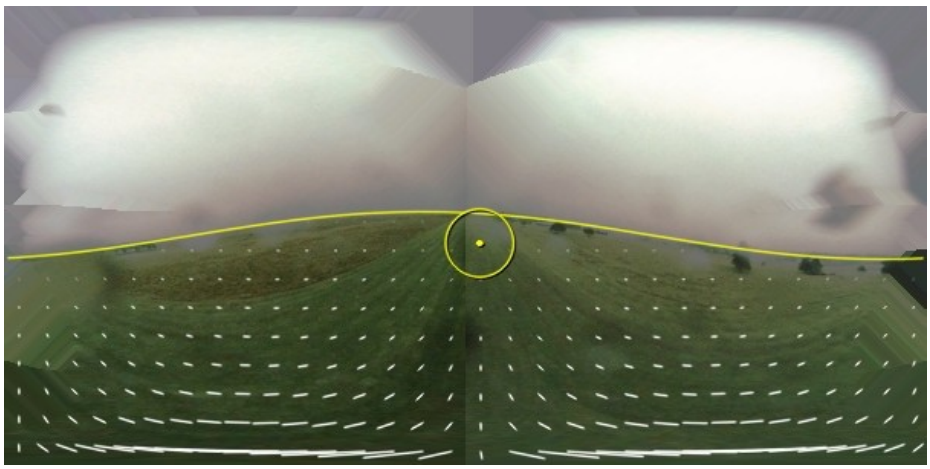


Figure 9.2: Flying in poor weather. Stitched image from the *iEye* vision system on board the aircraft showing estimated horizon (yellow line), estimated translation direction (yellow circle), and computed translational motion of the ground plane (white vectors). Note the poor visibility and rain drops on the lenses. This frame was captured during a successful automatic landing.

9.2.4 Broader horizons for bioinspired aviation

The eventual goal for the UAV and visual guidance systems described in this thesis is to be able to perform completely autonomous missions. The aircraft is currently able to fly specific tasks fully autonomously, after which a human controller on the ground pilots the aircraft back to the runway and engages an automatic landing procedure. To be able to complete entire missions without intervention from a human controller,

the UAV must also be able to take-off automatically and to navigate between the take-off and landing sites and the location of the task or mission. Investigating visual means enabling these capabilities is a topic of ongoing research.

Visual odometry (demonstrated using the *iEye* vision system in the accompanying video[1]) allows for precise navigation over flight distances up to a couple of kilometres. For longer missions, visual SLAM systems or other approaches capable of recognising landmarks must be used to counteract drift of the aircraft's estimated position, which occurs over time due to integration of noisy egomotion estimates. Many walking and flying insects, and honeybees in particular, demonstrate an amazing ability to return directly via the shortest path (*i.e.* a 'bee-line') to their nest site after foraging for many kilometres over very convoluted routes. It is thought that insects rely primarily on odometry, from visual input (*e.g.* [Srinivasan, 2011]) or proprioception (*e.g.* [Wehner, 1992]), and constrain drift of their position estimates using absolute orientation cues (from celestial polarisation patterns or sun position) and visual landmarks. Precise details of insects' navigational capabilities continue to be researched, however, and implementing such strategies on board small-scale UAVs may provide insight into the computational foundations and sensory requirements of these bioinspired algorithms as well as improve performance of visual odometry for flying robots.

In addition to navigation, biological organisms are far superior to present-day robotic systems in many aspects: from the remarkable agility and split-second path planning of the goshawk as it flies at high speed through dense undergrowth; to the apparent coordination and situational awareness of individuals within shoals of fish, flocks of migrating birds, or swarms of flying insects; or the circuit miniaturisation and robust fault tolerance of insects – exemplified by the fly that misses the rim of a teacup and lands in the tea, but is able to crawl out, shake itself off, and carry on none the worse for wear. Such remarkable capabilities suggest that we have only just begun to capitalise on nature's efforts in designing agile, robust, and autonomous systems; and that future generations of bioinspired UAVs will enable a range of exciting possibilities and applications.

[1] Attachment AV07 (available at http://youtu.be/R5t3zHNic3o). Note that this video is described in further detail in Section 6.6.

9.3 Conclusion

This thesis described two novel wide-angle vision systems that were designed specifically to simplify critical aspects of UAV guidance: detection and avoidance of obstacles, and robust estimation of attitude and egomotion. Several vision-based sensing and guidance algorithms were developed and implemented using these systems to display a range of complex behaviours. The research presented in this thesis has enabled manoeuvres such as automatic landing in an uncontrolled environment and visual tracking and interception of a moving vehicle to be performed for the first time using only visual input and computational resources available on board a small-scale UAV. The design methodologies proposed in this thesis represent an important step towards full autonomy for small-scale airborne platforms, thereby contributing towards exploitation of UAVs for civil and commercial applications and bringing autonomous UAVs a step closer to the remarkable capabilities of their biological counterparts.

Bibliography

M. Achtelik, M. Achtelik, S. Weiss, and R. Siegwart. Onboard IMU and monocular vision based control for MAVs in unknown in-and outdoor environments. In *Robotics and automation (ICRA), 2011 IEEE international conference on*, pages 3056–3063. IEEE, 2011.

G. Adiv. Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11 (5):477–489, 1989.

M. Agrawal, K. Konolige, and L. Iocchi. Real-time detection of independent motion using stereo. In *Proc. IEEE Workshop on Motion and Video Computing*, pages 207–214, Jan. 2005.

V. Aidala and S. Hammel. Utilization of modified polar coordinates for bearings-only tracking. *IEEE Transactions on Automatic Control*, 28(3):283–294, 1983.

R. R. Allen and S. S. Blackman. Implementation of an angle-only tracking filter. In *Proc. SPIE*, volume 1481, page 292, 1991.

O. Amidi, T. Kanade, and K. Fujita. A visual odometer for autonomous helicopter flight. *Robotics and Autonomous Systems*, 28:185–193, 1999.

F. Andert and F. Adolf. Online world modeling and path planning for an unmanned helicopter. *Autonomous Robots*, 27:147–164, 2009.

F. Andert, F. Adolf, L. Goormann, and J. Dittrich. Mapping and path planning in complex environments: An obstacle avoidance approach for an unmanned helicopter. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 745–750. IEEE, 2011.

