A HIDDEN MARKOV MODEL AND RELATIVE ENTROPY RATE APPROACH TO VISION-BASED DIM TARGET DETECTION FOR UAV SENSE-AND-AVOID

A DISSERTATION
SUBMITTED TO THE FACULTY OF BUILT ENVIRONMENT AND ENGINEERING,
SCHOOL OF ENGINEERING SYSTEMS OF THE QUEENSLAND UNIVERSITY OF TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

John Lai B. E. (Aerospace Avionics)
March 2010
Keywords

- Collision avoidance
- Dim target detection
- Filter bank
- Hidden Markov model
- Morphological filtering
- Relative entropy rate
- Sense-and-avoid
- Uninhabited/Unmanned aerial vehicle
Abstract

Uninhabited aerial vehicles (UAVs) are a cutting-edge technology that is at the forefront of aviation/aerospace research and development worldwide. Many consider their current military and defence applications as just a token of their enormous potential. Unlocking and fully exploiting this potential will see UAVs in a multitude of civilian applications and routinely operating alongside piloted aircraft. The key to realising the full potential of UAVs lies in addressing a host of regulatory, public relation, and technological challenges never encountered before.

Aircraft collision avoidance is considered to be one of the most important issues to be addressed, given its safety critical nature. The collision avoidance problem can be roughly organised into three areas: 1) Sense; 2) Detect; and 3) Avoid. Sensing is concerned with obtaining accurate and reliable information about other aircraft in the air; detection involves identifying potential collision threats based on available information; avoidance deals with the formulation and execution of appropriate manoeuvres to maintain safe separation.

This thesis tackles the detection aspect of collision avoidance, via the development of a target detection algorithm that is capable of real-time operation onboard a UAV platform. One of the key challenges of the detection problem is the need to provide early warning. This translates to detecting potential threats whilst they are still far away, when their presence is likely to be obscured and hidden by noise. Another important consideration is the choice of sensors to
capture target information, which has implications for the design and practical implementation of the detection algorithm. The main contributions of the thesis are: 1) the proposal of a dim target detection algorithm combining image morphology and hidden Markov model (HMM) filtering approaches; 2) the novel use of relative entropy rate (RER) concepts for HMM filter design; 3) the characterisation of algorithm detection performance based on simulated data as well as real in-flight target image data; and 4) the demonstration of the proposed algorithm’s capacity for real-time target detection. We also consider the extension of HMM filtering techniques and the application of RER concepts for target heading angle estimation.

In this thesis we propose a computer-vision based detection solution, due to the commercial-off-the-shelf (COTS) availability of camera hardware and the hardware’s relatively low cost, power, and size requirements. The proposed target detection algorithm adopts a two-stage processing paradigm that begins with an image enhancement pre-processing stage followed by a track-before-detect (TBD) temporal processing stage that has been shown to be effective in dim target detection. We compare the performance of two candidate morphological filters for the image pre-processing stage, and propose a multiple hidden Markov model (MHMM) filter for the TBD temporal processing stage. The role of the morphological pre-processing stage is to exploit the spatial features of potential collision threats, while the MHMM filter serves to exploit the temporal characteristics or dynamics.

The problem of optimising our proposed MHMM filter has been examined in detail. Our investigation has produced a novel design process for the MHMM filter that exploits information theory and entropy related concepts. The filter design process is posed as a mini-max optimisation problem based on a joint RER cost criterion. We provide proof that this joint RER cost criterion provides a bound on the conditional mean estimate (CME) performance of our MHMM filter, and this in turn establishes a strong theoretical basis connecting our filter
design process to filter performance. Through this connection we can intelligently compare and optimise candidate filter models at the design stage, rather than having to resort to time consuming Monte Carlo simulations to gauge the relative performance of candidate designs. Moreover, the underlying entropy concepts are not constrained to any particular model type. This suggests that the RER concepts established here may be generalised to provide a useful design criterion for multiple model filtering approaches outside the class of HMM filters.

In this thesis we also evaluate the performance of our proposed target detection algorithm under realistic operation conditions, and give consideration to the practical deployment of the detection algorithm onboard a UAV platform. Two fixed-wing UAVs were engaged to recreate various collision-course scenarios to capture highly realistic vision (from an onboard camera perspective) of the moments leading up to a collision. Based on this collected data, our proposed detection approach was able to detect targets out to distances ranging from about 400m to 900m. These distances, (with some assumptions about closing speeds and aircraft trajectories) translate to an advanced warning ahead of impact that approaches the 12.5 second response time recommended for human pilots. Furthermore, readily available graphic processing unit (GPU) based hardware is exploited for its parallel computing capabilities to demonstrate the practical feasibility of the proposed target detection algorithm. A prototype hardware-in-the-loop system has been found to be capable of achieving data processing rates sufficient for real-time operation. There is also scope for further improvement in performance through code optimisations.

Overall, our proposed image-based target detection algorithm offers UAVs a cost-effective real-time target detection capability that is a step forward in addressing the collision avoidance issue that is currently one of the most significant obstacles preventing widespread civilian applications of uninhabited aircraft. We also highlight that the algorithm development process has led to the discovery of a powerful multiple HMM filtering approach and a novel RER-based multiple
filter design process. The utility of our multiple HMM filtering approach and RER concepts, however, extend beyond the target detection problem. This is demonstrated by our application of HMM filters and RER concepts to a heading angle estimation problem.
# Contents

Keywords iii  
Abstract iv  
List of Figures xi  
List of Abbreviations xii  
Statement of Original Authorship xv  
Supervisory Group Details xvi  
Acknowledgements xvii  
List of Manuscripts xviii  
Proof of Submission and Publication of Manuscripts xix  
Acknowledgement of Joint Authorship of Papers xx

## 1 Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Background and Motivation</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Description of Research Problem</td>
<td>8</td>
</tr>
<tr>
<td>1.3 Overall Objectives of the Study</td>
<td>9</td>
</tr>
<tr>
<td>1.4 Specific Aims of the Study</td>
<td>10</td>
</tr>
</tbody>
</table>
2 Literature Review

2.1 Background Theory and Concepts

2.1.1 Morphological Filtering

2.1.2 Hidden Markov Model

2.1.3 Relative Entropy

2.1.4 Summary

2.2 Vision-Based UAV Sense-and-Avoid

2.3 Dim Target Detection

2.3.1 Image Pre-Processing

2.3.2 Temporal Filtering

2.4 Target Tracking

2.5 Hidden Markov Model Approximation

2.6 Summary

3 Morphological Pre-Processing (C1)

4 Hidden Markov Model Filter Banks (C2)

5 Hidden Markov Model Design (J1)

6 Airborne Collision-Detection System (J3)

7 Target Heading Angle Estimation (J2)

8 Conclusions

8.1 Summary of Key Contributions

8.2 Future Work

A Definition of Authorship and Contribution to Publication

B Manuscript J2 Proof of Submission
List of Figures

1.1 Relationship between sensing, detection, and avoidance functions. 5
1.2 A two-stage processing paradigm for dim target detection. . . . . . 11
1.3 Proposed two-stage processing approach for dim target detection. 13

5.1 Illustration of RER-based HMM design. . . . . . . . . . . . . . . 62
List of Abbreviations

ADS-B  Automatic Dependent Surveillance - Broadcast
AFRL  Air Force Research Laboratory
CME  conditional mean estimate
CME-TBHE  conditional mean estimate track-before-heading-estimation
CMO  close-minus-open
COTS  commercial-off-the-shelf
CPU  central processing unit
CSIRO  Commonwealth Scientific and Industrial Research Organisation
CUDA  compute unified device architecture
DoD  Department of Defense
DRA  Defense Research Associates
EKF  extended Kalman filter
EM  expectation-maximisation
ENU  east, north, up
EO  electro-optical
FAA  Federal Aviation Administration

FDSNR  false-alarm distinctness signal-to-noise-ratio

FPGA  field programmable gate array

GPS  Global Positioning System

GPU  graphic processing unit

HDOP  horizontal dilution of precision

HMM  hidden Markov model

IMM  interacting multiple-model

IMU  inertial measurement unit

IR  infrared

LIDAR  light detection and ranging

MAP-TBHE  maximum a priori track-before-heading-estimation

MHMM  multiple hidden Markov model

MITL  man-in-the-loop

PCI  peripheral component interconnect

PS  preserved-sign

PSNR  peak signal-to-noise ratio

QUT  Queensland University of Technology

RER  relative entropy rate

SIMD  single-instruction multiple-data
SNR  signal-to-noise ratio

TBD  track-before-detect

TBHE  track-before-heading-estimation

TCAS  Traffic Alert and Collision Avoidance System

TDSNR  target distinctness signal-to-noise-ratio

UAS  uninhabited/unmanned aircraft system

UAV  uninhabited/unmanned aerial vehicle

UERE  user equivalent range error

VDOP  vertical dilution of precision
Statement of Original Authorship

“The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.”

Signature _______________________

Date _______________________

xv
Supervisory Group Details

- Prof. Peter O’Shea (QUT); Principal Supervisor;  
  pj.oshea@qut.edu.au

- Prof. Rodney Walker (QUT); Associate Supervisor;  
  ra.walker@qut.edu.au

- Dr Jason J. Ford (QUT); Associate Supervisor;  
  j2.ford@qut.edu.au

- Dr Michael Bosse (CSIRO); Associate Supervisor;  
  Mike.Bosse@csiro.au
Acknowledgements

I would like to thank my supervisors for having faith in me, even when I had none in myself, and for showing me a light at the end of the tunnel, even when I could not see a way out. I would also like to thank my family for their unwavering support and encouragement throughout my studies, and for simply ‘putting up with me’. Finally, I would like to acknowledge the generous financial support provided by the Australian Government, QUT, and CSIRO via their scholarship programs.
List of Manuscripts

Journal Papers


Conference Papers


Proof of Submission and Publication of Manuscripts

Journal Papers

J1 Published and available from IEEEXplore
(http://ieeexplore.ieee.org/Xplore/guesthome.jsp)
Digital Object Identifier: 10.1109/TSP.2009.2028115

J2 Under peer review. See Appendix B for email confirming paper submission and assignment of manuscript number.

J3 Under peer review. See Appendix C for email confirming paper submission and assignment of manuscript number.

Conference Papers

C1 Published and available from the Australian Robotics & Automation Association website

C2 Published and available from IEEEXplore
(http://ieeexplore.ieee.org/Xplore/guesthome.jsp)
Digital Object Identifier: 10.1109/DICTA.2008.61
Acknowledgement of Joint Authorship of Papers

The following is a list of coauthors for one or more papers presented in this thesis:

- Dr Michael Bosse
- Dr Jason Ford
- Dr Luis Mejias
- Prof. Peter O’Shea
- Prof. Rodney Walker
Chapter 1

Introduction

1.1 Background and Motivation

On July 13, 2006, before the United States Senate Committee on Commerce, Science & Transportation, Nicholas Sabatini, the Federal Aviation Administration’s (FAA) Associate Administrator for Aviation Safety declared that “The development and use of unmanned aircraft systems (UAS) is the next great step forward in the evolution of aviation.” [1]. If, according to Nicholas Sabatini, unmanned or uninhabited aircraft are indeed the future of aviation, then the future is now right on our doorstep.

Simply stated, an uninhabited aerial vehicle (UAV) \(^1\) is any airborne device that operates without on-board crew or pilots. Such a device may fly autonomously without any human intervention, or otherwise be controlled by human operators at a remote location. They serve a multitude of purposes, from recreation and asset monitoring to surveillance and combat operations in hostile warzones. There is also a great variety in the size of UAVs; they can possess

\(^1\)The term UAV is often used to denote the physical aircraft itself only, whereas the term UAS encompasses the whole aircraft system that may include ground-based operators and support crews etc. In this thesis, we will often use the terms interchangably.
wingspans of about 15 centimeters to 75 meters, and can weigh from approximately 100 grams to over 11 tonnes [2].

The one thing we can be sure about with UAVs is that their numbers and usage are growing dramatically. In 2000 in the United States alone, the Department of Defense (DoD) had fewer than 50 uninhabited aircraft in their inventory; as of May 2008, they had more than 6,000 [3]. Furthermore, these military UAVs have also been extensively deployed. As of the end of May 2008, the US military services’ UASs had performed more than 230,000 flight hours in fiscal year 2008 [4].

Whilst most of the attention and activities to date have been focused on military oriented UAVs, worldwide interest in non-military UAV applications has been steadily gathering momentum over the last decade. Even so, investment in research and development for commercial and civil UAVs still pales in comparison to the vast sums being spent by governments and defence organisations on military UAVs. In the United States alone, the DoD budget plans for UASs total in excess of $17 billion for the 2008 to 2013 fiscal years [3]. It then comes as no surprise that of all UAV types, the vast majority (90 to 95 percent) are military [5], accounting for 90 percent of all funding for UAV systems worldwide [6].

As far as the development of civil and commercial UAVs is concerned, the current dominance of military oriented UAVs may only be temporary. This is because history is littered with examples of military endeavours that have ultimately become commercial successes in their own right. The Global Positioning System (GPS) and the World Wide Web/Internet are just a couple of the more prominent examples of technologies whose commercial and civil applications have now largely overshadowed their initial military exclusive uses. Similarly, this is the vision/expectation that many have for UAV technology. Although owing much of its current maturity and innovation to the guidance provided by the

\footnote{The total number represents the number of uninhabited aircraft, rather than UAS, and includes test and training assets.}
military, it is widely acknowledged that the true potential of UAVs can only be realised once that transition into the civil and commercial environment occurs [7]; that is, when uninhabited aircraft are granted the same freedom as traditional piloted aircraft, and when uninhabited aircraft are able to operate routinely among their piloted counterparts.

There are many issues that must be addressed before UAVs can operate safely and effectively in the civil and commercial environment. Integrating UAVs into the civilian airspace and air traffic management framework is an enormous challenge, not least because of the sheer number of parties and stakeholders involved in the UAS industry (legislators, aviation regulators, standards developing organisations, aircraft developers, manufacturers, operators, distributors etc.). A further complication lies in the nature of the existing aviation industry; it is arguably a legacy from the days when aviation implied piloted aviation. Thus, there is very little provision (if any) in the current regulations, standards, operation procedures etc. for ‘non-piloted’ aviation. According to [8] “It is not clear, . . . whether existing regulations that are based on a historical pairing of pilot and plane can be adapted to accommodate UASs, or whether UASs constitute a fundamentally different category of aircraft requiring their own set of regulations.” As a consequence, the FAA policy to date (and similarly amongst aviation authorities worldwide) has been to enforce highly restrictive operating requirements that involve segregating UAV operations from those of piloted aircraft, and confining the former to specially designated airspaces [2]. Although steps have been taken by the industry to develop the necessary framework to support non-piloted aviation, it is still very much a work in progress [8, 9].

The interaction between the uninhabited aircraft and all other parties (including piloted aircraft, other uninhabited aircraft, air traffic controllers, navigational aids etc.) lies at the heart of the integration problem. Spectrum management [10] (which deals with communication between aircraft, ground controllers etc.) and collision avoidance are two issues that are key to enabling uninhabited aircraft to
routinely fly among civilian, piloted aircraft. In fact, a special group formed in 2005 from the US military departments identified an automated ‘sense-and-avoid’ \(^3\) capability and secure, robust communication links as the two foremost challenges to achieving acceptance of uninhabited military aircraft routinely flying among civilian, piloted aircraft [4]. The collision avoidance or sense-and-avoid issue has been under the spotlight in recent years, not least because of the string of incidents involving UAVs and piloted aircraft since 2001. A notable incident occurred in 2004 which involved an Airlines Airbus A300B4 carrying over 100 passengers and a German army EMT Luna tactical UAV. The aircraft missed each other by less than 50 meters in the skies over Kabul in Afghanistan [11].

The significance of the collision avoidance issue is further highlighted in a paper by the MITRE Corporation, which reports that for UAVs “…collision avoidance has arguably become the most pressing safety concern, and, consequently, the focus of numerous studies by government, industry, universities, and research institutions worldwide.” [5]. Furthermore, a key objective discussed in the DoD Unmanned Systems Roadmap 2007-2032 [4] is to develop uninhabited systems that have the ability to autonomously sense and avoid other objects in order to provide a level of safety equivalent to comparable human piloted systems.

There are many challenges associated with UAS collision avoidance [5, 12]. Complex processes are involved in ensuring safety between two or more aircraft (or other airborne objects). It is anticipated that UAS collision avoidance will function in a ‘layered approach’ similar to the current framework in place for piloted aircraft [12]. This thesis will focus on a ‘pilotless version’ of the pilot-centric ‘see-and-avoid’ layer, often referred to as ‘sense-and-avoid’. See-and-avoid for piloted aircraft is often regarded as the critical ‘last-line-of-defence’ in the collision avoidance framework, to be relied upon when all other regular airspace procedures and separation services earlier in the chain have failed to trap and

\(^3\)For now, the terms collision avoidance and sense-and-avoid are used interchangeably. Later on in the discussion, sense-and-avoid will take on a more specific meaning.
avert a potential collision. Essentially, this involves the pilot visually scanning for potential collision threats and performing the necessary avoidance manoeuvres to prevent a collision [13]. Replicating the safety functionality of see-and-avoid for uninhabited aircraft requires many complex processes previously handled by a human pilot to be wholly or partly transferred to an automated system. The exact definition and terminology of these processes varies across authors and aviation experts; however, keeping aircraft safely apart in the air will generally involve the following functions:

1. Sensing: Obtaining accurate and reliable information about other aircraft in the air; sometimes referred to as gaining situational awareness.

2. Detection: Identifying and prioritising potential collision threats based on the information obtained via the sensors.

3. Avoidance: Determining manoeuvres necessary to maintain safe separation (if required), and engaging the aircraft control systems to carry out the required manoeuvres.

The general relationship between the sensing, detection, and avoidance functions is illustrated in Figure 1.1. The figure describes an iterative process where sensing is followed by detection, and if a threat is found, the avoidance function is then engaged (note that sensing and detection are performed continuously irrespective of whether a threat is found). Furthermore, the binary threat detection decision depicted in Figure 1.1 is a simplification of the type of information

Figure 1.1: Relationship between sensing, detection, and avoidance functions.
that can be supplied to the avoidance function. Typically, detection functions may also provide a probabilistic indication of the likelihood of a threat and also the relative position of the threat so that avoidance strategies can achieve outcomes of highest utility. As mentioned earlier, there are numerous equally valid variations on the above architecture and breakdown of functions (an example sense-and-avoid architecture is described in [14]). Many encapsulate the sensing and detection in one function, while others consider returning the aircraft to its prescribed course after an avoidance manoeuvre as part of the avoidance routine [12].

The detection function is arguably one of the most important aspects of the collision avoidance problem because of the critical information that it supplies to subsequent systems responsible for determining and executing appropriate avoidance manoeuvres. In general, detection can be achieved via cooperative or non-cooperative means, and the physical sensors that collect the information required may be classified as either active or passive.

Cooperative detection strategies rely on each aircraft broadcasting information (such as position, altitude and velocity data) about themselves (either continuously or only when interrogated by other aircraft) so that all aircraft are aware of each others’ location (the Traffic Alert and Collision Avoidance System (TCAS) and Automatic Dependent Surveillance - Broadcast (ADS-B) are two examples of cooperative systems in current use). Hence, for a cooperative scheme to be fully effective, all aircraft without exception must carry specialised transmitters to broadcast information and/or receivers to listen for telemetry from other aircraft. Unfortunately, a single aircraft that is not adequately equipped has the ability to compromise the entire system, since the non-equipped aircraft would effectively be blind to all other aircraft, and vice versa. In summary, the blanket requirement for all aircraft to carry specialised communications equipment significantly undermines the feasibility of cooperative systems in general, and will continue to do so until the installation of such equipment becomes more
affordable. As a result, much research has been conducted into alternative non-cooperative solutions.

In contrast to cooperative methods, non-cooperative detection derives information about other aircraft indirectly via sensors that may be classed as active (e.g. light detection and ranging (LIDAR)/laser and radar), or passive (e.g. electro-optical (EO) and infrared (IR) cameras). A key advantage of non-cooperative systems is that they can be made to function as stand-alone units that are independent of the equipment onboard other aircraft.

The development of non-cooperative detection systems for sense-and-avoid is an active area of research that is being pursued by universities, government organisations, and private enterprises worldwide. Much effort has been dedicated to developing a solution based on passive electro-optical sensors, due to their relatively low cost, weight, and power requirements, compared to the more traditional active sensors like radar. However, this passive non-cooperative approach to collision avoidance is regarded by the DoD Unmanned Systems Roadmap 2007-2032 [4] as one of the most challenging scenarios from a technical perspective. To date, many systems have been proposed, but none have demonstrated the level of maturity and effectiveness required to attain certification by aviation safety authorities. This is in part due to the lack of consensus within the community on a formal definition for the performance required by a collision avoidance system.

Thus, the research detailed in this thesis seeks to develop a robust collision threat detection capability for use in UAV sense-and-avoid based on a passive sensor in non-cooperative traffic conditions. The research in this thesis is also motivated by some earlier promising work carried out by Queensland University of Technology (QUT) and Commonwealth Scientific and Industrial Research Organisation (CSIRO) researchers in the area of airborne collision avoidance [16].

---

4Some consider that an acceptable level of risk is $1 \times 10^{-9}$, that is, one collision for every billion flight hours, based on a ‘catastrophic’ event that should be ‘extremely improbable’ from the FAA’s Safety Management System [12]. Others are attempting to quantify the safety requirements based on an equivalency to the human pilot ‘see-and-avoid’ capability [15].
1.2 Description of Research Problem

The research problem under investigation concerns the detection of potential collision-course objects in an airborne environment. This detection can be achieved by exploiting certain characteristics and dynamics that are particular to collision-course objects. The exact nature of these characteristics and dynamics depends largely on the type of sensor used to gather information about the environment. In this thesis, a computer-vision based solution to the detection problem is proposed. This approach is proposed because EO camera technology is widely accessible and cost effective, and the hardware itself generally has relatively low power, weight, and volume requirements compared with other sensor types [17]. Moreover, there is some analysis to suggest that a sense-and-avoid solution based on vision technology may offer the best chances for regulator approval [18].

Through a vision-based approach, it may be possible to detect collision course objects based on their physical appearance, or on the dynamics (motion) that they exhibit. The physical attributes of collision course objects, such as colour, shape, and size, depend largely on ambient lighting and atmospheric conditions, as well as the distance to the object. Given the complex nature of ambient lighting and atmospheric conditions, coupled with the large variety of aircraft paint schemes, colour is perhaps one of the weaker distinguishing characteristics of collision-course objects.

On the other hand, the size and (to a degree) shape of a collision-course object on the image plane of a camera is perhaps more useful. Given the 12.5 second reaction time prescribed for human pilots [13], a collision-course object must be detected at a distance of over 1 kilometer (assuming a closing velocity of 200 knots) to avoid a collision \(^5\). At this distance, aircraft and other objects of similar dimensions may take up an area anywhere from a few pixels to less than

\(^5\)In the interests of safety, it is desirable for targets to be detected as early (i.e. as far away) as possible.
one pixel on the image plane. It can be argued that a few pixels cannot really define any sort of ‘shape’, but at least it can be deduced that objects will tend to appear fainter and smaller the further away they are from the observer. Hence, in the interests of early detection (so that the pilot(s) may have as much time as possible to react), it is desirable to identify collision course objects whilst they appear as dim small point-like features.

Of all the characteristics of collision-course objects, their somewhat unique dynamics is perhaps the most suitable attribute to exploit. Objects on a collision course are characterised by their lack of motion when observed from the image plane of a fixed onboard vision sensor. Features that are moving across the image plane do not correspond to collision-course objects.

In this thesis we consider various filtering approaches to facilitate the detection of collision-course objects. In particular, we investigate filtering techniques that are able to exploit the physical and dynamic characteristic of collision-course objects to identify genuine collision threats from background noise and clutter.

1.3 Overall Objectives of the Study

The overall objective of the study is to develop a computer-vision based dim target detection algorithm in the context of a UAV sense-and-avoid application. As a secondary objective, we propose to concurrently explore information theory and other related concepts that may lead to a systematic theory-based algorithm design process. This is intended to provide a theoretical basis for our detection algorithm to complement experimental performance results based on computer generated data, as well as real target image data captured on-board UAVs recreating collision-course scenarios.

\(^6\)Depending on camera resolution, field of view etc.
1.4 Specific Aims of the Study

In this study, a vision-based solution to the dim target detection problem is considered, and specifically, an EO camera sensor is assumed to provide the visual target information. The typical output from an EO camera is a sequence of images (i.e. a video stream). Detecting collision-course objects is then a matter of searching for objects within the image sequence that possess the characteristics highlighted in the earlier problem description. There are many algorithms that have been proposed to perform this search, but two general approaches have emerged. One approach is to process each frame on an individual basis; only information from the current frame is used in the search for objects in that frame. Some of the processing options available afterwards include correlating the detections across image frames and forming tracks that describe the objects’ motion over time. Objects that do not have the motion profile matching the desired dynamic can then be eliminated as potential collision-course threats.

An alternative approach involves integrating the information over a number of image frames before deciding whether any objects of interest have been found. This approach is often referred to as a ‘track-before-detect’ (TBD) technique. A key advantage of the TBD approach is that the dynamics of the objects of interest are directly incorporated into the detection process, rather than being considered separately in some post-detection operation. Moreover, the integration of object information over time has been shown to be an effective strategy for detecting dim objects [19], which, in a collision avoidance context, equates to earlier detection.

We propose a two-stage processing approach for detecting collision-course objects that searches for targets on a ‘per-frame’ basis as well as over a sequence of frames. These two stages are: 1) an image pre-processing stage that, within each frame, highlights small point-like features; and 2) a subsequent TBD temporal filtering stage that exploits the dynamics of collision-course objects.
Figure 1.2 illustrates this two-stage processing approach. In the course of developing the detection algorithm, we aim to:

1. Identify a suitable image pre-processing approach (papers C1 and C2);
2. Identify a suitable TBD temporal filtering approach (papers C1 and C2);
3. Develop supporting tools to aid the design of filtering parameters to improve detection performance (paper J1);
4. Implement the algorithm on commercially available hardware and achieve real-time processing speeds (paper J3); and
5. Evaluate the algorithm’s detection performance using real image data captured on-board UAVs recreating collision-course scenarios (paper J3).

Figure 1.2: A two-stage processing paradigm for dim target detection.

1.5 Progression of Research Papers

The detection problem in UAV sense-and-avoid has been the overarching theme in our research program. In this thesis, we have attempted to address this problem directly through the development of a dim target detection algorithm. The algorithm development process has produced a mixture of application-based and theory-based results, and it is our desire to offer readers a balanced view of
CHAPTER 1. INTRODUCTION

these outcomes. To this end, we have divided the following account of our re-
search progress into two sections to highlight separately the application-based
and theory-based contributions of our papers.

From an application-based perspective, our papers provide an account of the
development of a dim target detection algorithm for the UAV sense-and-avoid
application. The proposed detection algorithm is based on a two-stage process-
ing paradigm that has been widely discussed in the literature: target information
captured by sensor hardware is firstly handled by an image pre-processing stage,
the output of which then undergoes further processing by a temporal filtering
stage. Chapter 3 (paper C1) compares two promising image morphology tech-
niques for use in the pre-processing stage. A ‘preserved-sign’ image morphology
technique was found to outperform a more well known ‘close-minus-open’ ap-
proach after extensive simulation. Having identified a suitable technique for
the pre-processing stage, Chapter 4 (paper C2) considers two candidate hidden
Markov model (HMM) based track-before-detect (TBD) approaches for carrying
out the temporal processing stage. Here, we found evidence that a bank of multi-
ple HMM filters offers an improvement in detection performance over a standard
single HMM filter. In particular a bank of multiple HMM filters exhibited a
higher detection rate for a range of false-alarm rates in various target scenar-
ios, and was less sensitive to the image signal-to-noise-ratio. This motivated
us to seek a suitable multiple HMM filtering approach for our target detection
algorithm (see Figure 1.3). In Chapter 5 (paper J1) a multiple HMM filtering
approach is proposed, and we considered the design of multiple HMM filters to
optimise detection performance. A novel multiple HMM filter design method is
proposed based on a joint relative entropy rate (RER) cost criterion that con-
nects filter parameters to detection performance. In summary, the results from
Chapters 3-5 culminate in the proposal of a detection algorithm comprising of
an image morphology pre-processing stage coupled to a multiple HMM temporal
filtering stage.
In the next stage of development we considered the practical deployment of our proposed detection algorithm onboard a UAV platform. Chapter 6 (paper J3) investigates 1) the impact of undesired camera sensor motion onboard the UAV platform (image jitter); 2) detection range performance based on real in-flight target image data; and 3) the capability for real-time operation based on commercially available hardware. It was found that detection distances ranging from 400m to 900m were possible based on vision from a fixed non-stabilised camera sensor that has been compensated for image jitter effects (without jitter compensation, detection distances are significantly reduced). The challenge of achieving real-time computational speeds was overcome by exploiting the parallel processing architectures of commercially available graphic processing unit (GPU) hardware. Overall, when combined with a suitable GPU-based device and camera sensor, our proposed detection algorithm can be used to deliver a real-time target detection capability ready for deployment onto a UAV platform.
We highlight that if a fixed non-stabilised camera is used, the addition of an image jitter compensation algorithm is necessary to maintain reasonable detection range performance.

On the other hand, an equally valid theory-based perspective is that our papers provide an account of the formulation and application of powerful RER based tools for model approximation, filter design, and parameter estimation. In Chapter 5 (paper J1) we exploit information theory and entropy related concepts to propose a RER based procedure for developing suitable HMM representations of models that are outside the class of HMMs and have uncertain dynamics. The utility of our proposed model approximation procedure is illustrated via a multiple HMM design problem where a finite set of HMMs is used to approximate a specified uncertain model. Our proposed model approximation procedure poses the HMM design problem as a mini-max optimisation problem based on a joint RER cost criterion. The optimisation process yields a set of HMMs with an associated design cost that may be used in a multiple HMM (MHMM) filtering approach. We show that the design cost remarkably establishes an overbound on the one-step-ahead MHMM filter conditional mean estimate error performance. This provides a useful connection between model design and filtering performance that can be exploited to gauge the relative performance of candidate model designs without the need for Monte Carlo simulations. These powerful RER based results encouraged us to seek other potential uses of RER concepts. Chapter 7 (paper J2) establishes a connection between RER and probabilistic distance concepts which provides an alternative path for tackling parameter estimation problems. A heading angle estimation problem is used to illustrate the utility of an RER-based parameter estimation approach.

Finally, the papers in this thesis may also be seen as demonstrating the utility of a multiple HMM filtering approach for various applications. In Chapter 4 and Chapter 5, multiple HMM filters are used in dim target detection. In Chapter 7 a target heading angle estimation application is considered.
Chapter 2

Literature Review

2.1 Background Theory and Concepts

The following sections contain some background theory and mathematical concepts that will be discussed throughout this thesis. Furthermore, these sections will also serve to introduce notation that will appear in later chapters.

2.1.1 Morphological Filtering

Having its basis in the theory of mathematical morphology [20], morphological filters represent a powerful class of filtering tools that have found a variety of important applications. This thesis exploits the utility of morphological filters in an image processing context; in particular, for processing discrete 2D image data quantised to a finite number of intensity or greyscale levels, such as might be expected from the output of an electro-optical sensor.

All morphological filters are based on two fundamental operations known as ‘dilation’ and ‘erosion’. Let $Y \oplus S$ and $Y \ominus S$ denote the dilation and erosion of a greyscale image $Y$ by a morphological structuring element $S$, respectively (see [20, 21] for formal definitions of the dilation and erosion operations). More sophisticated secondary filtering operations known as ‘opening’ and ‘closing’ can
be formed by combining the dilation and erosion operations. Let \( Y \circ S \) and \( Y \bullet S \) denote the opening and closing respectively of a greyscale image \( Y \) by a morphological structuring element \( S \). The opening operation is simply defined as an erosion followed by a dilation: \( Y \circ S = (Y \ominus S) \oplus S \). The reverse is true for the closing operation, that is \( Y \bullet S = (Y \oplus S) \ominus S \).

An intuitive understanding of the morphological opening procedure can be gained by visualising the operation as the darkening of locally bright regions (that are smaller than the structuring element size) to the values of their neighbourhood pixels. In a similar manner, morphological closing may be regarded as the process of brightening locally dark regions (that are smaller than the structuring element size) to match the values of neighbouring pixels.

It follows from the above concepts that taking the difference between an image and its morphological opening will produce an output identifying positively contrasting features (pixel groups that are brighter than their neighbours). This corresponds to what is referred to in the literature as a ‘top-hat’ transformation [22]. Let \( TH (Y, S) = Y - (Y \circ S) \) denote the top-hat transformation of image \( Y \). Its dual, the ‘bottom-hat’ transformation of image \( Y \), is defined as \( BH (Y, S) = (Y \bullet S) - Y \) [22] and accordingly will highlight negatively contrasting features (pixel groups that are darker than their neighbours). In both of these transformations, only features smaller than the structuring element are preserved, whilst larger features are suppressed. Thus, a judicious structuring element choice in combination with the above morphological transformations will allow features within an image to be identified based on their geometrical size.

In this thesis, two morphological filters will be considered that are based on the top-hat and bottom-hat transformations. The first is a ‘close-minus-open’ (CMO) filtering approach that is given by the sum of the top-hat and bottom-hat transformations:

\[
CMO (Y, S) = TH (Y, S) + BH (Y, S) \quad (2.1)
\]
As its name suggests, this combination of top-hat and bottom-hat transformations simplifies down to the difference between the closing and opening of an image. The CMO filtering technique, which has been referred to elsewhere in the literature as a self-complementary top-hat filtering approach [21], simultaneously preserves both positively and negatively contrasting features (both positive and negative contrasting features result in a non-negative output).

The second morphological filtering technique is termed a ‘preserved-sign’ (PS) approach [16], and is defined as the difference between the top-hat and bottom-hat transformations:

\[
PS(Y, S) = TH(Y, S) - BH(Y, S)
\]

\[
= [Y - (Y \circ S)] - [(Y \bullet S) - Y]
\]

\[
= 2Y - (Y \circ S) - (Y \bullet S).
\]

Note that the above definition shows that the PS filtering approach is in fact a variation on the top-hat contrast enhancement operator described in [21]. The distinguishing feature of the PS filtering approach is that the response to positively contrasting features is non-negative, whereas the response to negatively contrasting features is non-positive. This is in contrast with the CMO approach where any contrasting feature (positive or negative) is expressed as a non-negative output. Thus, unlike the CMO technique, the PS filtering approach identifies contrasting features and provides additional information regarding the contrast ‘polarity’.

In this thesis we consider the utility of morphological processing approaches for detecting small point-like features that may correspond to potential collision-course targets.
2.1.2 Hidden Markov Model

The utility and versatility of hidden Markov models (HMM) are evident from their widespread application in a multitude of technical disciplines, including non-linear stochastic control [23,24], signal and image processing [25–30], digital communications [31], and bioinformatics [32]. HMMs have found such a broad range of applications partly due to their well developed statistical formalism which enable parameters to be readily varied and accounted for. HMMs are particularly useful in inherently discrete-valued problems where an underlying quantity to be estimated cannot be observed directly, but may be inferred through another quantity to which it is related in a probabilistic sense. In other words, the hidden or unknown quantities to be estimated (i.e. the ‘states’ of the HMM) can be determined from related quantities that are directly observable (i.e. the ‘measurements’ of the HMM). For example, in a scenario where an imaging sensor is directed at a target, the true location of the target is usually corrupted and hidden by noise i.e. it is not directly observable. However, the noisy sensor output does provide some information (though imperfect) about the target location, and hence may be used to infer the true target location. Although HMMs can only be considered an optimal modelling approach in discrete-valued problems (under certain assumptions), they are often exploited in many continuous-valued problems with complex dynamics to simplify computations and allow tractable solutions to be developed.

A HMM is specified by its states, measurements, initial probabilities, transition probabilities, and measurement probabilities [33]. When a HMM is used to model a particular problem, the states correspond to the unknown quantity to be estimated, and the measurements correspond to the observed quantity related to the states that can be used to make inferences about the state. Let $X_k$ and $Y_k$ denote the state and measurement at time $k$, respectively. In a HMM, a first-order Markov assumption is usually applied, where the $P(X_k|X_{k-1},X_{k-2},\ldots,X_1) = \ldots$
2.1. BACKGROUND THEORY AND CONCEPTS

That is, the unknown quantity is modelled as a first-order Markov chain. The Markov assumption may or may not be appropriate depending on the situation, but it greatly simplifies the modelling by reducing the dependency on past values. The Markov assumption also has a direct impact on the relationship between state values. This relationship is often expressed in the form $A_{mn} = P(X_{k+1} = \text{state } m | X_k = \text{state } n)$ for $1 \leq m, n \leq N$, where $N$ denotes the total number of possible states. The elements of $A$ are known as the transition probabilities of the HMM. The initial probabilities of the HMM $\pi_m = P(X_1 = \text{state } m)$ for $1 \leq m \leq N$ are used to characterise the initial conditions of the Markov chain. Finally, to complete the parameterisation of the HMM, there are the measurement probabilities $B_m$ which provide the link between the measurements and the states of the HMM. Here, it is often assumed that the measurement at a particular time is dependent only on the state at that time, and is independent of all previous states and measurements; that is, $B_m = P(Y_k | X_k = \text{state } m) = P(Y_k | X_k = \text{state } m, X_{k-1}, X_{k-2}, \ldots, X_1, Y_{k-1}, Y_{k-2}, \ldots, Y_1)$ for $1 \leq m \leq N$.

To date, many filtering techniques have been developed to extract ‘optimal’ state estimates from a system modelled by a HMM. From a Bayesian Decision-Theoretic viewpoint [34], the a posteriori density function $p(X_k | Y_k, Y_{k-1}, \ldots, Y_1)$ contains all the information necessary to estimate the state. Depending on the criterion for optimality, the state estimate $\hat{X}_k$ may be computed from the a posteriori density function using the mode of $p(X_k | Y_k, Y_{k-1}, \ldots, Y_1)$ (which is the classical maximum likelihood estimate that maximises $P(\hat{X}_k = X_k)$), or the expected value $E(X_k | Y_k, Y_{k-1}, \ldots, Y_1)$ (which is the conditional mean estimate that minimises an error squared criterion $\int \left| X_k - \hat{X}_k \right|^2 p(X_k | Y_k, Y_{k-1}, \ldots, Y_1) dx$, provided that the second moment of $p(X_k | Y_k, Y_{k-1}, \ldots, Y_1)$ exists). These are perhaps two of the more popular Bayesian estimation techniques, although there are clearly many other estimation criteria that may be appropriate depending on the situation. While the techniques discussed so far are suitable for applications that require real-time online estimation of the state (that is, estimating the state
as the measurements are observed), other techniques exist that are appropriate for applications that require an entire sequence of state values to be estimated optimally. The Viterbi algorithm [35] is an example of a filtering technique that associates an optimal sequence of states to a given sequence of measurements. More specifically, the Viterbi algorithm is a formal technique based on dynamic programming methods that can be used to determine the single best state sequence in the sense of maximising \( P(\hat{X}_k, \hat{X}_{k-1}, \hat{X}_{k-2}, \ldots, \hat{X}_1 \mid Y_k, Y_{k-1}, Y_{k-2}, \ldots, Y_1) \) (this algorithm should not be confused with a ‘Viterbi-based’ filtering approach introduced later in the thesis in Chapter 3).