H. Badino, U. Franke, C. Rabe, and S. Gehrig. Stereo vision based detection of moving objects under strong camera motion. In *Proc. First International Conference on Computer Vision Theory and Applications*, volume 2, pages 253–260, Feb. 2006.

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

D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122, 1981.

D. B. Barber, J. D. Redding, T. W. McLain, R. W. Beard, and C. N. Taylor. Vision-based target geo-location using a fixed-wing miniature air vehicle. *Journal of Intelligent & Robotic Systems*, 47(4):361–382, 2006.

G. L. Barrows and C. Neely. Mixed-mode VLSI optic flow sensors for in-flight control of a micro air vehicle. In *Proc. SPIE*, volume 4109, pages 52–63, 2000.

G. L. Barrows, J. S. Chahl, and M. V. Srinivasan. Biologically inspired visual sensing and flight control. *The Aeronautical Journal*, 107(1069):159–168, 2003.

J. C. Bazin, C. Demonceaux, P. Vasseur, and I. S. Kweon. Motion estimation by decoupling rotation and translation in catadioptric vision. *Computer Vision and Image Understanding*, 114(2):254–273, 2010.

F. Belkhouche, B. Belkhouche, and P. Rastgoufard. Parallel navigation for reaching

a moving goal by a mobile robot. *Robotica*, 25(01):63–74, 2006.

R. Benosman and S. B. Kang. *Panoramic vision: sensors, theory, and applications*. Springer-Verlag New York Inc, 2001.

A. Beyeler. *Vision-based control of near-obstacle flight*. PhD thesis, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, May 2009.

A. Beyeler, C. Mattiussi, J.-C. Zufferey, and D. Floreano. Vision-based altitude and pitch estimation for ultra-light indoor aircraft. In *Proc. IEEE International Conference on Robotics and Automation*, pages 2836–2841, 2006.

A. Beyeler, J.-C. Zufferey, and D. Floreano. 3D vision-based navigation for indoor microflyers. In *Proc. IEEE International Conference on Robotics and Automation*, Roma, Italy, Apr. 2007.

A. Beyeler, J.-C. Zufferey, and D. Floreano. optipilot: Control of take-off and landing using optic flow. In *Proc. 2009 European Micro Air Vehicle Conference and Competition*, 2009a.

A. Beyeler, J.-C. Zufferey, and D. Floreano. Vision-based control of near-obstacle flight. *Autonomous Robots*, 27(3):201–219, 2009b.

D. Beymer and K. Konolige. Real-time tracking of multiple people using continuous detection. In *IEEE Frame Rate Workshop*, 1999.

N. Boeddeker, R. Kern, and M. Egelhaaf. Chasing a dummy target: smooth pursuit and velocity control in male blowflies. *Proceedings of the Royal Society B: Biological Sciences*, 270(1513):393, 2003.

G. R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Proc. of Intel Technology Journal, 1998*, 1998.

W. Burger and B. Bhanu. On computing a fuzzy focus of expansion for autonomous navigation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 563–568, 1989.

J. Byrne, M. Cosgrove, and R. Mehra. Stereo based obstacle detection for an unmanned air vehicle. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 2830–2835. Ieee, 2006.

P. Campoy, J. F. Correa, I. Mondragón, C. Martínez, M. Olivares, L. Mejías, and J. Artieda. Computer vision onboard UAVs for civilian tasks. *Unmanned Aircraft Systems*, pages 105–135, 2009.

N. E. Carey, J. J. Ford, and J. S. Chahl. Biologically inspired guidance for motion camouflage. In *5th Asian Control Conference*, volume 3, 2004.

V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

J. S. Chahl and M. V. Srinivasan. Visual computation of egomotion using an image interpolation technique. *Biological Cybernetics*, 74:405–411, 1996.

J. S. Chahl and M. V. Srinivasan. Reflective surfaces for panoramic imaging. *Applied Optics*, 36(31):8275–8285, 1997.

J. S. Chahl, M. V. Srinivasan, and S. W. Zhang. Landing strategies in honeybees

and applications to uninhabited airborne vehicles. *The International Journal of Robotics Research*, 23(2):101–110, Feb. 2004.

P. Chang and M. Hebert. Omni-directional visual servoing for human-robot interaction. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 1801–1807, 1998.

Y. Cheng. Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(8):790–799, 1995.

A. Cheung, S. Zhang, C. Stricker, and M. V. Srinivasan. Animal navigation: general properties of directed walks. *Biological cybernetics*, 99(3):197–217, 2008.

A. Cho, J. Kim, S. Lee, and C. Kee. Wind estimation and airspeed calibration using the uav with a single-antenna gnss receiver and airspeed sensor. In *Proceedings of the 19th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2006)*, pages 3050–3058, 2006.

T. S. Collett and A. J. King. Vision during flight. *The compound eye and vision of insects*, pages 437–466, 1975.

T. S. Collett and M. F. Land. Visual control of flight behaviour in the hoverfly syritta pipiens l. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 99(1):1–66, 1975.

T. S. Collett and M. F. Land. How hoverflies compute interception courses. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 125(3):191–204, 1978.

R. T. Collins. A space-sweep approach to true multi-image matching. In *Computer Vision and Pattern Recognition. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pages 358–363. IEEE, 1996.

D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 142–149. IEEE, 2000.

D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(5):564–577, 2003.

J. Conroy, G. Gremillion, B. Ranganathan, and J. Humbert. Implementation of wide-field integration of optic flow forautonomous quadrotor navigation. *Autonomous Robots*, 27:189–198, 2009.

P. Corke, P. Sikka, and J. Roberts. Height estimation for an autonomous helicopter. *Experimental Robotics VII*, pages 101–110, 2001.

T. D. Cornall and G. K. Egan. Measuring horizon angle from video on a small unmanned air vehicle. In *2nd International Conference on Autonomous Robots and Agents*, pages 339–344, 2004.

T. D. Cornall, G. K. Egan, and A. Price. Aircraft attitude estimation from horizon video. *Electronics Letters*, 42(13):744–745, 2006.

H. J. Dahmen, M. O. Franz, and H. G. Krapp. Extracting egomotion from optic flow: limits of accuracy and neural matched filters. *Motion vision-computational, neural,*

and ecological constraints. Berlin, Heidelberg, New York, Tokyo: Springer Verlag, pages 143–168, 2001.

T. Dang, C. Hoffmann, and C. Stiller. Fusing optical flow and stereo disparity for object tracking. In *Proc. IEEE International Conference on Intelligent Transportation Systems*, pages 112–117, 2002.

Defense Update, 2005. URL http://defense-update.com/products/s/seagull-uav.htm.

D. Demirdjian and R. Horaud. Motion-egomotion discrimination and motion segmentation from image-pair streams. *Computer Vision and Understanding*, 78: 53–68, 2000.

C. Demonceaux, P. Vasseur, and C. Pegard. Omnidirectional vision on UAV for attitude computation. In *Proc. IEEE International Conference on Robotics and Automation*, pages 2842–2847. IEEE, 2006.

C. Demonceaux, P. Vasseur, and C. Pegard. UAV attitude computation by omnidirectional vision in urban environment. In *Robotics and Automation, IEEE International Conference on*, pages 2017–2022, Apr. 2007.

G. N. DeSouza and A. C. Kak. Vision for mobile robot navigation: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(2), Feb. 2002.

Draganfly Innovations Inc., Jan. 2012. URL http://www.rctoys.com/rc-toys-and-parts/DF-SAVS/RC-HELICOPTERS.html.

R. Dudley. *The Biomechanics of Insect Flight: Form, Function, Evolution*. Princeton University Press, 2000.

H. Durrant-Whyte and T. Bailey. Simultaneous localization and mapping: Part I. *Robotics & Automation Magazine, IEEE*, 13(2):99–110, 2006.

D. Dusha, W. W. Boles, and R. Walker. Fixed-wing attitude estimation using computer vision based horizon detection. In *Proc. 12th Australian International Aerospace Congress*, pages 1–19, Melbourne, Australia, 2007.

G. Egan and B. Taylor. Characterisation of infrared sensors for absolute unmanned aerial vehicle attitude determination. *Monash Univ. TR MECSE-2007, Melbourne, Australia*, 2007.

S. M. Ettinger, M. C. Nechyba, P. G. Ifju, and M. Waszak. Towards flight autonomy: Vision-based horizon detection for micro air vehicles. In *Florida Conference on Recent Advances in Robotics*, volume 2002, 2002a.

S. M. Ettinger, M. C. Nechyba, P. G. Ifju, and M. Waszak. Vision-guided flight stability and control for micro air vehicles. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 2134–2140. IEEE, 2002b.

D. Eynard, P. Vasseur, C. Demonceaux, and V. Frémont. Real time UAV altitude, attitude and motion estimation from hybrid stereovision. *Autonomous Robots*, pages 1–16, 2012.

B. R. Fajen and W. H. Warren. Behavioral dynamics of intercepting a moving target. *Experimental Brain Research*, 180(2):303–319, 2007.

K. Fent and R. Wehner. Oceili: A celestial compass in the desert ant cataglyphis.

Science, 228(4696):192–194, 1985.

C. Fermuller and Y. Aloimonos. Ambiguity in structure from motion: Sphere versus plane. *International Journal of Computer Vision*, 28(2):137–154, 1998.

D. Floreano, J.-C. Zufferey, M. V. Srinivasan, and C. P. Ellington. *Flying Insects and Robots*. Springer-Verlag New York, 2009.

N. Franceschini. Visual guidance based on optic flow: A biorobotic approach. *Journal of Physiology*, 98:281–292, 2004.

U. Franke, C. Rabe, H. Badino, and S. Gehrig. 6d-vision: Fusion of stereo and motion for robust environment perception. *Lecture notes in computer science*, 3663:216, 2005.

U. Franke, S. Gehrig, H. Badino, and C. Rabe. Towards optimal stereo analysis of image sequences. *Lecture Notes in Computer Science*, 4931:43, 2008.

M. O. Franz and H. G. Krapp. Wide-field, motion-sensitive neurons and matched filters for optic flow fields. *Biological cybernetics*, 83(3):185–197, 2000.

M. O. Franz, J. S. Chahl, and H. G. Krapp. Insect-inspired estimation of egomotion. *Neural computation*, 16(11):2245–2260, 2004.

J. Frazier and R. Nevatia. Detecting moving objects from a moving platform. In *IEEE International Conference on Robotics and Automation*, pages 1627–1633, 1992.

Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.

N. Frietsch, O. Meister, C. Schlaile, J. Wendel, and G. F. Trommer. Detection and tracking of objects in an image sequence captured by a VTOL-UAV. In *Proc. SPIE*, volume 6561, 2007.

M. A. Garratt and J. S. Chahl. Vision-based terrain following for an unmanned rotorcraft. *Journal of Field Robotics*, 25:284–301, 2008.

Geometric Tools. Mathematics library, Aug. 2010. URL http://www.geometrictools.com/LibMathematics/LibMathematics.html.

K. Ghose, T. K. Horiuchi, P. S. Krishnaprasad, and C. F. Moss. Echolocating bats use a nearly time-optimal strategy to intercept prey. *PLoS Biol.*, 4(5), 2006.

J. J. Gibson. *The Perception of the Visual World*. Houghton Mifflin, 1950.

C. Gilbert and N. J. Strausfeld. The functional organization of male-specific visual neurons in flies. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 169(4):395–411, 1991.

J. Gluckman and S. K. Nayar. Ego-motion and omnidirectional cameras. In *IEEE International Conference on Computer Vision*, pages 999–1005, 1998.

W. E. Green. *A Multimodal Micro Air Vehicle for Autonomous Flight in Near-Earth Environments*. PhD thesis, Drexel University, Philadelphia, PA, June 2007.

W. E. Green and P. Y. Oh. Optic-flow-based collision avoidance. *Robotics & Automation Magazine, IEEE*, 15(1):96–103, 2008.

W. E. Green, P. Y. Oh, K. Sevcik, and G. Barrows. Autonomous landing for indoor

flying robots using optic flow. In *Proc. ASME International Mechanical Engineering Congress*, Washington, D.C., Nov. 2003.