In this thesis, HMMs are used to characterise the complex dynamics of a collision-course target as seen though an electro-optical imaging sensor. Furthermore, the estimation of target location is examined using the conditional mean estimate criterion.

### 2.1.3 Relative Entropy

One of the key building blocks of this thesis is the concept of relative entropy, which is a fundamental quantity in information theory, a rich area of study that has made significant contributions to a diverse range of fields such as communication theory, statistical physics (thermodynamics), computer science (algorithmic complexity), statistical inference, and probability and statistics. The concepts of entropy and relative entropy in information theory can be defined in terms of probability distributions. Entropy is a quantity that can be interpreted as the uncertainty of a single random variable. It can help to answer questions such as “What is the average length of the shortest description of the random variable?” On the other hand, relative entropy may be considered a measure of the ‘distance’ between two probability distributions. Consider two probability measures \( \mu \) and \( \nu \) on a measurable space \( (\Omega, \mathcal{F}) \). The relative entropy \( D(\mu \parallel \nu) \) of \( \mu \) with
respect to $\nu$ is defined by \[36\]

\[
\mathcal{D}(\mu \parallel \nu) \triangleq \begin{cases} 
\int_{\Omega} \left( \log \frac{d\mu}{d\nu} \right) d\mu, & \text{if } \mu \ll \nu \text{ and } \left| \log \left( \frac{d\mu}{d\nu} \right) \right| \text{ is integrable} \\
+\infty & \text{otherwise,}
\end{cases}
\] (2.3)

where \(\frac{d\mu}{d\nu}\) is the Radon-Nikodym derivative of $\mu$ with respect to $\nu$. Here, $\mu \ll \nu$ denotes that $\mu$ is absolutely continuous with respect to $\nu$, in the sense that $\mu = 0$ if $\nu = 0$. The relative entropy is not a true metric because it is non-symmetric and does not satisfy the triangle inequality. However, it does exhibit some of the properties of a true metric, being always non-negative and is zero if $\mu = \nu$. Therefore, it is often useful to think of the relative entropy as a pseudo-distance measure between $\mu$ and $\nu$.

In this thesis, relative entropy concepts are applied to hidden Markov models (although the theory developed can be generalised beyond HMMs). Due to the sequential nature of the Markov chain state process and measurement process, the relative entropy rate is often a more useful quantity that better captures the dynamics of HMMs. In general, the relative entropy rate (RER) of the measure $\mu$ with respect to $\nu$ is given by

\[
\mathcal{R}(\mu \parallel \nu) \triangleq \lim_{k \to \infty} \frac{1}{k} \mathcal{D}(\mu \parallel \nu). 
\] (2.4)

For two HMMs denoted by $\lambda$ and $\hat{\lambda}$, the usual RER is defined through the probability laws describing the measurement processes of the models. In contrast, of particular interest in this thesis is the RER between the probability laws describing the joint measurement and state processes. To make absolutely clear this distinction between the RER defined through the measurement probability laws and the RER defined through the joint measurement-state probability laws, we introduce an overbar notation such that the joint RER is denote by
\( \mathcal{R} (\lambda \| \hat{\lambda}) \) [37–39]. For example, if \( x_{[0, \infty]}, y_{[0, \infty]} \) and \( \hat{x}_{[0, \infty]}, \hat{y}_{[0, \infty]} \) denote state and measurement processes generated by two models \( \lambda \) and \( \hat{\lambda} \), respectively, then the standard RER is given by \( \mathcal{R} (\lambda \| \hat{\lambda}) = \mathcal{R} \left( p^\lambda \left( y_{[0, \infty]} \right) \big\| p^\hat{\lambda} \left( \hat{y}_{[0, \infty]} \right) \right) \), whereas the joint RER is given by \( \mathcal{R} (\lambda \| \hat{\lambda}) = \mathcal{R} \left( p^\lambda \left( x_{[0, \infty]}, y_{[0, \infty]} \right) \big\| p^\hat{\lambda} \left( \hat{x}_{[0, \infty]}, \hat{y}_{[0, \infty]} \right) \right) \), where \( p^\lambda \) and \( p^\hat{\lambda} \) are the probability laws corresponding to the two models.

This thesis exploits the above joint RER concepts to facilitate design of a finite set of HMMs to approximate a specified uncertain model, in the sense that each of the possible joint state and measurement behaviours representing the uncertainty is reasonably approximated by at least one of the HMMs.

2.1.4 Summary

In this thesis, we show that morphological pre-processing followed by HMM temporal filtering together form a potent combination for performing dim target detection. Our temporal filtering approach employs multiple HMM filters to process the morphological pre-processing output in parallel. This multiple filter approach is particularly useful when there is uncertainty in the target dynamics (for example an unknown heading angle). Consider a standard single filter approach, where the full range of target dynamics must be handled by the one filter model. The advantage of a multiple filter approach stems from each member filter only having to handle a fraction of the possible target dynamics. In our simulation studies we show that this flexibility of a multiple filter approach translates to improved detection performance under uncertain target dynamics.

Furthermore, in this thesis we demonstrate the novel use of RER concepts for filter design. In particular, we formulate a systematic filter design process based on a joint RER cost criterion, and provide a theoretical justification for this design process by showing that our proposed cost criterion establishes a bound on conditional mean estimate filtering error. We successfully applied the filter design process to the design of the multiple HMM filters in our proposed target.
2.2 Vision-Based UAV Sense-and-Avoid

One of the most significant challenges facing the integration of uninhabited aerial vehicles (UAVs) into civilian airspace is the need to develop a ‘sense-and-avoid’ capability for UAVs that can rival or exceed the ‘see-and-avoid’ functions performed by human pilots [14]. In piloted aircraft, see-and-avoid is a safety process that demands pilots to be vigilant and on the lookout for other objects in the sky, and to take action when necessary to avoid a collision [13]. Sense-and-avoid is designed to fulfill a similar collision avoidance role as see-and-avoid, but for ‘pilotless’ aircraft. The analogy between the human eye and man-made optics makes a machine-vision approach a natural candidate for addressing the sense-and-avoid problem. From a more pragmatic perspective, machine-vision represents a particularly attractive ‘sensing’ option because machine-vision sensors (i.e. electro-optical cameras) are commercially accessible and have relatively low power, weight, and volume requirements. These are important criteria in smaller UAV platforms. To date, many government organisations and research groups have been keen to exploit the appeal of vision sensors in the development of a sense-and-avoid capability [16, 40–45]. Perhaps the most advanced in terms of a ‘ready-to-fly’ system is described in a series of publications by the Defense Research Associates (DRA) and the Air Force Research Laboratory (AFRL) [41–43]. DRA and AFRL have jointly developed a sense-and-avoid system based on missile detection technology, which has undergone initial software and real-time hardware testing to be finally integrated onto a UAV platform. Other studies have focused on specific aspects of the sense-and-avoid problem. For example, the challenge of using image information to detect collision-course
targets has been investigated by numerous authors [16,40,44,45], and many candidate detection algorithms have been proposed that have exploited morphological filtering, dynamic programming, optical-flow, and Kalman filtering techniques.

One standout detection approach that has produced very encouraging results (and has heavily influenced the work in this thesis) exploits the combination of an image morphology pre-processing stage coupled to a Viterbi-based dynamic programming temporal filtering algorithm [16]. Using this approach, it was reported that targets could be detected at ranges 35-40\% greater than an alert human observer with prior knowledge of the approximate target location. Upon closer examination, the utility of morphological filters for extracting small point-like features (possibly corresponding to collision-course targets) has been demonstrated in other studies [46–48]. The Viterbi-based dynamic programming algorithm, on the other hand, appears to demonstrate reasonable performance despite lacking a sound statistical formalism and being reliant upon an ad-hoc process for ‘tuning’ algorithm parameters (for example, the assignment of the ‘memory factor’). Thus, an alternative temporal filtering technique may exist that offers superior detection performance. Hidden Markov model (HMM) temporal filtering approaches are a potential candidate as they possess a well established probabilistic framework that allows parameters to be intelligently designed [30]. Hence, the pairing of image morphology and HMM filtering techniques for target detection is worthy of further investigation.

Another challenge associated with using machine-vision for sense-and-avoid is the relatively high computational costs for processing image information that can be seen to be at odds with the requirement for real-time operation. The increasing complexity of image processing algorithms has meant that real-time operation is often beyond the capabilities of mainstream single and multi-core CPU based hardware implementations. As a result, special embedded high performance hardware such as dedicated digital signal processors and field programmable gate arrays (FPGAs) are often employed to enable real-time operation of vision-based
2.3. DIM TARGET DETECTION

sense-and-avoid systems [41, 42]. An alternative to the specialised embedded hardware are graphic processing unit (GPU) based hardware. GPU based hardware has an inherent parallel processing architecture that can be exploited to carry out computationally intensive functions in a sense-and-avoid system. This makes the implementation of a sense-and-avoid detection algorithm on GPU based hardware an appealing option to ensure that real-time operation can be realised.

2.3 Dim Target Detection

The ability to detect targets in naturally lit, high noise environments is becoming increasingly important. Significant challenges arise however in the use of computer vision for target detection because of the need to contend with not only the inherent noise of imaging sensors, but also with noise introduced by changing and unpredictable ambient conditions. The need to overcome these challenges has significantly driven the development of image filtering and processing techniques.

Over the last three decades, a two-stage processing paradigm has emerged for the detection of dim, sub-pixel sized targets [19, 40, 46, 49]. These two stages are: 1) an image pre-processing stage that, within each frame, highlights potential targets with attributes of interest; and 2) a subsequent temporal filtering stage that exploits target dynamics across frames. The latter temporal filtering stage is often based on a track-before-detect (TBD) processing concept where target information is collected and collated over a period of time before the detection decision is made.
CHAPTER 2. LITERATURE REVIEW

2.3.1 Image Pre-Processing

Generally, the goal of the image pre-processing stage is to enhance potential target features whilst suppressing background noise and clutter. There is an abundance of techniques and algorithms available which may be considered for this image processing role. In particular, non-linear spatial techniques such as median subtraction filters [50] have been widely discussed in the literature. Another non-linear image filtering approach that has received much attention over the last decade has its basis in mathematical morphology [20]. Numerous morphology-based filters have been proposed for the detection of small targets in infrared (IR) images [47, 51, 52]. Specific implementations of the morphological filtering approach include the Top-Hat filter [53], Close-Minus-Open filter [48], and the Hit-or-Miss filter [54]. The concept of a structuring element parameter (which can be likened to a template for matching objects of interest) is central to all of these specific filtering approaches. The top hat filter can be configured to identify objects either brighter or darker than the background (hot or cold objects, respectively) that are smaller than the structuring element size; the close-minus-open filter highlights both hot and cold objects in a single output; and the hit-or-miss employs a pair of structuring elements: one for matching the object and one for matching the object’s background. Although a large proportion of research has focused on IR images, there are recent examples of morphological filters being incorporated into target detection algorithms operating on EO (visual spectrum) images [16, 40, 46]. Moreover, a sign of the increasing popularity of morphological filters for small target detection is evident in the host of studies undertaken into the issue of parameter design [55, 56]. Finally, there have been efforts made to compare existing techniques with the morphology-based filters [46, 52, 57, 58], with the median filtering technique often featuring in the comparison studies. Of particular interest are the results in [46] which show a morphological filtering
2.3. DIM TARGET DETECTION

approach to be superior at supressing background clutter than a low-stop filtering technique. Moreover in [57], there is evidence to suggest that morphological filters are more suitable for detecting dim point targets than median filtering approaches.

Morphological filtering approaches are commonly employed in dim target detection problems to exploit the contrast between the intensity of target features and the intensity of the background. Depending on the relative brightness of the target with respect to the background, targets may be classified as either a positively contrasting feature (light target against darker background) or a negatively contrasting feature (dark target against lighter background). In [16], two candidate morphological processing approaches were proposed for the extraction of target features from image frames. The key difference is that one method retains information allowing the distinction between positive and negative features in the output, whereas the other method discards this ‘polarity’ information. An attempt was made in [16] to characterise the performance of these two candidate approaches. The studies conducted provided some evidence indicating that polarity information can be used to improve false-alarm performance, but the results were inconclusive given the limited data sets considered.

2.3.2 Temporal Filtering

The temporal filtering stage of the two-stage target detection paradigm is designed to extract image features that possess target-like temporal behaviour. For this role, two TBD filtering approaches have received much attention in the literature: Viterbi-based approaches and Bayesian-based approaches. Viterbi-based TBD approaches, also referred to as dynamic programming algorithms [19,25,46,49,59], have been a popular technique for dim target detection. This is in part due to the Viterbi algorithm’s utility in the context of tracking where, under a number of assumptions, it is able to efficiently determine the optimal
target track within a data sequence [35]. The detection performance of Viterbi
based TBD approaches have been analysed in [25, 26, 60], and modifications that
enhance the method’s ability to handle non-Gaussian clutter noise have been
proposed in [49]. An alternative TBD approach is based on Bayesian filtering
[27, 59, 61, 62]. In [61], the typical white Gaussian noise assumptions are re-
laxed, with consideration given to spatially correlated clutter. Moreover, in [27],
the modelling of clutter is expanded to encompass a variety of Gaussian and non-
Gaussian, correlated and uncorrelated clutter types, and the Bayesian algorithm
is extended to accommodate multiple targets that may feature randomly varying amplitudes or intensities. Some comparison between the Viterbi-based and
Bayesian-based TBD approaches has been made at the theoretical level in [27],
as well as on the practical level via Monte Carlo simulation trials in [59].

Hidden Markov model (HMM) filtering [30, 33] is a particular Bayesian fil-
tering approach that has been applied in a multitude of technical disciplines, in-
cluding non-linear stochastic control [23, 24], signal and image processing [25–30],
digital communications [31], and bioinformatics [32]. HMM filtering is an attrac-
tive candidate TBD approach for image-based dim target detection, as HMM
filtering algorithms are naturally suited to the discrete nature of image measure-
ments and flexible enough to handle the often highly non-linear dynamics. Competing approaches often lack the sound statistical formalism of HMMs and
the ability to propagate the target probability distribution over time (for exam-
ple dynamic programming [19]), or are relatively computationally expensive (for example particle filters [63]). In [59], the performance of a HMM-based TBD al-
gorithm is analysed in Monte Carlo simulations of linear target trajectories and
shown to be competitive against state-of-the-art approaches.

Despite the appeal of a HMM-based TBD approach, its application to an
image-based dim target detection problem is certainly not straightforward. There
are several aspects of the detection problem that conspire to make the design of
a suitable HMM TBD filter a challenge. In particular, potential targets may
possess a range of speeds and heading angles which introduce additional degrees of uncertainty in the target dynamics. Moreover, there is foreseeable difficulty in attempting to use a traditional HMM description (which is discrete-time and discrete-state by nature) to capture underlying target dynamics that are inherently continuous-valued. Such a design scenario not been adequately addressed in the literature.

Thus, it is unclear how a HMM filter should be designed to 1) account for uncertainty in target dynamics, and 2) adequately represent the dynamics from a different model class. Moreover, it is not known if a design criterion exists that can facilitate the comparison of candidate HMM filter configurations in terms of relative filtering performance. Several related bodies of work have been identified that may contribute to a potential solution. In particular, adaptive estimation algorithms and multiple model filtering approaches [64] have been proposed for dealing with systems that have time-varying or unknown model parameters. This may offer insight into how to address the case of uncertain target dynamics in the filter design problem. Furthermore, there are parallels between the HMM filter design problem and the HMM approximation problem described in Section 2.5, in the sense that filter design may be posed as seeking the closest HMM approximation of the true target dynamics. In a range of HMM estimation and approximation problems [37, 65–67], relative entropy concepts have been identified as an important design criteria. The possible applicability of these entropy related concepts to the problem of filter design is worthy of further investigation.

2.4 Target Tracking

Target tracking is a problem that combines elements of both estimation and statistical decision theory, and has a rich and extensive history [64]. From the early Kalman filtering technique (and its various extensions such as the extended
Kalman filter (EKF) [64]) to the more recent Monte Carlo or particle filtering methods [68], approaches to the target tracking problem have continued to evolve. Whilst the classical Kalman filtering techniques are still adequate for many existing problems, continued advances in computational capacity have encouraged researchers to tackle ever more complex target tracking scenarios.

The manoeuvring target tracking problem is substantially more complex than the basic constant velocity target tracking problem [64,69]. The history of manoeuvring target tracking shows adaptive Kalman filtering techniques as an early approach, followed by decision-based techniques and then multiple-model algorithms [70,71]. In particular, the good balance achieved between computational complexity and performance of the interacting multiple-model (IMM) technique has made it a popular manoeuvring target tracking approach [72]. Traditional IMM tracking approaches have tended to involve Kalman filtering techniques, but more recent filtering methods such as the particle filter are increasingly being used in IMM implementations [73,74].

For image-based target tracking problems, many authors have demonstrated the utility of hidden Markov model (HMM) based filtering techniques (for example [62]). HMMs are particularly well suited to exploit the discrete nature of image measurements and are flexible enough to handle non-linear dynamics [68]. Moreover, HMMs have a sound statistical formalism and an ability to propagate the target probability distribution over time, which are desirable properties often lacking in competing approaches (for example dynamic programming [49]). In addition, other competing approaches are often relatively computationally expensive (for example particle filters [63]). In [59], the performance of a HMM based track-before-detect (TBD) algorithm is analysed in Monte Carlo simulations of linear target trajectories and shown to be competitive against candidate state-of-the-art approaches such as dynamic programming, particle filtering methods, and the histogram probabilistic multihypothesis tracker.
2.5 Hidden Markov Model Approximation

Advances in hidden Markov model (HMM) signal processing tools over the last few decades have contributed to the widespread application of HMMs in a multitude of technical disciplines, including non-linear stochastic control [23, 24], signal and image processing [25–30], digital communications [31], and bioinformatics [32]. In particular, HMMs have been used to solve a variety of filtering problems in frequency tracking [29], speech recognition [30], character recognition [28], and dim target detection [25–27]. In many of these problems, HMMs were the key to circumventing the complex filtering solutions that would have resulted from directly modelling the underlying dynamics. Thus, the appeal and utility of a HMM-based approach lies in the powerful and well established suite of HMM processing tools that can be exploited to develop tractable solutions and simplify computations. This is so, however, only if the problem can be posed in a way that is consistent with the HMM framework, or an appropriate HMM representation of the dynamics can be found.

Several candidate approaches for determining HMM representations or approximations of a particular dynamic of interest have been proposed. One approach is through the application of classical data-based model inference techniques to infer the HMM that best matches the sample measurement and state sequences generated by the model dynamics under approximation. Data-based model inference or system identification is a classical signal processing problem that has been solved, over many decades, using a variety of techniques, including: information theory and entropy based techniques such as Akaike’s Information Criterion [75]; maximum likelihood based techniques (such as the EM algorithm [76]); and prediction error based techniques [77]. In the last few decades, many of the above techniques have been applied to a variety of HMM parameter estimation problems (see [30, 33, 65, 78–81]). The availability of data and the length of data required to yield representative parameters are key issues
that limit the applicability of data-based methods.

An alternative approach to constructing HMM approximations has also emerged from the HMM realisation problem [66, 67, 82–84]: given the output sequence probabilities, determine the HMM that has the same output sequence probabilities (if a valid HMM exists). In [82], existence and construction results for the HMM realisation problem were presented in terms of finite rank Hankel matrices and polyhedral cones (whose existence presupposes existence of a suitable HMM). This limitation was overcome in [83] through the introduction of an ultra-mixing property which guarantees existence of the necessary polyhedral cone, and allows a complete realisation result to be established. In [84], a different aspect of the realisation problem was examined through the use of subspace approaches to test equivalency between different HMM representations. Most recently, entropy related concepts have been used in [66, 67] to consider an approximate HMM realisation problem that addresses the question: given output string probabilities, determine the closest HMM (when no exact HMM realisation exists). However, the above realisation problems only seek HMMs capable of generating the prescribed output properties.

Despite the many approaches that have been proposed, the problem of determining a suitable HMM approximation/representation has not been completely resolved in a number of situations. As a result, there are many application-based examples in the literature where the design of HMM approximation models has been on an ad hoc basis rather than a systematic basis [19, 27, 72]. One particular situation that has not been adequately addressed is when a HMM approximation is sought for a system where the underlying true dynamics is not a member of the HMM class and there is uncertainty in the dynamics to be approximated. An example is the application of HMM filtering techniques to an image-based target detection problem (inherently non-HMM dynamics) where the target heading angle in the image plane is unknown (uncertain dynamics).
2.6 Summary

The key findings from our literature review are:

1. A recent study demonstrated the utility of two morphological filtering techniques for dim point-like target detection, but did not provide conclusive evidence supporting the choice of one technique over the other. This issue is addressed in C1, where we conduct a more extensive characterisation of the two promising morphological filtering techniques to find further evidence to aid the selection of a suitable technique for dim target detection;

2. The pairing of image morphology and HMM filtering techniques for dim target detection is an attractive combination that has not been thoroughly investigated. This is considered in C1 and C2, where we investigate the performance of a target detection algorithm based on a morphological filtering pre-processing stage followed by a HMM based temporal filtering stage. Moreover, C2 explores the possible benefits of using multiple HMM filters in the temporal filtering stage, in comparison to traditional single filter approaches;

3. There are particular situations where it is unclear how to determine suitable HMM approximations and design appropriate HMM filters, but the connections between model approximation and filter design suggests that a common set of tools may be applicable to both problems. In J1, we exploit powerful RER concepts to propose a solution to the HMM approximation problem for the case where there is uncertainty in the model under approximation and the model is additionally not a member of the HMM class. In J1, we also show that similar RER tools may be used to address the filter design problem, and in particular to propose a novel design process for the multiple HMM filtering approach considered in C2. Other applications of RER concepts and multiple HMM filtering are considered in J2; and
4. Realising real-time operation for an image-based detection algorithm can be a challenge, but GPU based hardware offer an alternative processing platform that can greatly outperform mainstream CPU based hardware and is generally more accessible than specialised computing devices such as embedded digital signal processors and FPGAs. In J3, we tap into the parallel processing capabilities of high performance GPU based hardware to demonstrate the capacity of our proposed detection algorithm to perform real-time target detection.
Chapter 3

Morphological Pre-Processing

(C1)

Our proposed target detection algorithm is based on a two-stage processing paradigm where an image pre-processing stage is followed by a temporal filtering stage. In this chapter, we compare two promising morphological filtering approaches for use in the image pre-processing stage: 1) a ‘close-minus-open’ filtering method and 2) a ‘preserved-sign’ filtering method. The results from our simulation studies show that the preserved-sign filtering approach offers the best tradeoffs for performance in terms of detection, time-to-detection, and false-alarm statistics.

In this chapter, we also take the first steps in determining a suitable temporal filtering approach that is compatible with morphologically processed image data to complete the two-stage detection paradigm. We compare the performance of two well published track-before-detect temporal filtering techniques: HMM filtering and a Viterbi-based filtering method (the Viterbi-based filtering approach considered here is inspired by the detection methods described in [16, 46], and should not be confused with the standard Viterbi algorithm [35], which is a fundamentally different filtering technique with different underlying assumptions and
mechanisms). Our results suggest that HMM filtering is superior in enhancing morphologically processed image data for target detection. We highlight that in these early comparison studies, the HMM filter transition probabilities are assigned empirically. This potential weakness is overcome by the development of a more systematic, theory-based approach later in Chapter 5.
Statement of Contribution of Co-Authors

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Digital Thesis database consistent with any limitations set by publisher requirements.

In the case of this chapter:

A Study of Morphological Pre-Processing Approaches for Track-Before-Detect Dim Target Detection

Status at time of writing:

Published in 2008

Details of Authorship and Contributions:

<table>
<thead>
<tr>
<th>Name and Order of Authors</th>
<th>Signatures of Authors</th>
<th>Area of Contribution Regarding Authorship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a) (i)</td>
<td>(a) (ii)</td>
</tr>
<tr>
<td>1. John Lai</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2. Jason J. Ford</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3. Peter O’Shea</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>4. Rodney Walker</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>5. Michael Bosse</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Name __________________________ Signature ______________ Date ______________

*See Appendix A for definition of authorship and criteria for contribution to publication
A Study of Morphological Pre-Processing Approaches for Track-Before-Detect Dim Target Detection

John Lai, Jason J. Ford, Peter O’Shea, and Rodney Walker
Queensland University of Technology, GPO Box 2434, Brisbane Qld 4001, Australia
j2.lai@student.qut.edu.au, {j2.ford, pj.oshea, ra.walker}@qut.edu.au

Michael Bosse
Autonomous Systems Laboratory, CSIRO ICT Centre QCAT, Brisbane Qld 4069, Australia
Mike.Bosse@csiro.au

Abstract
The track-before-detect processing technique has been employed in numerous computer vision based algorithms addressing the dim target detection problem. This processing technique has been shown to be effective under certain conditions; but in particularly noisy or highly cluttered environments, detection performance may be improved by introducing an image pre-processing stage to enhance the raw sensor measurements prior to integration. In this paper, we compare the ‘Close-Minus-Open’ (CMO) and ‘Preserved-Sign’ (PS) morphological image pre-processing techniques for suppressing unwanted noise and emphasizing target features in the measurement images. This investigation is motivated by the unmanned aerial vehicle “sense-and-avoid” application, where morphology-based filters have demonstrated a degree of success in the detection of small point-like features that may correspond to collision-course aircraft. For completeness, we also briefly examine two well published track-before-detect temporal filtering techniques which may be combined with the morphological pre-processing to detect dim, sub-pixel sized targets. Results from our simulation studies show that the PS approach achieves a higher detection rate than the CMO approach.

1 Introduction
The ability to detect and track targets in naturally lit, high noise environments is becoming increasingly important. Significant challenges arise however in the use of machine vision for target detection and tracking because of the need to contend with not only the inherent noise of imaging sensors, but also with noise introduced by changing and unpredictable ambient conditions. The need to overcome these challenges has significantly driven the development of image filtering and processing techniques. Over the last three decades, a two-stage processing paradigm has emerged for the simultaneous detection and tracking of dim, sub-pixel sized targets [Gandhi et al., 2006; Gandhi et al., 2003; Arnold et al., 1993; Barniv, 1985]. These two stages are: 1) an image pre-processing stage that, within each frame, highlights potential targets with attributes of interest; and 2) a subsequent temporal filtering stage that exploits target dynamics across frames. The latter temporal filtering stage is often based on a track-before-detect processing concept where target information is collected and collated over a period of time before the detection decision is made. In this paper, we are primarily interested in the image pre-processing stage of the above two-stage paradigm; but for completeness, we also briefly investigate the performance of two well known track-before-detect temporal filtering techniques which may be combined with the morphological pre-processing to detect slow dim sub-pixel sized targets.

Generally, the goal of the image pre-processing stage is to enhance potential target features whilst suppressing background noise and clutter. There is an abundance of techniques and algorithms available which may be considered for this image processing role. In particular, non-linear spatial techniques such as median subtraction filters [Deshpande et al., 1999] have been widely discussed in the literature. Another non-linear image filtering approach that has received much attention over the last decade has its basis in mathematical morphology [Dougherty and Lotufo, 2003]. Numerous morphology-based filters have been proposed for the detection of small targets in infrared (IR) images [Zhu et al., 2000; JiCheng et al., 1996; Tom et al., 1993]. Specific implementations of the morphological filtering approach include the Hit-or-Miss filter [Schafer and Casasent, 1995], Close-Minus-Open filter [Casasent and Ye, 1997], and the Top-Hat filter [Braga-Neto et al., 2004]. Although a large proportion of research has focused on IR images, there are recent examples of morphological filters being incorporated into target detection algorithms operating on video images [Carnie et al., 2006; Gandhi et al., 2006; Gandhi et al., 2003]. Moreover, a sign of the increasing popularity of morphological filters for small target detection is evident in the host of studies undertaken into the issue of parameter design [Zeng et al., 2006; Yu et al., 2003]. Finally, there have been efforts made to compare
existing techniques with the morphology-based filters [Gandhi et al., 2006; Wayen, 2002; Davey et al., 1993; Barnett et al., 1993], with the median filtering technique often featuring in the comparison studies.

The second focus of this paper concerns the temporal filtering stage that follows the image pre-processing. The temporal filter is designed to extract image features that possess target-like temporal behaviour. For this role, there are two particular filtering approaches that have received much attention in the literature: Viterbi based approaches and Bayesian based approaches.

The Viterbi algorithm has formed the basis of the temporal filtering stage in numerous track-before-detect algorithms [Davey et al., 2008; Gandhi et al., 2006; Tomissen and Evans, 1996; Arnold et al., 1993; Barniv, 1985]. This is in part due to its utility in the context of tracking where, under a number of assumptions, it is able to efficiently determine the optimal target track within a data sequence [Forney, 1973]. Some analysis of the Viterbi algorithm’s detection and tracking performance can be found in [Johnston and Krishnamurthy, 2000; Barniv and Kella, 1987; Tomissen and Evans, 1996], and modifications that enhance the algorithm’s tracking performance in the presence of non-Gaussian clutter noise have been proposed in [Arnold et al., 1993]. An alternative temporal filter design for track-before-detect algorithms is based on Bayesian filtering [Davey et al., 2008; Bruno, 2004; Bruno and Moura, 2001; Bruno and Moura, 1999]. In [Bruno and Moura, 1999], the traditional white Gaussian noise assumptions are relaxed, with consideration given to spatially correlated clutter. Moreover, in [Bruno and Moura, 2001], the modelling of clutter is expanded to encompass a variety of Gaussian and non-Gaussian, correlated and uncorrelated clutter types, and the Bayesian algorithm is extended to accommodate multiple targets that may feature randomly varying amplitudes or intensities.

Some comparison between the Viterbi and Bayesian approaches has been made at the theoretical level [Bruno and Moura, 2001], as well as on the practical level via Monte Carlo simulation trials [Davey et al., 2008]. However, conclusions about detection and false-alarm performance are beyond the scope of the theoretical analysis provided in [Bruno and Moura, 2001], while the simulation results of [Davey et al., 2008] are limited to a specific false-alarm rate (the detection tradeoffs for various false-alarms rates that would be useful to a system designer are not available).

The main aim of this paper is to investigate the use of two alternative morphological filtering approaches as the pre-processing stage for track-before-detect algorithms operating on image sequences. This investigation is motivated by the unmanned aerial vehicle “sense-and-avoid” application, where morphology-based filters have demonstrated a degree of success in the detection of small point-like features that may correspond to collision-course aircraft [Carmie et al., 2006]. Here, we compare the ‘Close-Minus-Open’ (CMO) approach with the less well characterised ‘Preserved-Sign’ (PS) technique. This comparison is performed in the context of a hidden Markov model (HMM) temporal filtering stage (which has recently been shown to be an effective choice for track-before-detect algorithms [Davey et al., 2008]). As a second objective, we briefly focus our attention on the temporal filtering aspect and consider the detection performance of a Viterbi-based approach as compared to a Bayesian HMM approach when using morphologically pre-processed image measurements as input.

We assess our detection algorithms via detection, time-to-detection, and false-alarm statistics that provide valuable insight into the tradeoffs in performance from using the two different pre-processing and temporal filtering approaches. The performance of the morphological pre-processing is examined under cross-tracking target scenarios featuring a range of target speeds and signal-to-noise ratios with different noise conditions. In our temporal filtering comparison, we consider not only cross-tracking targets, but also ‘emerging’ targets that gradually become more distinct over time, such as those that might be expected in an airborne collision avoidance scenario [BASI, 1991]. Our simulation studies show that a PS morphological pre-processing approach in combination with the HMM temporal filter provides the best tradeoff between detection and false-alarm performance.

2 Morphological Image Pre-Processing

In this section, we provide a brief review of the morphological filtering approaches compared in this paper.

2.1 Morphological Operations

We are concerned with greyscale morphological operations designed to be applied to discrete 2D image data quantised to a finite number of intensity or greyscale levels, such as might be expected from the output of an electro-optical sensor. Here, we combine two fundamental morphological operations known as ‘dilation’ and ‘erosion’ to create more sophisticated filtering operations for extracting small, point-like features that are present within an image frame.

Let $\text{Y} \oplus S$ and $\text{Y} \ominus S$ denote the dilation and erosion respectively of a greyscale image $\text{Y}$ by a morphological structuring element $S$ (see [Soule, 2003; Dougherty and Lotufo, 2003] for formal definitions of the dilation and erosion operations). The dilation and erosion operations can be combined to form secondary operations that play key roles in morphological image processing; these being ‘opening’ and its dual ‘closing’. Let $\text{Y} \circ S$ and $\text{Y} \bullet S$ denote the opening and closing respectively of a greyscale image $\text{Y}$ by a morphological structuring element $S$. The opening operation is simply defined as an erosion followed by a dilation

$$\text{Y} \circ S = (\text{Y} \ominus S) \ominus S,$$  

and the closing reverse – a dilation followed by an erosion

$$\text{Y} \bullet S = (\text{Y} \oplus S) \oplus S.$$  

An intuitive understanding of the morphological opening procedure can be gained by visualising the opening operation as the darkening of locally bright regions (which are smaller than the size of the structuring element) to the values of their neighbourhood pixels. In a similar manner, morphological closing may be regarded as the process of brightening locally dark regions (which are smaller than the structuring element) to match the values of neighbouring pixels.
It follows from the above concepts that taking the difference between an image and its morphological opening will produce an output identifying positively contrasting features (pixel groups that are brighter than their neighbours). This corresponds to what is referred to in the literature as a ‘top-hat’ transformation [Gonzalez, 2004]. Let \( TH(Y, S) = Y - (Y \ast S) \) denote the top-hat transformation of image \( Y \). Its dual, the ‘bottom-hat’ transformation of image \( Y \), is defined as \( BH(Y, S) = (Y \ast S) - Y \) [Gonzalez et al., 2004] and accordingly will highlight negatively contrasting features (pixel groups that are darker than their neighbours). In each case, only features smaller than the structuring element are preserved, whilst larger features are suppressed. Thus, via carefully chosen structuring elements, the above morphological transformations and operations represent a powerful set of image processing techniques for identifying features of interest based on their geometrical size.

The two morphological filters to be investigated in this paper are based on a combination of the top-hat and bottom-hat transformations. Both filtering approaches aim to differentiate between genuine and non-genuine features of interest based on size via appropriate design of the structuring elements.

### 2.2 Close-Minus-Open Filtering Approach

The Close-Minus-Open (CMO) filtering approach is given by the sum of the top-hat and bottom-hat transformations

\[
CMO(Y, S) = TH(Y, S) + BH(Y, S) = (Y - (Y \ast S)) + ((Y \ast S) - Y)
\]  

(3)

which simplifies down to the difference between the closing and opening of an image. We take advantage of this combination of secondary morphological operations, which has been referred to elsewhere in the literature as a self-complementary top-hat filtering approach [Soille, 2003], to simultaneously preserve both positively and negatively contrasting features with sizes that match potential features of interest (both positive and negative contrasting features result in a non-negative output).

### 2.3 Preserved-Sign Filtering Approach

An alternative to the CMO technique is the Preserved-Sign (PS) filtering approach, defined as the difference between the top-hat and bottom-hat transformations [Carnie et al., 2006]:

\[
PS(Y, S) = TH(Y, S) - BH(Y, S) = (Y - (Y \ast S)) - ((Y \ast S) - Y)
\]  

(4)

The above definition shows that the PS filtering approach is in fact a variation on the top-hat contrast enhancement operator described in [Soille, 2003].

The distinguishing feature of the PS filtering approach is that the response to positively contrasting features is non-negative, whereas the response to negatively contrasting features is non-positive. This is in contrast with the CMO approach where any contrasting feature (positive or negative) is expressed as a non-negative output. Thus, unlike the CMO technique, the PS filtering approach identifies contrasting features and provides additional information regarding the contrast ‘polarity’.

### 2.4 Proposed Filter Implementation

In this paper, we configure our CMO and PS filtering implementations to serve as powerful tools in the identification of small point-like features within the measurement image. For performance and computational reasons, we exploit a directional decomposition technique [Casasent and Ye, 1997] in our implementation of the CMO and PS morphological filters. In the case of the CMO approach, we take the minimum response from a pair of CMO filters using orthogonal 1D structuring elements. Here, one CMO filter operates exclusively in the vertical direction, while the other operates exclusively in the horizontal direction. Similarly, for the PS approach we implement two separate filters (one operating in the vertical direction and the other in the horizontal direction) using orthogonal 1D structuring elements, but take the minimum magnitude response from the pair of PS filters as the final output.

### 3 Temporal Filtering

In many electro-optical based detection problems, the existence of a target in a 3D volume of space must be determined from observations of a projection of the target space onto a 2D image plane. Here, target detection can be viewed as evaluating the likelihood of 2 alternate hypotheses, where \( H_i \) denotes the hypothesis that there is a single target present in the camera field of view, and \( H_k \) denotes the hypothesis that there is no target present. The temporal filtering approaches implemented in this paper will assume that under hypothesis \( H_k \), the projected target motion resides on a 2D plane fixed in space that is represented by the set of discrete 2D grid points \( \{(i, j) \mid 1 \leq i \leq N_h, 1 \leq j \leq N_v\} \), with vertical and horizontal resolutions \( N_h \) and \( N_v \) respectively. Let \( N = N_h N_v \) denote the total number of grid points. The measurements are provided by an electro-optical imaging sensor whose field of view is represented by a 2D grid of image pixel locations aligned with the target space and denoted \( \{(p, q) \mid 1 \leq p \leq N_h, 1 \leq q \leq N_v\} \).

#### 3.1 Hidden Markov Model Filtering

We will assume that, when present, the target is located within a particular pixel of the image frame at each time instant. Thus, each pixel \((i, j)\) represents a unique state of the HMM in our target detection problem. For notational convenience, we stack the columns of the image frame to form a vector of pixel locations. In this way, each state may be referenced by a single index, in the sense that if the target is at pixel location \((i, j)\), this corresponds to it being in the state \( m = (j-1)N_v + i \).

Let \( X_k \) denote the state (target location) at time \( k \). Between consecutive image frames the target may move to different pixel locations – that is, the target can transition between the states. The likelihood of state
transitions can be described by the HMM’s transition probabilities \( A^{m \to} = P(X_{t+1} = \text{state } m \mid X_t = \text{state } n) \) for \( 1 \leq m, n \leq N \), which is the probability of moving from any one pixel position (state) \( n \) to any other pixel position (state) \( m \). The transition probabilities can therefore be used to describe the expected mean target motion. For example, in the case of slow moving targets we tend to assign low probabilities to transitions between distant pixels. Moreover, initial probabilities \( \pi^m = P(X_0 = \text{state } m) \) for \( 1 \leq m \leq N \) are used to specify the probability that the target is initially located in state \( m \). Finally, to complete the parameterisation of the HMM, there are the measurement probabilities \( B^m(Y_t) = P(Y_t \mid X_t = \text{state } m) \) for \( 1 \leq m \leq N \) that are used to specify the probability of obtaining the observed measurement \( Y_t \), given that the target is actually in pixel location (state) \( m \) (see Elliott et al., 1995) for more details about the parameterisation of HMMs.