W. E. Green, P. Y. Oh, and G. L. Barrows. Flying insect inspired vision for autonomous aerial robot maneuvers in near-earth environments. In *Proc. IEEE International Conference on Robotics and Automation*, New Orleans, LA, Apr. 2004.

K. Han and G. N. DeSouza. Multiple targets geolocation using SIFT and stereo vision on airborne video sequences. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, St Louis, MO, Oct. 2009.

C. G. Harris and J. M. Pike. 3D positional integration from image sequences. *Image and Vision Computing*, 6(2):87–90, 1988.

D. J. Heeger and A. D. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7(2): 95–117, 1992.

L. Heng, G. H. Lee, F. Fraundorfer, and M. Pollefeys. Real-time photo-realistic 3D mapping for micro aerial vehicles. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 4012–4019. IEEE, 2011.

B. Herisse, S. Oustrieres, T. Hamel, R. Mahony, and F. X. Russotto. A general optical flow based terrain-following strategy for a vtol uav using multiple views. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 3341–3348. IEEE, 2010.

B. K. P. Horn, Y. Fang, and I. Masaki. Time to contact relative to a planar surface. In *Intelligent Vehicles Symposium, 2007 IEEE*, pages 68–74. IEEE, 2007.

S. Hrabar and G. Sukhatme. Vision-based navigation through urban canyons. *Journal of Field Robotics*, 26(5):431–452, 2009.

S. Hrabar and G. S. Sukhatme. Omnidirectional vision for an autonomous helicopter. In *Robotics and Automation. Proceedings. IEEE International Conference on*, volume 1, pages 558–563, 2003.

S. Hrabar, G. S. Sukhatme, P. Corke, K. Usher, and J. Roberts. Combined optic-flow and stereo-based navigation of urban canyons for a UAV. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, Edmonton, Canada, Aug. 2005.

A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy. Visual odometry and mapping for autonomous flight using an RGB-D camera. In *Int. Symposium on Robotics Research (ISRR),(Flagstaff, Arizona, USA)*, 2011.

Z. Hubert. zSim flight simulator, 2008. URL http://zsim.sourceforge.net.

M. Hwangbo. *Robust Monocular Vision-based Navigation for a Miniature Fixed-Wing Aircraft*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, Sept. 2009. Thesis proposal.

M. Irani and P. Anandan. A unified approach to moving object detection in 2D and 3D scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20 (6):577–589, 1998.

R. C. Jain. Segmentation of frame sequences obtained by a moving observer. *Pattern Analysis and Machine Intelligence, IEEE transactions on*, 6(5):624–629, 1984.

Z. Jia, A. Balasuriya, and S. Challa. Autonomous vehicles navigation with visual target tracking: Technical approaches. *Algorithms*, 1:153–182, 2008.

E. N. Johnson, A. J. Calise, Y. Watanabe, J. Ha, and J. C. Neidhoefer. Real-time vision-based relative aircraft navigation. *AIAA Journal of Aerospace Computing, Information, and Communication*, 4(4):707–738, 2007.

S. G. Johnson. The NLopt nonlinear-optimization package, July 2009. URL http://ab-initio.mit.edu/nlopt.

S. J. Julier and J. K. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, volume 3, page 26. Spie Bellingham, WA, 1997.

B. Jung and G. S. Sukhatme. Detecting moving objects using a single camera on a mobile robot in an outdoor environment. In *International Conference on Intelligent Autonomous Systems*, pages 980–987, 2004.

E. W. Justh and P. S. Krishnaprasad. Steering laws for motion camouflage. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 462(2076):3629, 2006.

Z. Kalal, J. Matas, and K. Mikolajczyk. Online learning of robust object detectors during unstable tracking. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 1417–1424. Ieee, 2009.

Z. Kalal, J. Matas, and K. Mikolajczyk. Pn learning: Bootstrapping binary classifiers by structural constraints. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 49–56. IEEE, 2010.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

J. Kang, I. Cohen, G. Medioni, and C. Yuan. Detection and tracking of moving objects from a moving platform in presence of strong parallax. In *Proc. IEEE International Conference on Computer Vision*, 2005.

J. Kannala and S. S. Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(8):1335–1340, 2006.

F. Kendoul, I. Fantoni, and K. Nonami. Optic flow-based vision system for autonomous 3D localization and control of small aerial vehicles. *Robotics and Autonomous Systems*, 57(6–7):591–602, 2009a.

F. Kendoul, K. Nonami, I. Fantoni, and R. Lozano. An adaptive vision-based autopilot for mini flying machines guidance, navigation and control. *Autonomous robots*, 27(3):165–188, 2009b.

F. Kendoul, Z. Yu, and K. Nonami. Guidance and nonlinear control system for autonomous flight of minirotorcraft unmanned aerial vehicles. *Journal of Field Robotics*, 27(3):311–334, 2010.

R. Kern, N. Boeddeker, L. Dittmar, and M. Egelhaaf. Blowfly flight characteristics are shaped by environmental features and controlled by optic flow information. *The Journal of Experimental Biology*, 215(14):2501–2514, 2012.

J. Kim and G. Brambley. Dual optic-flow integrated navigation for small-scale flying robots. In *Proc. of Australasian Conference on Robotics and Automation*, Brisbane, Australia, 2007.

J. Klappstein, T. Vaudrey, C. Rabe, A. Wedel, and R. Klette. Moving object segmentation using optical flow and depth information. *Proc. PSIVT (T. Wada, F. Huang, and S. Lin, editors), LNCS*, 5414:611–623, 2009.

G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234. IEEE, 2007.

J. J. Koenderink and A. J. Doorn. Facts on optic flow. *Biological cybernetics*, 56(4): 247–254, 1987.

H. G. Krapp and R. Hengstenberg. Estimation of self-motion by optic flow processing in single visual interneurons. *Nature*, 384(6608):463–466, 1996.

H. G. Krapp, B. Hengstenberg, and R. Hengstenberg. Dendritic structure and receptive-field organization of optic flow processing interneurons in the fly. *Journal of Neurophysiology*, 79(4):1902–1917, 1998.

F. Labrosse. The visual compass: Performance and limitations of an appearance-based method. *Journal of Field Robotics*, 23(10):913–941, 2006.

M. F. Land and T. S. Collett. Chasing behaviour of houseflies (Fannia canicularis). *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 89(4):331–357, 1974.