Detection Strategy

The HMM filtering approach performs temporal integration of the input measurements by recursively propagating \( \alpha^n_k \), an unnormalised probabilistic estimate of the target state \( X_\ell \), over time. This is achieved via the forward part of the forward-backward procedure described in [Rabiner, 1989], which can be decomposed into two stages: initialisation and recursion.

For \( 1 \leq m \leq N \)

1) Initialisation: Let \( \alpha^m_0 \) denote the probability \( P(Y_1, Y_2, \ldots, Y_k, X_1 = \text{state } m) \). Then

\[ \alpha^m_0 = \pi^m B^m(Y_1). \]

2) Recursion: At time \( k > 1 \), set

\[ \alpha^n_k = \sum_{m=1}^{N} \alpha^m_{k-1} A^{m \to} B^n(Y_k). \]

The forward procedure filtering result is closely related to the two probabilistic measures that we use to facilitate the detection of targets: 1) the probability of measurements up to time \( k \) assuming \( H_1 \), given by

\[ P(Y_1, Y_2, \ldots, Y_k \mid H_1) = \sum_{m=1}^{N} \alpha^m_k, \quad (5) \]

and 2) the conditional mean filtered estimate of the target state \( m \) given measurements up to time \( k \) and assuming \( H_1 \), given by

\[ \hat{X}_X^k = E[\hat{X}_X = \text{state } m \mid Y_1, Y_2, \ldots, Y_k, H_1] = \frac{\alpha^n_k}{\sum_{i=1}^{N} \alpha^n_i}, \quad (6) \]

where \( E[\cdot] \) denotes the mathematical conditional expectation operation (see [Billingsley, 1995] for more details). The probability \( P(Y_1, Y_2, \ldots, Y_k \mid H_1) \) may be interpreted as an indicator of target presence (following the probabilistic distance results of [Xie et al., 2005]), and the conditional mean estimate can be regarded an indicator of likely target locations.

In the interest of computational efficiency, we choose in this paper to evaluate the conditional mean estimate directly from the following expression [Elliott et al., 1995]:

\[ \hat{X}_X = \hat{X}_X^k, \quad (7) \]

where \( N_k \) is a scalar normalisation factor; \( B^n(Y_k) \) is a \( N \times N \) matrix where the main diagonal is occupied by the values of \( B^n(Y_k) \) for \( 1 \leq m \leq N \) and all other elements are zero; \( A \) is a \( N \times N \) matrix with elements \( A^{m \to} \); and \( \hat{X}_X \) is a \( N \times 1 \) vector consisting of elements \( \hat{X}_X^k \) for \( 1 \leq m \leq N \) that are equivalent to those given in (6). Moreover, we note the following relationship between the normalisation factor \( N_k \) and the probability of measurements up to time \( k \) assuming \( H_1 \):

\[ P(Y_1, Y_2, \ldots, Y_k \mid H_1) = \prod_{l=1}^{k} \frac{1}{N_k}. \quad (8) \]

HMM Filter Implementation

In this paper, we define the transition probabilities so that only a self-transition or a transition to any one of the 8-connected neighbours is possible from time \( k \) to \( k + 1 \). This leads to a sparse \( A \) matrix that is efficient to implement in practice. Figure 1 provides a visual representation of the possible state transitions in the HMM filter.

Figure 1. Possible state transitions of the HMM filter. The solid squares define the 8-connected neighbourhood of the state (depicted here as the central cell), with the arrows indicating the transition possibilities from time \( k \) to \( k + 1 \). Thus, the state may only undergo a self-transition (dark cell) or a transition to any one of its 8-connected neighbours (light cells).

Furthermore, we note that our implementation of the HMM filter exploits the following probabilistic relationship between target location \( X_t \) and the pre-processed measurements \( Y_t \):
The Viterbi-based algorithm performs temporal integration of the input measurements by recursively generating a set of intermediate images $a$ for each velocity cell $(u,v)$ that is considered. This process can be divided into two stages: initialisation and recursion.

For all $(u,v)$, $1 \leq i \leq N_i$, and $1 \leq j \leq N_j$, 

1) Initialisation: Let $a^i_{ji}(u,v)$ denote the $ij$th pixel of the intermediate image frame at time $k$ for velocity cell $(u,v)$. Then $a^i_{ji}(u,v) = \theta$.

2) Recursion: At time $k > 1$, set 

$$a^i_{ji}(u,v) = \begin{cases} (1-\beta)y^i_{ji} + \beta \cdot \max_{\gamma \in \gamma^i_{ij}} \{a^i_{ji}(u,v)\}, \\
\end{cases}$$

where $y^i_{ji}$ is the $ij$th pixel of the pre-processed image at time $k$, and $\beta$ represents a memory factor that can vary between zero and one.

At anytime $T$ when a detection decision is required, we take the maximum output across corresponding pixels of the intermediate image frames belonging to each velocity cell:

$$a^i_{\max} = \max_{(u,v)} \{a^i_{ji}(u,v)\},$$

for $1 \leq i \leq N_i$ and $1 \leq j \leq N_j$. This final image $a_{\max}$ that consolidates target information from across all velocity cells then serves as the basis for declaring detections.

### Viterbi-Based Filter Implementation

The Viterbi-based filter implemented here mirrors those seen in [Carnie et al., 2006] and [Gandhi et al., 2006], where four velocity cells are used to detect targets that move with constant velocity in any direction, but are limited to a maximum speed of 1 pixel per frame. Nonmaximal suppression is applied to the output of the filter to reduce an undesirable “dilation” effect where pixels in the neighbourhood of the target also attain significantly large values [Gandhi et al., 2006].

### Performance Characterisation

In our main study, we compare the performance of the two alternative morphological pre-processing techniques:

- A Close-Minus-Open filter, and
- A Preserved-Sign filter

in the context of the track-before-detect problem by applying them to a large number of image sequences containing cross-tracking targets having a variety of intensity and speed attributes. This pre-processed data is then sent to a HMM temporal filtering stage. The HMM filtering stage is individually optimised for each particular pre-processing technique to ensure that neither approach is unfairly disadvantaged.

As a second study, we compare the HMM temporal filtering approach from above with a Viterbi-based approach by applying them to a large number of morphologically pre-filtered image sequences containing two different target types. In particular, we consider cross-tracking (constant size and intensity) and looming targets. The cross-tracking target case allows us to compare results with those of existing studies [Gandhi et al., 2006]; whereas the looming target case allows us to examine...
performance in a new and important detection scenario.

In the following subsections, we describe our metrics for quantifying performance, our procedure for generating image sequences, specific filtering parameters, as well as the presentation of results. These aspects comprise the simulation framework of our comparison studies.

4.1 Performance Metrics

In the detection of cross-tracking targets, we are interested in detection versus false-alarm statistics, whereas in the detection of looming targets we consider time-to-detection versus false-alarm statistics. If a target is present, the track-before-detect algorithm is considered to have achieved a detection if the algorithm correctly identifies the target’s presence and locates it to within two pixels of the true position. We define the detection rate as the number of detections divided by the maximum number of possible detections. A false-alarm event occurs if the track-before-detect algorithm incorrectly declares the presence of a target. We define the false-alarm rate as the number of false-alarms divided by the total number of possible false-alarm events. The time-to-detection is defined as the frame when detection is first achieved.

For the HMM filtering approach, let \( \eta \), our test statistic for declaring the presence of a target at time \( k \), be given by

\[
\eta = \frac{1}{k} \log \left( \prod_{i=1}^{k} \frac{1}{N_i} \right)
\]  

(11)

(in practice, a statistically equivalent recursive moving average would instead be implemented). When \( \eta \) exceeds a predefined threshold, the HMM track-before-detect algorithm considers a target to be present and located at state

\[
y = \arg \max_{s} \hat{X}_k^s
\]  

(12)

at time \( k \). Our definition of \( \eta \) and \( y \) is motivated by the detection strategy discussed in Section 3.1.

For the Viterbi-based filtering approach, our test statistic \( \lambda \) for declaring the presence of a target at time \( k \) is given by

\[
\lambda = \max_{s} \left( q_k^s \right)
\]  

(13)

When \( \lambda \) exceeds a predefined threshold, the Viterbi-based algorithm considers a target to be present and located at state

\[
\varsigma = \arg \max_{o} \left( o_k^o \right)
\]  

(14)

at time \( k \). The definition of \( \lambda \) and \( \varsigma \) follow from the interpretation of the Viterbi-based algorithm’s output as an image, where the pixel values correspond to target signal strength. Finally, we note that detection and false-alarm statistics are based on the final temporal filtering output after the last image sequence frame is processed.

4.2 Image Sequence Generation

As much as possible, we attempt to include image sequences in our simulation studies that are sourced from authentic data; however our access to this type of data is very limited. Thus, in order to carry out the large number of trials our simulation studies require, we make use of synthetically generated image sequences. The image frames that comprise our synthetic image sequences is formed by adding a noise component and a target signature (if required) to a uniform background image set at an arbitrary greyscale level of 128.

Consider any finite sized target with a non-zero velocity. As this target traverses across the image, the physical extent of the target is likely to overlap multiple pixels at any time. We model the target signature in each pixel as being proportional to the amount the target overlaps the pixel. For example, if the target occupies half the area of a particular pixel, then the target signature in that pixel is assigned half the value of the target intensity.

The basic target model is that of a cross-tracking target that has fixed target size and intensity for the duration of the image sequence. Conversely, our looming target model represents the scenario of a fixed size object approaching the imaging sensor from a very distant location at constant speed. Based on simple point light source and thin lens optics assumptions, we find that the looming target scenario can be constructed by maintaining a fixed target intensity and increasing the target size exponentially at a rate that is consistent with the closing speed and target distance that is being simulated.

Our synthetic image frames will incorporate either one of two types of noise: 1) Zero-mean Uncorrelated Gaussian Noise, or 2) Spatially Correlated Noise. We model the spatially correlated noise as a Gauss-Markov random field (GMRF) (see [Moura and Balram, 1993; Moura and Balram, 1992] for more details) because it is believed to be representative of an important component of the noise present in electro-optical sensors [Bruno and Moura, 1999]. For computational reasons, a first-order, spatially homogenous field with free or Dirichlet boundary conditions is assumed, parameterised by vertical and horizontal field potentials or interactions.

Signal-to-noise ratio (SNR) is used to provide an indication of how distinct a target is (i.e. how well it stands out from the background, and in turn the ease with which it may be detected). Here, we define SNR as

\[
\text{SNR} = 20 \log_{10} \left( \frac{I}{\sigma} \right) \text{ dB},
\]  

where \( I \) is the maximum target signature value and \( \sigma \) the noise standard deviation. The upper bound for the SNR, the peak SNR (PSNR), is achieved when \( I \) is equal to the target intensity. In our simulation studies, the tendency of the SNR to vary with time leads us to quote the PSNR instead as a means of signifying how distinct the target is. We note that the PSNR is only indicative of the target’s distinctiveness and provides an upper bound on the average SNR for cross-tracking targets only (for looming targets, the SNR grows with time and hence the PSNR is not defined).

4.3 Filtering Parameters

Morphological Pre-Processing

Both the CMO and PS approaches employ the same pair of structuring elements given by \( a = [1,1,1,1,1] \) and \( s = [1,1,1,1,1] \) to filter in the horizontal and vertical...
Temporal Filtering
For the Viterbi-based filter, we let $\beta = 0.75$, which has been demonstrated in [Carne et al., 2006] to be a reasonable choice for the memory factor.

For the HMM filter, we assign a probability of $7/15$ to self-transitions and a probability of $1/15$ for transitions to adjacent pixels. We highlight that performance was not overly sensitive to these parameter choices (the particular values selected here were found to give reasonable performance). Furthermore, it is assumed that the target may be initially located anywhere within the image frame with equal probability, and hence we set $\pi^* = [1/N]$ for $1 \leq m \leq N$.

To construct our measurement probability matrix $B_{m} (Y_{k})$, we are required to estimate the probabilities $P(Y_m^k \mid X_k \neq m)$ and $P(Y_m^k \mid X_k = m)$. The former describes our prior knowledge about the distribution of pixel values in the absence of a target (i.e. the noise and clutter distribution), while the latter captures our prior knowledge about the distribution of values at pixels containing a target. We estimate the required probabilities for $B_{m} (Y_{k})$ directly from data. The probability $P(Y_m^k \mid X_k \neq m)$ is estimated as the average frequency that each pixel value resulted from a non-target location. Using a similar procedure, $P(Y_m^k \mid X_k = m)$ is estimated as the average frequency that each pixel value measurement resulted from a target location.

Finally, for convenience, we choose in both temporal filtering approaches to process positively and negatively contrasting targets separately when dealing with the PS filtering output.

4.4 Performance Tradeoff Curves
In this paper, our preferred method of presenting the simulation study results is via graphical performance curves that concisely illustrate the tradeoffs between two suitable performance metrics (such as detection vs. false-alarm or time-to-detection vs. false-alarm) independent of threshold values. For this approach it is essential that the performance metrics be generated from a fixed set of threshold values evaluated against the test statistics.

5 Performance Comparison Studies
In our first study, we evaluate the performance of the two morphological filtering techniques by allowing their corresponding track-before-detect algorithms featuring a common HMM temporal stage to process a large number of image sequences containing cross-tracking targets with various target attributes and noise properties. In the second study, we briefly compare the HMM and Viterbi-based temporal filtering approaches using morphologically pre-processed image sequences containing cross-tracking and looming targets.

5.1 Study 1
The CMO and PS pre-processing techniques are compared in two simulation scenarios using synthetically generated image sequences. In scenario 1, we examine the performance of the pre-processing techniques in uncorrelated Gaussian noise, whereas in scenario 2 spatially correlated noise is considered.

Scenario 1
The aim of this scenario is to investigate the performance of the morphological filters in uncorrelated Gaussian noise (with zero-mean; standard deviation of 1) for a selection of target speeds and PSNRs. Specifically, we consider a $1 \times 1$ pixel cross-tracking target travelling at speeds of 0.1, 0.2, and 0.3 pixels/frame and with intensities corresponding to PSNRs of approximately 8, 9.5, and 11 dB. Image frames of size $N_1 = 111$ and $N_0 = 147$ comprise image sequences that are 151 frames in length. For each combination of target speed and PSNR, detection counts are gathered from 10³ separate image sequences each containing only a single target (all targets converged towards the centre of the image frame, but began at initial locations evenly distributed about the centre). The same number of image sequences (without targets) is used in the calculation of false-alarm counts.

The simulation results show the PS filtering approach to be superior to the CMO approach for all combinations of PSNRs and target speeds. For example, consider the detection vs. false-alarm tradeoff curves illustrated in Figure 2, which is for a target speed of 0.1 pixels/frame at 8 dB PSNR. Here, the detection rates associated with the PS approach are significantly higher than those for the CMO approach.

![Figure 2. Detection vs. false-alarm performance for CMO and PS morphological filtering (uncorrelated Gaussian noise; target speed 0.1 pixels/frame; PSNR 8 dB)](image)

Table 1. Detection rates for CMO and PS morphological filtering under Gaussian noise at a false-alarm rate of $10^{-3}$.

<table>
<thead>
<tr>
<th>PSNR</th>
<th>Target Speed</th>
<th>CMO</th>
<th>PS</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.1</td>
<td>0.19</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.07</td>
<td>0.70</td>
</tr>
<tr>
<td>9.5</td>
<td>0.3</td>
<td>0.01</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.14</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.35</td>
<td>0.96</td>
</tr>
<tr>
<td>11</td>
<td>0.3</td>
<td>0.11</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.98</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.87</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.50</td>
<td>0.97</td>
</tr>
</tbody>
</table>

A comparison of the detection rates at a false-alarm rate of $10^{-3}$ is provided in Table 1.
Scenario 2
The simulation parameters of the previous scenario are applied here, with the exception that spatially correlated noise is used instead of uncorrelated Gaussian noise. The spatially correlated noise is modelled by a GMRF having a horizontal and vertical interaction factor of 0.12 and driven by a zero mean Gaussian signal with a standard deviation of 1.

The simulation results show the PS filtering approach to be superior to the CMO approach for all combinations of PSNRs and target speeds. An example of the detection versus false-alarm tradeoff is illustrated in Figure 3, while in Table 2 a comparison of the detection rates at a false-alarm rate of $10^{-3}$ is provided. Comparing Figures 2 and 3 as well as Tables 1 and 2 reveals little difference in the performance trends of the morphological filtering approaches under the two noise types, with detection perhaps slightly better in correlated noise. However, we anticipate that different results will be obtained with correlated noise parameterised by a higher interaction factor.

Figure 3. Detection vs. false-alarm performance for CMO and PS morphological filtering (GMRF noise; target speed 0.1 pixels/frame; PSNR 8 dB)

<table>
<thead>
<tr>
<th>PSNR</th>
<th>Target Speed</th>
<th>Detection Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.1</td>
<td>0.22 0.94</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.99 0.12</td>
</tr>
<tr>
<td>9.5</td>
<td>0.1</td>
<td>0.79 0.99</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.42 0.96</td>
</tr>
<tr>
<td>11</td>
<td>0.1</td>
<td>0.98 1.00</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>0.90 0.99</td>
</tr>
</tbody>
</table>

Table 2. Detection rates for CMO and PS morphological filtering under GMRF noise at a false-alarm rate of $10^{-3}$.

5.2 Study 2
The HMM and Viterbi-based temporal filtering approaches are compared in three simulation scenarios. Scenario 1 simulates a cross-tracking target in spatially correlated noise, whereas scenarios 2 and 3 investigate detection performance for a looming target under spatially correlated noise and real image noise conditions, respectively. We choose to pre-process the image measurements using the PS morphological filtering approach based on its superior performance results from our first study.

Scenario 1
The performance of the HMM and Viterbi-based temporal filtering approaches in spatially correlated noise is compared in this scenario. The spatially correlated noise is modelled by a GMRF having a horizontal and vertical interaction factor of 0.12 and driven by a zero mean Gaussian signal with a standard deviation of 1. Image frames of size $N_v = 111$ and $N_h = 147$ comprise image sequences that are 151 frames in length. Detection counts are gathered from $5 \times 10^4$ separate image sequences each containing only a single cross-tracking target (of $1 \times 1$ pixel size, travelling at 0.2 pixels/frame, and with a PSNR of approximately 9.5 dB). The initial locations of the targets are uniformly distributed about the centre of the image. For the calculation of false-alarm counts, the same number of image sequences but without targets is used.

Figure 4 illustrates the detection vs. false-alarm curves for the two temporal filtering approaches. This figure shows the HMM approach demonstrating dramatically better performance than the Viterbi-based approach. We note that this unexpectedly large performance difference is in contrast to recently published results [Davey et al., 2008], which show the HMM (Bayesian) and Viterbi-based (Dynamic Programming) techniques to be quite competitive when morphological pre-processing is not used. We acknowledge that the morphological pre-processing may be more suited to the HMM approach, and that the use of “backtracking” may improve the performance of the Viterbi-based approach implemented here.

Figure 4. Detection vs. false-alarm performance for HMM and Viterbi-based temporal filtering (GMRF noise; target speed 0.2 pixels/frame; PSNR 9.5 dB)

Scenario 2
In this scenario, we focus on filtering performance for looming targets as opposed to cross-tracking targets. Simulation parameters from the previous scenario are carried over to this scenario, with the exception that image sequences are lengthened to 201 frames for looming targets, and these targets being set to move at 0.02 pixels/frame. The first 20 frames are reserved for algorithm initialisation, leaving the remaining 181 frames for characterising time-to-detection performance. Each image sequence begins with a sub-pixel sized target of...
low intensity (approximately 0 dB) during the initialisation frames, which is then allowed to gradually become larger and more distinct such that it models the scenario of a 1 square metre profile object approaching at approximately 103 metres per second from a distance of around 5 kilometres (the time interval between frames is 0.2 second).

Figure 5 shows the time-to-detection vs. false-alarm curves for the two temporal filtering approaches. This figure illustrates that the HMM approach has better looming target detection performance than the Viterbi-based approach (i.e. earlier detection for a specified false-alarm rate). For example, at a false-alarm rate of $10^{-3}$, the HMM algorithm detects about 32 frames earlier than the Viterbi-based algorithm. For illustrative purposes, we present in Figure 6 samples of the typical image pre-processing and temporal filtering outputs.

Scenario 3
For scenario 3, the same simulation parameters as the previous scenario are used, except that the synthetic noise is replaced by noisy image backgrounds obtained from real data sequences. Due to the limited availability of real data, we are unable to provide a comprehensive comparison study of the two algorithms. Processing the real data sets did however provide some useful insight into the existence of non-target semi-persistent features that can delay target detection. These artefacts appear to impact the HMM approach more than the Viterbi-based technique. A possible method of mitigating the detection delays is by introducing extra processing layers that recognise persistent non-target features and reinitialises the filter accordingly. These issues are beyond the scope of this paper and are the subject of ongoing research.

6 Conclusion
In this paper, we compared the Close-Minus-Open (CMO) and Preserved-Sign (PS) morphological image pre-processing techniques in the context of track-before-detect dim target detection. Furthermore, we briefly examined how a HMM temporal filtering approach compares with a competing Viterbi-based algorithm when morphologically processed image data is used. The results from our simulation studies show that the PS pre-processing technique in combination with a HMM temporal filtering approach offers the best tradeoffs for performance in terms of detection, time-to-detection, and false-alarm statistics.

However, it is likely that further refinement of the algorithms is necessary in order to attain a level of performance that would be acceptable in a practical collision avoidance system for UAVs. Our ongoing work includes improving the detection vs. false-alarm tradeoffs through the development of a theoretical basis for the design of HMM filtering parameters (such as the measurement and transition probabilities), as well as addressing persistent non-target features present in noisy real data backgrounds.

Acknowledgements
The authors would like to thank R. Carnie for providing the real image sequences used in this paper. Computational resources and services used in this work were provided by the HPC and Research Support Unit, Queensland University of Technology, Brisbane, Australia.

References


Chapter 4

Hidden Markov Model Filter
Banks (C2)

Our proposed target detection algorithm is based on a two-stage processing paradigm where an image pre-processing stage is followed by a temporal filtering stage. In the previous chapter we considered the selection of a suitable morphological filtering approach for use in the image pre-processing stage. Furthermore, we found evidence that a HMM temporal filtering technique was more compatible with morphological pre-processing than a Viterbi-based method. This chapter examines whether the general HMM filtering technique can be improved through the use of multiple filters. In particular, we compare two candidate HMM filtering approaches for use in the temporal filtering stage: 1) a single HMM filter and 2) multiple HMM filters (HMM filter bank).

The results from our simulation studies show that a multiple filter approach provides superior detection performance compared with a standard single filter implementation under a range of target speed and signal-to-noise ratio scenarios. We highlight that the performance difference between the candidate filtering approaches is most evident for higher target speeds. However, the advantage afforded by multiple filters diminishes as the target speed is reduced, to the point
where both filtering approaches have similar performance for a near stationary target. This behaviour is not unexpected given that a stationary target represents a limiting case where more models/filters will not contribute any additional information to the description of the target dynamics.

The results also show a multiple filter approach to be less sensitive to variations in target intensity; i.e. degradation in detection performance for dimmer targets is less dramatic for the multiple filter approach than the single filter approach. Overall, we believe that the superior performance of the multiple filter approach can be attributed largely to the collective ability of multiple filter models to provide a more precise description of the target dynamics.

Finally, in our studies so far we have often used empirical approaches to establish appropriate HMM parameter values. This potential weakness in our design process will be addressed in the next chapter, where we will describe a systematic method for optimising particular HMM filter parameters to enhance target detection performance.
Statement of Contribution of Co-Authors

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the Australasian Digital Thesis database consistent with any limitations set by publisher requirements.

In the case of this chapter:

Hidden Markov Model Filter Banks for Dim Target Detection from Image Sequences

Published in 2008

Details of Authorship and Contributions:

<table>
<thead>
<tr>
<th>Name and Order of Authors</th>
<th>Signatures of Authors</th>
<th>Area of Contribution Regarding Authorship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(a) (i) (a) (ii) (b) (i) (b) (ii) (c)</td>
</tr>
<tr>
<td>1. John Lai</td>
<td>✓ ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>2. Jason J. Ford</td>
<td>✓ ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>3. Peter O'Shea</td>
<td>✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>4. Rodney Walker</td>
<td>✓ ✓ ✓</td>
<td></td>
</tr>
</tbody>
</table>

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
</table>

*See Appendix A for definition of authorship and criteria for contribution to publication*
(This page is intentionally left blank)
Hidden Markov Model Filter Banks for Dim Target Detection from Image Sequences

John Lai, Jason J. Ford, Peter O’Shea and Rodney Walker
Queensland University of Technology, GPO Box 2434, Brisbane Qld 4001, Australia
j2.lai@student.qut.edu.au, {j2.ford, pj.oshea, ra.walker}@qut.edu.au

Abstract

The track-before-detect processing technique has been employed in numerous computer vision based algorithms to address the dim target detection problem. In this processing approach, target information (as often provided by an image processing stage that has emphasised target features or suppressed unwanted noise) is integrated over a period of time before the detection decision is made. In this paper, we compare two Hidden Markov Model (HMM) based track-before-detect temporal filtering approaches for dim target detection that use image data pre-processed with a Preserved-Sign morphological filter. The two compared temporal filtering approaches are: a standard HMM filter (recent studies have shown this to be close to the state-of-the-art) and a novel HMM filter bank approach.

Results from our simulation study involving various combinations of target speeds and signal-to-noise ratios show that the proposed novel HMM filter bank approach achieves a higher detection rate than the standard HMM approach.

1. Introduction

The ability to detect and track targets in naturally lit, high noise environments is becoming increasingly important. Significant challenges arise however in the use of machine vision for target detection and tracking because of the need to contend with not only the inherent noise of imaging sensors, but also with noise introduced by changing and unpredictable ambient conditions. The need to overcome these challenges has significantly driven the development of image filtering and processing techniques.

Over the last three decades, a two-stage processing paradigm has emerged for the simultaneous detection and tracking of dim, sub-pixel sized targets [1-4]. These two stages are: 1) an image pre-processing stage that, within each frame, highlights potential targets with attributes of interest; and 2) a subsequent temporal filtering stage that exploits target dynamics across frames. The latter temporal filtering stage is often based on a track-before-detect processing concept where target information is collected and collated over a period of time before the detection decision is made.

In this paper, we are interested in the temporal filtering stage of the above two-stage paradigm, and in particular track-before-detect approaches for the detection of slow dim sub-pixel sized targets. This filtering stage is designed to enhance image features that possess target-like temporal behaviour. While an abundance of techniques and algorithms may be considered for this role, there are two particular approaches that have received much attention in the literature: Viterbi based approaches and Bayesian based approaches.

The Viterbi algorithm has formed the basis of the temporal filtering stage in numerous track-before-detect algorithms [1, 2, 4-6]. This is in part due to its utility in the context of tracking where, under a number of assumptions, it is able to efficiently determine the optimal target track within a data sequence [7]. Some analysis of the Viterbi algorithm’s detection and tracking performance can be found in [5, 8, 9], and modifications that enhance the algorithm’s tracking performance in the presence of non-Gaussian clutter noise have been proposed in [2]. An alternative temporal filter design for track-before-detect algorithms is based on Bayesian filtering [6, 10-14]. In [10], the typical white Gaussian noise assumptions are relaxed, with consideration given to spatially correlated clutter. Moreover, in [13], the modelling of clutter is expanded to encompass a variety of Gaussian and non-Gaussian, correlated and uncorrelated clutter types, and the Bayesian algorithm is extended to accommodate multiple targets that may feature randomly varying amplitudes or intensities.

The aim of this paper is to investigate the use of two alternative hidden Markov model (HMM) filtering techniques.
approaches as the temporal processing stage for track-
before-detect algorithms operating on images 
sequences. This investigation is principally motivated 
by recent results demonstrating the utility of HMM-
based filters for target detection and tracking. In 
particular, the detection performance of a standard 
HMM-based filtering approach is shown to be close to 
the state-of-the-art under certain conditions [6]. Hence, 
we compare a traditional single HMM filter with a new 
approach involving a bank of independent HMM 
filters. This comparison is performed in the context of 
a Preserved-Sign (PS) morphological pre-processing 
stage, which has been shown to be an effective choice 
for track-before-detect algorithms [15]. 

We assess our detection algorithms via detection 
and false-alarm statistics that, in turn, provide valuable 
insight into tradeoffs in performance from using the 
two different temporal filtering approaches. This 
assessment is conducted via simulation studies of slow 
dim sub-pixel targets, such as those that might be 
expected in an airborne collision avoidance scenario 
[16]. Further, the performance of the algorithms is 
examined for a range of target speeds and signal-to-
noise ratios. The results from the simulation studies 
show that the multiple-HMM filtering bank approach 
is superior to the single HMM filter by providing better 
detection rates over a range of false-alarm rates. 

2. Morphological Image Pre-Processing

The preserved-sign morphological filter [15] is 
based on image morphology operations known as top-
hat and bottom-hat transformations [17]. It is a 
variation on the top-hat contrast enhancement operator 
described in [18]. The effect of the top-hat 
transformation is to identify positively contrasting 
features within an image that are smaller than a certain 
size (the cut-off size is specified through filtering 
kernels known as structuring elements), while the 
bottom-hat transformation performs a similar function 
but instead targets negatively contrasting features. It 
can be shown that subtracting the bottom-hat 
transformation from the top-hat transformation of the 
same image, which defines the preserved-sign 
morphological filtering operation, simultaneously 
identifies both positively and negatively contrasting 
features, where the response to positively contrasting 
features is non-negative and the response to negatively 
contrasting features is non-positive.

We thus take advantage of the preserved-sign 
morphological filter to preserve features with sizes that 
match potential targets and to provide a means of 
differentiating between positively and negatively 
contrasting features. For performance and 
computational reasons, we exploit a directional 
decomposition technique [19] in our implementation of 
the morphological filter, whereby we take the 
iminimum magnitude response from a pair of 
preserved-sign filters using orthogonal 1D structuring 
elements. Here, one preserved-sign filter operates 
exclusively in the vertical direction, while the other 
operates exclusively in the horizontal direction.

The following section details the general HMM 
temporal filtering approach that is applied to the pre-
processed image data. This filtering stage is designed 
to exploit the differences between the temporal 
behaviour of genuine and false targets.

3. Hidden Markov Model Filtering

We will assume that, when present, the target is 
located within a particular pixel of the image frame at 
each time instant. The pixels of an image frame thus 
represent the states of the HMM used in our target 
detection problem. Between consecutive image frames 
the target may move to different pixel locations – that 
is, the target can transition between the states. The 
likelihood of state transitions can be described by the 
HMM’s transition probability matrix $A$, where each 
element $A_{ij}$ is the probability of moving from any one 
pixel position (state) $i$ to any other pixel position (state) $j$ [20]. The matrix $A$ can therefore be used to 
describe the expected mean target motion. For 
example, in the case of slow moving targets we tend to 
assign low probabilities to transitions between distant 
 pixels. Moreover, an initial probability matrix $\pi$ is 
used to specify the probability of the target initial 
location [20]. Finally, to complete the parameterisation 
of the HMM, there is the measurement probability 
matrix $B_k(Y_k)$, with elements $B_k(y_j)$ that are used to 
specify the probability of obtaining the observed 
measurement $Y_k$, given that the target is actually in 
pixel location (state) $i$ [20].

Using this HMM model, target detection can be 
viewed as evaluating the likelihood of 2 alternate 
 hypotheses. Let $H_1$ denote the hypothesis that there is 
a single target present in the camera field of view, and 
let $H_0$ denote the hypothesis that there is no target 
present. The proposed HMM can be used to develop 
conditional mean estimates and infer the level of 
evidence in support of the hypothesis that the target is 
present (following the results of [21]).

Appropriate HMM parameters describing the target 
detection problem are defined in the following 
sections.
3.1. Markov Chain State Process

In many electro-optical based detection problems, the existence of a target in a 3D volume of space must be determined from observations of a projection of the target space onto a 2D image plane. In this paper, we use a first-order discrete-time discrete-state Markov chain to model the projected motion of a possible target on this 2D image plane.

Under hypothesis \( H_i \), the projected target motion is assumed to reside on a 2D plane fixed in space that is represented by the set of discrete 2D grid points \( \{(i, j)|1 \leq i \leq N_x, 1 \leq j \leq N_y\} \), with vertical and horizontal resolutions \( N_x \) and \( N_y \) respectively. Let \( N = N_xN_y \) denote the total number of grid points. Without loss of generality, it is notionally convenient to introduce an indicator vector representation of the target location \( X_k \) by stacking the columns of the 2D grid to form a single column vector. Hence, if the target is at grid location \((i, j)\), the \( N \times 1 \) target indicator vector \( X_k \) is zero everywhere except for the \([(j-1)N_x + i]^\text{th} \) element, which is assigned the value of 1.

The dynamics of the target can be modelled as a first-order discrete-time Markov chain with transition probabilities described by a matrix \( A \) composed of elements \( A_{ij} = P(X_{k+1} = e_j | X_k = e_i) \), where \( e_i = (0, \ldots, 0, 1, 0, \ldots, 0) \) with 1 in the \( i \)th position. We also use the initial probability matrix \( \pi \) to denote the \( N \) initial (prior) probabilities, where \( \pi_i = P(X_0 = e_i) \) for \( 1 \leq i \leq N \). Under hypothesis \( H_i \), our target estimation problem can be considered a HMM filtering problem.

3.2. Measurement Process

The measurements are provided by an electro-optical imaging sensor whose field of view is represented by a 2D grid of image pixel locations, which are denoted \( \{(p, q)|1 \leq p \leq N_x, 1 \leq q \leq N_y\} \). We model these image measurements as noisy observations of the target Markov chain. Let \( Y_k \) denote the pre-processed image measurement at time \( k \). For notational convenience, \( Y_k \) is an \( N \times 1 \) vector formed by stacking together the columns of the pre-processed image frame. Under hypothesis \( H_i \), at time \( k \), we assume the target measurement satisfies the equation:

\[
Y_k = LX_k + w_k, \quad (1)
\]

where \( L \) is a scalar quantity representing the target intensity, and \( w_k \) is a \( N \times 1 \) vector containing the additive noise component, assumed to be i.i.d (independent and identically distributed) and having density \( \psi(\cdot) \) (for example, a Gaussian density). We will denote a sequence of pre-processed image frames from time \( \ell \) to \( k \) as \( Y_{\ell:k} = \{Y_\ell, Y_{\ell+1}, \ldots, Y_k\} \).

The probabilistic relationship between target location \( X_k \) and the pre-processed measurement \( Y_k \) is described by a \( N \times N \) measurement probability matrix \( B (Y_k) \), where the \( ij \)th element \( B_{ij} = p(Y_k | X_k = e_j) \) if \( i = j \) and \( B_{ij} = 0 \) elsewhere, for \( 1 \leq i, j \leq N \). Under the following assumptions:

1) The statistical properties of pixel values within an image are spatially independent; that is,

\[
p(Y_k | X_k = e_{m}) = p(Y_k | X_k = e_{n}) \quad (2)
\]

for all \( i, j \) and \( m, n \), and 2) Individual pixels do not allow the opportunity of perfect detection, in the sense that \( p(Y_k | X_k = e_i) > 0 \) whenever \( p(Y_k | X_k = e_i) > 0 \), it can be shown that a quantity proportional to \( B_i (Y_k) \), which we denote as \( \overline{B}_i (Y_k) \), can be given as

\[
\overline{B}_i (Y_k) = \begin{cases} 
p(Y_k | X_k = e_i) & \text{if } i = j \\
p(Y_k | X_k \neq e_i) & \text{if } i \neq j \\
0 & \text{otherwise} \end{cases} \quad (3)
\]

for \( 1 \leq i \leq N \). We highlight the computational advantage that \( \overline{B}_i (Y_k) \) affords, given that \( p(Y_k | X_k = e_i) \) and \( p(Y_k | X_k \neq e_i) \) can each be determined on a single-pixel basis (rather than requiring the probability of a whole image). Admittedly, the above result would not hold in the presence of extended (multi-pixel) targets, or spatially correlated noise.

Remark

We note that even when the assumption of spatial independence of pixel statistical properties does not strictly hold (as in the case of extended targets or
spatially correlated noise), we have found that HMM filtering performance using the above result for $F_k(Y_k)$ to be acceptable in the sense that detection performance is competitive with other candidate (non-HMM based) detection algorithms.

### 3.3. Detection Strategy

Detection is the process of recognising the presence of a target as well as determining its location. We exploit powerful HMM filtering algorithms to provide us with two probabilistic measures to facilitate the assessment of target presence and target location: 1) $P(Y_k | H_1)$, the probability of measurements up to time $k$ assuming $H_1$, and 2) $\hat{X}_k = E[X_k | y_k, H_1]$, the conditional mean filtered estimate of the target state $X_k$ given measurements up to time $k$ and assuming $H_1$, where $E[\cdot | .]$ denotes the mathematical conditional expectation operation (see [22] for more details).

Given our HMM target model, the conditional mean estimate can be recursively calculated as [20]

$$\hat{X}_k = N_k B_k(Y_k) A \hat{X}_{k-1}, \quad (4)$$

where $N_k$ is a normalisation factor. The conditional mean estimate may be interpreted as an indicator of likely target locations.

The likelihood of a particular model (our $H_1$ hypothesis) is related to the product of normalisation factors, as demonstrated in the probabilistic distance results of [21]. Hence, the value of $P(Y_k | H_1) = \prod_{i=1}^k 1/N_i$ can be used to detect target presence.

### 4. Proposed Temporal Filters

In this paper, we implement a standard HMM filter (competitive with state-of-the-art detection approaches) and a HMM filter bank consisting of four filters. Each HMM filter in the bank uses the same pre-processed image data, but otherwise operates independently of all other filters. The filter bank approach is less well characterised than the standard single HMM filter, and its application has not been prevalent in the context of dim-target detection from imaging sensors. We do however acknowledge that the general concept of using multiple filters has been studied previously (e.g. [23]).

The preserved-sign morphological pre-processing output provides the opportunity for our HMM filtering implementations to process both positively and negatively contrasting targets simultaneously. However, for convenience, we choose in this paper to process positively and negatively contrasting targets separately.

The transition probabilities of the standard HMM filter are defined so that only a self-transition or a transition to any one of the 8-connected neighbours is possible from time $k$ to $k+1$. This leads to a sparse $A$ matrix. Figure 1 provides a visual representation of the possible state transitions in the standard HMM filter.

![Figure 1. Possible state transitions of the standard HMM filter. The solid squares define the 8-connected neighbourhood of the state (depicted here as the central cell), with the arrows indicating the transition possibilities from time $k$ to $k+1$. Thus, the state may only undergo a self-transition (dark cell) or a transition to any one of its 8-connected neighbours (light cells).](image)

In contrast, for each HMM filter in the filter bank, we limit the possible transitions to only three of the 8-connected neighbours, in addition to self-transitions. The three pixels in the 8-connected neighbourhood to which transitions are possible are selected so that the filters in combination cover all the possible transitions in the standard HMM filter. Figure 2 provides a visual representation of four possible state transition schemes, each of which is assigned to a different filter of the HMM filter bank.
Figure 2. State transitions of the HMM filter bank. (a) – (d) Separate schemes each assigned to a different filter of the filter bank.

5. Performance Characterisation

We compare the performance of the two alternative HMM temporal filtering techniques

- A single HMM filter, and
- A bank of four HMM filters

by applying them to a large number of morphologically pre-filtered image sequences containing targets having a variety of intensity and speed attributes. In the following subsections, we describe our metrics for quantifying performance, our procedure for generating image sequences, as well as the presentation of results. These aspects comprise the simulation framework of our comparison study.

5.1. Performance Metrics

Comparisons between the algorithms are made on the basis of detection versus false-alarm statistics evaluated on sets of data of length $T$. If a target is present, the track-before-detect algorithm is considered to have achieved a detection if the algorithm correctly identifies the target’s presence and locates it to within two pixels of the true position. We define the detection rate as the number of detections divided by the maximum number of possible detections. A false-alarm event occurs if the track-before-detect algorithm incorrectly declares the presence of a target. We define the false-alarm rate as the number of false-alarms divided by the total number of possible false-alarm events.

Let $\alpha$, our test statistic for declaring the presence of a target, be given by

$$\alpha = \frac{1}{T} \log \left( \prod_{t=1}^{T} \frac{1}{N_t} \right).$$

(5)

When $\alpha$ exceeds a predefined threshold, the track-before-detect algorithm considers a target to be present and located at

$$\beta = \arg \max \left( \hat{x}_t \right).$$

(6)

Our definition of $\alpha$ and $\beta$ is motivated by the detection strategy discussed in Section 3.3.

In the case of the filter bank, detection and false-alarm events may be triggered by any filter. Where multiple filters are involved, decisions revolve around the dominant filter – that is, the filter with the maximum $\alpha$ test statistic. In particular, if a target is present and this presence is declared in multiple filters, we look to the dominant filter’s $\beta$ to provide an estimate of the target location. On the other hand, if a target is not present, the $\alpha$ test statistic of the dominant filter exceeding the threshold is sufficient for a false-alarm event to have occurred, even though the threshold may not be exceeded by all filters.

5.2. Image Sequence Generation

Due to our limited access to authentic data, synthetically generated image sequences are used in conducting the large number of trials our comparison study required. The image frames that comprise our synthetic image sequences consist of three elements: background, noise, and the target signature.

The background component consists of a uniform image which is set at an arbitrary greyscale level of 128. This simply forms the base level image intensity.
to which the noise and target elements are added. To create the noise component of the image frame, we form a noise image consisting of random samples from a zero-mean Gaussian distribution. The noise is spatially and temporally uncorrelated. We plan to examine a more extensive range of noise types in future studies.

Consider any finite sized target with a non-zero velocity. As this target traverses across the image, the physical extent of the target is likely to overlap multiple pixels at any time. We can form an image which we refer to as the ‘target signature’ by assigning to each pixel a value calculated as the target intensity scaled by the amount of target overlap. For example, if the target occupies half the area of a particular pixel, then that pixel as part of the target signature is assigned half the value of the target intensity. Accordingly, if the target does not overlap a particular pixel, it is assigned the value of zero. We add the target signature to the noise and image components discussed earlier to complete the process of embedding a target into a synthetic image frame.

5.3. Signal-to-Noise Ratio

The concept of a signal-to-noise ratio (SNR) quantity can provide an indication of how distinct a target is (i.e. how well it stands out from the background, and in turn the ease with which it may be detected). Here, we define SNR as $20 \log_{10} (I/ \sigma)$ dB, where $I$ is the maximum target signature value and $\sigma$ the noise standard deviation. The upper bound for the SNR, the peak SNR (PSNR), is achieved when $I$ is equal to the target intensity. In our simulation studies, the tendency of the SNR to vary with time leads us to quote the PSNR instead as a means of signifying how distinct the target is. We note that the PSNR is only indicative of the target’s distinctiveness and provides an over bound on the average SNR.

5.4. Detection versus False-Alarm Curves

We apply a range of threshold values to the $\alpha$ test statistic in order to capture as best as possible a representative dynamic range of detection and false-alarm rates. The results may be presented in separate graphs that illustrate the detection and false-alarm rates as a function of their respective threshold values. Alternatively, in the case that the same threshold range is applied to both the evaluation of detections and false-alarms, the performance information may be consolidated into a single detection vs. false-alarm graph that concisely illustrates the tradeoffs between the two performance metrics independent of threshold values. In this paper, the latter approach is preferred in the presentation of performance comparison results.

6. Performance Comparison Study

The aim of this study is to investigate the performance of the candidate temporal filters in uncorrelated Gaussian noise (with zero-mean; standard deviation of 1) for a selection of target speeds and PSNRs. Specifically, we consider a $1 \times 1$ pixel target travelling at speeds of 0.1, 0.2, and 0.3 pixels/frame and with intensities corresponding to PSNRs of approximately 8, 9.5, and 11 dB. Image frames of size $N_x = 111$ and $N_y = 147$ comprise image sequences that are 151 frames in length. For each combination of target speed and PSNR, detection counts are gathered from $10^7$ separate image sequences each containing only a single target (all targets converged towards the centre of the image frame, but began at initial locations evenly distributed about the centre). The same number of image sequences (without targets) is used in the calculation of false-alarm counts. The horizontal and vertical structuring elements of the morphological pre-processing filter are given by $s_h = [1, 1, 1, 1, 1]'$ and $s_v = [1, 1, 1, 1, 1]'$, respectively.

6.1. HMM Filter Parameters

For the standard single HMM filter, we assign a probability of 7/15 to self-transitions and a probability of 1/15 for transitions to adjacent pixels. In the case of the filter bank, each member filter has self-transitions with probability 7/10 and adjacent pixel transitions with probability 1/10. We highlight that performance was not overly sensitive to these parameter choices (the particular values selected here were found to give reasonable performance). A theoretical basis for the design of the transition probabilities is ongoing work we intend to report in later papers.

To construct our measurement probability matrix $\mathbf{B}(Y_i)$, we are required to estimate the densities $p(Y_i' \mid X_j \neq e_i)$ and $p(Y_i' \mid X_j = e_i)$. The former density describes our prior knowledge about the distribution of pixel values in the absence of a target (i.e. the noise and clutter distribution), while the latter density captures our prior knowledge about the distribution of values at pixels containing a target.

We estimate the required densities for $\mathbf{B}(Y_i)$ directly from data. The density $p(Y_i' \mid X_j \neq e_i)$ is estimated as the average frequency that each pixel
value resulted from a non-target location. Using a similar procedure, \( p(x'_i | x_i = e) \) is estimated as the average frequency that each pixel value measurement resulted from a target location.

### 6.2. Results

In this study, we found the HMM filter bank to be superior to the single filter approach for all combinations of PSNRs and target speeds. The performance difference is most evident in the lowest PSNR (8 dB), highest target speed (0.3 pixels/frame) scenario, as illustrated by the detection rate versus false-alarm rate curves of Figure 3. Here, the HMM filter bank has significantly better detection performance than the near state-of-the-art single HMM approach. For instance, at a false-alarm rate of \( 10^{-3} \), the HMM filter bank has a 92% detection rate compared with a rate of 26% for the single filter approach.

![Figure 3. Detection vs. false-alarm performance comparison (target speed 0.3 pixels/frame; PSNR 8 dB).](image)

In figures 4 and 5, we illustrate the detection performance of the two temporal filters as a function of PSNR and target speed, respectively. For both figures the false-alarm rate is fixed at \( 10^{-3} \). In Figure 4 where results for a target speed of 0.3 pixels/frame are shown, we observe that the performances of the two temporal filters converge as we approach higher PSNRs. In Figure 5 where a PSNR of 9.5 is applicable, performance for both filtering approaches tend to improve for slower moving targets. Although not shown here, similar trends to the above exist for other false-alarm and PSNR/target speed combinations.

![Figure 4. Detection performance comparison as a function of PSNR (target speed 0.3 pixels/frame; false-alarm rate \( 10^{-3} \)).](image)

![Figure 5. Detection performance comparison as a function of target speed (PSNR 9.5 dB; false-alarm rate \( 10^{-3} \)).](image)

### 7. Conclusion

In this paper we have compared the performance of two HMM temporal filtering approaches in the context of track-before-detect dim target detection. Our simulation study has yielded promising results showing that a filter bank approach provides superior detection performance compared with the near state-of-the-art standard single filter implementation.

We plan to consider a more extensive range of noise and target types and introduce authentic data into the performance characterisation process in future comparison studies.

### 8. References


Chapter 5

Hidden Markov Model Design

(J1)

In the previous chapter, we showed that an improvement in detection performance could be achieved by using multiple HMM filters in the temporal filtering stage, as opposed to just using a single HMM filter. The performance of the temporal filtering stage can be further enhanced, as we show in this chapter, by intelligently designing the multiple HMM filters to optimise detection performance. We propose a novel filter design process that is posed as a mini-max optimisation problem based on a joint RER cost criterion. The underlying strategy in our filter design process is to select a collection of filter models so that a sound representation of the target dynamics can be achieved at all times, particularly in the presence of uncertainty or inevitable variations in the expected target dynamics. We achieve this by quantifying each candidate collection’s worst-case performance through a max-RER design cost metric, and then selecting the collection that has the lowest design cost. For example, consider a hypothetical situation in which there are four candidate model collections under consideration: candidates ‘A’, ‘B’, ‘C’, and ‘D’. Figure 5.1 illustrates the max-RER design
costs associated with each candidate collection. In this case, it is clear that candidate ‘C’ has the lowest design cost, and hence this is the collection that will be chosen. In this chapter we also propose a conditional-RER cost metric that may be used in place of the max-RER cost when a priori information is known about the expected target dynamics.
Statement of Contribution of Co-Authors

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Digital Thesis database consistent with any limitations set by publisher requirements.

In the case of this chapter:

Relative Entropy Rate based Multiple Hidden Markov Model Approximation

Status at time of writing: Published January 2010

Details of Authorship and Contributions:

<table>
<thead>
<tr>
<th>Name and Order of Authors</th>
<th>Signatures of Authors</th>
<th>Area of Contribution Regarding Authorship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(a) (i) (a) (ii) (b) (i) (b) (ii) (c)</td>
</tr>
<tr>
<td>1. John Lai</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>2. Jason J. Ford</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓ ✓</td>
</tr>
</tbody>
</table>

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Name __________________ Signature __________________ Date __________________

*See Appendix A for definition of authorship and criteria for contribution to publication
(This page is intentionally left blank)
Relative Entropy Rate based Multiple Hidden
Markov Model Approximation

John Lai*, Jason J. Ford

Abstract

This paper proposes a novel relative entropy rate (RER) based approach for multiple HMM (MHMM) approximation of a class of discrete-time uncertain processes. Under different uncertainty assumptions, the model design problem is posed either as a min-max optimisation problem or stochastic minimisation problem on the RER between joint laws describing the state and output processes (rather than the more usual RER between output processes). A suitable filter is proposed for which performance results are established which bound conditional mean estimation performance and show that estimation performance improves as the RER is reduced. These filter consistency and convergence bounds are the first results characterising multiple HMM approximation performance and suggest that joint RER concepts provide a useful model selection criteria. The proposed model design process and MHMM filter are demonstrated on an important image processing dim-target detection problem.

Index Terms

Relative Entropy Rate, hidden Markov models, Markov processes, Detection.

EDICS Category(s): ASP-ANAL, SSP-FILT, SSP-IDEN

I. INTRODUCTION

Advances in hidden Markov model (HMM) signal processing tools over the last few decades have contributed to the widespread application of HMMs in a multitude of technical disciplines, including...
non-linear stochastic control [1], [2], signal and image processing [3]–[8], digital communications [9], and bioinformatics [10]. In particular, HMMs have been used to solve a variety of filtering problems in frequency tracking [7], speech recognition [8], character recognition [6], and dim target detection [3]–[5]. In these filtering problems, approximate HMM representations of the complex dynamics are often exploited to simplify computations and to allow tractable solutions to be developed in situations where other non-linear filtering tools are sometimes inadequate.

Despite the existence of an extensive suite of HMM processing tools and their successful application to a variety of situations, the problem of determining a suitable HMM approximation has not been completely resolved in a number of situations. As a result, there are many application-based examples in the literature where the design of HMM approximation models has been on an ad hoc basis rather than a systematic basis [5], [11], [12].

One apparent model approximation approach is through the application of classical data-based model inference techniques to infer the HMM that best matches the sample measurement and state sequences generated by the model under approximation. Data-based model inference or system identification is a classical signal processing problem that has been solved, over many decades, using a variety of techniques, including: information theory and entropy based techniques (such as Akaike’s Information Criterion [13]); maximum likelihood based techniques (such as the EM algorithm [14]); and prediction error based techniques [15]. In the last few decades, many of the above techniques have been applied to a variety of HMM parameter estimation problems, see [8], [16]–[21].

An alternative approach to HMM approximation has also emerged from the HMM model realization problem [22]–[26]: given the output string probabilities, determine the HMM that has the same output string probabilities (if a valid HMM exists). In [24], existence and construction results for the HMM realization problem was presented in terms of finite rank Hankel matrices and polyhedral cones (whose existence presupposes existence of a suitable HMM). This limitation was overcome in [23] through the introduction of an ultra-mixing property which guarantees existence of the necessary polyhedral cone, and allows a complete realization result to be established. In [22], a different aspect of the realization problem was examined through the use of subspace approaches to test equivalency between different HMM representations. Most recently, entropy related concepts have been used in [25], [26] to consider an approximate HMM realization problem that addresses the question: given output string probabilities, determine the closest HMM (when no exact HMM realization exists). However, the above realization problems only seek HMMs capable of generating the prescribed output properties.

In other work, relative entropy concepts have been identified as an important design criteria in a range
of HMM estimation and approximation problems [17], [25]–[27]. For example, it has been recently shown that the relative entropy rate (RER) between the joint state and output processes of two HMMs allows model parameters to be related to the relative likelihood that two compared HMMs produced a particular output sequence [27], [28]. That is, RER concepts allow HMM parameter values to be directly interpreted in terms of expected filter performance [29]. For these reasons, RER concepts seem to be a natural design criteria that allows comparison between two candidate HMM representations.

In this paper, we exploit RER concepts to facilitate design of a finite set of HMMs to approximate a specified uncertain model, in the sense that each of the possible joint state and output behaviours representing the uncertainty is reasonably approximated by at least one of the HMMs. This set of HMMs will be termed a HMM collection. Our approach is partially motivated by the multiple model hybrid system filtering approaches used in target tracking applications, where maneuver uncertainty is approximated by a finite set of continuous-valued dynamic models that, together, represent the range of possible target behaviours [30], [31].

The two main contributions of this paper are the proposal of a multiple hidden Markov model (MHMM) filter for uncertain dynamics, and the proposal of joint output-state RER (as distinct from output-only RER) based design procedures for the introduced MHMM filter. Furthermore, we establish MHMM filter consistency results, and are also able to show that the proposed joint RER based design approach leads to some upper bounding results on filter estimation error.

The benefits of our joint RER based design procedures are illustrated through simulation of a detection problem involving dim targets in image sequences that has previously been approached using dynamic programming and HMM based approaches [3], [5], [11], [32]–[34].

This paper is organized as follows: In Section II, our nominal model dynamics are defined, and our basic HMM approximation model, our HMM collection concept, and our MHMM filter are all introduced. In Section III, various RER concepts are introduced and several important relationships presented. In Section IV, we consider the design of HMM collections, and establish some some useful filtering performance properties for our MHMM filter. In Section V, some implementaion issues are briefly discussed, before a dim target image processing application is introduced in Section VI, and some simulation studies are presented that illustrate the application of our RER-based MHMM design approach. Finally, in Section VII, some conclusions are made.
II. DYNAMICS, MODELS AND REPRESENTATIONS

This section introduces our nominal dynamics, our concept of a HMM collection, and then presents a suitable filtering algorithm.

A. Nominal Dynamics

All processes will be defined on an abstract complete probability space \((\Omega, \mathcal{F}, P)\). For \(k > 0\), consider the following state process,

\[
x_{k+1} = f(x_k, w_k, \theta), \quad x_0 \in \mathbb{R}^n
\]

where \(x_k \in \mathbb{R}^n\) is a measurable function called the state with initial value \(x_0\) and density \(\phi_{x_0}(\cdot)\), \(w_k\) is a noise process with density \(\phi_w(\cdot)\), and \(\theta \in S_\theta\) is an unknown quantity that parameterizes a range of possible dynamics, with true value \(\theta^0 \in S_\theta\). Here \(S_\theta\) is a set describing the possible values of the uncertain quantity. We will restrict \(f, \phi_{x_0}, \) and \(\phi_w\) so that the state process is bounded to a subset of \(\mathbb{R}^n\), in that \(x_k \in S_x \subset \mathbb{R}^n\) for all \(k\), and that the process is both stationary and ergodic, see [30, pp. 62-70].

For example, an angular random walk (with a drift bias) would naturally have bounded state space \(S_x\) and might be expressed in the form \(x_{k+1} = \text{Proj}_{[0, 2\pi]}(x_k + \theta + w_k)\), where \(x_k\) denotes angle, \(w_k\) is a Gaussian noise, and \(\text{Proj}_{[0, 2\pi]}\) is a projection operation into \([0, 2\pi]\). Note this process is bounded, stationary and ergodic. Moreover, the drift bias \(\theta\) might represent an uncertain quantity taking possible values in the set \(S_\theta\).

We will assume that the state is observed through a measurement process \(y_k \in \mathbb{R}^m\), described by the output mapping process

\[
y_k = c(x_k) + v_k
\]

where \(v_k\) is a noise process that is independent of the process \(x_k\) with density \(\phi(\cdot)\). Without loss of generality, the relationship between measurement \(y\) and state \(x\) in (2) will be described by the probability law \(p^0(y|x) > 0\) for all \(x \in S_x, y \in \mathbb{R}^m\). Finally, for \(k \geq 0\), we assume that \(x_k\) and \(y_k\) are \(\mathcal{F}\)-measurable functions.

We will denote the true model of \(x_k\) and \(y_k\) processes as \(\lambda^0(\theta^0)\). When our knowledge of \(\theta\) is uncertain, we will denote our uncertain model as \(\lambda^0(\theta)\) or simply as \(\lambda^0\). In this paper, we address two key questions to illustrate how the uncertain model \(\lambda^0\) can be approximated by a collection of HMMs: how to select a suitable HMM collection, and how to construct filter estimates from a specified HMM collection.
B. A Single HMM

We first consider approximation of a known non-linear model \( \lambda^0 (\theta^0) \) by a suitable HMM. To construct such a HMM approximation, we first introduce a bounded \( n \)-dimensional spatial discretization of the state-space \( S_x \). Without loss of generality, let us introduce the spatial grid
\[
G_h = \{ (x, y, z, \ldots) : x = \pm m_x h_x, y = \pm m_y h_y, z = \pm m_z h_z, \ldots \}
\]
that approximates the space \( S_x \) with some spacing parameter \( h = [h_x, h_y, h_z, \ldots] \). Let \( N_h \) denote the number of grid points.

Let \( \epsilon_i = (0, \ldots, 0, 1, 0, \ldots, 0) \) denote a vector with 1 in the \( i \)th position, and zero elsewhere. At time instant \( k \), and with spacing parameter \( h \), we will let \( X_{k|h} \in \{ e_1, e_2, \ldots, e_{N_h} \} \) be an indicator vector that denotes the state of a Markov chain process defined on grid \( G_h \); that is, \( X_{k|h} \) refers to a specific grid location on \( G_h \) which will be denoted by the function \( G_h \left( X_{k|h} \right) \).

The Markov state process is assumed to have transition probabilities described by a matrix \( A_h \), with \( i \)th element \( A_h^{ij} = p^{\lambda_h}(X_{k+1|h} = e_i | X_{k|h} = e_j) \), where \( p^{\lambda_h}(\cdot) \) will denote the probability law describing our HMM. Given that (1) is stationary and ergodic, we expect that our approximation task will lead to irreducible and aperiodic chains [40, pp. 50]. We use the initial probability matrix \( \pi_h \) to denote the \( N_h \) \textit{a priori} probabilities, where \( \pi_h^i = p^{\lambda_h}(X_0|h = e_i) \) for \( 1 \leq i \leq N_h \). Throughout, we will assume that the initial probabilities \( \pi_h \) are known.

The Markov chain is considered to provide an approximation in the following sense. Let \( C_h(X_{k|h}) \) denote a \( h \)-sized \( n \)-dimensional box or cell containing grid location \( G_h \left( X_{k|h} \right) \). The cell \( C_h(X_{k|h}) \) will be used to describe the set of possible locations in \( S_x \) represented by state \( X_{k|h} \). We assume that the boxes \( C_h(.) \) completely cover \( S_x \) in the sense that for all \( x \in S_x \), \( x \in C_h(X_{k|h}) \) for some \( X_{k|h} \), and we assume that boundaries between adjacent boxes are not shared. For example, we could enumerate the boxes in some manner and assign the shared boundary between two adjacent boxes to be part of the box with higher enumeration (but other boundary handling methods are possible).

Related to this Markov chain, we also consider the discrete-time approximating process \( x_{k|h} \in C_h \left( X_{k|h} \right) \) for all \( k \geq 0 \), where \( x_{k|h} \) has uniform distribution over the cell \( C_h \left( X_{k|h} \right) \). We assume that the cell is chosen so that each grid point \( G_h \left( X_{k|h} \right) \) is centred in a corresponding cell \( C_h \left( X_{k|h} \right) \) in the sense that \( E \left[ x_{k|h} | X_{k|h} \right] = G_h \left( X_{k|h} \right) \). With this construction, we can consider \( x_{k|h} \) to be a “blurred” version of the chain process.

We assume the measurements related to this process are characterized by the probability law
\[
p^{\lambda_h}(y_{k|h} | x_{k|h}) \in C_h(e_i) = p^{\lambda_h}(y_{k|h} | X_{k|h} = e_i) \quad \text{for} \quad 1 \leq i \leq N_h \quad \text{and all possible values of} \ y_{k|h} .
\]
For notational convenience in later matrix equations, these measurement probabilities can be arranged into a
\( N_h \times N_h \) diagonal measurement matrix \( B \left( y_{k|h} \right) \), where the \( rs \)th element is defined as

\[
B_{rs} \left( y_{k|h} \right) = \begin{cases} 
p^{\lambda_h} \left( y_{k|h} | X_{k|h} = e_s \right) & \text{if } r = s \\
0 & \text{otherwise,} 
\end{cases}
\]

for \( 1 \leq r, s \leq N_h \). Throughout, we will assume that the measurement laws \( p^0(y|x) \) and \( p^{\lambda_h}(y|x) \) are absolutely continuous with respect to each other (that is, these measurement laws are equivalent), for all \( x \in \mathcal{S}_x \) (see [35, p. 422] for the definition of absolutely continuous).

We will use \( \lambda_h = (A_h, B \left( y_{k|h} \right), \pi_h) \) to denote a candidate HMM based approximation with corresponding discrete-time approximating processes \( x_{k|h} \) and \( y_{k|h} \). We will use the short hand \( x_{[a,b]} \) to denote the process \( x_k \) with \( k \in [a,b] \). We similarly define \( y_{[a,b]}, x_{[a,b]|h}, y_{[a,b]|h} \) for processes \( y_k, x_{k|h} \) and \( y_{k|h} \) with \( k \in [a,b] \).

Next, we will introduce our concept of a HMM collection. Later on in the paper, we will illustrate the advantages of using a HMM collection (instead of a single HMM model) in approximating our nominal system.

### C. A HMM Collection

Let \( \mathcal{M}_{q|h} = \{ \lambda_{1|h}, \lambda_{2|h}, \ldots, \lambda_{q|h} \} \) denote a particular collection of \( q \) HMM models on grid \( G_h \), where \( \lambda_{i|h} \) is the \( i \)th HMM, with Markov process \( X_{k|h,i} \), parameterized by transition probability matrix \( A_{i|h} \), the common initial probability vector \( \pi_h \) and output matrix \( B_{i|h} \left( y_{k|h} \right) \). Note, for each \( p = 1, 2, \ldots, q \), \( A_{p|h} \in S_{A,h} \), where

\[
S_{A,h} = \left\{ A_h : 0 \leq A_{i|h} \leq 1 \text{ for all } i, j \text{ and } \sum_{i=1}^{N_h} A_{ij|h} = 1 \text{ for all } j \text{ and } A_h \text{ is irreducible and aperiodic} \right\}
\]

denotes the set of all valid transition probability matrices with the required stationarity and ergodicity properties.

Furthermore, let \( A_{q|h} = \{ A_{1|h}, A_{2|h}, \ldots, A_{q|h} \} \) denote the corresponding collection of transition matrices associated with a particular \( \mathcal{M}_{q|h} \). We similarly define \( B_{q|h} \). Finally, let \( S_{\mathcal{M}_{q|h}} \) denote the set of all valid HMM collections \( \mathcal{M}_{q|h} \) (we similarly define sets \( S_{A_{q|h}} \) and \( S_{B_{q|h}} \)).

A HMM collection will be used to define an approximation process that takes values in the cells \( C_h \left( X_{k|h,i} \right) \), for some choice of \( i \in \{1, \ldots, q\} \). That is, one of the models in the HMM collection will be used to define an approximating state process. A precise definition of this multiple HMM approximation process will be given in Section IV, once we have developed the necessary technical framework.
D. Our Proposed HMM Filter

Given a single HMM, the conditional mean estimate of $X_k|\mathbf{h}$ given measurements up to time $k$, $y_{[0,k]}|\mathbf{h}$, is [16]:

$$\hat{X}_k|\mathbf{h} = N_k B \left( y_k|\mathbf{h} \right) A_h \hat{X}_{k-1}|\mathbf{h},$$

(4)

where $N_k$ is a scalar normalisation factor and $\hat{X}_k|\mathbf{h} = E[X_k|y_{[0,k]}|\mathbf{h}]$. Alternatively, the closely related unnormalised conditional mean estimate $\alpha_k|\mathbf{h}$ is given by the following recursion [8]:

$$\alpha_{k,i} = B_{ii} \left( y_k|\mathbf{h} \right) \left[ \sum_{j=1}^{N_h} \alpha_{k-1,j} A_{ij} \right]$$

for $1 \leq i \leq N_h$.

where we note that $\hat{X}_k|\mathbf{h}$ and $\alpha_k|\mathbf{h}$ are related as follows:

$$\hat{X}_k|\mathbf{h} = \frac{\alpha_k|\mathbf{h}}{\sum_{i=1}^{N_h} \alpha_{k,i}}.$$

Given a HMM collection $\mathcal{M}_{q|\mathbf{h}}$, we can consider the $q$ parallel autonomous HMM filters

$$\hat{X}_k|\mathbf{h},i = N_i B_{i} \left( y_k|\mathbf{h} \right) A_{i|\mathbf{h}} \hat{X}_{k-1}|\mathbf{h},i,$$

(5)

for $i = 1, \ldots, q$, where $\hat{X}_k|\mathbf{h},i = E[X_k|\mathbf{h},i,y_{[0,k]}|\mathbf{h}]$. Let $W_i^k \triangleq p^{\lambda_{ih}}(y_{[0,k]},X_{[0,k]}|\mathbf{h},i)$ denote the $i$th joint likelihood at time $k$. Then let us propose the HMM collection (or MHMM) based filter estimate

$$\hat{x}_k|\mathbf{h} \triangleq \sum_{j=1}^{N_h} \hat{X}_{k|\mathbf{h},i}^j G_h(e_j)$$

(6)

where $i^*$ is selected so that $W_i^k |\mathbf{h} \geq W_i^* |\mathbf{h}$ for all $i$. Determination of $i^*$ in (6) will be discussed later in Section V.

Later on in Section IV, we will discuss some interesting performance results for our MHMM filter (5), (6). Next, we introduce some relative entropy concepts that are key to our model selection method.

III. Relative Entropy Rate

Consider two probability measures $\mu$ and $\nu$ on a measurable space $(\Omega, \mathcal{F})$. The relative entropy $D(\mu||\nu)$ of $\mu$ with respect to $\nu$ is defined by [36]

$$D(\mu||\nu) \triangleq \begin{cases} 
\int_{\Omega} \left( \log \frac{d\mu}{d\nu} \right) d\mu, & \text{if } \mu \ll \nu \text{ and } \log \left( \frac{d\mu}{d\nu} \right) \text{ is integrable} \\
+\infty, & \text{otherwise},
\end{cases}$$

(7)

where $(d\mu/d\nu)$ is the Radon-Nikodym derivative of $\mu$ with respect to $\nu$. Here, $\mu \ll \nu$ denotes that $\mu$ is absolutely continuous with respect to $\nu$, in the sense that $\mu = 0$ wherever $\nu = 0$. The relative
entropy $D(\mu \| \nu)$ provides a pseudo-distance measure between $\mu$ and $\nu$ (not a true distance because it is non-symmetric and does not satisfy the triangle inequality).

When interested in dynamic systems, the relative entropy rate (RER) is often more useful. The RER $\mathcal{R}$ of the measure $\mu$ with respect to $\nu$ is given by

$$\mathcal{R}(\mu \| \nu) \triangleq \lim_{k \to \infty} \frac{1}{k} D(\mu \| \nu).$$

(8)

The usual RER between two models $\lambda$ and $\hat{\lambda}$, $\mathcal{R}(\lambda \| \hat{\lambda})$, is defined through the probability laws describing the output processes of the models (See Remark at end of section). In this paper, we are also interested in the RER between the joint probability laws describing the joint output and state processes. To make absolutely clear this distinction between the RER defined through the output probability laws and the RER defined through the joint output-state probability laws, we introduce an overbar notation such that the joint RER is denote by $\overline{\mathcal{R}}(\lambda \| \hat{\lambda})$ [27], [28], [37]. For example, if $x_{[0,\infty]}$, $y_{[0,\infty]}$ and $\hat{x}_{[0,\infty]}$, $\hat{y}_{[0,\infty]}$ denote state and output processes generated by two models $\lambda$ and $\hat{\lambda}$, respectively, then the standard RER is given by $\mathcal{R}(\lambda \| \hat{\lambda}) = \mathcal{R}(p^\lambda(y_{[0,\infty]}) \| p^\lambda(\hat{y}_{[0,\infty]}))$, whereas the joint RER is given by $\overline{\mathcal{R}}(\lambda \| \hat{\lambda}) = \overline{\mathcal{R}}(p^\lambda(x_{[0,\infty]}; y_{[0,\infty]}) \| p^\lambda(\hat{x}_{[0,\infty]}; \hat{y}_{[0,\infty]}))$, where $p^\lambda$ and $p^\lambda$ are the probability laws corresponding to the two models.

Corollary 2 and Theorem 3 of [27] show that

$$\overline{\mathcal{R}}(\lambda \| \hat{\lambda}) \geq \lim_{k \to \infty} \frac{1}{k} \log \left( \frac{\sum_{i=1}^{N} \alpha_i^\lambda_{k|h}}{\sum_{i=1}^{N} \alpha_i^\hat{\lambda}_{k|h}} \right).$$

This illustrates that the joint RER between two HMMs with common $B(\hat{y}_{h|h})$ overbounds the ratio of expected likelihoods. We now introduce a fairly intuitive but important result.

**Lemma 3.1:** The joint RER upper bounds the standard RER in the sense that

$$\overline{\mathcal{R}}(\lambda \| \hat{\lambda}) \geq \mathcal{R}(\lambda \| \hat{\lambda}).$$

**Proof:** The chain rule of relative entropy between two probability functions $p$ and $q$ states [36, Ch. 2] $D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x))$ for two random variables $x$ and $y$. Note $D(p(y|x)||q(y|x)) = \int_x p(x) \int_y p(y|x) \log \frac{p(y|x)}{q(y|x)} \, dy \, dx$ is a conditional relative entropy. Here, because relative entropies and conditional relative entropies are non-negative, that is, $D(\|, \) \geq 0$, we obtain that $D(p(x,y)||q(x,y)) \geq D(p(x)||q(x))$. Using the definition of RER gives the equivalent result $\mathcal{R}(p(x,y)||q(x,y)) \geq \mathcal{R}(p(x)||q(x))$. Using our joint RER notation gives the Lemma statement.

Lemma 3.1 establishes that models close in a joint RER sense must be close in an standard RER sense. However, an important realization in the context of this paper is that the opposite does not hold. That is, it is possible for models to be close in a standard RER sense, but have unbounded joint RER.
Lemma 3.2: Consider a Markov state process \( x_{\theta} \) that is ergodic, a measurement process \( y_{\theta} \) described by (2), and assume that \( \mathcal{D}(p(x_0, y_0) \parallel q(x_0, y_0)) \) is finite. The joint RER is related to the conditional relative entropy between a posteriori densities through the expression

\[
\mathcal{R} \left( p(x_{[0, \infty]}, y_{[0, \infty]}), q(x_{[0, \infty]}, y_{[0, \infty]}) \right) = E \left[ \mathcal{D} \left( p(x_k | x_{k-1}, y_k) \parallel q(x_k | x_{k-1}, y_k) \right) \right] + E \left[ \mathcal{K}_k \right] \quad \text{for } k \geq 1
\]

where \( \mathcal{K}_k = \mathcal{D}(p(y_k | x_{k-1}) \parallel q(y_k | x_{k-1})) \geq 0 \).

Proof:

The chain rule of relative entropy [36, Ch. 2] can then be applied to \( \mathcal{D}(p(x_{[0, \infty]}, y_{[0, \infty]}), q(x_{[0, \infty]}, y_{[0, \infty]})) \) to give

\[
\mathcal{D} \left( p(x_{[0, \infty]}, y_{[0, \infty]}), q(x_{[0, \infty]}, y_{[0, \infty]}) \right) = \mathcal{D} \left( p(x_{[0, k-1]}, y_{[0, k-1]}), q(x_{[0, k-1]}, y_{[0, k-1]}) \right) \\
+ E \left[ \mathcal{D} \left( p(y_k | x_{[0, k-1]}, y_{[0, k-1]}), q(y_k | x_{[0, k-1]}, y_{[0, k-1]}) \right) \right] \\
+ E \left[ \mathcal{D} \left( p(x_k | y_{[0, k-1]}, x_{[0, k-1]}), q(x_k | y_{[0, k-1]}, x_{[0, k-1]}) \right) \right].
\]

Now noting that for Markov processes \( p(x_k | y_{[0, k-1]}, x_{[0, k-1]}) = p(x_k | y_k, x_{k-1}) \) and \( p(y_k | y_{[0, k-1]}, x_{[0, k-1]}) = p(y_k | x_{k-1}) \) we can write

\[
\mathcal{D} \left( p(x_{[0, \infty]}, y_{[0, \infty]}), q(x_{[0, \infty]}, y_{[0, \infty]}) \right) = \mathcal{D}(p(x_0, y_0) \parallel q(x_0, y_0)) \\
+ E \left[ \sum_{\ell=1}^{k} \mathcal{D}(p(x_{\ell} | y_{\ell}, x_{\ell-1}), q(x_{\ell} | y_{\ell}, x_{\ell-1})) \right] + \mathcal{K}_k
\]

The lemma result follows from ergodicity of the processes involved, and from the definition of RER.

The above lemma illustrates that the joint RER is related to an important conditional relative entropy between a posteriori densities. In the next section, this result is used to propose a model selection approach for our HMM collections and to establish interesting filtering performance results.

Remark: We highlight that we interpret the joint RER between model \( \lambda^0(\theta) \) and the HMM \( \lambda_h \) in terms of the joint probability law generating the processes \( x_k \) and \( y_k \) and the joint probability law generating the approximating processes \( x_{\theta} \) and \( y_{\theta} \). Here our assumptions that 1) the measurement laws are absolutely continuous, 2) that \( x_k \in S_x \), and 3) that the \( h \)-sized \( n \)-dimensional boxes \( C_h(.) \) cover \( S_x \); together ensure that the joint RER \( \mathcal{R}(\lambda^0(\theta) \parallel \lambda_h) \) yields a finite value (note that \( \mathcal{R}(\lambda_h \parallel \lambda^0(\theta)) \) may not be finite).
IV. MHMM FILTERING AND MODEL SELECTION FOR UNCERTAIN DYNAMICS

Multiple model filter approaches have been used since the mid 1960s for systems with uncertain or varying dynamics, see [31] for a survey. These multi-model filter approaches have handled model uncertainty by seeking a set of models that provides good approximation of the system’s behaviour at all times.

In this section we establish an interesting connection between the quantities estimated by our proposed MHMM filter (5), (6) and the conditional mean estimate (CME) based on the true model (1), (2). Furthermore, we introduce the problem of selecting suitable HMM collections for use in the MHMM filter. We then show some filtering performance results.

A. MHMM Filter and Conditional Mean Estimate

We begin our introduction of the model selection problem by noting that members of our HMM collection \( M_{q|h} \) will be selected on the basis of the joint RER between HMM members and the nominal model; that is, we use \( \bar{R} \left( \lambda^0(\theta) \| \lambda_{i|h} \right) \) as our criterion for model selection, where \( \lambda_{i|h} \gg \lambda^0(\theta) \).

Note that the RER is non-symmetric, and it will later become apparent that the above order of RER arguments is required so that we can handle uncertainty in \( \theta \). We now establish some basic properties of our proposed MHMM filter. Let \( i^0 \) denote the closest model in the sense that

\[
\bar{R} \left( \lambda^0(\theta) \| \lambda_{i^0|h} \right) = \min_{i \in \{1, q\}} \bar{R} \left( \lambda^0(\theta) \| \lambda_{i|h} \right).
\]

**Definition 4.1:** Let us define an approximating state process for our HMM collection \( M_{q|h} \) as \( x_{k}^M \in C_h(X_{k|h,i}^{0}) \) for each \( k \geq 0 \), where \( x_{k}^M \) has uniform distribution within \( C_h(X_{k|h,i}^{0}) \). That is, \( x_{k}^M \) is a uniformly “blurred” version of the \( i^0 \) model with the property that

\[
E \left[ x_{M}^k \mid y_{[0,k]} \right] = G_h \left( X_{k|h,i}^{0} \right).
\]

**Lemma 4.1:** For sufficiently large \( k \), the MHMM filter estimate is a conditional mean estimate (CME) of the defined process \( x_{k}^M \) in the sense that

\[
\hat{x}_{k|h} = E \left[ x_{k}^M \mid y_{[0,k]} \right] \ a.s.. \]

**Proof:** From [38], [39] it can be shown that likelihood of the HMM model closest to \( \lambda^0(\theta^0) \) (in a joint relative entropy sense) will almost surely dominate the other models in a joint likelihood sense, for large \( k \), so that \( i^* = i^0 \) and

\[
\hat{x}_{k|h} = \sum_{j=1}^{N_h} \hat{X}_{j|h,i}^{0} G_h(e_j) \ a.s.. \]

Then, we note from the tower property of conditional probabilities [16, Appendix A] that the CME can be written as

\[
E \left[ x_{k}^M \mid y_{[0,k]} \right] = E \left[ E \left[ x_{k}^M \mid y_{[0,k]}, X_{k|h,i}^{0} \right] \mid y_{[0,k]} \right] \ a.s.. \]

(9)
Further, from our measurement model assumptions we note that

\[ E \left[ x_k^M \mid y_{[0,k]}, X_{k|h,i} \right] = E \left[ x_k^M \mid X_{k|h,i} \right] = G_h \left( X_{k|h,i} \right), \]

where the 2nd line follows from our “blurring” condition \( E \left[ x_k^M \mid X_{k|h,i} \right] = G_h \left( X_{k|h,i} \right) \). This result substituted into (9) then allows us to write

\[ E \left[ x_k^M \mid y_{[0,k]} \right] = E \left[ G_h(X_{k|h,i}) \mid y_{[0,k]} \right]. \]

The lemma result then follows by noting, for large enough \( k \), that

\[ E \left[ G_h(X_{k|h,i}) \mid y_{[0,k]} \right] = \sum_{j=1}^{N_h} G_h(e_j) E \left[ X_{k|h,i}^j \mid y_{[0,k]} \right] = \hat{x}_{k|h} \text{ a.s.} \]

We highlight that the above result is not dependent on any specific choice of \( M_{q|h} \). Next, we propose a joint RER based approach for selecting HMM collections that provide good approximations for our nominal system, and then show some interesting performance results.