S. Lange, N. Snderhauf, P. Neubert, S. Drews, and P. Protzel. Autonomous corridor flight of a UAV using a low-cost and light-weight RGB-D camera. In U. Rckert, S. Joaquin, and W. Felix, editors, *Advances in Autonomous Mini Robots*, pages 183–192. Springer Berlin Heidelberg, 2012.

J. W. Langelaan, N. Alley, and J. Neidhoefer. Wind field estimation for small unmanned aerial vehicles. *Journal of Guidance, Control, and Dynamics*, 34(4): 1016–1030, 2011.

G. W. Larson, H. Rushmeier, and C. Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *Visualization and Computer Graphics, IEEE Transactions on*, 3(4):291–306, 1997.

B. Leibe, N. Cornelis, K. Cornelis, and L. Van Gool. Dynamic 3D scene analysis from a moving vehicle. *CVPR07*, 2007.

N. Liebsch and M. V. Srinivasan. Sticking the head into the wind: The effect of wind and the role of vision on the landing performance of honey bees. In *International Congress of Neuroethology*, Salamanca, Spain, Aug. 2010. Poster abstract #387.

J. P. Lindemann, R. Kern, J. H. van Hateren, H. Ritter, and M. Egelhaaf. On the computations analyzing natural optic flow: quantitative model analysis of the blowfly motion vision pathway. *The Journal of neuroscience*, 25(27):6435–6448,

2005.

H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

V. Mahadevan and N. Vasconcelos. Background subtraction in highly dynamic scenes. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–6. Ieee, 2008.

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

A. Mallet, S. Lacroix, and L. Gallo. Position estimation in outdoor environments using pixel tracking and stereovision. In *Proc. IEEE International Conference on Robotics and Automation*, San Francisco, CA, Apr. 2000.

I. R. Manchester and A. V. Savkin. Circular navigation guidance law for precision missile/target engagements. In *Proc. IEEE 41st Conference on Decision and Control*, volume 2, 2002.

I. R. Manchester, E. M. P. Low, and A. V. Savkin. Vision-based interception of a moving target by a mobile robot. In *IEEE International Conference on Control Applications*, pages 397–402, 2007.

R. Mandelbaum, G. Salgian, and H. Sawney. Correlation-based estimation of ego-motion and structure from motion and stereo. In *Proc. International Conference on Computer Vision*, Kerkyra, Greece, Sept. 1999.

L. Matthies and S. Shafer. Error modeling in stereo navigation. *Robotics and Automation, IEEE Journal of*, 3(3):239–248, June 1987.

T. G. McGee, R. Sengupta, and K. Hedrick. Obstacle detection for small autonomous aircraft using sky segmentation. In *Proc. IEEE International Conference on Robotics and Automation*, pages 4679–4684. IEEE, 2005.

S. Merhav. *Aerospace sensor systems and applications*. Springer Verlag, 1996.

P. Merrell, D.-J. Lee, and R. Beard. Obstacle avoidance for unmanned air vehicles using optical flow probability distributions. In *Proc. SPIE Conference Series*, volume 5609, pages 13–22, Dec. 2004.

M. G. J. Minnaert. *The nature of light & colour in the open air*. Dover Pubns, 1954.

A. Mizutani, J. S. Chahl, and M. V. Srinivasan. Motion camouflage in dragonflies. *Nature*, 423(6940):604, 2003.

I. F. Mondragón, P. Campoy, C. Martinez, and M. Olivares. Omnidirectional vision applied to Unmanned Aerial Vehicles (UAVs) attitude and heading estimation. *Robotics and Autonomous Systems*, 58(6):809–819, 2010a.

I. F. Mondragón, M. A. Olivares-Méndez, P. Campoy, C. Martínez, and L. Mejias. Unmanned aerial vehicles UAVs attitude, height, motion estimation and control using visual systems. *Autonomous Robots*, 29(1):17–34, 2010b.

R. J. D. Moore, S. Thurrowgood, D. Bland, D. Soccol, and M. V. Srinivasan. A

stereo vision system for UAV guidance. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, pages 3386–3391, St Louis, MO, Oct. 2009.

R. J. D. Moore, S. Thurrowgood, D. Bland, D. Soccol, and M. V. Srinivasan. UAV altitude and attitude stabilisation using a coaxial stereo vision system. In *Proc. IEEE International Conference on Robotics and Automation*, Anchorage, AK, May 2010.

R. J. D. Moore, S. Thurrowgood, D. Bland, D. Soccol, and M. V. Srinivasan. A fast and adaptive method for estimating UAV attitude from the visual horizon. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, San Francisco, CA, Sept. 2011a.

R. J. D. Moore, S. Thurrowgood, D. Bland, D. Soccol, and M. V. Srinivasan. A bio-inspired stereo vision system for guidance of autonomous aircraft. In A. Bhatti, editor, *Advances in Theory and Aplications of Stereo Vision*, pages 305–326. InTech, Rijeka, Jan. 2011b.

R. J. D. Moore, S. Thurrowgood, D. Soccol, D. Bland, and M. V. Srinivasan. A method for the visual estimation and control of 3-DOF attitude for UAVs. In *Proc. ARAA Australasian Conference on Robotics and Automation*, Melbourne, VIC, Dec. 2011c.

R. J. D. Moore, S. Thurrowgood, and M. V. Srinivasan. Vision-only estimation of wind field strength and direction from an aerial platform. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, Vilamoura, Algarve, Oct. 2012.

M. I. Mote and R. Wehner. Functional characteristics of photoreceptors in the compound eye and ocellus of the desert ant, cataglyphis bicolor. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 137:63–71, 1980.

K. Nakayama and J. M. Loomis. Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception*, 3(1):63–80, 1974.

G. Nalbach and R. Hengstenberg. The halteres of the blowfly calliphora. *Journal of Comparative Physiology A*, 175(6):695–708, 1994.

S. C. Nardone, A. Lindgren, and K. Gong. Fundamental properties and performance of conventional bearings-only target motion analysis. *Automatic Control, IEEE Transactions on*, 29(9):775–787, 1984.

S. K. Nayar. Catadioptric omnidirectional camera. In *Computer Vision and Pattern Recognition. Proceedings., IEEE Computer Society Conference on*, pages 482–488, June 1997.