**B. Min-Max RER HMM Collection Design**

As a stepping stone, we first consider the case of a single HMM design when \( \theta^0 \) is known, and let \( S_{\lambda,h} \) denote the set of all HMMs of size \( N_h \). We seek to approximate the dynamics \( \lambda^0(\theta^0) \) by a \( \lambda_h^* \in S_{\lambda,h} \) which is the HMM that minimizes the joint RER with respect to the dynamics \( \lambda^0(\theta^0) \), in the sense that

\[ \overline{R} \left( \lambda^0(\theta^0) \parallel \lambda_h^* \right) = \min_{\lambda_h \in S_{\lambda,h}} \overline{R} \left( \lambda^0(\theta^0) \parallel \lambda_h \right). \]  

(10)

From [27] and the result of [38], [39], we see that a small RER implies a large likelihood. Hence, we would expect \( \lambda_h^* \) to provide reasonable filtering performance.

We now extend the above approach to the case when \( \theta^0 \) is unknown. This uncertainty can be handled by using our proposed HMM collection representation. We propose a HMM collection \( M_{q|h} \) be chosen so that, for every value of \( \theta \in S_{\theta^0} \), \( \lambda^0(\theta) \) is close in a joint RER sense to at least one model \( \lambda_{i|h} \) in the collection. That is, our model selection strategy is to choose a collection \( M_{q|h} \) so that for each \( \theta \in S_{\theta^0} \), \( \overline{R} \left( \lambda^0(\theta) \parallel \lambda_{i|h} \right) \) is small for at least one choice of \( i \). Hence, a natural design measure is the max-RER cost

\[ J^W \left( M_{q|h} \right) = \max_{\theta \in S_{\theta^0}} \left[ \min_{i \in [1,q]} \overline{R} \left( \lambda^0(\theta) \parallel \lambda_{i|h} \right) \right]. \]  

(11)
We highlight that the order of RER arguments in (11) is important to reflect the motivation for our model selection approach.

Now, our selection problem to find a suitable MHMM design can be posed as the min-max problem of finding a candidate $M_{q|h}^W \in S_{M_{q|h}}$ such that

$$J^W(M_{q|h}^W) = \min_{M_{k:k} \in S_{M_{k:k}}} J^W(M_{q|h}).$$

(12)

We now let $p^0(\cdot)$ and $p^M(\cdot)$ denote laws corresponding to models $\lambda^0(\theta^0)$ and $M_q^W$, respectively. We let $\hat{x}_{k|k-1} = E^0[x_k|x_{k-1}, y_k]$ and $\hat{x}_{k|k-1,h} = E^M[x_k|x_{k-1}, y_k]$ denote the one-step ahead CMEs (estimate of current state value given last state value and current measurement), respectively under each model.

We now present an interesting result about MHMM filter performance.

**Theorem 4.1:** Consider the state process $x_{[0,k]}$ and the measurement process $y_{[0,k]}$. Then the min-max design cost $J^W(M_{q|h}^W)$ provides a bound on one-step-ahead CME performance, in the sense that

$$J^W(M_{q|h}^W) \geq B E \left[ \left| \hat{x}_{k|k-1} - \hat{x}_{k|k-1,h} \right|^2 \right]$$

where $B > 0$ is a finite constant. Moreover, if we additionally assume $x_k$ is a strictly stationary process, initialised with its stationary distribution and that we have consistent initialization $\hat{x}_{0|h} = \hat{x}_0^0$, then $J^W(M_{q|h}^W) = 0$ implies that $\hat{x}_{k|h} = \hat{x}_k^0$ a.s..

**Proof:** We first note that

$$\hat{x}_{k|k-1} - \hat{x}_{k|k-1,h} = \int_x \left( p^0(x|x_{k-1}, y_k) - p^M(x|x_{k-1}, y_k) \right) x \ dx$$

Taking the expectation of the magnitude squared we obtain

$$E \left[ \left| \hat{x}_{k|k-1} - \hat{x}_{k|k-1,h} \right|^2 \right]$$

$$= E \left[ \int_x \left( p^0(x|x_{k-1}, y_k) - p^M(x|x_{k-1}, y_k) \right) x \ dx \right]^2$$

$$\leq E \left[ \int_x \left( p^0(x|x_{k-1}, y_k) - p^M(x|x_{k-1}, y_k) \right) dx \right]^2$$

$$= E \left[ \left( \int_x \left| p^0(x|x_{k-1}, y_k) - p^M(x|x_{k-1}, y_k) \right| dx \right)^2 \right]$$

$$\leq \left| x_{max} \right|^2 E \left[ \left( \int_x \left| p^0(x|x_{k-1}, y_k) - p^M(x|x_{k-1}, y_k) \right| dx \right)^2 \right]$$

$$\leq \left| x_{max} \right|^2 B_1 E \left[ D \left( p^0(x|x_{k-1}, y_k) \left\| p^M(x|x_{k-1}, y_k) \right) \right] \right.$$

for some positive constant $B_1$. 


In the 4th line we have used \( x_k \in S_x \) to imply the existence of a finite upper bound \( x_{\text{max}} \), and in the 5th line we have used [36, Lemma 11.6.1] which shows the relative entropy bounds the \( L_1 \) norm of the difference between laws. Then noting that \( \mathcal{R} \left( \lambda^0(\theta^0) \| \lambda_{i|h} \right) = \min_{i \in [1,q]} \mathcal{R} \left( \lambda^0(\theta^0) \| \lambda_{i|h} \right) \) from our definition of \( x_k^M \) lets us apply Lemma 3.2 to give the first part of the theorem result.

The second part of the theorem then follows from noting that if \( J^W(M_{q|h}) = 0 \) can be achieved under any choice of \( q \), then using that relative entropies are non-negative we obtain that \( p^0(x_k|x_{k-1}, y_k) = p^M(x_k|x_{k-1}, y_k) \) a.s. and \( p^0(y_k|x_{k-1}) = p^M(y_k|x_{k-1}) \) a.s.. Recall that Bayes’ rule lets us write
\[
p(x_k, x_{k-1}, y_k) = p(x_k|x_{k-1}, y_k)p(y_k|x_{k-1})p(x_{k-1}).
\]
Then, using marginalization integrals applied to \( p(x_k, x_{k-1}, y_k) \) over \( y_k \), followed by Bayes’ rule, and stationary process properties, we can show that \( p^0(x_k|x_{k-1}) = p^M(x_k|x_{k-1}) \) a.s.. Similarly, we can show that \( p^0(y_k|x_k) = p^M(y_k|x_k) \) a.s.. We highlight that the irreducible and aperiodic Markov chains used in \( M_{q|h} \) define stationary processes [40]. Finally, using \( \hat{x}_{0|h} = \hat{x}^0_{0|h} \) with an inductive argument on the conditional mean estimate forward recursion [30, Ch. 10]
\[
p(x_k|y_{[0,k]}) = B_k p(y_k|x_k) \int p(x_k|x_{k-1})p(x_{k-1}|y_{[0,k-1]}) dx_{k-1}
\]
where \( B_k \) is a normalization constant, gives the second result.

When combined with exponential forgetting properties of HMM filters, Theorem 4.1 illustrates that our min-max MHMM design process based on joint RER is useful because it bounds the estimation error induced by the modeling approximation. That is, the above HMM collection design process can be used to develop MHMM filters offering reasonable performance.

C. Conditional RER HMM Collection Design

Assume \( \theta \in S_\theta \) is an unknown model parameter that is \textit{a priori} described by the probability \( P(\theta) \).

Let us define the conditional joint RER (a natural extension of the conditional relative entropy defined in [36, Ch. 2]) as
\[
\mathcal{R}^C \left( \lambda \| \lambda^\hat{\lambda} \right) = \int_{\theta \in S_\theta} P(\theta) \mathcal{R} \left( \lambda(\theta) \| \lambda^\hat{\lambda} \right) d\theta
\]
where \( \mathcal{R} \left( \lambda(\theta) \| \lambda^\hat{\lambda} \right) \) is the RER under the assumption \( \theta \) is known with certainty.

We can then introduce the conditional RER criteria for choosing a collection \( M_{q|h} \in S_{M_{q|h}} \),
\[
J^C \left( M_{q|h} \right) = \min_{i \in [1,q]} \mathcal{R}^C \left( \lambda^0 \| \lambda_{i|h} \right)
\]
\[
= \int_{\theta \in S_\theta} P(\theta) \min_{i \in [1,q]} \mathcal{R} \left( \lambda^0(\theta) \| \lambda_{i|h} \right) d\theta.
\]
The candidate $M_{q|h}^C$ will be a conditional RER-MHMM model if
\[
J^C(M_{q|h}^C) = \min_{M_{q|h} \in S_{M_{q|h}}} J^C(M_{q|h}) \tag{13}
\]

D. Monotonicity of RER HMM Collections

We now establish a result that shows that min-max or conditional joint RER MHMM collections are sensible in another sense. Let
\[
J^{U,i}_h(q) = \min_{M_{q|h} \in S_{M_{q|h}}} J^i(M_{q|h})
\]
where $i \in \{W,C\}$ denotes the either the min-max RER design criteria or the conditional RER design criteria, from a $q$ sized collection of HMMs.

**Theorem 4.2:** The RER design criteria $J^{U,i}_h(q)$ is monotonic in the number of models $q$, in the sense that
\[
J^{U,i}_h(q) \leq J^{U,i}_h(q - 1)
\]
for all $q$ and both $i \in \{W, C\}$.

**Proof:** Note that we can write $M_{q|h} = \{M_{q-1|h}, \lambda_{q|h}\}$. First consider the case $i = W$, and let us define a bad model $\lambda^B$ as
\[
\max_{\theta \in S_{\theta}} \left[ \bar{R}(\lambda^0(\theta) \| \lambda^B) \right] = \max_{\lambda} \max_{\theta \in S_{\theta}} \left[ \bar{R}(\lambda^0(\theta) \| \lambda) \right].
\]
We can then write
\[
J^{U,W}_h(q) = \min_{M_{q|h} \in S_{M_{q|h}}} J^W(M_{q|h})
\]
where in the 4th line we note that $\lambda^B$ is never the minimising argument. A similar argument proves the $i = C$ case.

V. IMPLEMENTATION ISSUES

This section discusses two implementation issues.
A. RER Approximation

The HMM collection design approach described in the previous section is reliant on being able to evaluate \( \mathcal{R} \left( \lambda^0(\theta) \| \lambda_{ij|h} \right) \) for a range of values \( \theta \in S_\theta \) and for different candidate models \( \lambda_{ij|h} \). Yet, a suitable close-form expression for the joint RER between a locally consistent chain \( \lambda \) assume a common measurement model \( B \) transitions, roughly in the sense that the model Markov chain, using the results of [1]. This involves selecting a Markov chain so that the chain and filter recursion (5). This is the strategy we use in our simulation example.

Here we propose to approximate the model \( \lambda^0(\theta) \), for specific values of \( \theta \), with a locally consistent Markov chain, using the results of [1]. This involves selecting a Markov chain so that the chain and the model \( \lambda^0(\theta) \) have consistent local statistical properties (including equivalent expected mean state transitions, for all \( x \in S_x \).

Let \( \lambda^{LC,\theta} \) denote a chain that is “locally consistent” to the state component of the model \( \lambda^0(\theta) \). If we assume a common measurement model \( B \left( \eta_{ij|h} \right) \), we can then take advantage of an existing result from [27] in which the joint RER between a locally consistent chain \( \lambda^{LC} = (A,B \left( \eta_{ij|h} \right) ,\pi) \) and another HMM \( \lambda = (A,B \left( \eta_{ij|h} \right),\pi) \), is given by

\[
\mathcal{R} \left( \lambda^{LC} \| \lambda \right) = \sum_{r=1}^{N} \sum_{s=1}^{N} \left[ \pi_r A^{sr} \log \left( \frac{A^{sr}}{A^{sr(\theta)}} \right) \right],
\]

where \( N \) is the size of the HMMs. Thus, we propose to approximate the joint RER \( \mathcal{R} \left( \lambda^0(\theta) \| \lambda_{ij|h} \right) \) as

\[
\mathcal{R} \left( \lambda^0(\theta) \| \lambda_{ij|h} \right) \approx \sum_{r=1}^{N} \sum_{s=1}^{N} \left[ \pi_r(\theta) A^{sr(\theta)} \log \left( \frac{A^{sr(\theta)}}{A^{sr|h}} \right) \right],
\]

where \( A(\theta) \) and \( \pi(\theta) \) are the \( \theta \)-dependent parameters of the locally consistent approximation of \( \lambda^{LC,\theta} \).

B. Selection of \( i^* \)

In practice, the joint likelihoods \( W^i_k \) are not available, but \( i^* \) can be determined using any strategy that ensures selection of the model closest in joint RER sense. For example, let us introduce \( L^i_k = p^{\lambda^{LC,\theta}} (\eta_{ij|h}) = \left( \prod_{r=1}^{k} N_{ij}^{r} \right)^{-1} \) which denotes the \( i \)th likelihood (in the usual output-sense) at time \( k \), and let \( j^* \) denote the model with the highest likelihood (selected so that \( L^i_k \geq L^j_k \) for all \( j \)). Then one strategy for selecting \( i^* \) is based on only allowing HMM collections \( \mathcal{M}_{q|h} \) that have \( j^* = i^* \) a.s.. If \( \mathcal{M}_{q|h} \) is constrained this way then \( j^* \) can be calculated (and hence \( i^* \)) from the \( N^i_k \) determined in the MHMM filter recursion (5). This is the strategy we use in our simulation example.
VI. APPLICATION: DIM TARGET DETECTION PROBLEM IN IMAGE SEQUENCES

Detection and tracking of unknown target trajectories is an estimation problem where the possible presence of an underlying state process (target) must be detected from an observation sequence.

We will restrict our attention to detection of (approximately) constant velocity target trajectories by considering the following time-invariant model

\[
x_k = x_{k-1} + \begin{bmatrix} v_x \\ v_y \end{bmatrix},
\]

(16)

where \(x_k = [d_x^k, d_y^k]'\) denotes the target’s pixel co-ordinates in the image plane and \((v_x, v_y)\) denotes the target’s velocity in pixels/frame. The target is assumed to be observed through noisy imaging measurements from an electro-optical sensor, whose field of view is represented by a 2D grid of image pixel locations \(G = \{(i, j) | 1 \leq i \leq N^u, 1 \leq j \leq N^v\}\). We let \(y_{ih}^k\) be the \(N^u \times N^v\) pixel image measurement at the \(k\)th sample instant. Moreover, we let \(y_{ih}^k\) denote the value at the \(i\)th pixel, under some enumeration scheme of pixel locations.

We highlight that the stated dynamics (16) do not exhibit the bounded state space requirement of (1), nor satisfy the ergodicity properties that are required to establish Theorem 4.1. However, the bounded, stationary, and ergodic state properties are satisfied by the following noisy model with projection

\[
x_k = \text{Proj}_G \left( x_{k-1} + \begin{bmatrix} v_x \\ v_y \end{bmatrix} + w_k \right),
\]

(17)

where \(w_k\) is a \(N(0, \sigma_w^2)\) zero-mean Gaussian noise process (\(\sigma_w^2 > 0\) is assumed small) and \(\text{Proj}_G\) is an operation that projects any target crossing the boundary of \(G\) back inside \(G\). In our study, we use (17) as a model for (16), and allow \(\text{Proj}_G\) to project the target to the opposite side of \(G\) (and still travelling in the same direction) whenever the target encounters a boundary. Note that this projection choice does not have a significant impact on observed filtering performance in our application, because we are not concerned about scenarios that involve tracking targets near the boundaries of \(G\). Finally, note that our design assumed that targets have approximately constant velocity with small \(\sigma_w^2\), but our simulated data contained targets with perfectly constant velocity i.e. \(\sigma_w^2 = 0\).

We let \(v = \sqrt{(v_x)^2 + (v_y)^2}\) denote the velocity magnitude, let \(\psi = \tan^{-1}(v_y/v_x)\) denote the heading angle, and let \(\theta = (v, \psi)\). Then \(\lambda^0(\theta)\) denotes the target model with velocity \(v\) and heading \(\psi\). The uncertainty in the target dynamics is described by a finite bounded set of possible velocity and heading angles \(S_\theta = \{(v, \psi) | v_{\min} \leq v \leq v_{\max}, \psi_{\min} \leq \psi \leq \psi_{\max}\}\).
A. Basic HMM Representation

We now develop discrete Markov chain models to represent the continuous motion $x_k$ of a target as projected onto a discrete 2D image plane grid $G$ parameterised by a grid spacing parameter $h$. Let $N_h = N^u N^v$ denote the total number of grid locations on the image grid $G$.

We now introduce the following assumptions:

1) Individual pixels do not allow the opportunity of perfect detection, in the sense that $p^\lambda_h(y^i_k|X_k|h \neq e_i) > 0$ whenever $p^\lambda_h(y^i_k|X_k|h = e_i) > 0$; and

2) The statistical properties of pixel values within an image are spatially independent, in the sense that

$$p^\lambda_h(y^r_k,y^s_k|X_k|h = e_i) = p^\lambda_h(y^r_k|X_k|h = e_i) p^\lambda_h(y^s_k|X_k|h = e_i)$$

for all $r, s$ and $i$.

Under these assumptions, elements of $B(y_k|h)$ can be determined from the simplification

$$p^\lambda_h(y^i_k|X_k|h = e_i) = c \left[ \frac{p^\lambda_h(y^i_k|X_k|h = e_i)}{p^\lambda_h(y^i_k|X_k|h \neq e_i)} \right]$$

where $c$ is a proportionality constant (see [29] for a justification of this simplification).

We highlight that the calculation of the measurement probabilities is greatly simplified under the above assumptions because $p^\lambda_h(y^i_k|X_k|h = e_i)$ and $p^\lambda_h(y^i_k|X_k|h \neq e_i)$ can each be determined on a single-pixel basis, rather than requiring the probability of a whole image. In this study, we assume that the measurement model is known, and we determine $p^\lambda_h(y|x) = p^0(y|x)$ empirically from given observation data. Therefore, constructing a suitable HMM representation is simplified to the problem of determining suitable transition probability matrices.

B. Simulation Studies

In the following sections, we will apply the joint RER MHMM design strategy to the problem of detecting dim targets traveling with unknown heading $\psi$ from morphologically pre-processed image sequences [41]. We will compare min-max RER MHMM designs with the ad hoc MHMM designs we proposed in previous work [29].
1) Transition Probability Matrix Design: In the following simulation studies, we will consider slow moving targets traversing across an image plane. We simplify our transition probability design problem by exploiting some a priori information about the dynamics under approximation. Firstly, the Proj\(_G\) operation dictates certain transition behaviour from chain states at the boundary of \(S_x\). Secondly, at chain values inside the boundaries, we can assume transitions occur only to neighbouring locations or are self-transitions. Hence, there is a maximum of 9 non-zero transition probabilities from any internal state. Moreover, we can assume behaviour is uniform in the sense that these neighbourhood transition properties do not depend on location in the grid. This allows us to assume an overall sparse matrix structure and simplifies our search space to the 9 probabilities describing the neighbourhood transition properties. We highlight that an irreducible and aperiodic \(A_h\) can be achieved within this sparse matrix structure.

2) Optimisation: The computational aspects of optimisation with respect to \(A_{q|h}\in S_{A_{q|h}}\) and \(\theta \in S_\theta\) are beyond the scope of this paper; however it is likely that existing standard optimisation routines can be applied should an efficient numerical solution be sought. In our studies, we preferred to approximate search spaces by finite sets and use “brute-force” optimisation or manual search methods. For example, the maximum operation over \(\theta \in S_\theta\) is approximated by a brute-force maximisation over a finite set approximation of the range of \(S_\theta\).

Results: In the following studies, we simulate targets observed through discrete measurements from imaging sensor of dimension \(111 \times 147\). Measurement noise is modeled as a Gauss Markov random field (GMRF) parameterised by vertical and horizontal interaction factors of 0.12 and driven with \(N(0,1)\) Gaussian noise (see [5], [29]).

The MHMM designs assume uniform a priori state probabilities \(\pi_h\), and will employ a common measurement model \(B(Y_k)\) that is empirically determined from observation data.

In all studies below, target detection is based on a threshold test of the magnitude of \(L^j\), with the target location being estimated as the location of the maximum of \(\hat{X}_{k|h,j}\). This estimated location corresponds to the location of maximum probability in the underlying MHMM filters.

a) Min-Max RER Design of 1-Model MHMM Filter: Before introducing our more sophisticated studies involving multiple HMMs, we first present a simpler special case of the MHMM filter design where only a single HMM is used in the filter. In this example, we consider the scenario where the target moves at a constant velocity \(v = 0.2\) pixels/frame. We compare the detection performance of an ad hoc single-model filter (based on a Markov chain with \(A_{ii} = 7/15\) for self-transitions, and \(A_{ij} = 1/15\) for all adjacent state transitions, as reported in [29]) with that of a min-max RER single-model filter (based
on a Markov chain with $A_{ii} = 0.8$ for self-transitions, $A_{ij} = 4.2 \times 10^{-5}$ for diagonal state transitions and $A_{ij} = 0.05$ for vertically and horizontally adjacent state transitions). The max RER design cost (11) achieved by the two filter designs is shown in Table I.

<table>
<thead>
<tr>
<th>MHMM Design</th>
<th>Max RER Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-model Ad-Hoc</td>
<td>0.0723</td>
</tr>
<tr>
<td>1-model Min-Max</td>
<td>0.0308</td>
</tr>
<tr>
<td>4-Model Ad-Hoc</td>
<td>0.0273</td>
</tr>
<tr>
<td>4-Model Min-Max</td>
<td>0.0154</td>
</tr>
</tbody>
</table>

The two HMM filters are compared on the basis of their detection performance on $5 \times 10^4$ image sequence test cases, where each sequence had a length of 151 frames. The bottom two curves in Fig.1 illustrate the average detection and false-alarm performance of the two single-model MHMM filters (as the detection threshold is varied). The two filter designs can be compared by observing that for any false-alarm rate, the min-max RER based design provided better detection performance. In a rough sense, across a range of false-alarm rates, the RER based design provides 6% improvement in detection performance compare to our base ad hoc design.

b) Min-Max RER Design of 4-Model MHMM Filter: In this example, we consider the performance of a 4-model MHMM filter designed using the min-max RER criterion. We evaluate the 4-model filter in two ways: against the min-max RER-based single-model MHMM filter given in the previous study, and against an ad hoc 4-model MHMM filter examined in [29].

For design purposes, we assumed that the heading angle uncertainty could be divided into four equal quadrants to coincide with the 4 available HMMs in the MHMM design (this also forces a $i^* = j^* \ a.s.$ design constraint). Due to the symmetry in the problem, we then assumed each HMM design will have the same transition probability structure. The manual search for transition probability parameters that minimised our max RER cost (11) yielded a MHMM design with self-transition probabilities of 0.8, a probability of transition to diagonally adjacent states of $9.5 \times 10^{-6}$ (within the quadrant), and a probability of transition to vertically and horizontally adjacent states of 0.1 (within the quadrant). The ad hoc 4-model MHMM reported in [29] has self-transition probabilities of 0.7, and a probability of transition to all other adjacent states (within the quadrant) of 0.1. The max RER design cost (11) achieved by the two
Firstly, we will illustrate the significant advantage of the multiple model approach over the single-model design, using the same simulation parameters described in the previous example. The first and third curves (from the top) of Fig.1 show the detection rate vs. false-alarm rate tradeoff for the 4-model and single-model min-max RER designs, respectively. The multiple model design demonstrates superior performance compared to the single-model design. For example, at a false-alarm rate of $10^{-2}$, the detection rate of the 4-model MHMM filter is nearly 0.99, whereas for the single-model filter it is about 0.86. The superiority of the multiple model approach is also evident for the ad-hoc designs.

The top two curves in Fig.1 show the detection vs. false-alarm performance tradeoffs for the min-max RER and ad hoc based 4-model MHMM designs. The min-max RER design demonstrates a consistent detection rate improvement of between 5% and 6% over the ad hoc design across a range of false-alarm rates.

c) Min-Max RER Design with Variable Velocity: We now consider a problem where the target velocity may take on one of three possible values $v \in [0.1, 0.2, 0.3]$ pixels/frame.

We compare two 4-model MHMM designs: one that explicitly accounts for all the possible variations in
target velocity \((v \in [0.1, 0.2, 0.3] \text{ pixels/frame})\); and one designed to the average target velocity \((v = 0.2 \text{ pixels/frame})\). The min-max RER variable velocity MHMM design has a probability of 0.696 for self-transitions, a probability of 0.0002 for diagonally adjacent state transitions, and finally a probability of 0.1501 for vertically and horizontally adjacent state transitions. The fixed velocity MHMM design has the same transition probabilities as the 4-filter min-max RER design in the previous example.

Table II shows detection rates and corresponding RER costs for the two filter designs as a function of various target velocity cases (for a fixed false-alarm rate of \(10^{-3}\)). The overall performance results given in the final row of the table show that the variable design has a higher detection rate. We also examined filter performance on simulation subsets. Although the fixed velocity MHMM design achieves a slightly better detection rate in two out of the three target velocity cases, the variable velocity MHMM design boasts a superior minimum detection performance across the three velocity cases (although not shown, this trend is also observed across a range of other false-alarm rates). The detection rate for the fixed velocity MHMM drops to as low as 0.8992, whereas the detection rate for the variable velocity MHMM does not fall below 0.9485. Finally, we observe that the RER costs seem to provide a good indication of the relative detection performance of the filters in all velocity cases.

### Table II

<table>
<thead>
<tr>
<th>Target (v)</th>
<th>4-MHMM for (v = 0.2)</th>
<th>4-MHMM for (v \in [0.1, 0.2, 0.3])</th>
</tr>
</thead>
<tbody>
<tr>
<td>(v = 0.1)†</td>
<td>0.9905, 0.0118</td>
<td>0.9639, 0.0207</td>
</tr>
<tr>
<td>(v = 0.2)†</td>
<td>0.9865, 0.0154</td>
<td>0.9768, 0.0183</td>
</tr>
<tr>
<td>(v = 0.3)†</td>
<td>0.8992, 0.0434</td>
<td>0.9485, 0.0231</td>
</tr>
<tr>
<td>Variable†</td>
<td>0.9587, 0.0434</td>
<td>0.9631, 0.0231</td>
</tr>
</tbody>
</table>

† (pixels/frame), \(5 \times 10^4\) simulations each case. ‡ \(1.5 \times 10^5\) simulations.

3) **Computational Effort:** Our algorithms were executed in MATLAB (version 7.1.0.246 (R14) Service Pack 3) on a Pentium IV dual core cpu @ 3GHz with 1 GB of memory. The average cpu time taken (image size 111 \(\times\) 147 pixels) for a 1-model MHMM filter is 0.0234 seconds/iteration, and for a 4-model MHMM filter it is 0.0383 seconds/iteration.
VII. CONCLUSION

This paper developed a new joint relative entropy rate (RER) based approach for multiple HMM (MHMM) approximation of uncertain processes. A MHMM filter was introduced, and a RER-based model selection or design approach was proposed. Performance results were established that bound filter performance and show that the filter converges to the true CME as the RER between the uncertain system and the MHMM is reduced. The model design process and MHMM filter approach were demonstrated on an important image processing dim-target detection problem.

ACKNOWLEDGMENT

Computational resources and services used in this work were provided by the HPC and Research Support Unit, Queensland University of Technology, Brisbane, Australia.

REFERENCES


Chapter 6

Airborne Collision-Detection System (J3)

In this chapter, we investigate several key issues concerning the deployment of our proposed target detection algorithm onboard a UAV platform. In particular, we examine the robustness of our detection algorithm to inherent camera sensor motion, and consider compensation techniques that may be used to mitigate the resultant image ‘jitter’ effects. Moreover, we characterise the detection range of our proposed detection algorithm based on real in-flight recorded image data of collision-course UAV targets. Here, two fixed-wing UAVs were engaged to recreate various collision-course scenarios to capture highly realistic vision (from an onboard camera perspective) of the moments leading up to a collision. Based on this collected data, our proposed detection approach was able to detect targets out to distances ranging from about 400m to 900m. These distances, (with some assumptions about closing speeds and aircraft trajectories) translate to an advanced warning ahead of impact that approaches the 12.5 second response time recommended for human pilots. Finally, we explore the potential of GPU-based computing architectures in enabling real-time execution of the target detection algorithm under the severely limited volume, power, and weight budgets of a
UAV platform. We identify a candidate GPU device suitable for integration onto UAV platforms that can be expected to handle real-time processing of 1024 by 768 pixel image data at a rate of approximately 30Hz. We consider this rate of processing more than adequate for real-time target detection.
Statement of Contribution of Co-Authors

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student’s thesis and its publication on the Australasian Digital Thesis database consistent with any limitations set by publisher requirements.

In the case of this chapter:

**Airborne Vision-based Collision-Detection System**

Status at time of writing:

Submitted 10 March 2010

Details of Authorship and Contributions:

<table>
<thead>
<tr>
<th>Name and Order of Authors</th>
<th>Signatures of Authors</th>
<th>Area of Contribution Regarding Authorship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(a) (i) (a) (ii) (b) (i) (b) (ii) (c)</td>
</tr>
<tr>
<td>1. John Lai</td>
<td>✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>2. Luis Mejias</td>
<td>✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
<tr>
<td>3. Jason J. Ford</td>
<td>✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
</tr>
</tbody>
</table>

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

____________________  __________________  __________________
Name                        Signature                      Date

*See Appendix A for definition of authorship and criteria for contribution to publication*
Airborne Vision-based Collision-Detection System

John Lai∗
Australian Research Centre for Aerospace Automation (ARCAA)
Queensland University of Technology
GPO Box 2434, Brisbane Queensland 4001
js.lai@qut.edu.au

Luis Mejias
Australian Research Centre for Aerospace Automation (ARCAA)
Queensland University of Technology
GPO Box 2434, Brisbane Queensland 4001
luis.mejias@qut.edu.au

Jason J. Ford
Australian Research Centre for Aerospace Automation (ARCAA)
Queensland University of Technology
GPO Box 2434, Brisbane Queensland 4001
j2.ford@qut.edu.au

Abstract

Machine vision represents a particularly attractive solution for sensing and detecting potential collision-course targets due to the relatively low cost, size, weight, and power requirements of vision sensors (as opposed to radar and TCAS). This paper describes the development and evaluation of a real-time vision-based collision detection system suitable for fixed-wing aerial robotics. Using two fixed-wing UAVs to recreate various collision-course scenarios, we were able to capture highly realistic vision (from an onboard camera perspective) of the moments leading up to a collision. This type of image data is extremely scarce and was invaluable in evaluating the detection performance of two candidate target

∗Corresponding author; Tel.: +61 7 3138 9808; Fax: +61 7 3138 1516.
detection approaches. Based on the collected data, our detection approaches were able to detect targets at distances ranging from 400m to about 900m. These distances (with some assumptions about closing speeds and aircraft trajectories) translate to an advanced warning of between 8-10 seconds ahead of impact, which approaches the 12.5 second response time recommended for human pilots. We overcame the challenge of achieving real-time computational speeds by exploiting the parallel processing architectures of graphics processing units found on commercially-off-the-shelf graphics devices. Our chosen GPU device suitable for integration onto UAV platforms can be expected to handle real-time processing of 1024 by 768 pixel image frames at a rate of approximately 30Hz. Integration of our system onto a UAV platform is currently underway and flight trials where all processing is performed onboard will be conducted in the near future.

1 Introduction

The problem of unmanned aerial vehicle (UAV) collision avoidance or ‘sense-and-avoid’ has been identified as one of the most significant challenges facing the integration of UAVs into the national airspace (Unmanned Aircraft Systems Roadmap 2007–2032, 2007; DeGarmo, 2004). The full potential of UAVs can never be realised unless the sense-and-avoid issue is adequately addressed. To this end, machine vision has emerged as a promising means of addressing the ‘sense’ and ‘detect’ aspects of collision avoidance, which demands the ability to automatically detect and track targets in naturally lit, high noise environments.

Machine vision represents a particularly attractive solution for sensing and detecting potential collision-course targets due to the relatively low cost, size, weight, and power requirements of the sensors involved (Maroney et al., 2007). Man-in-the-loop (MITL) solutions have been prototyped and demonstrated (Bryner, 2006). Furthermore, an automated detection system developed by Defense Research Associates was implemented and flight tested, where the detection approach is based on target pixel movement (or energy flow) in the image frame (Utt et al., 2005; Utt et al., 2004).

There are many hurdles that must be overcome in the use of machine vision for target detection and tracking. We need to contend with not only the inherent noise of imaging sensors, but also with noise introduced by changing and unpredictable ambient conditions. The desire to overcome these challenges has driven the development of specialized image filtering and processing techniques that are optimized for the type of dim-target characteristics that are experienced in near collision events between fixed-wing aircraft or fixed-wing
Over the last three decades, a two-stage processing paradigm has emerged for the simultaneous detection and tracking of dim, sub-pixel sized targets (Gandhi et al., 2006; Gandhi et al., 2003; Arnold et al., 1993; Barniv, 1985). These two stages are: 1) an image pre-processing stage that, within each frame, highlights potential targets with attributes of interest; and 2) a subsequent temporal filtering stage that exploits target dynamics across frames. The latter temporal filtering stage is often based on a track-before-detect processing concept where target information is collected and collated over a period of time before the detection decision is made.

Generally, the goal of the image pre-processing stage is to enhance potential target features whilst suppressing background noise and clutter. There is an abundance of techniques and algorithms available which may be considered for this image processing role. In particular, non-linear spatial techniques such as median subtraction filters (Deshpande et al., 1999) have been widely discussed in the literature. Another non-linear image filtering approach that has received much attention over the last decade has its basis in mathematical morphology (Dougherty and Lotufo, 2003). Numerous morphology-based filters have been proposed for the detection of small targets in infrared (IR) images (Zhu et al., 2000; JiCheng et al., 1996; Tom et al., 1993). Specific implementations of the morphological filtering approach include the Hit-or-Miss filter (Schaefer and Casasent, 1995), Close-Minus-Open filter (Casasent and Ye, 1997), and the Top-Hat filter (Braga-Neto et al., 2004). Although a large proportion of research has focused on IR images, there are recent examples of morphological filters being incorporated into target detection algorithms operating on visual spectrum images (Carnie et al., 2006; Gandhi et al., 2006; Gandhi et al., 2003). Moreover, a sign of the increasing popularity of morphological filters for small target detection is evident in the host of studies undertaken into the issue of parameter design (Zeng et al., 2006; Yu et al., 2003). Finally, there have been efforts made to compare existing techniques with the morphology-based filters (Gandhi et al., 2006; Warren, 2002; Tom et al., 1993; Barnett et al., 1993), with the median filtering technique often featuring in the comparison studies.

On the other hand, the temporal filtering stage that follows the image pre-processing is designed to extract image features that possess target-like temporal behaviour. For this role, there are two particular filtering approaches that have received much attention in the literature: Viterbi based approaches and Bayesian based approaches. The Viterbi algorithm has formed the basis of the temporal filtering stage in numerous track-before-detect algorithms (Davey et al., 2008; Gandhi et al., 2006; Tonissen and Evans, 1996; Arnold et al., 1993; Barniv, 1985). The popularity of the Viterbi algorithm is in part due to its utility in the context
of tracking where, under a number of assumptions, it is able to efficiently determine the optimal target track within a data sequence (Forney, 1973). Some analysis of the Viterbi algorithm’s detection and tracking performance have been conducted (Johnston and Krishnamurthy, 2000; Tonissen and Evans, 1996; Barniv and Kella, 1987), and modifications that enhance the algorithm’s tracking performance in the presence of non-Gaussian clutter noise have been proposed (Arnold et al., 1993). An alternative temporal filter design for track-before-detect algorithms is based on Bayesian filtering (Davey et al., 2008; Bruno, 2004; Bruno and Moura, 2001; Bruno and Moura, 1999). Advances in this filtering approach have considered the relaxation of typical white Gaussian noise assumptions and spatially correlated clutter (Bruno and Moura, 1999). Moreover, the modeling of clutter has been expanded to encompass a variety of Gaussian and non-Gaussian, correlated and uncorrelated clutter types, and the Bayesian algorithm is extended to accommodate multiple targets that may feature randomly varying amplitudes or intensities (Bruno and Moura, 2001). Finally, some comparison between the Viterbi and Bayesian approaches has been made at the theoretical level (Bruno and Moura, 2001), as well as on the practical level via Monte Carlo simulation trials (Davey et al., 2008).

A survey of potential technologies for unmanned aerial vehicle (UAV) sense-and-avoid concluded that the visual/pixel based technology offered the best chances for regulator approval (Karhoff et al., 2006). It is interesting to note that some studies have shown that it is actually difficult to detect and avoid a collision using the human visual system (Limitations of the See-and-Avoid Principle, 1991). To date, public domain hardware implementation of vision-based sense-and-avoid systems have been limited to a small number. Arguably, the most significant developments have been made by (Utt et al., 2005), where a combination of field programmable gate array chips and microprocessors using multiple sensors were tested in a twin-engine Aero Comander aircraft. A challenge that faces any vision-based sense-and-avoid system is the requirement of real-time operation. Motivated by this fact, we exploit the capabilities of data-parallel arithmetic architectures such as Graphics Processing Units (GPUs), which have proven to be very capable parallel processing devices that can outperform current CPUs by up to an order of magnitude (Owens et al., 2005).

The key contributions of this paper are: 1) demonstration of the coordinated flight of fixed-wing UAVs performing collision-course scenarios for collection of suitable test image data; 2) application of HMM and Viterbi-based target detection approaches to a computer vision sense-and-avoid problem 3) analysis of the target detection approaches in terms of detection range and sensitivity to undesired camera motion (image jitter); and 4) implementation of the target detection approaches using GPU-based hardware and the demonstration of real-time detection capabilities. The use of fixed-wing UAVs for data collection presented
unique challenges, particularly in relation to the synchronisation of aircraft flights to ensure a realistic scenario is simulated, and to maximise the target aircraft’s time spent in the camera field of view. Given that image jitter can have a significant impact on performance (Utt et al., 2005), it was also important for us to characterise the effect of varying degrees of jitter on our target detection approaches.

This paper is structured as follows. Section 2 introduces the basic characteristics of collision behaviour that make vision-based collision detection a difficult problem. Section 3 separately introduces the HMM and Viterbi based detection approaches examined in this paper. Section 4 evaluates the performance of the proposed collision detection system in three key areas: 1) the resilience of the detection algorithms to undesired camera motion (image jitter); 2) the target detection capability of the detection algorithms; and 3) the real-time processing capacity of a GPU-based hardware implementation. Section 5 describes some of the lessons learnt and future work planned.

2 Characteristics of Potential Collision Risks

It may be possible to detect collision course objects based on their physical appearance or the dynamics that they exhibit. The physical attributes of collision course objects, such as colour, shape, and size, depend largely on ambient lighting and atmospheric conditions, as well as the distance to the object. Given the complex nature of ambient lighting and atmospheric conditions, coupled with the large variety of aircraft paint schemes, colour is perhaps one of the weaker distinguishing characteristics of collision-course objects. On the other hand, the size and (to a degree) shape of a collision-course object on the image plane of a camera is perhaps the most useful. Given the 12.5 second reaction time recommended for human pilots (FAA Advisory Circular: Pilots’ role in collision avoidance, 1983), a collision-course object must be detected at a distance of over 1 kilometer (assuming a closing velocity of 100 m/s) to avoid a collision. At this distance, aircraft and other objects of similar dimensions may take up an area anywhere from a few pixels to less than one pixel on the image plane. It can be argued that a few pixels cannot really define any sort of ‘shape’, but at least it can be deduced that in the context of early detection, collision-course objects will tend to be small, point-like features, becoming smaller and dimmer the earlier that detection is required. Of all the characteristics of collision-course objects, their somewhat unique dynamics is perhaps the most suitable attribute to exploit. Objects on a collision course appear at the output of a fixed onboard vision sensor as relatively stationary features on the image plane (Limitations of the See-and-Avoid Principle, 1991).

Even though an automated system is likely to require less time to recognise threats and take evasive action, it can be argued that in the interests of safety, it is best to detect targets as early (i.e. as far away) as possible. Depending on camera resolution, field of view etc.
Features that are moving across the image plane do not correspond to collision-course objects.

The typical output from an computer vision sensor is a sequence of images (i.e. a video stream). Detecting collision-course objects is then a matter of searching for objects within the image sequence that possess the above characteristics; that is, dim sub-pixel size targets that are slowly moving in the image frame. These target properties correspond to the two stage detection paradigm that is described in the next section: morphological filtering to detect pin-like targets, and temporal filtering to detect persistent or almost stationary features.

3 Detection Algorithms

The two candidate detection algorithms that will be considered in this paper are designed to process the sensor measurements in two stages: 1) image pre-processing, followed by 2) track-before-detect temporal filtering. Both algorithms will share a common ‘Close-Minus-Open’ morphological image pre-processing stage, but one approach will be coupled to a hidden Markov model temporal filter, whereas the other approach will be coupled to a Viterbi-based temporal filter.

3.1 Morphological Image Pre-Processing

This paper considers an image pre-processing technique that exploits greyscale morphological operations in order to process discrete 2D image data quantised to a finite number of intensity or greyscale levels, such as might be expected from the output of an electro-optical sensor. In particular, the Close-Minus-Open (CMO) morphological filter used is based on image morphology operations known as top-hat and bottom-hat transformations (Gonzalez et al., 2004). The effect of the top-hat transformation is to identify positively contrasting (brighter than background) features within an image that are smaller than a certain size (the cut-off size is specified through filtering kernels known as structuring elements), while the bottom-hat transformation performs a similar function but instead targets negatively contrasting (darker than background) features. It can be shown that summing the top-hat and bottom-hat transformations of the same image, which defines the Close-Minus-Open filtering operation, simultaneously identifies both positively and negatively contrasting features. This combination of morphological operations has been referred to elsewhere in the literature as a self-complementary top-hat filtering approach (Soille, 2003).

In this paper, the CMO filter is configured to serve as a powerful tool in the identification of small point-
like features within the measurement image. For performance and computational reasons, the CMO filter implemented in this paper exploits a directional decomposition technique (Casasent and Ye, 1997). More specifically, the minimum response from a pair of CMO filters using orthogonal 1D structuring elements is used. Here, one CMO filter operates exclusively in the vertical direction, while the other operates exclusively in the horizontal direction. The vertical and horizontal structuring elements of the CMO morphological preprocessing filter are given by $s_v = [1, 1, 1, 1, 1]'$ and $s_h = [1, 1, 1, 1, 1]'$, respectively. An example of the output after CMO morphological preprocessing is illustrated in Fig. 3b.

### 3.2 Temporal Filtering

In many machine vision based detection problems, the existence of a target in a 3D volume of space must be determined from observations of a projection of the target space onto a 2D image plane. Here, target detection can be viewed as evaluating the likelihood of two alternate hypotheses, where $H_1$ denotes the hypothesis that there is a single target present in the camera field of view, and $H_2$ denotes the hypothesis that there is no target present (it is also possible to consider a multiple target hypothesis; however, for clarity and ease of presentation, only the single and no target cases are discussed as this is sufficient to demonstrate the key concepts behind the temporal filtering approaches).

The temporal filtering approaches implemented in this paper will assume that under hypothesis $H_1$, the projected target motion resides on a 2D plane fixed in space that is represented by the set of discrete 2D grid points $\{(i,j) | 1 \leq i \leq N_v, 1 \leq j \leq N_h\}$, with vertical and horizontal resolutions $N_v$ and $N_h$ respectively. Let $N = N_v \times N_h$ denote the total number of grid points. The measurements are provided by an imaging sensor whose field of view is represented by a 2D grid of image pixels locations aligned with the target space and denoted $\{(p,q) | 1 \leq p \leq N_v, 1 \leq q \leq N_h\}$.

#### 3.2.1 Hidden Markov Model Filtering

It will be assumed that, when present, the target is located within a particular pixel of the image frame at each time instant. Thus, each pixel $(i,j)$ represents a unique state of the HMM in the target detection problem. For notational convenience, the columns of the image frame are stacked to form a vector of pixel locations. In this way, each state may be referenced by a single index, in the sense that if the target is at pixel location $(i,j)$, this corresponds to it being in the state $m = [(j-1)N_v + i]$.

Let $x_k$ denote the state (target location) at time $k$. Between consecutive image frames the target may move
to different pixel locations; that is, the target can transition between the states. The likelihood of state transitions can be described by the HMM’s transition probabilities \( A_{mn} = P(x_{k+1} = \text{state } m | x_k = \text{state } n) \) for \( 1 \leq m, n \leq N \), which is the probability of moving from any one pixel position (state) \( n \) to any other pixel position (state) \( m \). The transition probabilities can therefore be used to describe the expected mean target motion. For example, in the case of slow moving targets low probabilities tend to be assigned for transitions between distant pixels. Moreover, initial probabilities \( \pi_m = P(x_1 = \text{state } m) \) for \( 1 \leq m \leq N \) are used to specify the probability that the target is initially located in state \( m \). Finally, to complete the parameterisation of the HMM, there are the measurement probabilities \( B_m (Y_k) = P(Y_k | x_k = \text{state } m) \) for \( 1 \leq m \leq N \) that are used to specify the probability of obtaining the observed image measurement \( Y_k \in R^{Nv \times Nh} \), given that the target is actually in pixel location (state) \( m \) (see (Elliott et al., 1995) for more details about the parameterisation of HMMs).

**HMM Detection Strategy** The HMM filtering approach performs temporal integration of the input measurements by recursively propagating \( \alpha^m_1 \), an unnormalised probabilistic estimate of the target state \( x_1 \), over time. This is achieved via the forward part of the forward-backward procedure (Rabiner, 1989), which can be decomposed into two stages: initialisation and recursion.

For \( 1 \leq m \leq N \)

1. Initialisation: Let \( \alpha^m_1 \) denote the probability \( P(Y_1, Y_2, \ldots, Y_k, x_k = \text{state } m) \). Then \( \alpha^m_1 = \pi^m B^m (Y_1) \).
2. Recursion: At time \( k > 1 \), set \( \alpha^m_k = \frac{\sum_{n=1}^N \alpha^n_{k-1} A^{mn}}{\sum_{n=1}^N \alpha^n_k} B^m (Y_k) \).

The forward procedure filtering result is closely related to the two probabilistic measures that facilitate the detection of targets: 1) the probability of measurements up to time \( k \) assuming \( H_1 \), given by

\[
P(Y_1, Y_2, \ldots, Y_k | H_1) = \sum_{m=1}^N \alpha^m_k, \tag{1}
\]

and 2) the conditional mean filtered estimate of the target state \( m \) given measurements up to time \( k \) and assuming \( H_1 \), given by

\[
\hat{x}^m_k = E[x_k = \text{state } m | Y_1, Y_2, \ldots, Y_k, H_1] = \frac{\alpha^m_k}{\sum_{m=1}^N \alpha^m_k}, \tag{2}
\]
where $E[.|.]$ denotes the mathematical conditional expectation operation (Billingsley, 1995). The probability $P(Y_1, Y_2, \ldots, Y_k | H_1)$ may be interpreted as an indicator of target presence (following the probabilistic distance results of (Xie et al., 2005)), and the conditional mean estimate can be regarded as an indicator of likely target locations.

In the interest of computational efficiency, the conditional mean estimate is evaluated directly from the following expression (Elliott et al., 1995):

$$
\hat{x}_k = N_k B_k(Y_k) A \hat{x}_{k-1},
$$

where $N_k$ is a scalar normalisation factor; $B_k(Y_k)$ is a $N \times N$ matrix where the main diagonal is occupied by the values of $B^m(Y_k)$ for $1 \leq m \leq N$ and all other elements are zero; $A$ is a $N \times N$ matrix with elements $A^{mn}$; and $\hat{x}_k$ is a $N \times 1$ vector consisting of elements $\hat{x}_m^k$ for $1 \leq m \leq N$ that are equivalent to those given in (3). Moreover, note the following relationship between the normalisation factor $N_k$ and the probability of measurements up to time $k$ assuming $H_1$:

$$
P(Y_1, Y_2, \ldots, Y_k | H_1) = \prod_{l=1}^{k} \frac{1}{N_l}.
$$

For the HMM filtering approach, let $\eta_k$, the test statistic for declaring the presence of a target, be given by the following exponentially weighted moving average filter with a window length of $L$:

$$
\eta_k = \left( \frac{L-1}{L} \right) \eta_{k-1} + \left( \frac{1}{L} \right) \log \left( \frac{1}{N_k} \right).
$$

In the later performance evaluation studies, we found a window length of $L = 10$ produced good detection results. A shorter window length may give earlier detections, but this is likely to be at the expense of increased false alarms. When $\eta_k$ exceeds a predefined threshold, the HMM detection algorithm considers the target to be present and located at state

$$
\gamma_k = \arg \max_m (\hat{x}_m^k)
$$

at time $k$. The definition of $\eta_k$ and $\gamma_k$ is motivated by the filtering quantities discussed earlier.

**HMM Filter Bank** In this paper, a HMM filter bank consisting of four filters is implemented (Lai et al., 2008). Each HMM filter in the bank uses the same pre-processed image data, but otherwise operates...
independently of all other filters. The filter bank approach is less well characterised than the standard single
HMM filter, and its application has not been prevalent in the context of dim-target detection from imaging

sensors.

The transition probability parameters of each filter in the HMM filter bank are designed to handle a range of
slow target motion. For example, Fig. 1 illustrates an 3-by-3 patch of the possible transitions as seen in the
image plane (which could handle target motion within 2 pixels per frame). The possibility of larger motion
can be handled by larger patch sizes. These type of target motions correspond to transition probability
matrices that only have non-zero probabilities for self-transitions and transitions to states nearby in the
image plane (all other transitions have zero probability).

![Figure 1: Transition patch of size 3-by-3 corresponding to only 9 non-zero transition probabilities.]

An ad hoc approach to assigning the specific probability values within the patch has been used in this paper;
better performance might be possible by designing the parameters according to relative entropy rate based

strategies (Lai and Ford, 2010).

Furthermore, note that the implemented HMM filter exploits the following probabilistic relationship between
target location $x_k$ and the pre-processed measurements $Y_k$:

$$B^m(Y_k) = \frac{P(Y_k^m| x_k = \text{state } m)}{P(Y_k^m| x_k \neq \text{state } m)}$$

(7)

for $1 \leq m \leq N$. Note the computational advantage that (7) affords, given that $P(Y_k^m| x_k = \text{state } m)$ and
$P(Y_k^m| x_k \neq \text{state } m)$ can each be determined on a single-pixel basis (rather than requiring the probability
of a whole image (Lai et al., 2008)).
To construct the measurement probability matrix $B_k(Y_k)$, estimates of the probabilities $P(Y_m^k|x_k \neq \text{state } m)$ and $P(Y_m^k|x_k = \text{state } m)$ are required. The former describes the prior knowledge about the distribution of pixel values in the absence of a target (i.e., the noise and clutter distribution), while the latter captures the prior knowledge about the distribution of values at pixels containing a target. The required probabilities for $B_k(Y_k)$ are trained directly from sample data. The probability $P(Y_m^k|x_k \neq \text{state } m)$ is estimated as the average frequency that each pixel value resulted from a non-target location; one way this can be calculated is by sampling pixel values from image sequences without a target. Using a similar procedure, $P(Y_m^k|x_k = \text{state } m)$ is estimated as the average frequency that each pixel value measurement resulted from a target location, and this can be calculated based on image sequences containing a target. We acknowledge that in using an empirical approach for estimating the measurement probabilities, the amount of data available will influence the quality of the estimates. In general, estimates of $P(Y_m^k|x_k \neq \text{state } m)$ tend to be easier to obtain, due to the relative abundance of non-target image data compared with those containing targets. From our experiences, probabilities estimated from at least 100 image frames have provided reasonable performance.

Note that in the HMM filter bank, detection events may be triggered by any filter. In particular, if a target is present and this presence is declared in more than one filter, then look to the dominate filter (the one with the highest $\gamma_k$) to provide the estimate of target position.

Remark: Strictly speaking, the right hand side of (7) is proportional to $B_m(Y_k)$ (the applicable scaling factor may be absorbed by $N_k$ in the implementation of (3)). Moreover, this relationship only holds under the following assumptions: 1) the statistical properties of pixel values within an image are spatially independent, and 2) individual pixels do not allow the opportunity of perfect detection, in the sense that $P(Y_m^k|x_k \neq \text{state } m) > 0$ whenever $P(Y_m^k|x_k = \text{state } m) > 0$. Admittedly, the presence of extended (multi-pixel) targets or spatially correlated noise would violate the above assumptions, but has been found to only have moderate impact on performance.

3.2.2 Viterbi-Based Filtering

A Viterbi-based temporal filtering approach that is based on a dynamic programming algorithm (Carnie et al., 2006; Gandhi et al., 2006) was also examined. In this approach, pre-processed image frames are integrated over time along possible target trajectories in order to improve the signal-to-noise ratio. The output is an image, with large pixel values indicating likely target locations.
The motion of the target is modelled in terms of discrete velocity cells, where each cell \((u,v)\) encompasses a range of possible target velocities. Assuming constant target velocity (i.e. no transitions between velocity cells) and a velocity cell resolution of one pixel/frame, it can be shown that if the target is at any particular pixel \((i,j)\) at time \(k\), there exists a neighbourhood of four connected pixels within which the target will be located at time \(k+1\) (Tonissen and Evans, 1996). This neighbourhood of four connected pixels is denoted by \(Q^{ij}\) and collectively referred to as the *forward state transitions*. By symmetry, if the target is at any particular pixel \((i,j)\) at time \(k\), there exists a neighbourhood of four connected pixels within which the target was located at time \(k - 1\). This neighbourhood of four connected pixels is denoted by \(\bar{Q}^{ij}\) and collectively referred to as the *backward state transitions*. Hence, for each velocity cell \((u,v)\) there exists a unique set of forward state transitions \(Q^{ij}_{uv}\) and backward state transition \(\bar{Q}^{ij}_{uv}\) corresponding to each particular pixel \((i,j)\). This model for target dynamics can be modified to allow for transitions covering more than four pixels at a time by adjusting the resolution of the velocity cells. However, previous analysis (Tonissen and Evans, 1996) showed that better performance is achieved for smaller numbers of possible transitions, with four transition cell possibilities considered to be a reasonable choice for slow, non-manoeuvring targets (Gandhi et al., 2006).

**Viterbi-based Detection Strategy**  The Viterbi-based algorithm performs temporal integration of the input measurements by recursively generating a set of intermediate images \(D\) for each velocity cell \((u,v)\) that is considered. This process can be divided into two stages: initialisation and recursion.

For all \((u,v), 1 \leq i \leq N_v, \text{ and } 1 \leq j \leq N_h\)

1. **Initialisation:** Let \(D^\ast_{ij}(u,v)\) denote the \(ij\)th pixel of the intermediate image frame at time \(k\) for velocity cell \((u,v)\). Then \(D^\ast_{ij}(u,v) = 0\).

2. **Recursion:** At time \(k > 1\), set

\[
D^{ij}_k (u,v) = \left[ (1 - \beta) Y^{ij}_k \right] + \beta \max_{(r,s) \in Q^{ij}_{uv}} \left[ D^{rs}_{k-1} (u,v) \right],
\]

where \(Y^{ij}_k\) is the \(ij\)th pixel of the pre-processed image at time \(k\), and \(\beta\) represents a memory factor that can vary between zero and one.

At any time \(k\) when a detection decision is required, take the maximum output across corresponding pixels.
of the intermediate image frames belonging to each velocity cell:

\[ D_{ij}^{\text{max},k} = \max_{(u,v)} \left[ D_{ij}^{k}(u,v) \right], \quad (8) \]

for \(1 \leq i \leq N_v\) and \(1 \leq j \leq N_h\). This final image \(D_{\text{max},k}\) that consolidates target information from across all velocity cells then serves as the basis for declaring detections.

For the Viterbi-based filtering approach, a test statistic \(\lambda_k\) for declaring the presence of a target at time \(k\) is given by

\[ \lambda_k = \max_{ij} \left( D_{ij}^{\text{max},k} \right). \quad (9) \]

When \(\lambda_k\) exceeds a predefined threshold, the Viterbi-based algorithm considers a target to be present and located at state

\[ \varsigma_k = \arg \max_{ij} \left( D_{ij}^{\text{max},k} \right) \quad (10) \]

at time \(k\). The definition of \(\lambda_k\) and \(\varsigma_k\) follow from the interpretation of the Viterbi-based algorithm’s output as an image, where the pixel values correspond to target signal strength.

**Viterbi-Based Filter Implementation** The Viterbi-based filter implemented here mirrors those described in the existing literature (Carnie et al., 2006; Gandhi et al., 2006), where four velocity cells are used to detect targets that move with constant velocity in any direction, but are limited to a maximum speed of 1 pixel per frame (this is analogous to the 4 filters in the HMM filter bank). Non-maximal suppression is applied to the output of the filter to reduce an undesirable “dilation” effect where pixels in the neighbourhood of the target also attain significantly large values (Gandhi et al., 2006).

Furthermore, for the Viterbi-based filter, let \(\beta = 0.75\), which has been demonstrated to be a reasonable choice for the memory factor (Carnie et al., 2006).

### 4 Collision-Detection System Evaluation

We begin by evaluating our two candidate target detection approaches in terms of robustness to image jitter and target detection performance. We then discuss our proposed hardware implementation of the detection system, and evaluate the capacity of this system to support real-time execution of the candidate detection algorithms. The candidate detection algorithms are:
• CMO image pre-processing with HMM temporal filtering (patch size 5-by-5); and
• CMO image pre-processing with Viterbi-based temporal filtering.

To facilitate the evaluation of the detection algorithms, we will consider two types of signal-to-noise-ratio (SNR) quantities: 1) a target distinctness SNR (TDSNR), and 2) a false-alarm distinctness SNR (FDSNR). The TDSNR provides a quantitative measure of the detection capability of the algorithm, and is defined as:

\[
\text{TDSNR} = 10 \log_{10} \left( \frac{P_T}{P_N} \right)^2,
\]

where \(P_T\) is the average target pixel intensity and \(P_N\) is the average non-target pixel intensity at the filtering output. In general, the more conspicuous the target is at the output of the filter, the higher the TDSNR value. On the other hand, FDSNR measures the tendency of the algorithm to produce false-alarms, and is defined as:

\[
\text{FDSNR} = 10 \log_{10} \left( \frac{P_F}{P_N} \right)^2,
\]

where \(P_F\) is the average of the highest non-target pixel intensity and \(P_N\) is the average non-target pixel intensity at the filtering output. Strong filter responses away from the true target location will tend to increase the FDSNR value. For convenience, we will let \(\Delta \text{DSNR} = \text{TDSNR} - \text{FDSNR}\) denote the difference between the two SNR metrics.

4.1 Impact of Image Jitter

In this paper, we consider image jitter as the apparent motion of the image background between two consecutive frames of an image sequence. This apparent motion is caused by displacement of the camera relative to the objects in the scene, and gives the illusion that objects in an image are moving when in fact they are stationary in the environment. Image jitter is inherently present in all image data recorded on moving platforms. Passive motion-dampening devices or actively stabilised camera mounts may serve to reduce the effects of image jitter; however, jitter cannot be completely eliminated as no platform can be perfectly stable. In many cases, the effects of jitter can be quite severe, particularly for small UAV platforms that are more susceptible to unpredictable disturbances caused by wind gusts and air turbulence.

In addition to the physical mechanisms that can be employed to reduce jitter effects whilst images are being recorded, there are ego-motion compensation techniques that can be used after the image is captured to produce a jitter-corrected version of the raw image. Ego-motion compensation techniques typically rely on
estimating the amount of camera displacement either indirectly via gyroscopes, accelerometers, and other inertial sensors (Jorge and Jorge, 2004), or directly through tracking salient features (Jung and Sukhatme, 2005).

In this study, we are concerned with the inherent robustness of the detection algorithms to jitter affected data that have not undergone ego-motion compensation. Characterising this inherent robustness of the detection algorithms is important for two main reasons. Firstly, particularly in an airborne environment, it would not be uncommon to encounter ‘blue-sky’ conditions where the background is more or less uniform, rendering feature-based compensation techniques much less effective, if not useless. Moreover, no compensation technique is perfect, and hence it would be highly desirable for any practical detection algorithm to at least possess some level of resilience to image jitter in the absence of ego-motion compensation. We will characterise the jitter performance of the two candidate detection algorithms by applying them to some test image sequences containing varying degrees of jitter. The filtering output of the candidate detection algorithms will then be evaluated in terms of the two types of SNR quantities, TDSNR and FDSNR, defined earlier.

4.1.1 Test Data

Figure 2a provides a sample of the type of data that was used to characterise the detection algorithms. The typical target size and background contrast can be seen in a close-up region of interest depicted in Fig. 2b. Another close-up sample is provided in Fig. 3a. Here, the target occupies around 6-10 pixels, is quite distinct, and may be considered fairly easy to detect. The relatively high target signal-to-noise ratio does not diminish

(a)  
(b)

Figure 2: (a) Sample image from jitter characterisation test data set; (b) Region of interest containing target.
the significance and practical relevance of the performance characterisation because investigating the impact of image jitter on algorithm performance is the main interest of this study, as opposed to investigating dim target detection performance (this is considered separately in the next subsection).

In the data sets, a range of jitter characteristics has been artificially introduced, ranging from low to moderate to extreme. A low amount of jitter is defined as involving apparent inter-frame background motion of between 0 and 1 pixels per frame. A moderate amount of jitter is defined as involving apparent inter-frame motion of between 1 and 3 pixels per frame. Finally, an extreme amount of jitter is defined as involving apparent inter-frame motion is greater than 4 pixels per frame.

4.1.2 Jitter Test Results

Tables 1 and 2 illustrate the typical performance of the HMM and Viterbi-based temporal filtering approaches under various jitter scenarios. In the presence of a low level of jitter both detection approaches were able to successfully track the target (successful tracking implies that the target is able to be located within two pixels of its true position). Figure 3 (c) and (d) illustrate the typical output of the HMM and Viterbi-based filter, respectively, under low jitter conditions. We highlight that in this case the target in the HMM filter output is considerably more distinct than in the Viterbi-based filter output, and this is reflected in the larger ∆DSNR value for the HMM filter. However, under moderate jitter conditions, there were periods when the HMM filter failed to track the target successfully. This suggests that the HMM filter is more sensitive to the effects of jitter. Part of this sensitivity may be attributed to several modelling assumptions built into the HMM filter that are violated by image jitter. On the other hand, the ‘ad-hoc’ nature of the Viterbi-based filter tends to make it slightly less sensitive to image jitter. Finally, in the presence of extreme jitter, both filtering approaches demonstrated poor tracking capabilities and exhibited correspondingly low ∆DSNR values.

Overall, the ∆DSNR values seem to provide a reasonable indication of the detection algorithms’ jitter tracking performance. Using as a baseline the value from the low jitter scenario, which we denote by ∆DSNR₀, the results suggest that as a rough rule-of-thumb successful tracking can be accomplished under a particular jitter scenario x when:

\[ \Delta \text{DSNR}_x > 0.5 \Delta \text{DSNR}_0, \]  

where \( \Delta \text{DSNR}_x \) is the ∆DSNR value corresponding to jitter scenario x.

Although neither the HMM nor Viterbi-based approaches handle extreme jitter, we note that if the target signal-to-noise ratio is high enough it may be possible to detect potential collision-course targets from the
### Table 1: Jitter performance of HMM temporal filtering

<table>
<thead>
<tr>
<th>Jitter Level</th>
<th>Tracking Outcome</th>
<th>SNR Value (dB)</th>
<th>TDSNR</th>
<th>FDSNR</th>
<th>ΔDSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Success</td>
<td>72.84</td>
<td>40.71</td>
<td>32.13</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>Success</td>
<td>71.12</td>
<td>40.93</td>
<td>30.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fail</td>
<td>-58.61</td>
<td>38.10</td>
<td>-96.71</td>
<td></td>
</tr>
<tr>
<td>Extreme</td>
<td>Fail</td>
<td>-57.60</td>
<td>35.61</td>
<td>-93.21</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Jitter performance of Viterbi-based temporal filtering

<table>
<thead>
<tr>
<th>Jitter Level</th>
<th>Tracking Outcome</th>
<th>SNR Value (dB)</th>
<th>TDSNR</th>
<th>FDSNR</th>
<th>ΔDSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Success</td>
<td>36.20</td>
<td>33.28</td>
<td>2.92</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>Success</td>
<td>35.33</td>
<td>32.30</td>
<td>3.03</td>
<td></td>
</tr>
<tr>
<td>Extreme</td>
<td>Fail</td>
<td>30.00</td>
<td>29.60</td>
<td>0.40</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Sample of (a) test data frame; (b) CMO image pre-processing output; (c) HMM filter bank output (dominant member filter); and (d) Viterbi-based filter output, after processing 41 frames.
output of the morphological image pre-processing stage alone.

4.2 Target Detection Performance

In a potential collision scenario it is desirable to identify the other aircraft as early as possible to ensure enough time is available to plan and carry out appropriate actions. For instance, evasive manoeuvres may take a period of time to execute as the aircraft cannot be expected to respond instantaneously to control inputs. In general, increasing the range at which the threat can be detected will lead to earlier detection and hence more time to react. Thus, detection range may be considered a suitable statistic for characterising the detection performance of the candidate filtering algorithms.

In this study, we will firstly quantify detection performance in terms of the maximum range at which consistent detection occurs. We define consistent detection to have occurred when a simple threshold on the output of the filter allows the target to be detected in 10 successive frames with zero false-alarms. A consistent detection, however, does not preclude the existence of strong filter responses at non-target locations which are undesirable as they may be mistaken for the true target. Hence, $\Delta$DSNR values (averaged across the 10 consistent detection frames) will be presented together with the detection range statistics to provide an indication of the detection confidence.

We highlight that special experiments involving two fixed-wing UAVs flying near collision-course trajectories were conducted to collect suitable test data for offline post-processing by the detection algorithms. We elaborate on the details of these data collection experiments in the following paragraphs.

4.2.1 Test Data

Image sequences depicting aircraft on collision-course naturally serve as ideal test cases for evaluating collision detection algorithms; however, due to the inherent risk of flying aircraft on converging paths, this type of image data is extremely scarce. This motivated us to conduct our own special flight experiments to collect suitable test data.

Two fixed-wing UAVs were deployed to collect suitable test data: 1) a Flamingo UAV (Silvertone UAV, www.silvertone.com) measuring 2.9m from nose to tail with a wingspan of 4m; and 2) a Boomerang 60 model airplane (Phoenix Models) measuring 1.5m from nose to tail with a wingspan of 2.1m. The Flamingo was powered by a 26cc 2-stroke Zenoah engine driving a 16 by 6 inch propeller. The powerplant for the
Boomerang was an O.S. 90 FX engine driving a 15 by 8 inch propeller.

The avionics payload of the Flamingo included a MicroPilot® MP2128g flight controller, Microhard radio modems, an Atlantic Inertial SI IMU04 inertial measurement unit (IMU), a separate NovAtel OEMV-1 GPS device (in addition to the standard GPS capability provided by the MicroPilot® board), and an extensively customised PC104 mission flight computer. In contrast, the Boomerang only possessed a basic setup that featured a MicroPilot® MP2028g flight controller and Microhard radio modems.

In our experiments, the Flamingo served as the image data acquisition platform and was further equipped with a fixed non-stabilised Basler Scout Series scA1300-32fm/fc camera fitted with a Computar H0514-MP lens that offered 51.4 degree horizontal, 39.5 degree vertical, and 62.3 degree diagonal angles of view. The camera could be turned on and off remotely from the ground control station, and was configured to record 1024 by 768 pixel resolution image data at a rate of 15Hz with a constant shutter speed to maintain consistent lighting in the image frames. A solid-state hard-disk was used to store the recorded image data, as opposed to conventional mechanical disk drives which may be susceptible to vibrations during flight.

Figure 4 shows our UAV platforms (the Boomerang is in front of the Flamingo) configured for data collection.

![UAV platforms](image_url)
We highlight the streamlined pod just forward of the Flamingo’s wing that houses the camera along with the IMU. Our goal was to use the Flamingo, denoted as the camera aircraft, to record images of the Boomerang, which acted as the target aircraft, whilst both aircraft were flown autonomously along preset waypoints that defined near collision-course paths (adequate height and time separation was maintained at all times for safety reasons). In this way we could then capture realistic examples of what would be observed by an aircraft in the moments leading up to a collision. In particular, we sought image data that depicted the target aircraft gradually emerging from a state of being imperceptible to the naked eye to being clearly distinguishable (as opposed to data where the target aircraft suddenly enters into the field of view). We considered this type of data to be ideal for evaluating the detection range capability of the candidate filtering algorithms. The captured image data was timestamped during flight so that it could be later correlated with inertial and GPS-based position logs from both aircraft in order to estimate the detection range.

We conducted numerous flights at an altitude of approximately 450m above mean sea level over agricultural and livestock fields in the town of Kingaroy, about 210km north-west of the city of Brisbane, Australia. Our testing site covered a 4 by 4 kilometer area centered on an unsealed rural airstrip. Apart from obtaining permission from the landowner, no additional approvals or waivers were required as our flight operations fell under existing civil aviation safety authority regulations ((Civil Aviation Safety Regulations 1998, 2009); specifically, Part 101, subpart 101.F). Figure 5 illustrates our general approach to recreating a head-on collision scenario (not to scale). Typically after take-off, both aircraft were placed in holding patterns until they were ready to be released to follow predefined waypoints that brought the aircraft onto converging paths (a similar strategy was used in the simulation of other collision geometries). Apart from take-offs and landings, the aircraft were flown autonomously throughout the collection of data. Figure 6 illustrates the aircraft trajectories in an actual head-on engagement scenario that was enacted.

One of the key challenges of this approach involved timing the release of the aircraft from their respective holding patterns in a way that maximised the time spent by the target aircraft in the field of view of the camera aircraft. In the case of the head-on collision scenario, this meant having both aircraft simultaneously in the shaded region of Figure 5 for as long as possible. However, it was difficult to perform this synchronisation, and some of the data collected was not considered suitable for analysis. In spite of this, we were able to locate useful image sequences from three separate engagement scenarios featuring two different types of collision geometries. In two out of the three scenarios, the camera aircraft and the target aircraft are converging head-on; and in one scenario the aircraft are converging at right-angles to each other. Figures 6 to 8 illustrate overhead views of the collision geometries flown in each of the three engagement scenarios;
Figure 5: Typical flight plan for enacting a head-on collision scenario (not to scale). After exiting from their respective holding patterns (HP), the paths of the camera aircraft (solid line) and the target aircraft (dashed line) followed predefined waypoints (×) before converging at the designated area for capturing data (shaded region). The aircraft are approximately 2km apart at the start of their passing runs.

Also shown is a zoomed in sample of the typical vision obtained of the target aircraft for each scenario. The aircraft paths are plotted in a local East, North, Up (ENU) Cartesian coordinate system relative to a runway used for take-offs and landings (not shown).

Analysis of the data in all three scenarios revealed the presence of extreme jitter magnitudes exceeding 4 pixels per frame. This level of jitter was not completely unexpected; the usual jitter anticipated from the use of a non-stabilised camera was exacerbated by the relatively lightweight UAV platforms (compared with manned aircraft), which are more liable to external disturbances such as wind gusts and air turbulence.

Given the presence of extreme jitter and in light of the results from the earlier jitter performance characterisation, we considered it appropriate to undertake some level of jitter compensation of the image data before applying the detection algorithms. As this paper is not primarily concerned with the subject of camera-
Figure 6: Engagement scenario 1. (a) Camera aircraft (□) and target aircraft (○) positions leading up to point of detection (solid line) and shortly after detection (dashed line); (b) sample image frame containing target aircraft.

Figure 7: Engagement scenario 2. (a) Camera aircraft (□) and target aircraft (○) positions leading up to point of detection (solid line) and shortly after detection (dashed line); (b) sample image frame containing target aircraft.
motion compensation, we implemented a basic image correction procedure involving a combination optical flow (Lucas and Kanade, 1981) and template matching (Brunelli, 2009) techniques. Although this image correction procedure only accounted for translational motion, it was sufficient to reduce the overall effect of jitter to a moderate level. Image sequences with this moderate level of ‘residual’ jitter were then processed by the detection algorithms.

4.2.2 Detection Results

Table 3 illustrates the detection performance of the HMM and Viterbi-based filtering approaches in each of the three engagement scenarios. The detection ranges are calculated based on the camera and target aircraft GPS position information corresponding to the image frame that detection is declared (given that no unusual satellite geometries were encountered during the collection of the data, we estimate the 1-σ 3D position error of the GPS positions to be approximately 17 meters, based on standard values for HDOP, VDOP, and UERE (Parkinson and Spilker, 1996)).

In all three scenarios detection occurred a few frames earlier for the Viterbi-based algorithm when compared with the HMM algorithm. This is reflected in the slightly greater detection ranges for the Viterbi-based approach. The approximate positions of the aircraft at the time of detection are marked on the flight paths...
Table 3: Target detection performance of HMM and Viterbi-based temporal filtering

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Geometry</th>
<th>HMM Filter</th>
<th>Viterbi-based Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Detection Range (m)</td>
<td>∆DSNR</td>
</tr>
<tr>
<td>1</td>
<td>Head-On</td>
<td>412</td>
<td>38.04</td>
</tr>
<tr>
<td>2</td>
<td>Head-On</td>
<td>562</td>
<td>31.94</td>
</tr>
<tr>
<td>3</td>
<td>Right-Angle</td>
<td>881</td>
<td>19.55</td>
</tr>
</tbody>
</table>

given in Figures 6(a), 7(a), and 8(a) (due to the small differences in detection ranges, the aircraft positions corresponding to when detection is declared are not separately marked for the two algorithms). Overall, we obtained detection ranges that are generally consistent with the results reported in an earlier study (Carnie et al., 2006), which applied a detection algorithm similar to the Viterbi-based approach discussed in this paper. The earlier study attempted to detect a manned Cessna aircraft using image data captured by a fixed ground-based camera, and reported detection ranges out to about 6km. This range is roughly in proportion to our results in Table 3, considering that a Cessna aircraft is around 6-7 times the size of our Boomerang target aircraft. While we do not claim that a linear relationship exists between target size and detection distance, this comparison of detection range results reinforces the intuitive notion that larger targets should be able to be detected at greater ranges.

Furthermore, Table 3 shows the ∆DSNR values corresponding to the two target detection approaches. The ∆DSNR values for the HMM filter are consistently higher than the values for the Viterbi-based filter. This suggests that the HMM filter may be more effective at suppressing non-target responses (whilst maintaining a strong response at the true target location), and as a consequence, we may expect in general a lower incidence of false-alarms for the HMM filter than the Viterbi-based filter. We may also interpret the ∆DSNR values in terms of the empirical SNR criterion (13) established earlier for successful tracking. For the HMM filtering approach, the SNR criterion is easily satisfied, with the ∆DSNR values in all three scenarios clearly greater than $16.065 = 0.5 \times 32.13$. On the other hand, the results for the Viterbi-based filtering approach are mixed. Scenario 1 is a borderline case, and in Scenario 2 the ∆DSNR value is just below the threshold of $1.46 = 0.5 \times 2.92$. This suggests that the consistent detections established by the Viterbi-based filter in scenarios 1 and 2 are perhaps less reliable, in the sense that the filter may possibly be on the verge of losing track of the target.

We conjecture that residual jitter effects present in the image sequences after compensation (after all, no compensation technique is perfect) may be partly responsible for the difference in detection range performance. This is because residual jitter would tend to affect the HMM filter slightly more so than the Viterbi-based
We also highlight the significantly larger detection ranges reported for the right-angle collision geometry compared with the head-on cases. One possible explanation for this is that the target aircraft tends to appear as a larger object from the right-angle perspective (cross-section area approximately 0.25m$^2$) than from head-on (cross-section area approximately 0.1m$^2$).

We have used detection range as a metric for comparing the performance of the two candidate target detection algorithms. However, from an operational point of view it is the time-to-impact from the point of detection that is perhaps more informative, rather than the absolute distance to the conflicting aircraft. This is because, analogous to the time needed by human pilots to respond to a collision situation, an automated system also requires an interval of time to plan and carry out appropriate actions. Depending on the closing speed of the aircraft involved, a particular detection range may or may not provide sufficient time to accommodate both processing needs and aircraft response lag.

The Federal Aviation Administration has issued an advisory circular that provides guidance on the time required for a pilot to recognise an approaching aircraft and execute an evasive manoeuvre (FAA Advisory Circular: Pilots’ role in collision avoidance, 1983). According to the advisory circular, as a general rule, a conflicting aircraft must be detected at least 12.5 seconds prior to the time of impact. This motivates us to undertake some further analysis to gauge approximately how much time ahead of impact the system was able to detect the target. We chose to conduct our analysis on the worst-case scenarios i.e. the head-on geometries of scenarios 1 and 2. In our study we projected, from the point of detection, hypothetical trajectories for each aircraft corresponding to the shortest possible collision-course path (this path is defined by a straight line drawn between the camera and target aircraft positions at the point of detection). Each aircraft’s average speed in the last 5 seconds leading up to the point of detection was used as the speed travelled along the projected trajectories. Under these circumstances, the time-to-impact at the point of detection in Scenario 1 was estimated to be around 8 seconds (based on a combined closing speed of 51m/s), and for Scenario 2 the time was around 10 seconds (based on a combined closing speed of 53m/s). It is clear that these times are below the recommended 12.5 seconds; however, our results must be considered in the appropriate context.