P. P. Neumann, S. Asadi, A. J. Lilienthal, M. Bartholmai, and J. H. Schiller. Autonomous gas-sensitive microdrone: Wind vector estimation and gas distribution mapping. *Robotics Automation Magazine, IEEE*, 19(1):50–61, march 2012.

T. Neumann and H. H. Bulthoff. Insect inspired visual control of translatory flight. In *Proc. 6th European Conference on Artificial Life (ECAL'01)*, Prague, Czech Republic, Sept. 2001.

T. Neumann and H. H. Bulthoff. Behaviour oriented vision for biomimetic flight

control. In *Proc. EPSRC/BBSRC International Workshop on Biologically Inspired Robotics*, Bristol, UK, Aug. 2002.

D. Nister. An efficient solution to the five-point relative pose problem. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–195. IEEE, 2003.

D. Nister, O. Naroditsky, and J. Bergen. Visual odometry. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, D.C., July 2004.

F. O'Gorman and M. B. Clowes. Finding picture edges through collinearity of feature points. *Computers, IEEE Transactions on*, 100(4):449–456, 1976.

P. Y. Oh, W. E. Green, and G. L. Barrows. Neural nets and optic flow for autonomous micro-air-vehicle navigation. In *Proc. ASME International Mechanical Engineering Congress and Exposition*, Anaheim, CA, Nov. 2004.

R. M. Olberg. Object-and self-movement detectors in the ventral nerve cord of the dragonfly. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 141(3):327–334, 1981.

R. M. Olberg, A. H. Worthington, and K. R. Venator. Prey pursuit and interception in dragonflies. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 186(2):155–162, 2000.

C. F. Olson, L. H. Matthies, M. Schoppers, and M. W. Maimone. Rover navigation using stereo ego-motion. *Robotics and Autonomous Systems*, 43:215–229, 2003.

H. J. Palanthandalam-Madapusi, A. Girard, and D. S. Bernstein. Wind-field reconstruction from flight data using an unbiased minimum-variance unscented filter. *Transactions of the Institute of Measurement and Control*, 33(6):718–733, 2011.

M. J. D. Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, Reino Unido*, 2009.

W. Premerlani. Imu wind estimation, Dec. 2009. URL http://diydrones.com/forum/topics/wind-estimation-without-an.

A. A. Proctor, E. N. Johnson, and T. B. Apker. Vision-only control and guidance for aircraft. *Journal of Field Robotics*, 23(10):863–890, 2006.

C. Rabe, U. Franke, and S. Gehrig. Fast detection of moving objects in complex scenarios. In *IEEE Intelligent Vehicles Symposium*, pages 398–403, 2007.

P. V. Reddy, E. W. Justh, and P. S. Krishnaprasad. Motion camouflage in three dimensions. In *Decision and Control, 2006 45th IEEE Conference on*, pages 3327–3332. IEEE, 2006.

W. Ribi, E. Warrant, and J. Zeil. The organization of honeybee ocelli: Regional specializations and rhabdom arrangements. *Arthropod Structure & Development*, 40(6):509–520, 2011.

J. M. Roberts, P. I. Corke, and G. Buskey. Low-cost flight control system for a

small autonomous helicopter. In *Proc. Australasian Conference on Robotics and Automation*, Auckland, New Zealand, Nov. 2002.

J. M. Roberts, P. I. Corke, and G. Buskey. Low-cost flight control system for a small autonomous helicopter. In *Proc. IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, Sept. 2003.

J. Rogers, M. Costello, and D. Hepner. Roll orientation estimator for smart projectiles using thermopile sensors. *Journal of Guidance Control and Dynamics*, 34(3):688, 2011.

J. Rohac. Accelerometers and an aircraft attitude evaluation. In *IEEE Conference on Sensors*, page 6. IEEE, 2005.

R. Rosales and S. Sclaroff. 3D trajectory recovery for tracking multiple objects and trajectory guided recognition of actions. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 117–123. IEEE, 1999.

P. E. Ross. When will software have the right stuff? *IEEE Spectrum*, 48(12):38–43, Dec. 2011. International edition.

F. Ruffier and N. Franceschini. Optic flow regulation: the key to aircraft automatic guidance. *Robotics and Autonomous Systems*, 50:177–194, 2005.

D. E. Rumelhart. Learning internal representations by error propagation. Technical report, Institute for Cognitive Science, University of California San Diego, 1985.

G. Salgian, J. Bergen, S. Samarakekera, and R. Kumar. Moving target indication from a moving camera in the presence of strong parallax. In *Proc. of the 25th Army Science Conference*, Orlando, FL, 2006.

S. Saripalli and G. S. Sukhatme. Landing a Helicopter on a Moving Target. In *IEEE International Conference on Robotics and Automation*, pages 2030–2035, 2007.

S. Saripalli, J. F. Montgomery, and G. S. Sukhatme. Visually guided landing of an unmanned aerial vehicle. *IEEE Trans. Robot. Autom.*, 19(3):371–380, 2003.

D. Scaramuzza and R. Siegwart. Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *Robotics, IEEE Transactions on*, 24(5):1015–1026, 2008.

H. Scharr. *Optimal operators in digital image processing.* PhD thesis, Heidelberg University, 2000.

S. Scherer, S. Singh, L. Chamberlain, and S. Saripalli. Flying fast and low among obstacles. In *Proc. IEEE International Conference on Robotics and Automation*, Roma, Italy, Apr. 2007.

F. Schill, R. Mahony, P. Corke, and L. Cole. Virtual force feedback teleoperation of the InsectBot using optical flow. In *Proc. Australasian Conference on Robotics & Automation*. ARAA, 2008.

C. Schilstra and J. H. van Hateren. Blowfly flight and optic flow. i. thorax kinematics and flight dynamics. *Journal of experimental biology*, 202(11):1481–1490, 1999.

H. Schuppe and R. Hengstenberg. Optical properties of the ocelli of calliphora

erythrocephala and their role in the dorsal light response. *Journal of Comparative Physiology A*, 173:143–149, 1993.

S. Schwarz, L. Albert, A. Wystrach, and K. Cheng. Ocelli contribute to the encoding of celestial compass information in the australian desert ant melophorus bagoti. *Journal of Experimental Biology*, 214(6):901–906, 2011.

R. Seidl and W. Kaiser. Visual field size, binocular domain and the ommatidial array of the compound eyes in worker honey bees. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 143(1):17–26, 1981.

M. Shimizu and M. Okutomi. Significance and attributes of subpixel estimation on area-based matching. *Systems and Computers in Japan*, 34(12), 2003.

P. A. Shoemaker, A. M. Hyslop, and J. S. Humbert. Optic flow estimation on trajectories generated by bio-inspired closed-loop flight. *Biological cybernetics*, 104 (4):339–350, 2011.

P. J. Simmons and F. C. Rind. Orthopteran dcmd neuron: a reevaluation of responses to moving objects. ii. critical cues for detecting approaching objects. *Journal of neurophysiology*, 68(5):1667–1682, 1992.

G. M. Siouris. *Missile guidance and control systems*. Springer Verlag, 2004.

D. Soccol, S. Thurrowgood, and M. V. Srinivasan. A vision system for optic-flow-based guidance of UAVs. In *Proceedings, Ninth Australasian Conference on Robotics and Automation, Brisbane*, pages 10–12, 2007.

X. Song and R. Nevatia. A model-based vehicle segmentation method for tracking. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1124–1131. IEEE, 2005.

L. Spacek and C. Burbridge. Instantaneous robot self-localization and motion estimation with omnidirectional vision. *Robotics and Autonomous Systems*, 55 (9):667–674, 2007.

M. V. Srinivasan. How insects infer range from visual motion. *Reviews of Occulomotor Research*, 5:139–156, 1993.

M. V. Srinivasan. Honeybees as a model for the study of visually guided flight, navigation, and biologically inspired robotics. *Physiological Reviews*, 91(2):413–460, 2011.

M. V. Srinivasan and M. Davey. Strategies for active camouflage of motion. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 259 (1354):19, 1995.

M. V. Srinivasan and M. Lehrer. Temporal acuity of honeybee vision: behavioural studies using moving stimuli. *Journal of Comparitive Physiology*, 155:297–312, 1984.

M. V. Srinivasan and S. Zhang. Visual motor computations in insects. *Annual Review of Neuroscience*, 27:679–696, 2004.

M. V. Srinivasan and S. W. Zhang. Visual control of honeybee flight. *Orientation and Communication in Arthropods*, 84:95–113, 1997.

M. V. Srinivasan, M. Lehrer, W. H. Kirchner, and S. W. Zhang. Range perception through apparent image speed in freely-flying honeybees. *Visual Neuroscience*, 6: 519–535, 1991.

M. V. Srinivasan, S. W. Zhang, and K. Chandrashekara. Evidence for two distinct movement-detecting mechanisms in insect vision. *Naturwissenschaften*, 80:38–41, 1993.

M. V. Srinivasan, S. W. Zhang, M. Lehrer, and T. S. Collett. Honeybee navigation en route to the goal: Visual flight control and odometry. *Journal of Experimental Biology*, 199:237–244, 1996.

M. V. Srinivasan, S. W. Zhang, J. S. Chahl, E. Barth, and S. Venkatesh. How honeybees make grazing landings on flat surfaces. *Biological Cybernetics*, 83(3): 171–183, Aug. 2000.

M. V. Srinivasan, S. W. Zhang, J. S. Chahl, G. Stange, and M. Garratt. An overview of insect inspired guidance for application in ground and airborne platforms. *Proc. Inst. Mech. Engnrs. Part G*, 218:375–388, 2004.

M. V. Srinivasan, S. Thurrowgood, and D. Soccol. An optical system for guidance of terrain following in UAVs. In *Proc. IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 51–56, Sydney, Australia, 2006.

M. V. Srinivasan, S. Thurrowgood, and D. Soccol. Competent vision and navigation systems: From flying insects to autonomously navigating robots. *IEEE Robot. Autom. Mag.*, 16(3):59–71, Sept. 2009.

M. V. Srinivasan, R. J. D. Moore, S. Thurrowgood, D. Soccol, and D. Bland. From biology to engineering: insect vision and applications to robotics. In F. G. Barth, J. A. C. Humphrey, and M. V. Srinivasan, editors, *Frontiers in Sensing*, pages 19–39. Springer-Verlag, Wien, 2012.

G. Stange and J. Howard. An ocellar dorsal light response in a dragonfly. *Journal of Experimental Biology*, 83(1):351–355, 1979.

G. P. Stein, O. Mano, and A. Shashua. A robust method for computing vehicle ego-motion. In *Proc. IEEE Intelligent Vehicles Symposium*, pages 362–368, Dearborn, MI, Oct. 2000.

R. H. Stone. *Configuration design of a canard configured tail sitter unmanned air vehicle using multidisciplinary optimization.* PhD thesis, University of Sydney, Sydney, NSW, 1999.

W. Sturzl, N. Boeddeker, L. Dittmar, and M. Egelhaaf. Mimicking honeybee eyes with a 280 degree field of view catadioptric imaging system. *Bioinspiration & Biomimetics*, 5:036002, 2010.

T. Suzuki and T. Kanade. Measurement of vehicle motion and orientation using optical flow. In *Proc. IEEE International Conference on Intelligent Transportation Systems*, pages 25–30, Tokyo, Japan, Oct. 1999.

A. Talukder and L. Mathies. Real-time detection of moving objects from moving vehicles using dense stereo and optical flow. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, Sendai, Japan, Sept. 2004.

L. F. Tammero and M. H. Dickinson. The influence of visual landscape on the free flight behavior of the fruit fly drosophila melanogaster. *Journal of Experimental Biology*, 205(3):327–343, 2002.

J. P. Tardif, Y. Pavlidis, and K. Daniilidis. Monocular visual odometry in urban environments using an omnidirectional camera. In *Intelligent Robots and Systems. IEEE/RSJ International Conference on*, pages 2531–2538. IEEE, 2008.

C. P. Taylor. Contribution of compound eyes and ocelli to steering of locusts in flight: I. behavioural analysis. *Journal of Experimental Biology*, 93(1):1–18, 1981.

G. K. Taylor and H. G. Krapp. Sensory systems and flight stability: What do insects measure and why? *Insect Mechanics and Control*, page 231, 2007.