The advisory circular information pertains to human pilots, with the 12.5 seconds broken down into the time taken to perform various sub-tasks. Over half of the time is allocated to the tasks of recognising the existence of the collision situation, making a decision to perform an avoidance manoeuvre, and manipulating the aircraft controls to execute the manoeuvre. An automated system must also carry out similar tasks, but there is potential for the tasks to be completed much more quickly given the rapidly improving processing...
capacities of computing hardware and the near instantaneous actuation of flight controls. Hence, we expect an equivalent safe detection time ahead of impact for automated systems to be less than 12.5 seconds.

4.3 Proposed System Hardware Configuration and Performance

Image processing is traditionally a very computationally intensive task. Hence, we considered it a challenge to achieve real-time performance for our proposed vision-based collision detection system. The increasing complexity of image processing algorithms is pushing at the processing limits of modern CPU-based computing architectures, even with the use of optimized computer vision libraries. Highly specialised hardware such as field programmable gate arrays (FPGAs) and dedicated digital signal processors have been used to overcome the limitations of the CPU; however, such hardware are typically not readily accessible. In this paper, we investigate the potential of widely available commercial-off-the-shelf (COTS) Graphics Processing Units (GPUs) for running our candidate target detection algorithms. The parallel processing (single-instruction, multiple-data (SIMD)) capability of GPUs is ideally suited to computer vision tasks, which often require the same calculations to be performed repeatedly on each individual pixel (or group of pixels) over an entire image.

Figure 9a illustrates the high-level hardware architecture of our collision detection system. Particular attention was paid to the interaction between hardware components to enhance processing speed. The key design strategies that we followed included: 1) the GPU module should carry out all the computationally intensive image processing tasks; 2) memory transfers between the host computer and the GPU module should be minimized; and 3) the host computer should be kept free to perform other non-image related tasks. By adhering to these strategies, we produced a system whereby each image frame captured by the vision sensor was directly copied from the host computer to the GPU module for processing. This allowed the entire detection algorithm to be executed on the GPU module and ensured that the GPU was utilized efficiently (in addition, offloading all the image processing to the GPU module eliminated time consuming memory transfers that would otherwise be required if the processing was shared between the host and GPU).

Once processing is completed on the GPU, only one additional memory transfer was needed to return the processing result (for example, a detection probability) to the host computer for further analysis. This entire process is repeated at the capture of the next image frame.

Recently, the parallel processing capabilities of GPUs have been exploited by other authors for tracking applications (Ohmer and Redding, 2008). Figure 9b illustrates the general way in which we exploited the GPU for processing images. At a high level, this involved dividing the image frame into blocks, with each
block containing a number of threads. We aimed to allocate one thread to each pixel of the image frame so
that operations could be carried out simultaneously on all pixels, greatly enhancing the processing speed.

Overall, as a minimum standard, the system was required to handle real-time processing of image data
(1024 by 768 pixel image frames, encoded at 8 bits per pixel) captured at a rate of 30Hz. As a first step to
characterizing the performance of a GPU based system, we bench-tested an implementation of the system
architecture in Figure 9a using the following desktop computer components:

- Host Computer: Intel Pentium IV 3.2GHz CPU (with Hyper-threading); 1Gb SDRAM at 666MHz;
  Linux Ubuntu 32 Bit operating system
- GPU Module: NVIDIA GeForce GTX 280, 1Gb GDDR3 RAM (GV-N28-1GH-B)

The NVIDIA GeForce GTX 280 has 30 multiprocessors and a compute capability of 1.3 (NVIDIA CUDA
the GPU module was facilitated by NVIDIA’s Compute Unified Device Architecture (CUDA) framework
configuration we were able to achieve an average frame rate of approximately 150Hz (or an average of 7.47ms
to process each frame) for the Viterbi-based detection algorithm, and a frame rate of approximately 130Hz
(or an average of 6.51ms to process each frame) for the HMM algorithm (these processing times include file
input/output operations). It is clear that the processing rates for both algorithms are well in excess of our earlier 30Hz requirement.

Our bench-test with desktop components has demonstrated the enormous potential of GPUs for realising real-time performance in vision-based systems. Motivated by these encouraging results, we are in the process of implementing a flight-ready system architecture using the following components:

- **Host Computer**: Intel Core 2 Duo Merom 800MHz CPU; 2Gb SDRAM at 800MHz; Linux Debian 32 Bit operating system
- **GPU Module**: NVIDIA GeForce 9600 GT, 512Mb GDDR3 RAM (GV-N96TGR-512I)

This flight configuration is to be integrated on a Shadow Mk-1 UAV (Integrated Dynamics), with a wingspan of 5.2m and measuring 2.95m in length, as shown in Figure 10.

![Figure 10: Shadow UAV.](image)

The NVIDIA GeForce 9600 GT was selected as it offered a good balance between processing performance, power consumption, and volume requirements. Specifically, it was the highest performing (in terms of the number of multiprocessors) GPU device that did not require an external power connection to supplement...
supply from the PCI bus. Figure 11 illustrates the GeForce 9600 GT (left) alongside the host computer. The GPU has 8 multiprocessors, a compute capability of 1.1, and consumes only 59 Watts of power (low-power version).

From experience, our implementation of the detection algorithms on the GPU scales roughly linearly with the number of multiprocessors. Hence, we anticipate processing rates on our flight hardware configuration to be approximately four times less than those reported for the bench-tests. This translates to a frame rate of around 37Hz for the Viterbi-based algorithm and a frame rate of around 32Hz for the HMM approach, which both still satisfy our 30Hz requirement. We highlight that there is still scope for further improvement in processing speeds, as we have yet to exploit advanced GPU code optimisation techniques (such as pipelining and dynamic memory allocation methods).

Figure 11: Flight hardware GPU module (left) and host computer (right).
5 Lessons Learnt and Future Work

Our long term goal extends beyond the ability to simply detect potential collision threats. We aim to establish a well-refined target detection capability, before progressively evolving this into a fully autonomous closed-loop collision avoidance system. This closed-loop system will include decision-making and control modules so that the UAV platform may automatically and actively respond to any detected threats. In the near term, our plans are to conduct further testing and refinement of the detection system, as discussed in the following sections.

5.1 Hardware Implementation and Evaluation

Implementation of the flight hardware configuration is currently in progress. The next step is to verify the anticipated processing speed of the flight hardware against the actual speed obtained once the implementation is completed. Should the actual speed fall below our expectations or our minimum requirements, we may consider employing advanced code optimisation techniques to improve processing speed. Alternatively, we may choose to upgrade to a more powerful GPU device at the expense of increased energy consumption.

5.2 Target Detection

Based on the detection range and image jitter performance results, neither the HMM nor the Viterbi-based algorithm stood out as the decidedly better detection approach. Admittedly, only three data sets were used in the analysis of detection range; hence, we intend to collect more data for analysis once the ability to achieve a more precise synchronisation of flights can be demonstrated (see next section).

Moreover, our image jitter analysis results combined with the inter-frame motion observed in the collected image data suggests that jitter compensation is necessary for our target detection algorithms to perform effectively on a UAV platform using an unstabilised camera. In addition to the simple image-based compensation approach used in this paper, we are testing more advanced image correction techniques based on aircraft attitude and inertial measurements. If our compensation techniques prove to be inadequate, we may consider the use of a stabilised camera and/or increasing the data capture rate to further mitigate jitter effects in the captured data.

Once the flight hardware has been integrated onto the UAV platform, we intend to carry out flight trials similar to those used in the collection of collision-course image data, with the exception that all processing
is performed onboard during flight. We will then compare the detection times and ranges to our earlier results and those recommended for human piloted aircraft (FAA Advisory Circular: Pilots’ role in collision avoidance, 1983).

5.3 Data Collection

We encountered many challenges in the collection of suitable test data, particularly in our flight trials that simulated collision-course scenarios. These experiments required the aircraft flights to be synchronised in time so that the target aircraft is kept in the field of view of the camera aircraft for as long as possible. We found it particularly difficult to achieve a precise synchronisation of the flights. On reflection, we have identified several factors that have contributed to this: 1) The inability to colocate the ground control stations of each aircraft due to the range limits of aircraft communication links (our suboptimal approach to overcoming this was to station personnel at separate ground control stations and to coordinate the release of the aircraft from their holding patterns via hand-held radios); 2) The use of fixed-wing platforms which are more difficult to position accurately as they must be kept in constant motion (aircraft positions were monitored by sight and through the MicroPilot® Horizon software); 3) The inability to make adjustments to flight paths once the aircraft are released from their holding patterns. This is due to the aircraft being operated in an autonomous mode.

To improve the synchronisation of flights in future trials, we are considering upgrades to antennas and cabling to improve signal gain so that the ground controls stations can be colocated. This will allow the flights to be better coordinated. It may also be possible to improve flight synchronisation by introducing more flexibility into the flight plan; for example, by adding a small segment of manual control just after the aircraft’s release from the holding pattern so that adjustments and corrections can be made to improve timing if necessary (remotely piloting the aircraft for the entire experiment is not considered feasible as it is too difficult to maintain a consistent altitude for an extended period of time). An alternative is to develop new software that will allow both aircraft to be controlled by a single control station. This will open up the opportunity to have the aircraft flights automatically synchronised without human intervention (this is perhaps the most sophisticated approach and a longer term solution). Overall, the field and operational experiences gained from our flight trials will be invaluable in fine-tuning future data collection experiments.
6 Conclusion

This paper proposes a vision-based collision detection system for aerial robotics. It considers two target
detection approaches and a GPU-based hardware implementation. Special flight experiments were conducted
to collect suitable test data for evaluating the detection capabilities of a HMM-based detection algorithm and
a Viterbi-based algorithm. The detection algorithms were able to detect targets at distances ranging from
400m to around 900m (depending on the collision geometry). Based on aircraft closing speeds leading up
to the point of detection and a hypothetical extrapolation of aircraft trajectories, these detection distances
would provide an advanced warning about 8 to 10 seconds prior to impact. These results are not too far from
the 12.5 second response time recommended for human pilots, and are an encouraging sign for vision-based
collision detection approaches. We exploit the parallel processing capabilities of GPUs to enable our detection
algorithms to execute in real time, and anticipate our flight-ready hardware configuration to achieve frame
rates of approximately 30Hz, based on earlier bench-testing results.

Acknowledgments

This research is part of the Smart Skies Project and is supported, in part, by the Queensland State Gov-
ernment Smart State Funding Scheme. The authors also gratefully acknowledge the contribution of Scott
McNamara and Angus Thomson for support with the GPU experimentations.

References


and Data Processing of Small Targets, Orlando, FL.

Aerospace and Electronic Systems, AES-21:144-156.

Barniv, Y. and Kella, O. (1987). Dynamic programming solution for detecting dim moving targets part II:


Chapter 7

Target Heading Angle

Estimation (J2)

In the previous chapters, we have successfully applied HMM filtering techniques and RER concepts to the problem of target detection. This encouraged us to explore the potential utility of HMM filters and RER concepts in other areas of the collision avoidance problem. Target heading angle is useful information that could assist not only with devising appropriated avoidance manoeuvres, but also with the post-evasion monitoring of targets after an evasive manoeuvre has been executed to ensure collision threats have been safely avoided. In this chapter, we demonstrate an RER-based approach to the heading angle estimation problem that exploits HMM filtering outputs.
Statement of Contribution of Co-Authors

The authors listed below have certified that:

1. they meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
2. they take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
3. there are no other authors of the publication according to these criteria;
4. potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
5. they agree to the use of the publication in the student's thesis and its publication on the Australasian Digital Thesis database consistent with any limitations set by publisher requirements.

In the case of this chapter:

RER based Target Heading Angle Estimation using HMM Filters

Status at time of writing:

Submitted 10 March 2010

Details of Authorship and Contributions:

<table>
<thead>
<tr>
<th>Name and Order of Authors</th>
<th>Signatures of Authors</th>
<th>Area of Contribution Regarding Authorship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(a) (i) (a) (ii) (b) (i) (b) (ii) (c)</td>
</tr>
<tr>
<td>John Lai</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Jason J. Ford</td>
<td>✓</td>
<td>✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Peter O'Shea</td>
<td>✓</td>
<td>✓ ✓ ✓</td>
</tr>
<tr>
<td>Luis Mejias</td>
<td></td>
<td>✓ ✓ ✓</td>
</tr>
</tbody>
</table>

Principal Supervisor Confirmation

I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Name ___________________________ Signature ___________________________ Date ___________

*See Appendix A for definition of authorship and criteria for contribution to publication
RER based Target Heading Angle Estimation using HMM Filters

John Lai*, Jason J. Ford, Peter O’Shea, and Luis Mejias

Abstract

This paper investigates the use of HMM filters for image-based heading angle estimation. An interesting relationship between relative entropy rate and probabilistic distance concepts is established that motivates the formulation of a novel heading angle estimator based on HMM filtering outputs. Results from simulation studies illustrate the superior performance of the proposed heading angle estimator over track-before-heading-estimation approaches. Good heading angle estimation performance is also observed under realistic noise conditions using image data captured from a non-stabilized camera.

Index Terms

Angle estimation, hidden Markov models, parameter estimation, relative entropy rate, target tracking.

EDICS Category(s): ASP-ANAL, SSP-FILT, SSP-IDEN

I. INTRODUCTION

Advances in hidden Markov model (HMM) signal processing tools over the last few decades have contributed to the widespread application of HMMs in a multitude of technical disciplines, including non-linear stochastic control [1], [2], signal and image processing [3]–[8], digital communications [9], and bioinformatics [10]. In particular, HMMs have been used to solve a variety of filtering problems in frequency tracking [7], speech recognition [8], character recognition [6], and dim target detection [3]–[5]. In these filtering problems, approximate HMM representations of the complex dynamics are often
exploited to simplify computations and to allow tractable solutions to be developed in situations where other non-linear filtering tools are sometimes inadequate. In this paper, HMM techniques are applied to a heading angle estimation problem.

The problem of determining current target heading angle in the image plane can be posed as the parameter estimation problem of inferring the heading angle parameter best describing the observed sample measurements. The discrete quantized nature of image measurements introduces an additional source of noise that can make accurate estimation of heading angle a challenge. Parameter estimation techniques have evolved over many decades of study, and traditionally have involved the optimization of a performance criteria such as maximizing the likelihood (e.g. expectation-maximization (EM) algorithm [16]) or minimizing the sum of squared residuals (least squares method [17]). These classical parameter estimation techniques have been applied in conjunction with HMM methods to a variety of parameter estimation problems, see [8], [18]–[21]. Parameter estimation approaches exploiting both HMM and information theory related concepts (such as mutual information, cross entropy, and the Kullback-Leibler distance measure) have also been developed [8], [22], [23]. In this paper, we propose a new parameter estimation approach for the target heading estimation problem which incorporates HMM techniques along with relative entropy rate concepts.

The main contribution of this paper is the proposal of an online target heading angle estimator based on HMM filter outputs that is robust in the presence of image quantization noise. We first establish a fundamental connection between joint relative entropy rate (RER) and probabilistic distance measures. Then we propose a heading angle estimator that exploits this connection. We also illustrate the performance of our heading angle estimator through simulated linear target trajectories, and compare this to the performance of a track-before-heading-estimation approach under various signal-to-noise ratios. Finally, we illustrate the heading angle estimator’s ability to estimate object heading angle under realistic noise conditions using real image data collected from a non-stabilized camera.

The paper is organized as follows: In Section II we introduce our general parameterized target model and image-based measurement process. In Section III we present a HMM approximation of the target dynamics and then present our HMM filtering approach. In Section IV we introduce important RER concepts and our proposed heading angle estimator. In Section V we examine the performance of our proposed heading angle estimator. Finally, in Section VI, conclusions are presented.
We will begin by introducing a general target model so that we can establish some important results. Later in Section IV, we will introduce specific dynamics for the target heading estimation problem that constitute the focus of this paper. Consider a target model where all processes will be defined on an abstract complete probability space \((\Omega, \mathcal{F}, P)\). For \(k > 0\), consider the following target state process on Cartesian coordinates,

\[
x_{k+1} = f(x_k, \theta) + w_k, \quad x_0 \in \mathbb{R}^2
\]

where \(x_k \in \mathbb{R}^2\) is a measurable function called the state (one state value for each coordinate) with initial value \(x_0\) and density \(\varphi_{x_0}(\cdot)\), \(w_k\) is a noise process with density \(\varphi_w(\cdot)\), and \(\theta \in S_\theta\) is a quantity that parameterizes a range of possible system dynamics (later, we will consider the specific case where \(\theta\) represents the target heading angle). We will restrict \(f\), \(\varphi_{x_0}\), and \(\varphi_w\) so that the state process is bounded to a subset of \(\mathbb{R}^2\), in that \(x_k \in S_x \subset \mathbb{R}^2\) for all \(k\), and that the process is both stationary and ergodic, see [11, pp. 62-70].

We will assume that the target state is observed through noisy image measurements from an electro-optical sensor. At each time \(k\), the observed image measurement consists of a grid with \(N_u\) rows and \(N_v\) columns of image pixel values (i.e. there are a total of \(m = N_u N_v\) pixels). Let \(y_k \in \mathbb{R}^m\) denote the measurement at time \(k\), and \(y^i_k\) denote the value at the \(i\)th pixel, under some enumeration scheme of pixel locations. The measurement process can be described by the following mapping process

\[
y^i_k = c^i(x_k) + v^i_k,
\]

where \(c^i(x)\) is the target intensity at pixel location \(i\) when the target is in state \(x\), and \(v^i_k\) are noise processes that are independent of the process \(x_k\) and mutually independent of \(v^j_k\) for \(i \neq j\) with density \(\phi(i)\). We highlight that these \(\phi(i)\) correspond to the noise density at each pixel, and are typically characterized in an experimental manner. Without loss of generality, the relationship between measurement \(y\) and state \(x\) in (2) will be described by the probability law \(p^0(y|x)\) for all \(x \in S_x, y \in \mathbb{R}^m\). We will also use the shorthand \(x_{[a,b]}\) to denote the process \(x_k\) with \(k \in [a,b]\); we similarly define \(y_{[a,b]}\) for process \(y_k\). Finally, for \(k \geq 0\), we assume that \(x_k\) and \(y_k\) are \(\mathcal{F}\)-measurable functions, and we will denote the model (1) (2) as \(\lambda(\theta)\).

In this paper, our aim is to determine parameter \(\theta\) from the image measurements \(y_k\). This may be classed as an adaptive filtering problem that can be difficult to solve due to the continuous nature of \(\theta\) in combination with the discrete multi-dimensional image measurements that are highly non-linear functions.
of the target state. The discrete nature of the input measurements motivates us to investigate a hidden Markov model approximation of the target dynamics and measurement process. Hidden Markov models have a well established statistical formalism incorporating powerful filtering tools that we will exploit. Furthermore, our approach will also incorporate multiple model techniques that are useful in non-linear problems.

III. HIDDEN MARKOV MODEL APPROXIMATION AND FILTERING

A. Hidden Markov Model Approximation

To construct our HMM approximation for the target model, we first introduce a bounded 2D spatial discretization of the state-space $S_x$. Without loss of generality, let us introduce the spatial grid $G_h = \{(x_1, x_2) : x_1 = \pm m_1 h_1, x_2 = \pm m_2 h_2\}$ that approximates the space $S_x$ with some spacing parameter $h = [h_1, h_2]$, where $m_1, m_2$ are non-negative integers. Let $N$ denote the number of grid points.

Let $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)'$ denote a vector with 1 in the $i$th position, and zero elsewhere. At time instant $k$, and with spacing parameter $h$, we will let $X_k \in \{e_1, e_2, \ldots, e_N\}$ be an indicator vector that denotes the state of a Markov chain process defined on grid $G_h$; that is, $X_k$ refers to a specific grid location on $G_h$ which will be denoted by the function $G_h(X_k)$. In other words, $G_h(\cdot)$ is a function that maps an indicator vector onto a grid location. We will use $\lambda = (A, B(y_k), \pi)$ to denote a candidate HMM based approximation, where $A, B$ and $\pi$ are discussed in the following paragraphs. Also, let $p^\lambda(\cdot)$ denote the probability law describing $\lambda$.

The Markov state process is assumed to have transition probabilities described by a $N \times N$ matrix $A$, with $ij$th element $A^{ij} = p^\lambda(X_{k+1} = e_i | X_k = e_j)$. We assume that transitions occur only to neighboring locations or are self-transitions (this corresponds typically to the behavior of a slow-moving target). Thus, away from the boundaries of the grid, there is a maximum of 9 non-zero transition probabilities from any internal state (later, we show how $\theta$ can impact on the choice of values for $A^{ij}$). Moreover, we assume that the transition behavior is uniform in the sense that these neighborhood transition properties do not depend on the location in the grid. At the grid boundaries, we allow transitions from states at the boundaries of the grid that would otherwise fall outside the grid to “wrap-around” onto the opposite side of the grid, so that boundary states also have 9 possible non-zero transition probabilities. This choice of boundary behavior helps with some ergodicity and relative entropy properties that will be used later in the paper.

Given that (1) is stationary and ergodic, we expect that our approximation task will lead to irreducible and aperiodic chains [24, pp. 50]. Also, we use the initial probability matrix $\pi$ to denote the $N a priori$
probabilities. Throughout, we will assume that the initial probabilities $\pi$ are known.

To allow us to apply later introduced relative entropy rate concepts in a consistent manner, we need to introduce a specific interpretation of our HMM approximation so that our approximation completely covers the state space occupied by the original model (1). Let $C_h(x_k)$ denote a $h$-sized 2D box or cell containing grid location $G_h(x_k)$. The cell $C_h(x_k)$ will be used to describe the set of all possible locations in $S_x$ represented by state $X_k$. We assume that the boxes $C_h(.)$ completely cover $S_x$ in the sense that for all $x \in S_x, x \in C_h(X_k)$ for some $X_k$, and we assume that boundaries between adjacent boxes are not shared. For example, we could enumerate the boxes in some manner and assign the shared boundary between two adjacent boxes to be part of the box with higher enumeration (but other boundary handling methods are possible). Related to this Markov chain, we also consider the discrete-time approximating process $x_k \in C_h(X_k)$ for all $k \geq 0$, where $x_k$ has uniform distribution over the cell $C_h(X_k)$. We assume that the cell is chosen so that each grid point $G_h(X_k)$ is centered in a corresponding cell $C_h(X_k)$ in the sense that $E[x_k|X_k] = G_h(X_k)$. With this construction, we can consider $x_k$ to be a “blurred” version of the chain process.

We assume the measurements related to this process is characterized by the probability law $p^\lambda(y_k|x_k \in C_h(e_i)) = p^\lambda(y_k|X_k = e_i)$ for $1 \leq i \leq N$ and all possible values of $y_k$. For notational convenience in later matrix equations, these measurement probabilities can be arranged into a $N \times N$ diagonal measurement matrix $B(y_k)$, where the $r$th element is defined as

$$B^{rs}(y_k) = \begin{cases} p^\lambda(y_k|X_k = e_s) & \text{if } r = s \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

for $1 \leq r, s \leq N$. Throughout, we will assume that the measurement laws $p^\lambda(y|x)$ and $p^\lambda(y|x)$ are absolutely continuous with respect to each other (that is, these measurement laws are equivalent), for all $x \in S_x$ (see [25, p. 422] for the definition of absolutely continuous).

As an extension to the basic HMM approximation described above, let us consider a set of $q$ HMMs $M_q = \{A_1, \lambda_2, \ldots, \lambda_q\}$ on grid $G_h$, where $\lambda_i$ is the $i$th HMM with Markov process $X_{k,i}$, parameterized by transition probability matrix $A_i$, output matrix $B_i(y_k)$, and the common initial probability vector $\pi$. Note, for each $i = 1, 2, \ldots, q, A_i \in S_A$, where $S_A = \{A : 0 \leq A^{ij} \leq 1 \text{ for all } i, j \text{ and } \sum_{i=1}^{N} A^{ij} = 1 \text{ for all } j\}$ denotes the set of all valid transition probability matrices with the required stationarity and ergodicity properties.

Furthermore, let $A_q = \{A_1, A_2, \ldots, A_q\}$ denote the corresponding set of candidate transition matrices associated with a particular set of HMMs $M_q$. We similarly define $B_q$ to represent the set of candidate
measurement matrices. Finally, let $S_{M,q}$ denote the collection of all valid HMM sets $M_q$ with $q$ models (we similarly define sets $S_{A,q}$ and $S_{B,q}$).

B. Our Proposed HMM Filter

In this section we will present our HMM filtering approach that is based on the multiple hidden Markov model (MHMM) filter bank described in [15]. Let us begin by describing a filter for a single HMM. If $\lambda$ is the true model, then the conditional mean estimate of $X_k \in \mathbb{R}^N$ given measurements up to time $k$ is [19]:

$$\hat{X}_k = N_k B(y_k) A \hat{X}_{k-1}$$  \hspace{1cm} (4)

where $\hat{X}_k = E[X_k|y_{[0,k]}]$ is a $N \times 1$ vector and $N_k = 1^T A \hat{X}_{k-1}$ is a scalar normalizing factor, with $1^T = [1,1,1,\ldots,1]$ denoting a $1 \times N$ row vector of all ones. We will use $\hat{X}_k^i$ to denote the $i$th element of $\hat{X}_k$. Alternatively, the closely related unnormalized conditional mean estimate $\alpha_k$ is given by the following recursion [8]:

$$\alpha_k^i = B^i(y_k) \left[ \sum_{j=1}^{N} \alpha_{k-1}^j A^{ij} \right] \text{ for } 1 \leq i \leq N,$$

where we note that $\hat{X}_k$ and $\alpha_k$ are related as follows:

$$\hat{X}_k = \frac{\alpha_k}{\sum_{i=1}^{N} \alpha_k^i}.$$  

This single model filter forms the basis of the MHMM filter bank, which we discuss next.

A MHMM filter bank is based on a set of HMMs, and can be described as a number of single HMM filters running independently in parallel. Consider a set of $q$ HMMs, $\lambda_i \in M_q$. For $i = 1, 2, \ldots, q$, we can consider the $q$ parallel filters

$$\hat{X}_{k,i} = N_k^i B_i(y_k) A_i \hat{X}_{k-1,i},$$  \hspace{1cm} (5)

where $\hat{X}_{k,i} = E[X_{k,i}|y_{[0,k]}]$ corresponds to the estimate for model $\lambda_i$. Let $L_k^i \triangleq p^{\lambda_i}(y_{[0,k]}) = \left( \prod_{\tau=1}^{k} N_{\tau}^i \right)^{-1}$ denote the likelihood of the $i$th model $\lambda_i$ at time $k$. We can now introduce the MHMM based filter estimate

$$\hat{x}_k \triangleq \sum_{j=1}^{N} \hat{X}_{k,i}^j G_h(e_j),$$  \hspace{1cm} (6)

where $i^*$ is selected so that $L_k^{i^*} \geq L_k^i$ for all $i$.

As an example of how the MHMM filter bank might be useful in the context of a parameter estimation problem, consider the following situation. Consider $q$ HMMs approximating dynamics for (1) for a discrete set of possible parameter values $\theta_i \in S_{\theta}, i = 1, 2, \ldots, q$, and let these approximation HMMs be collected...
in the set $\mathcal{M}_q$. That is, one HMM approximation $\lambda_i$ is created for each parameter value $\theta_i$ that is considered. When the MHMM filter bank is applied to an observation sequence, one model $\lambda_i^*$ will be most likely, and this model’s corresponding parameter $\theta_i^*$ provides a rough estimate of the true parameter $\theta^0$. Later, we solve a heading angle estimation problem using a variation of this idea, together with RER concepts that we introduce in the next section.

IV. TARGET HEADING ANGLE ESTIMATION

A. Relative Entropy Rate Concepts

Consider two probability measures $\mu$ and $\nu$ on a measurable space $(\Omega, \mathcal{F})$. The relative entropy or Kullback-Leibler divergence $D_{KL}(\mu \parallel \nu)$ of $\mu$ with respect to $\nu$ is defined by \[ D_{KL}(\mu \parallel \nu) \triangleq \begin{cases} \int_{\Omega} \left( \log \frac{d\mu}{d\nu} \right) d\mu, & \text{if } \mu \ll \nu \text{ and } \left| \log \frac{d\mu}{d\nu} \right| \text{ is integrable} \\ +\infty & \text{otherwise,} \end{cases} \] (7)

where $(d\mu/d\nu)$ is the Radon-Nikodym derivative of $\mu$ with respect to $\nu$. Here, $\mu \ll \nu$ denotes that $\mu$ is absolutely continuous with respect to $\nu$, in the sense that $\mu = 0$ wherever $\nu = 0$. The relative entropy $D_{KL}(\mu \parallel \nu)$ provides a pseudo-distance measure between $\mu$ and $\nu$ (not a true distance because it is non-symmetric and does not satisfy the triangle inequality). In the following entropy related concepts, we will use the convention $0 \log 0 = 0$.

When interested in dynamic systems, the relative entropy rate (RER) is often more useful. If $\mu_k, \nu_k$ denote measures on finite sequences of length $k$, and $\mu, \nu$ denote the measures obtained using Kolmogorov’s extension theorem [25], then the RER $\mathcal{R}$ of the measure $\mu$ with respect to $\nu$ is given by

$$\mathcal{R}(\mu \parallel \nu) \triangleq \lim_{k \to \infty} \frac{1}{k} D_{KL}(\mu_k \parallel \nu_k).$$

(8)

We now highlight that the usual RER between two models $\lambda$ and $\tilde{\lambda}$, $\mathcal{R}(\lambda \parallel \tilde{\lambda})$, is defined through the probability laws describing the output processes of the models (See remark at end of section). In this paper, we are also interested in the RER between the joint probability laws describing the joint output and state processes. To make absolutely clear this distinction between the RER defined through the output probability laws and the RER defined through the joint output-state probability laws, we introduce an overbar notation such that the joint RER is denoted by $\bar{\mathcal{R}}(\lambda \parallel \tilde{\lambda})$ [28]-[30]. For example, let $x_{[0,\infty]}; y_{[0,\infty]}$ denote the state and output processes generated by model $\lambda$, respectively. Then, the standard RER is given by $\mathcal{R}(\lambda \parallel \tilde{\lambda}) = \mathcal{R}(p^\lambda(y_{[0,\infty]}) \parallel p^\lambda(y_{[0,\infty]}))$, whereas the joint RER is given by $\bar{\mathcal{R}}(\lambda \parallel \tilde{\lambda}) = \mathcal{R}(p^\lambda(x_{[0,\infty]}; y_{[0,\infty]}) \parallel p^\lambda(x_{[0,\infty]}; y_{[0,\infty]}))$, where $p^\lambda$ and $p^\lambda$ are the probability laws
corresponding to the two models. To further clarify the meaning of relative entropy rate in our problem, we highlight that the joint RER can also be expressed as an integral:

$$\bar{R} (\lambda \parallel \lambda) = \lim_{k \to \infty} \frac{1}{k} \int_{\mathbb{R}^{m+k}} p^A (x_{[0:k]}^k, y_{[0:k]}^k) \log \left( \frac{p^A (x_{[0:k]}^k, y_{[0:k]}^k)}{p^\lambda (x_{[0:k]}^k, y_{[0:k]}^k)} \right) dx_{[0:k]} dy_{[0:k]},$$  \(9\)

where \(m\) is the dimension of the measurement \(y_k\). Next we introduce a formula for evaluating the joint RER between two HMMs under certain assumptions \(28\). Assume \(\lambda_1 = (A_1, B(y_k), \pi_1)\) and \(\lambda_2 = (A_2, B(y_k), \pi_2)\) are two HMMs that share a common measurement model \(B(y_k)\) and are of the same size. The joint RER between \(\lambda_1\) and \(\lambda_2\) is given by

$$\bar{R} (\lambda_1 \parallel \lambda_2) = \sum_{r=1}^N \sum_{s=1}^N \pi_r^A A_{sr}^p \log \left( \frac{A_{sr}^p}{A_{sr}^q} \right).$$  \(10\)

where \(N\) is the size of the HMMs. Also, we highlight that only when \(\lambda_1 \ll \lambda_2\) will the joint RER \(\bar{R}\) yield meaningful values.

We now present an interesting RER triangle result between three measures.

**Theorem 4.1:** Let \(P, M^1\) and \(M^2\) be three measures on a space \((\Omega, \mathcal{F})\), where \(M^1 \gg P\) and \(M^2 \gg P\). Suppose that \(G\) is a sub-\(\sigma\)-field and that \(P_G, M^1_G\), and \(M^2_G\) are the respective restrictions. Then the following result holds

$$D_{KL} (P \parallel M^2) - D_{KL} (P \parallel M^1) = E \left[ D_{KL} (P_G \parallel M^2_G) - D_{KL} (P_G \parallel M^1_G) \right].$$  \(11\)

Moreover,

$$\bar{R}^2_2 (P \parallel M^1, M^2) \triangleq \bar{R} (P \parallel M^2) - \bar{R} (P \parallel M^1) = E \left[ \bar{R} (P_G \parallel M^2_G) - \bar{R} (P_G \parallel M^1_G) \right].$$  \(12\)

**Proof:** From Lemma 5.2.5 in [31, p. 85], for two measures \(M \gg P\) on a space \((\Omega, \mathcal{G})\) with sub-\(\sigma\)-field \(\mathcal{G}\) we have that

$$D_{KL}(P \parallel M) = D_{KL}(P_G \parallel M_G) + D_{KL}(P \parallel S)$$

where \(S\) is a measure with the property that \(S(G) = P(G)\) for all \(G \in \mathcal{G}\).
Using this result twice, we have that
\[
E \left[ D_{KL} \left( P \parallel M^2 \right) - D_{KL} \left( P \parallel M^1 \right) \right] = E \left[ D_{KL} \left( P_G \parallel M^2_\theta \right) + D_{KL} \left( P \parallel S^2 \right) \right] - E \left[ D_{KL} \left( P_G \parallel M^1_\theta \right) + D_{KL} \left( P \parallel S^1 \right) \right] = E \left[ D_{KL} \left( P \parallel S^2 \right) - D_{KL} \left( P \parallel S^1 \right) \right].
\]

Now noting that \( E \left[ D_{KL} \left( P \parallel S^2 \right) - D_{KL} \left( P \parallel S^1 \right) \right] = E \left[ E \left[ D_{KL} \left( P \parallel S^2 \right) - D_{KL} \left( P \parallel S^1 \right) \mid G \right] \right] = E \left[ 1_G D_{KL} \left( P \parallel S^2 \right) - 1_G D_{KL} \left( P \parallel S^1 \right) \right] = 0 \)
and that \( E[D_{KL}(P \parallel M)] = D_{KL}(P \parallel M) \) gives the first result. We highlight that the second line follows from the tower property of expectations, and that in the third line \( 1_G \) is the indicator function on \( G \). The final equality follows because \( S^1(G) = S^2(G) = P(G) \) for all \( G \in \mathcal{G} \).

The second theorem result follows from the definition of the relative entropy rate.

We now introduce the concept of probabilistic distance between models so that we can relate our RER triangle results of Theorem 4.1 to filter outputs. The probabilistic distance between two models \( \lambda \) and \( \hat{\lambda} \) is defined as [28]:
\[
D(\lambda, \hat{\lambda}) \triangleq \lim_{k \to \infty} \frac{1}{k} \left[ \log(p^\lambda(y[0:k])) - \log(p^{\hat{\lambda}}(y[0:k])) \right].
\] (13)

To connect our Theorem 4.1 result with our probabilistic distance concept, let \( P \) and \( M \) be the probability measures obtained from the two models \( \lambda \) and \( \hat{\lambda} \) (using Kolomgorv’s extension theorem, see [25] for details), and let us consider \( G \) to be the \( \sigma \)-field generated by an observation sequence \( y[0:k] \). In a slight abuse of notation, it can be shown that the probabilistic distance is equal to the joint RER in the sense that [28]
\[
D(P, M) = \bar{R}(P_G \parallel M_G) \ P\text{-a.s.}
\] (14)
(here, ‘a.s.’ stands for ‘almost surely’, and denotes that the equation holds with probability 1; see [25] for more information).
Then Theorem 4.1 can be used to relate joint RER differences to probabilistic distance differences in the sense that
\[
\bar{R}_2^2 \left( P \| M^1, M^2 \right) = E \left[ D \left( P, M^2 \right) - D \left( P, M^1 \right) \right].
\]  
(15)

We highlight that when considering the special case of HMMs, a specific formula is available to calculate the probabilistic distance between two HMMs. Consider two HMMs \( \lambda_a \) and \( \lambda_b \) with \( N \) states. Let \( y_{[0,k]}^a \) denote the measurement sequence generated by \( \lambda_a \). Now consider the unnormalized conditional mean estimates \( \alpha_{a,k} \left( y_{[0,k]}^a \right) \) and \( \alpha_{b,k} \left( y_{[0,k]}^b \right) \) for \( \lambda_a \) and \( \lambda_b \), respectively, based on measurement sequence \( y_{[0,k]}^a \). In [28] it is shown that the probabilistic distance \( D \left( \lambda_a, \lambda_b \right) \) can be calculated from measurement data as
\[
D \left( \lambda_a, \lambda_b \right) = \lim_{k \to \infty} \frac{1}{k} \log \left( \frac{\sum_{i=1}^N \alpha_{a,k} \left( y_{[0,k]}^a \right)}{\sum_{i=1}^N \alpha_{b,k} \left( y_{[0,k]}^b \right)} \right). 
\]  
(16)

### B. Directional HMMs

Our HMM approximation in Section III can be used to describe the motion of an object on a discrete 2D grid; this interpretation follows naturally by recalling that \( G_h \left( X_k \right) \) describes a mapping of HMM states \( X_k \) onto specific locations in the 2D grid \( G_h \). One way to characterize the motion of an object is to consider its displacement on \( G_h \) between consecutive time instants. In terms of our HMM approximation, this can be expressed as an expected one-step grid displacement vector \( H \in R^2 \), defined as
\[
H = E \left[ G_h \left( X_{k+1} \right) - G_h \left( X_k \right) \right].
\]  
(17)

We can rewrite the vector \( H \) in a polar representation \((v, \psi)\), where \( v = |H| \) and \( \psi = \angle H \) denote the magnitude and angle of \( H \), respectively. The magnitude \( v \) relates to the mean speed of the object, and the angle \( \psi \) is associated with the mean direction of travel or heading angle of the object. In the following paragraphs, we will show how HMMs describing motion with particular \( v \) and \( \psi \) values (which we denote as directional HMMs) can be used in conjunction with the general relationships in Theorem 4.1, (14), and (16) to solve our target heading angle estimation problem.

We begin by considering how directional HMMs can be constructed based on HMMs that have particular state transition structures. To do this, we will introduce two types of directional HMMs: cardinal direction HMMs and general direction HMMs.

We first introduce HMMs that are characterized by a restricted state transition structure where transitions from each state are limited to only six other states on the grid \( G_h \). These six valid transition states form a 2 by 4 block of adjacent cells on the grid, as illustrated in Figure 1(a), with elements parameterized by the probabilities \( C^S \), \( C^H \), \( C^V \) and \( C^D \).
Fig. 1. State transition structure patches for (a) cardinal direction HMMs in set $S_C$, and (b) general direction HMMs in set $S_Q$.