C. Teuliere, L. Eck, and É. Marchand. Chasing a moving target from a flying UAV. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4929–4934, 2011.

S. Thakoor, J. Chahl, M. V. Srinivasan, L. Young, F. Werblin, B. Hine, and S. Zornetzer. Bioinspired engineering of exploration systems for NASA and DoD. *Artificial life*, 8(4):357–369, 2002.

S. Thakoor, S. Zornetzer, B. Hine, J. Chahl, and G. Stange. Bioinspired engineering of exploration systems: a horizon sensor/attitude reference system based on the dragonfly ocelli for mars exploration applications. *Journal of Robotic Systems*, 20 (1):35–42, 2003.

W. B. Thompson and T. C. Pong. Detecting moving objects. *International Journal of Computer Vision*, 4(1):39–57, 1990.

S. Thurrowgood, W. Stuerzl, D. Soccol, and M. V. Srinivasan. A panoramic stereo imaging system for aircraft guidance. In *Proc. Ninth Australasian Conference on Robotics and Automation*, Brisbane, Australia, Dec. 2007.

S. Thurrowgood, D. Soccol, R. J. D. Moore, D. Bland, and M. V. Srinivasan. A vision based system for attitude estimation of UAVs. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, St Louis, MO, Oct. 2009.

S. Thurrowgood, R. J. D. Moore, D. Bland, D. Soccol, and M. V. Srinivasan. UAV attitude control using the visual horizon. In *Proc. Australasian Conference on Robotics and Automation*, Brisbane, Australia, Dec. 2010.

C. L. Tisse, H. Durrant-Whyte, and R. A. Hicks. An optical navigation sensor for micro aerial vehicles. *Computer Vision and Image Understanding*, 105(1):21–29, 2007a.

C. L. Tisse, O. Frank, and H. Durrant-Whyte. Hemispherical depth perception for slow-flyers using coaxially aligned fisheye cameras. In *Proc. International Symposium on Flying Insects and Robots*, page 123, Ascona, Switzerland, Aug. 2007b.

S. Todorovic and M. C. Nechyba. A vision system for intelligent mission profiles of micro air vehicles. *Vehicular Technology, IEEE Transactions on*, 53(6):1713–1725, 2004.

S. Todorovic, M. C. Nechyba, and P. G. Ifju. Sky/ground modeling for autonomous

MAV flight. In *Proc. IEEE International Conference on Robotics and Automation*, volume 1, pages 1422–1427. IEEE, 2003.

P. H. S. Torr, A. Zisserman, and D. W. Murray. Motion clustering using the trilinear constraint over three views. In *In Workshop on Geometrical Modeling and Invariants for Computer Vision. Xidian*, 1995.

UAS Roadmap. Unmanned aircraft systems roadmap 2005–2030. Technical report, Office of the Secretary of Defense, Aug. 2005.

K. P. Valavanis. *Advances in Unmanned Aerial Vehicles: State of the Art and the Road to Autonomy*. Springer, 2007.

A. van den Kroonenberg, T. Martin, M. Buschmann, J. Bange, and P. Vörsmann. Measuring the wind vector using the autonomous mini aerial vehicle M2AV. *J. Atmos. Oceanic Technol.*, 25:1969–1982, 2010.

J. H. van Hateren and C. Schilstra. Blowfly flight and optic flow. ii. head movements during flight. *Journal of Experimental Biology*, 202(11):1491–1500, 1999.

A. Verri and E. Trucco. Finding the epipole from uncalibrated optical flow. *Image and Vision Computing*, 17(8):605–609, 1999.

A. Viquerat, L. Blackhall, A. Reid, S. Sukkarieh, and G. Brooker. Reactive collision avoidance for unmanned aerial vehicles using doppler radar. In C. Laugier and R. Siegwart, editors, *Field and Service Robotics*, volume 42, pages 245–254. Springer Berlin / Heidelberg, 2008.

R. Wehner. Arthropods. *Animal homing*, 2:43–144, 1992.

S. Weiss, D. Scaramuzza, and R. Siegwart. Monocular-SLAM-based navigation for autonomous micro helicopters in GPS-denied environments. *Journal of Field Robotics*, 28(6):854–874, 2011.

W. G. Wellington. Motor responses evoked by the dorsal ocelli of sarcophaga aldrichi parker, and the orientation of the fly to plane polarized light. *Nature*, 172:1177–1179, 1953.

W. G. Wellington. Bumblebee ocelli and navigation at dusk. *Science*, 183(4124): 550–551, 1974.

N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor. Omni-directional vision for robot navigation. In *Omnidirectional Vision, 2000. Proceedings. IEEE Workshop on*, pages 21–28, 2000.

R. Yanushevsky. *Modern missile guidance*. CRC, 2007.

Y. Yasushi. Omni-directional sensing and its applications. *IEEE Trans. Inf. & Syst., E*, 82, 1999.

A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Computing Surveys (CSUR)*, 38(4), 2006.

L. C. L. Yuan. Homing and navigational courses of automatic target-seeking devices. *Journal of Applied Physics*, 19:1122, 1948.

D. Zachariah and M. Jansson. Self-motion and wind velocity estimation for small-scale UAVs. In *Robotics and Automation (ICRA), 2011 IEEE International Conference*

on, pages 1166–1171. IEEE, 2011.

L. A. Zadeh. Fuzzy sets*. *Information and control*, 8(3):338–353, 1965.

J.-C. Zufferey and D. Floreano. Fly-inspired visual steering of an ultralight indoor aircraft. *IEEE Trans. Robot.*, 22:137–146, 2006.

J.-C. Zufferey, A. Klaptocz, A. Beyeler, J.-D. Nicoud, and D. Floreano. A 10-gram microflyer for vision-based indoor navigation. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, Beijing, China, Oct. 2006.

J.-C. Zufferey, A. Beyeler, and D. Floreano. Near-obstacle flight with small UAVs. In *Proc. International Symposium on Unmanned Aerial Vehicles*, Orlando, FL, June 2008.

J.-C. Zufferey, A. Beyeler, and D. Floreano. Autonomous flight at low altitude with vision-based collision avoidance and GPS-based path following. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 3329–3334, may 2010.