We highlight the symmetry imposed on the state transition structure, in that the two diagonal transition probabilities $C^D$ are equal and the two horizontal transition probabilities $C^H$ are equal. Due to this symmetry and from our definition of $H$ (17), we can see that the patch in Figure 1(a) corresponds to a HMM with mean heading angle $\psi = \pi/2$ radians. Moreover, HMMs with heading angles $\psi \in \{0, \pi, 3\pi/2\}$ radians can also be generated by considering state transition structures corresponding to $\pi/2$ rotations of this patch. We will refer to HMMs with the structure in Figure 1(a) (and also $\pi$, and $3\pi/2$ rotations) as cardinal HMMs because they can only represent 4 different directions (the set of all possible cardinal HMMs will be denoted $S_C$).

We now describe a second type of state transition structure for directional HMMs in which the transition possibilities for each state are restricted to only four other states and form a 2 by 2 block of adjacent cells on the grid $G_h$, as illustrated in Figure 1(b). Here, the four valid transition states are parameterized by the probabilities $Q^S$, $Q^H$, $Q^V$, and $Q^D$. The advantage of this state transition structure is that it allows the construction of more general directional HMMs that can represent motion from a continuous range of mean heading angles and mean speeds (as opposed to the cardinal heading angle representations in $S_C$). From (17) and through the application of simple trigonometry, directional HMMs adhering to the state transition structure described in Figure 1(b) can be used to represent motion characteristics $\psi \in \left[0, \pi/2\right]$ radians and $0 \leq v \leq 1$ by assigning transition probabilities according to

$$
Q^S (\psi, v) = [1 - v \cos (\psi)] [1 - v \sin (\psi)]
$$

$$
Q^H (\psi, v) = v \cos (\psi) [1 - v \sin (\psi)]
$$

(a) 

(b)
\[ Q^V(\psi, v) = v \sin(\psi) [1 - v \cos(\psi)] \]  
\[ Q^D(\psi, v) = v \cos(\psi) v \sin(\psi) \]

Representations for mean heading angles \( \psi > \frac{\pi}{2} \) radians can be described by \( \frac{\pi}{2} \), \( \pi \), and \( \frac{3\pi}{2} \) radian rotations of the state transition structure depicted in Figure 1(b). Let \( S_Q \) denote the set of HMMs with state transition structures described by Figure 1(b) (and \( \frac{\pi}{2}, \pi, \) and \( \frac{3\pi}{2} \) rotations) and with transition probabilities given by (18)-(21).

We now state an interesting result that relates joint RER differences to angular differences.

To summarize, we have introduced the concept of directional HMMs, which are HMMs with particular state transition structures that can be used to describe motion (17) with a mean heading angle \( \psi \) and mean speed \( v \). Let \( \lambda_{\psi,v} \in S_Q \) denote a directional HMM from the set \( S_Q \). Similarly, let \( \lambda_{\psi,v}^C \in S_C \) denote a directional HMM from the set \( S_C \). Although other types of directional HMMs are possible, these two sets are sufficient for the purposes of this paper.

We will show that the joint relative entropy rate between directional HMMs belonging to sets \( S_C \) and \( S_Q \) exhibit some interesting properties that result from the symmetry in their transition structures. For this purpose, let us consider 3 HMMs \( \lambda_{\psi,v} \in S_Q \) and \( \lambda_{\psi_1,v_1}^C, \lambda_{\psi_2,v_2}^C \in S_C \), where \( \lambda_{\psi,v} \) has transition probabilities denoted by \( Q^C, Q^H, Q^V, Q^D \); \( \lambda_{\psi_1,v_1}^C \) has transition probabilities denoted by \( C^1,S, C^1,H, C^1,V, C^1,D \); and \( \lambda_{\psi_2,v_2}^C \) has transition probabilities denoted by \( C^2,S, C^2,H, C^2,V, C^2,D \). Furthermore, consider the following difference of joint RERs:

\[ R^2_1(\lambda_{\psi,v} \parallel \lambda_{\psi_1,v_1}^C, \lambda_{\psi_2,v_2}^C) = R(\lambda_{\psi,v} \parallel \lambda_{\psi_1,v_1}^C) - R(\lambda_{\psi,v} \parallel \lambda_{\psi_2,v_2}^C) \]

We now state an interesting result that relates joint RER differences to angular differences.

**Lemma 4.1:** Consider three directional HMMs \( \lambda_{\psi,v} \in S_Q \) and \( \lambda_{\psi_1,v_1}^C, \lambda_{\psi_2,v_2}^C \in S_C \), where \( \psi_2 = \psi_1 + \frac{\pi}{2} \) and \( v_2 = v_1 \). If we introduce \( \Delta \psi = \psi - \psi_1 \), then when \( \psi_1 = 0 \) or \( \pi \) radians and \( 0 \leq \Delta \psi \leq \frac{\pi}{2} \), the following result holds:

\[ R^2_1(\lambda_{\psi,v} \parallel \lambda_{\psi_1,v_1}^C, \lambda_{\psi_2,v_2}^C) = v \sqrt{2} \log \left( \frac{C^2 V}{C^2 H} \right) \cos \left( \Delta \psi + \frac{\pi}{4} \right) \]  

Moreover, when \( \psi_1 = \frac{\pi}{2} \) or \( \frac{3\pi}{2} \) radians and \( 0 \leq \Delta \psi \leq \frac{\pi}{2} \), the following result holds:

\[ R^2_1(\lambda_{\psi,v} \parallel \lambda_{\psi_1,v_1}^C, \lambda_{\psi_2,v_2}^C) = v \sqrt{2} \log \left( \frac{C^2 V}{C^2 H} \right) \sin \left( \Delta \psi - \frac{\pi}{4} \right) \]

**Proof:** First we consider the case where \( \psi_1 = 0 \), so that \( \Delta \psi = \psi \) and \( \psi_2 = \frac{\pi}{2} \). Using (10) and exploiting the symmetry of HMM sets \( S_C \) and \( S_Q \), the joint RER between HMMs \( \lambda_{\psi,v} \) and \( \lambda_{\psi_1,v_1}^C \) is...
given by
\[ \bar{\mathcal{R}}(\lambda \psi, v \| \lambda_c \psi_1, v_1) = Q^S(\psi, v) \log\left( \frac{Q^S(\psi, v)}{C_1, S} \right) + Q^H(\psi, v) \log\left( \frac{Q^H(\psi, v)}{C_1, H} \right) + Q^V(\psi, v) \log\left( \frac{Q^V(\psi, v)}{C_1, V} \right) + Q^D(\psi, v) \log\left( \frac{Q^D(\psi, v)}{C_1, D} \right). \]

A similar expression can be obtained for \( \bar{\mathcal{R}}(\lambda \psi, v \| \lambda_c \psi_2, v_2) \). The difference in joint RERs \( \bar{\mathcal{R}}^2(\lambda \psi, v \| \lambda_c \psi_1, \lambda_c \psi_2) \) can then be written as
\[ \bar{\mathcal{R}}^2(\lambda \psi, v \| \lambda_c \psi_1, \lambda_c \psi_2) = \bar{\mathcal{R}}(\lambda \psi, v \| \lambda_c \psi_2) - \bar{\mathcal{R}}(\lambda \psi, v \| \lambda_c \psi_1), \]
\[ = Q^S(\psi, v) \log\left( \frac{C_1^S}{C_2^S} \right) + Q^H(\psi, v) \log\left( \frac{C_1^H}{C_2^H} \right) + Q^V(\psi, v) \log\left( \frac{C_1^V}{C_2^V} \right) + Q^D(\psi, v) \log\left( \frac{C_1^D}{C_2^D} \right) \]
\[ = Q^H(\psi, v) \log\left( \frac{C_1^H}{C_2^H} \right) + Q^V(\psi, v) \log\left( \frac{C_1^V}{C_2^V} \right) \]
\[ = \log\left( \frac{C_1^V}{C_2^V} \right) \left[ Q^H(\psi, v) - Q^V(\psi, v) \right] \]
\[ = v \log\left( \frac{C_1^V}{C_2^V} \right) \left[ \cos(\psi) - \sin(\psi) \right] \]
\[ = v \sqrt{2} \log\left( \frac{C_1^V}{C_2^V} \right) \cos\left( \psi + \frac{\pi}{4} \right). \]

We obtain the second line by applying logarithmic identities and canceling terms. The fourth and fifth lines follow from exploiting the symmetry between \( \lambda_c \psi_1, v_1 \) and \( \lambda_c \psi_2, v_2 \). Finally, a trigonometric identity is used in the last line. Hence, the first lemma result holds when \( \psi_1 = 0 \) by noting that \( \psi = \Delta \psi \) when \( \psi_1 = 0 \). By symmetry, the first result also holds when \( \psi_1 = \pi \) radians. By antisymmetry, the second result holds when \( \psi_1 = \frac{\pi}{2} \) or \( \frac{3\pi}{2} \) radians.

This lemma, in combination with Theorem 4.1, (14), and (16) establish relationships that we can exploit to estimate target speed and heading angle parameters. We next consider the target heading angle estimation application, although the technique described may be easily adapted to estimate target speed.

Remark: We highlight that the patches depicted in Figure 1 are abstract representations of the HMM transition probability matrix; they should not be confused with the actual structure of individual state transition matrices, which depend on the enumeration of states. We also highlight that the patches in Figure 1 are only representative of the transition behavior of internal states within the grid boundaries;
transitions from boundary states are described by the “wrap-around” rule described earlier in Section III.A.

C. Relative Entropy Rate based Heading Angle Estimation

Let us consider a target heading angle estimation problem based on image information. Our attention will be restricted to linear target trajectories by considering the following target model (which is consistent with our model description (1))

$$x_k = x_{k-1} + \begin{bmatrix} V \cos(\psi) \\ V \sin(\psi) \end{bmatrix},$$  \hspace{1cm} (25)

where $x_k = [d^x_k, d^y_k]'$ denotes the target’s pixel coordinates in the image plane, $V$ denotes the target’s speed in pixels/frame, and $\psi \in S_\psi$ denotes the target heading angle in the image plane. Both the target’s speed and heading angle are constant for all $k$ (static target model), and we denote their true values by $V^0$ and $\psi^0$, respectively. In this application example, we assume that the target is slow-moving ($|V| \leq 1$), and that the true speed $V^0$ is known. Our aim is to estimate the true heading angle $\psi^0$ (this corresponds to the parameter $\theta$ in (1)) based on noisy image measurements from an electro-optical sensor, whose field of view is represented by a 2D grid of image pixel locations $G = \{(i, j) | 1 \leq i \leq N_u, 1 \leq j \leq N_v\}$. We let $y_k$ be the $N_u \times N_v$ pixel image measurement at the $k$th sample instant, and assume this model is consistent with description (2). Moreover, we let $y^i_k$ denote the value at the $i$th pixel, under some enumeration scheme of pixel locations. We will let $\lambda^0(\psi^0)$ denote the true model of the target dynamics.

The problem of estimating heading angle $\psi$ can be difficult due to the combination of underlying dynamics that are continuous, coupled with the multi-dimensional discrete measurement input, as well as the highly non-linear relationship between measurements and the unknown quantity $\psi$. The quantized nature of the measurements (discrete pixels and grey-level intensities) motivates us to seek a HMM based approximation of the target dynamics $\lambda^0(\psi)$.

One method of constructing HMM approximations exploits what are termed locally consistent properties [1]. In this paper, we propose to approximate the model $\lambda^0(\psi)$, for specific values of $\psi$, with a locally consistent Markov chain, using the results of [1]. This involves selecting a Markov chain so that the chain and $\lambda^0(\psi)$ have consistent local statistical properties. Specifically, suitable locally consistent approximations for our heading angle estimation problem may be chosen from the HMM set $S_Q$ (see Remark 1 at end of section).

Our locally consistent HMM approximation of the target dynamics helps us to reduce the complexity of the heading angle estimation problem, so that we may approach the problem using existing tools in
the form of our MHMM filter bank (5) whilst exploiting the results from Theorem 4.1 and Lemma 4.1 applicable to HMMs. Key to our approach is the relationship between RER and probabilistic distance described in (15), which when combined with the link between RER and angular motion established in Lemma 4.1 allows us to determine an estimate for heading angle \( \psi \) directly from the filtering output of the MHMM filter bank.

We now present our approach to the heading angle estimation problem considering our HMM approximation of the target dynamics. Let us assume that a set of HMMs
\[
\mathcal{M}_q = \{ \lambda_1, \lambda_2, \ldots, \lambda_q : \lambda_i \in \mathcal{S}_C \text{ for } i = 1, 2, \ldots, q \}
\]
satisfying the conditions in Lemma 4.1 is available for use in our MHMM filter bank (see Remark 2 at end of section). Also, let \( \lambda_{\psi_1,v_1}^c, \lambda_{\psi_2,v_2}^c \in \mathcal{M}_q \) be the two HMMs in the set that are closest to \( \lambda^0 (\psi^0) \), in the sense of having the highest likelihoods \( L_k^i \) as \( k \to \infty \). From \([26],[32]\), it is clear that the directions of these two HMMs, \( \psi_1 \) and \( \psi_2 \), will only be different by \( \frac{\pi}{2} \), and that \( \psi^0 \) will be between them (except for trivial cases when \( V^0 = 0 \)). Without loss of generality, let us reorder labels so that \( \psi_2 \) is the larger angle, and let us denote the transition probabilities of \( \lambda_{\psi_2,v_2} \) by \( C^2, S, C^2, H, C^2, V, \) and \( C^2, D \). Finally, let \( y_{[0,k]}^v \) denote the measurements generated by \( \lambda^0 (\psi^0) \), and let \( \alpha_{1,k} \left( y_{[0,k]}^v \right) \) and \( \alpha_{2,k} \left( y_{[0,k]}^v \right) \) denote the unnormalized conditional mean estimates of \( \lambda_{\psi_1,v_1} \) and \( \lambda_{\psi_2,v_2} \) at time \( k \), respectively, based on \( y_{[0,k]}^v \).

We now propose a two-stage ‘coarse-to-fine’ method for obtaining our RER-based heading angle estimate \( \hat{\psi} \). Initially, we look to the angular sector bounded by \( \psi_1 \) and \( \psi_2 \) to provide a coarse estimate of heading angle, in the sense that we narrow the range of angle possibilities down to the interval \([\psi_1, \psi_2]\) radians. We then proceed to refine our initial estimate with a finer resolution estimate

\[
\Delta \hat{\psi} \triangleq \begin{cases} 
\cos^{-1} \left( \frac{D^2 \left( y_{[0,k]}^v \left\| \lambda_{\psi_1,v_1}^c \lambda_{\psi_2,v_2}^c \right\| \right)}{V^0} \right) - \frac{\pi}{4}, & \text{if } \psi_1 = 0, \pi \\
\sin^{-1} \left( \frac{D^2 \left( y_{[0,k]}^v \left\| \lambda_{\psi_1,v_1}^c \lambda_{\psi_2,v_2}^c \right\| \right)}{V^0} \right) + \frac{\pi}{4}, & \text{if } \psi_1 = \frac{\pi}{2}, \frac{3\pi}{2},
\end{cases}
\]

(26)

where \( V = V^0 \sqrt{2} \log \left( C^2, V / C^2, H \right) \) and \( D^2 \left( y_{[0,k]}^v \left\| \lambda_{\psi_1,v_1}^c \lambda_{\psi_2,v_2}^c \right\| \right) \) is given by

\[
D^2 \left( y_{[0,k]}^v \left\| \lambda_{\psi_1,v_1}^c \lambda_{\psi_2,v_2}^c \right\| \right) = \lim_{k \to \infty} \frac{1}{K} \log \left( \sum_{i=1}^{N} \alpha_{1,k} \left( y_{[0,k]}^v \right) \right).
\]

(27)

We highlight that in (27), the true model is generating the measurements used to create the estimates \( \alpha_{1,k} \) and \( \alpha_{2,k} \) (see Remark 3 at end of section). To illustrate the connection between (26) and our results in Lemma 4.1, we highlight that (15) can be rewritten as

\[
\mathcal{R}^2 \left( \lambda^0 (\psi^0) \left\| \lambda_{\psi_1,v_1}^c \lambda_{\psi_2,v_2}^c \right\| \right) = E \left[ D^2 \left( y_{[0,k]}^v \left\| \lambda_{\psi_1,v_1}^c \lambda_{\psi_2,v_2}^c \right\| \right) \right].
\]

(28)
Finally, our RER-based heading angle estimate \( \hat{\psi} \) is obtained by combining our coarse and fine estimates:

\[
\hat{\psi} \triangleq \psi_1 + \Delta \hat{\psi}.
\]  

Equation (29)

The estimate \( \hat{\psi} \) is meaningful in the following sense:

**Lemma 4.2:** Consider two cardinal HMMs \( \lambda_{\psi_1,v_1}^c, \lambda_{\psi_2,v_2}^c \in S_C \), where \( \psi_2 = \psi_1 + \frac{\pi}{2}, \psi_1 \leq \psi^0 \leq \psi_2 \), and \( v_2 = v_1 \). If there is a \( \lambda_{\psi,v} \in S_Q \) with heading angle \( \psi = \psi^0 \) and speed \( v = V^0 \) which generates the measurements \( y_{[0,k]}^\psi \), then

\[
E \left[ \cos \left( \Delta \hat{\psi} \right) \right] = \cos \left( \Delta \psi^0 \right) \text{ when } \psi_1 = 0, \pi \text{ or } (30)
\]

\[
E \left[ \sin \left( \Delta \hat{\psi} \right) \right] = \sin \left( \Delta \psi^0 \right) \text{ when } \psi_1 = \frac{\pi}{2}, \frac{3\pi}{2} \text{, (31)}
\]

where \( \Delta \psi^0 = \psi^0 - \psi_1 \).

**Proof:** The lemma result follows from the application of Theorem 4.1 and Lemma 4.1. □

In the next section, we will evaluate the performance of our proposed heading angle estimator.

**Remark:**

1) For local consistency to hold as described in [1], both the mean and variance of state transition properties must be considered. Elements from the set \( S_Q \) do not strictly satisfy the variance conditions required. However, in [15] a similar approximation approach was used and resulted in good filtering performance.

2) In [15], methods for selecting \( M_q \) so that they conform to the conditions of Lemma 4.1 are described.

3) Equation (27) should not be confused with the probabilistic distance quantity (16), which takes on a similar form. The subtle difference is that in (16), one of the conditional mean estimates corresponds to the model that generated the measurement data. However, we highlight that (16) can be considered a specific case of (27), in the sense that

\[
D \left( \lambda_a, \lambda_b \right) = D_{\alpha} \left( y_{[0,k]}^\alpha \| \lambda_a, \lambda_b \right),
\]

where \( y_{[0,k]}^\alpha \) is the measurement sequence generated by \( \lambda_a \).

V. HEADING ANGLE ESTIMATION PERFORMANCE

The problem of estimating heading angle from image-based measurements is difficult because of the inherent quantization in pixel locations and intensities. These quantization effects may be considered as an additional noise source obscuring the underlying true heading angle. In this section, we first illustrate
the relationship between RER and probabilistic distance differences, which is exploited by our proposed heading angle estimator. This is followed by two studies designed to evaluate the performance of our proposed target heading angle estimator (one on simulated data, one on real data).

A. Illustration of Connection between RER and Probabilistic Distance

Consider the directional HMMs $\lambda_{\psi_1,v_1}, \lambda_{\psi_2,v_2} \in S_C$, where $\psi_1 = 0$, $\psi_2 = \frac{\pi}{2}$, and $v_1 = v_2$. Let $y_{0,k}^{\psi}$ denote the measurements generated by $\lambda^0 (\psi^0)$ with $V^0 = 0.2$, and let $\lambda_{\psi,v} \in S_Q$, where $\psi = \psi^0$ and $v = V^0$, denote the locally consistent HMM approximation of $\lambda^0 (\psi^0)$. Figure 2 illustrates the relationship between $D_2^1 \left( y_{0,k}^{\psi} \left\| \lambda_{\psi_1,v_1}^c, \lambda_{\psi_2,v_2}^c \right\| \right)$ and $\bar{R}_2^1 \left( \lambda_{\psi,v} \parallel \lambda_{\psi_1,v_1}^c, \lambda_{\psi_2,v_2}^c \right)$ for a range of target heading angles $\psi^0 \in [0, \frac{\pi}{2}]$ (values for $D_2^1 \left( y_{0,k}^{\psi} \left\| \lambda_{\psi_1,v_1}^c, \lambda_{\psi_2,v_2}^c \right\| \right)$ are based on MHMM filtering outputs after processing 240 frames). We highlight that similar relationships are observed for $\psi^0 > \frac{\pi}{2}$. This close link between $D_2^1 \left( y_{0,k}^{\psi} \left\| \lambda_{\psi_1,v_1}^c, \lambda_{\psi_2,v_2}^c \right\| \right)$ and $\bar{R}_2^1 \left( \lambda_{\psi,v} \parallel \lambda_{\psi_1,v_1}^c, \lambda_{\psi_2,v_2}^c \right)$ demonstrates the property illustrated in (28) and motivates our use of Lemma 4.1 in the construction of our heading angle estimator. Figure 2 also illustrates the key concept behind our proposed heading angle estimator, in the sense that it provides a mapping of MHMM filter outputs to particular target heading angles $\psi^0$.

![Fig. 2. Relationship between quantity $D_2^1 \left( y_{0,k}^{\psi} \left\| \lambda_{\psi_1,v_1}^c, \lambda_{\psi_2,v_2}^c \right\| \right)$ and joint RER difference $\bar{R}_2^1 \left( \lambda_{\psi,v} \parallel \lambda_{\psi_1,v_1}^c, \lambda_{\psi_2,v_2}^c \right)$ for case when $V^0 = 0.2$ pixel/frame.](image-url)
B. Heading Angle Estimation Case Studies

We now evaluate the performance of the proposed heading angle estimator in two studies. In both studies we implement a MHMM filter bank that uses a 4-element HMM set we denote as $\mathcal{M}_4$, where each element $\lambda_i$ is based on the HMM representation in Section III.A and shares a common known measurement model. Let the first element in the set be denoted by $\lambda_1 \in S_C$ with transition probabilities $C^S = 0.757$, $C^H = 0.038$, $C^V = 0.153$, and $C^D = 0.007$ (min-max RER optimized design [15]). The other elements in the set $\mathcal{M}_4 = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ are given by $\frac{\pi}{2}$ radian rotations. Image morphology techniques are used to pre-process image data in both studies prior to HMM filtering [33].

For comparison purposes, we introduce a track-before-heading-estimation (TBHE) technique (we acknowledge that there is no clear standard heading angle estimation approach, but a TBHE technique is an intuitive candidate approach). In TBHE, the gradient of a line connecting the estimated start and end points of the target track provides an estimate of heading angle. If the estimated target track is based on conditional mean estimate (CME) information, we term this a CME-TBHE approach. Similarly, if we estimate target track using maximum a priori (MAP) information, we term this a MAP-TBHE approach.

1) Heading Angle Estimation using Simulated Data: This first study uses simulated target imagery, where targets are observed through discrete measurements from an imaging sensor of dimension 111 by 147 pixels. The image measurement noise is modeled as a Gauss Markov random field (GMRF) parameterized by vertical and horizontal interaction factors of 0.12 and driven with $N(0,1)$ Gaussian noise (for more details see [33]). We examine the performance of our proposed RER-based heading angle estimator (29) compared to the two TBHE approaches for a range of peak signal-to-noise ratio (PSNR) cases. We define the PSNR quantity as, $\text{PSNR} = 20 \log_{10} \left( \frac{\eta}{\sigma} \right)$, where $\eta$ is the peak target intensity and $\sigma$ the noise standard deviation, to characterize how distinct the simulated targets are. We consider image sequences of simulated targets with linear dynamics (25); that is, constant target speed $V$ and constant heading angle $\psi$ (we assume that $V$ is known). The image sequences are processed with a HMM filter bank (5) using the set $\mathcal{M}_4$, from which we obtain estimates of target heading angle via the CME-TBHE, MAP-TBHE, and our RER-based approaches.

At each PSNR = 10, 12, 14, 16, 18, and 20 dB, we simulated 500 target engagements sequences at a heading angle $\psi^0 = \frac{\pi}{6}$ radians and $V^0 = 0.2$ pixel/frame. The angle error of the CME-TBHE, MAP-TBHE, and our RER-based estimates were recorded over the last 125 frames of each image sequence (to minimize initial filter transient effects). Figure 3 illustrates the typical convergence rate of our RER-based estimate to the true heading angle for target engagement sequences at PSNR = 16 dB (we observed
similar convergence rates at other PSNRs). Figure 4 shows the average squared angle error per frame of the CME-TBHE, MAP-TBHE, and our RER-based angle estimation approaches. We highlight that the squared error of our RER-based approach is consistently below the squared error of the two TBHE approaches across the range of PSNRs considered.

![Figure 4. Average squared angle error for CME-TBHE, MAP-TBHE, and RER-based heading angle estimation (500 cases of target dynamics $\psi^0 = \frac{\pi}{6}$ radians and $V^0 = 0.2$ pixel/frame simulated at each PSNR).](image-url)

Fig. 4. Average squared angle error for CME-TBHE, MAP-TBHE, and RER-based heading angle estimation (500 cases of target dynamics $\psi^0 = \frac{\pi}{6}$ radians and $V^0 = 0.2$ pixel/frame simulated at each PSNR).

Fig. 3. Heading angle estimate $\hat{\psi}$ for target dynamics $\psi^0 = \frac{\pi}{6}$ radians, $V^0 = 0.2$ pixel/frame, and PSNR = 16 dB.
2) Heading Angle Estimation Using Real Data: Finally, we consider estimation of the heading angle of a remotely operated helicopter in real image sequences. The image data is captured using a non-stabilized camera; hence, we firstly compensate the image frames for undesirable camera self-motion using a procedure based on a combination of optical-flow [34] and image template matching [35] techniques. Apart from this minor post-processing of image frames, our approach for heading angle estimation is the same as in the previous study; that is, we used (26), (27), and (29) to calculate our angle estimate ̂\(\psi\).

We applied our heading estimation technique to data from two scenarios. In Scenario 1, the helicopter moved at approximately 0.43 pixels/frame with a heading angle of approximately 0.71 radians (these ‘truth’ values are based on the TBHE approach discussed in the previous study). Figure 5 shows our heading angle estimate for Scenario 1 converging to the true angle. In Scenario 2, the helicopter moved at approximately 0.18 pixels/frame with a heading angle of approximately 3.64 radians. Our heading angle estimate for this scenario also converges to the true angle, as illustrated in Figure 6. These results suggest that our estimator is not overly sensitive to noise and other artifacts typically present in real image data.

![Graph](image)

Fig. 5. Scenario 1 helicopter heading angle estimate ̂\(\psi\) versus frame number. From visual inspection of image sequence, ̂\(\psi\) ≈ 0.71 radians and \(V_0\) ≈ 0.43 pixel/frame.
Fig. 6. Scenario 2 helicopter heading angle estimate $\hat{\psi}$ versus frame number. From visual inspection of image sequence, $\psi_0 \approx 3.64$ radians and $V_0 \approx 0.18$ pixel/frame.

VI. CONCLUSION

In this paper we have applied HMM filtering techniques to the problem of target heading angle estimation. Specifically, we have proposed a novel target heading angle estimator that exploits an interesting connection between relative entropy and probabilistic distance concepts. Our simulation studies demonstrate the superiority of our proposed heading angle estimator over track-before-heading-estimation approaches.

REFERENCES


Chapter 8

Conclusions

In this thesis we have proposed a target detection algorithm for the UAV sense-and-avoid application based on a two-stage processing paradigm. An image morphology pre-processing stage is coupled to a multiple hidden Markov model (HMM) temporal filtering stage to detect stationary dim sub-pixel size features in the image plane that are characteristic of potential collision-course targets. We have characterised the performance of the proposed algorithm using real in-flight target image data, and demonstrated the practical feasibility of the proposed algorithm via a graphic processing unit (GPU) based hardware implementation.

The target detection process is a vital function of any collision avoidance system, and our detection algorithm can be considered a step forward in addressing the airspace integration challenges for non-piloted aircraft. When combined with a suitable GPU-based device and camera sensor, our proposed detection algorithm can be used to deliver a real-time target detection capability ready for deployment onto a UAV platform (should a fixed non-stabilised camera be used however, an image jitter compensation technique is required to correct the raw sensor data prior to detection). It is anticipated that the proposed algorithm may be interfaced with appropriate sensing and avoidance modules to form a prototype end-to-end collision avoidance system, with a view to conducting flight trials.
whereby the performance of the entire system may be validated. Success here will contribute significantly to the removal of one of the most important obstacles hindering the integration of UAVs into civil airspace, where arguably the greatest potential of UAVs lie.

We also highlight that the development of the target detection algorithm has occurred alongside the discovery of several powerful relative entropy rate (RER) concepts. For example, the design of our multiple HMM temporal filter is aided by joint RER concepts that allow the filter design process to be cast as a min-max optimisation problem based on a joint RER cost criterion. This joint RER cost criterion is shown to bound the conditional mean estimate performance of the filter, establishing an important connection between the filter design process and expected filtering performance. As a consequence, there exists a unique insight into the relative filtering performance of candidate HMM filter designs simply through their respective joint RER design costs. We argue that this is a much more elegant and time-efficient approach than traditional methods of evaluating competing filter designs through Monte Carlo simulations.

Furthermore, we highlight that multiple HMM filter design is but one of many possible applications of the RER concepts that we have uncovered. The joint RER concepts that were so readily applied to the multiple filter design problem are easily transferrable to a related but more general HMM approximation problem. The HMM approximation problem often arises when HMMs are needed to approximate complex system dynamics to simplify computations and to allow tractable solutions to be developed. We show that a joint RER cost criterion may be used to determine suitable HMM approximation models, even when the system under approximation does not inherently obey the constraints of HMMs and there is uncertainty in the system dynamics. We have also established a connection between joint RER and probabilistic distance quantities that allows an unknown system parameter to be estimated from HMM filtering outputs. This is illustrated in a heading angle estimation problem using specially
designed HMM filters.

Throughout our papers we have demonstrated our RER ideas in the context of HMMs and HMM filtering techniques. However, the information theory body of knowledge which underlies all the RER related theorems and connections that we establish is not specific to any particular class of models or processes. From this general perspective is where we can gain a real appreciation of the significant RER concepts that we have unearthed. In this thesis, we have presented filter design, model approximation, and parameter estimation concepts that may be applied to models from any class, not just HMMs. We do concede that the application of these concepts in practice is subject to the existence and ability to evaluate the RER between different models and processes. When there is no closed-form solution for the RER in a particular scenario, we may exploit Monte Carlo or numerical approximations of the RER quantity.

8.1 Summary of Key Contributions

In this thesis, our proposed image-based target detection algorithm offers UAVs a much sought after target detection capability that can play a significant role in resolving the collision avoidance issue for uninhabited aircraft. In the development of the target detection algorithm, we have:

- Identified suitable image processing techniques for dim target detection (C1 and C2);
- Illustrated the benefits of a multiple model filtering approach (C2, J1 and J2);
- Proposed a novel RER-based filter design process (J1);
- Characterised detection performance based on real in-flight target image data (J3); and
8.2. Future Work

We have identified some possible areas for further consideration and future research:

- In this thesis, we have mainly dealt with RER concepts for HMMs. The generalisation of our RER concepts to other model classes is a promising avenue of future research.

- We have shown that our joint RER cost criterion establishes a conditional mean estimate performance bound for our multiple HMM filtering approach. Further analysis can be conducted to possibly determine the ‘tightness’ of the performance bound.

- For our target detection algorithm, we have found that employing multiple HMM filters in the temporal processing stage provided an improvement.

Furthermore, through the novel use of RER concepts, we have:

- Proposed general solutions to important problems in model approximation and filter design (J1);

- Applied general results to the class of HMMs to address HMM approximation and HMM filter design problems (J1);

- Established performance bounds for an RER-based HMM filter design process (J1); and

- Demonstrated the versatility of RER-based techniques and its potential in other applications (J2).
in detection performance compared with using just a single HMM filter. Although we implemented a bank of four parallel HMM filters in our simulation studies, we do not claim this to be the ‘optimal’ number of filters under any circumstances. We anticipate a diminishing return in performance gain with increasing filter numbers, but the exact relationship is not known. Characterising this relationship can help determine the optimal number of filters to employ in particular detection scenarios.
Appendix A

Definition of Authorship and Contribution to Publication

• The following is an extract from QUT’s “Statement of Authorship and Location of Data” form
ADDITIONAL INFORMATION

Authorship:
The Joint National Health and Medical Research Council/Australian Vice Chancellors’ Committee Statement and Guidelines on Research Practice states that:

“Authorship is substantial participation, where all the following conditions are met:

(a) (i) conception and design, or 
(ii) analysis and interpretation of data; and

(b) (i) drafting the article or 
(ii) revising it critically for important intellectual content; and

(c) final approval of the version to be published.

Participation solely in the acquisition of funding or the collection of data does not justify authorship. General supervision of the research group is not sufficient for authorship.”

Data storage and retention:
Wherever possible, original data should be retained in the school. Data should be safely held for as long as readers of publications might reasonably expect to be able to raise questions that require reference to it. This should be at least five years, or in the case of clinical research, 15 years may be more appropriate. Where it is impossible or impracticable to hold data, a written indication of the location of the data, or key information regarding its location (eg the way in which it is called up from a limited access database), must be kept in the school.

Further considerations:
- These guidelines are in agreement with those set down in the QUT MANUAL OF POLICIES AND PROCEDURES and the AVCC/NHMRC Guidelines on Authorship.
- A “Statement of authorship and location of data” form should be completed for each manuscript stating the area of contribution of each of the authors. Circulation of the draft manuscript plus this form within the research group should provide all potential authors with the opportunity to make a case for their inclusion as an author.
- Conception of the project idea relates to the topic of the manuscript to be published and would relate to the original major project plan or to major project redirections during the course of the work.
- Analysing and interpreting the data would normally relate to regular review and interpretation of the data, rather than review and input at a single time point. An exception to this might be epidemiological papers where this aspect of the work is extensive and critical.
- Revising the article critically means making significant changes to the manuscript to alter its impact. It does not mean simply reading the manuscript and offering some useful comments.
- Successfully obtaining grant funds to carry out the project in no way grants authorship. However, getting a grant should indicate a significant input to original project ideas and should allow for continued input into the project.
- When the manuscript makes use of a technique which was developed by a previous worker, it would be usual practice to acknowledge that person for their contribution rather than include them as an author.
- Resolving disputes: preferably, all authors with the first author taking a significant role should discuss authorship openly at an early stage. However, if agreement cannot be reached within the research group, then the corresponding author may attempt to resolve any authorship disputes via the School/Faculty Research Committee Chair. If dispute still exists, QUT MANUAL OF POLICIES AND PROCEDURES dispute resolution procedures should be followed.
Appendix B

Manuscript J2 Proof of Submission

- The following email confirms paper submission and assignment of manuscript number.
From: onbehalfof+z.kowalski+ieee.org@manuscriptcentral.com on behalf of z.kowalski@ieee.org
Sent: Wednesday, 10 March 2010 4:05 PM
To: John Lai; JOHN LAI
Cc: John Lai; JOHN LAI; Jason Ford; Peter O'Shea; Luis Mejias Alvarez; z.kowalski@ieee.org
Subject: Original manuscript submitted - T-SP-09875-2010 "RER based Target Heading Angle Estimation using HMM Filters"

10-Mar-2010

Dear Mr. John Lai,

Your manuscript, T-SP-09875-2010, entitled "RER based Target Heading Angle Estimation using HMM Filters" has been successfully submitted online and is presently being given full consideration for publication in the IEEE Transactions on Signal Processing as a Regular Paper.

Author's Guide link: http://mcv3help.manuscriptcentral.com/tutorials/Author.pdf

Please mention the above manuscript ID in all future correspondence or when contacting the IEEE Signal Processing Society for questions. If there are any changes in your account information or e-mail address, please log in to Manuscript Central at http://mc.manuscriptcentral.com/tsp-ieee and edit your user information as appropriate.

You can also view the status of your manuscript at any time by checking your Author Center after logging in to http://mc.manuscriptcentral.com/tsp-ieee.

NOTE: The review process will not begin until the all required items listed below are provided. If they were not provided they can be sent to Mr. Ziggy Kowalski at z.kowalski@ieee.org or your manuscript will be unsubmitted.

1) A Single Column Double Spaced version (30 pages or less for regular papers, 12 pages or less for correspondence papers). See the extract of the Information for Authors below;

2) A Double Column Single Spaced version;

3) You will need to email me a .pdf file for each of the references listed in your paper that have NOT YET been published or otherwise hard to access for reviewers. Once received, I will upload them for you on Manuscript Central as supporting information so that it can be accessible to reviewers.

IMPORTANT:
Please name the files according to how they are numbered in your list of references:

For example, files for the following imaginary references:
[2] Author name, Title, Journal, etc.
[14] Author name, Title, Journal, etc.
Would be named:
REFERENCE_2.pdf
REFERENCE_11.pdf
REFERENCE_14.pdf

The Transactions on Signal Processing has dedicated itself to providing the best possible publication opportunity. To that end, all individuals connected with the review and publication process have been requested to do their utmost to maintain the quality of the journal. We appreciate you selecting the Transactions on Signal Processing.

1
Sincerely,

Mr. Ziggy Kowalski
Coordinator Society Publications
IEEE Signal Processing Society
z.kowalski@ieee.org

(From Information for Authors at http://www.signalprocessingsociety.org/publications/periodicals/tsp/tsp-author-info/):

Electronic manuscript: Prepare a PDF file containing your manuscript in double-spaced format (one full blank line between lines of type) using a font size of 11 points or larger, having a margin of at least 1 inch on all sides. For a regular paper, the manuscript may not exceed 30 double-spaced pages, including title; names of authors and their complete contact information; abstract; text; all images, figures and tables; and all references. (If you are using LaTeX you may need to use \documentclass[11pt,onecolumn,draftcls]{IEEEtran} as the first line of the manuscript source file)

Double-column version of manuscript: You are required to also submit a roughly formatted version of the manuscript in single-spaced, double column IEEE format (10 points for a regular submission or 9 points for a Correspondence) using the IEEE style files (it is allowed to let long equations stick out). This version will serve as a confirmation of the approximate publication length of the manuscript at submission, and gives an additional confirmation of your understanding that overlength page charges will be paid when billed.
Appendix C

Manuscript J3 Proof of Submission

- The following email confirms paper submission and assignment of manuscript number.
10-Mar-2010

Dear Mr. Lai,

Your manuscript entitled "Airborne Vision-based Collision-Detection System" has been successfully submitted online and is presently being given full consideration for publication in Journal of Field Robotics.

Your manuscript number is ROB-10-0043. Please mention this number in all future correspondence regarding this submission.

You can view the status of your manuscript at any time by checking your Author Center after logging into http://mc.manuscriptcentral.com/rob. If you have difficulty using this site, please click the 'Get Help Now' link at the top right corner of the site.

Thank you for submitting your manuscript to Journal of Field Robotics.

Sincerely,

Journal of Field Robotics Editorial Office
Reference List


