Detection of Aircraft Below The Horizon for Vision-Based Detect And Avoid in Unmanned Aircraft Systems

Timothy L. Molloy, Jason J. Ford, and Luis Mejias

School of Electrical Engineering and Computer Science

Queensland University of Technology (QUT)

Brisbane, Queensland, Australia 4001

{t.molloy,j2.ford,luis.mejias}@qut.edu.au

Abstract

Vision-based aircraft detection technology may provide a credible sensing option for automated detect and avoid in small to medium size fixed-wing unmanned aircraft systems (UAS). Reliable vision-based aircraft detection has previously been demonstrated in sky-region sensing environments. This paper describes a novel vision-based system for detecting aircraft below the horizon in the presence of ground clutter. We examine the performance of our system on a data set of 63 near collision encounters we collected between a camera-

equipped manned aircraft, and a below-horizon target. In these 63 encounters,

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/rob.21719

Figure 1: A camera frame showing an aircraft (marked with a rectangle and in bottom right insert) below the horizon and in front of structured ground clutter including roads, buildings, trees, hills, fields, and dams. The aircraft is approximately 1.26 km from the camera.

our system successfully detects all aircraft, at an average detection range of 1890m (with a standard error of 43m and no false alarms in 1.1 hours). Furthermore, our system does not require access to inertial sensor data (which significantly reduces system cost), and operates at over 12 frames per second.

1 Introduction

A key barrier to the routine operation of unmanned aircraft systems (UAS) in civilian airspace is the current lack of an automated replacement of the human pilot’s ability to see and avoid other aircraft (Clothier et al., 2015). The main challenge in achieving autonomous detect and avoid is reliably detecting aircraft with sensors that can be carried on UAS due to cost, size, weight, and power considerations. For small to medium sized fixed-wing UAS, vision is the leading sensing technology for detecting aircraft that are uncooperative (in the sense that they do not share information) (Mcfadyen and Mejias, 2016). Whilst vision-based techniques for reliably detecting aircraft against sky backgrounds have been experimentally demonstrated, only partial progress has been made in the detection of aircraft against structured backgrounds such as the ground clutter in Figure 1 (Cho et al., 2013; Nussberger et al., 2014, 2015; Zarandy et al., 2015). In this paper, we propose a vision-based system for reliably detecting potential collision threats that appear below the horizon (against ground clutter).

Over the last decade, several research and industry groups have conducted flight experiments that demonstrate the feasibility of detecting potential collision threats using vision [see (Pellebergs, 2010; Mcfadyen and Mejias, 2016) and references therein]. Most recently, real-time vision-based aircraft detection has been demonstrated on unmanned aircraft (Cho et al., 2013; Zarandy et al., 2013, 2014), and large scale experiments have been conducted using manned aircraft as surrogate UAS platforms for data capture (Lai et al., 2013; Fasano et al., 2014; Nussberger et al., 2014). These recent experimental studies have highlighted that vision-based aircraft detection systems need to be able to detect collision aircraft in three distinct visual sensing environments: the unstructured sky-region above the horizon; the structured ground clutter below the horizon; and, the interface between the two regions near the horizon. Furthermore, these experiments have demonstrated that potential collision threats are often visually inconspicuous in the sense that they are slow moving, and occupy a small number of pixels in images from current camera systems (Lai et al., 2011). Detecting potential collision threats with vision at ranges that allow for the execution of collision avoidance manoeuvres is therefore a challenging problem.

Considerable effort has recently been dedicated to developing vision-based systems for detecting potential collision threats that appear from unstructured sky backgrounds (Gandhi et al., 2006; Carnie et al., 2006; Lai et al., 2011, 2013; Dey et al., 2011; Nussberger et al., 2014, 2015). Many of these systems have been built on a two-stage filtering paradigm involving both spatial filtering (to improve the contrast of small image features), and temporal filtering (to emphasise slowly moving objects that persist over multiple frames). By exploiting morphological spatial filters and hidden Markov model (HMM) temporal filters, systems based

on the two-stage paradigm now offers low false alarm rates and detection ranges suitable for autonomous mid-air collision avoidance in the sky-region (Lai et al., 2013).

Several groups have recently made partial progress in developing vision-based systems for detecting potential collision threats when they appear below the horizon against structured ground clutter (Cho et al., 2013; Nussberger et al., 2014, 2015; Zarandy et al., 2015). These systems have predominantly used frame differencing techniques to detect collision threats that have apparent motion with respect to background ground clutter. Although these techniques have had their initial feasibility demonstrated, detection systems based on frame differencing may fail during true collision encounters when collision threats have small apparent motion between consecutive frames (Zarandy et al., 2015). Furthermore, frame differencing is often insufficient for rejecting ground clutter, and so current state-of-the-art systems in (Nussberger et al., 2014) and (Nussberger et al., 2015) also use digital terrain maps. Reliably detecting potential collision threats below the horizon solely on the basis of vision is therefore a significant open problem.

The main contributions of this paper are: (i) the proposal of a vision-based system for detecting aircraft below the horizon in the presence of structured ground clutter; (ii) the application of HMM temporal filters to detect targets in ground clutter; and (iii) an extensive study of the detection range and missed detection performance of our system in 63 head-on collision-like encounters we collected during flight experiments with two manned aircraft. Although our experimental evaluation is insufficient for drawing strong conclusions about the false alarm performance of our system, it is more extensive than all previous efforts to characterise vision-based collision threat detection below the horizon. Indeed, the volume

of our experimental data enables us to access information about the reliability of detection (e.g., the mean detection range as well as the standard error of the mean detection range).

The remainder of this paper is structured as follows. In Section 2, we provide a brief background of below-horizon aircraft detection by describing the characteristics of potential collision threats, and computer vision techniques that have previously been applied in related applications. In Section 3, we present our system for detecting potential collision threats in the presence of structured ground clutter. In Section 4, we describe the implementation of our system (including its computational speed and complexity). In Section 5, we describe our flight experiments and report the results of several studies we conducted. Finally, in Section 6, we discuss lessons learnt before presenting conclusions in Section 7.

2 Background and Related Work

In this section, we provide a brief background discussing the visual characteristics of potential collision threats, and we review several existing approaches to vision-based object detection.

2.1 Visual Characteristics of Below-Horizon Potential Collision Threats

Visually detecting potential collision threats below the horizon is challenging due to their subtle visual characteristics, and the presence of ground clutter (Nussberger et al., 2014). Specifically, aircraft on collision courses subtend small visual angles until collision is imminent, and maintain a near-constant relative bearing (Morris, 2005). However, in many cases, potential collision threats have (small) apparent motion relative to the ground clutter

(a)
(b)
(c)
(d)

Figure 2: Example motion of (a) a potential collision threat, (b) a collision threat, (c) a ground object, and (d) a collision threat with relative motion similar to that of a ground object. Figure adapted from (Nussberger et al., 2014).

(even when they are established on a collision course geometry, cf. Figures 2a, 2b, and 2c) (Nussberger et al., 2014). In these cases, the key challenges associated with visually detecting potential collisions threats from a moving aircraft are: (i) detecting image features that move relative to ground clutter; (ii) rejecting stationary ground objects (as in Figure 2c); and, (iii) rejecting moving ground objects (e.g., cars on roads). In this paper, we will focus on methods for overcoming challenges (i) and (ii), and we will later discuss the difficulties of rejecting moving ground objects.

2.2 Related Work

Recent vision-based systems for detecting potential collision threats below the horizon have been heavily inspired by techniques for detecting and tracking moving ground objects (e.g., people, vehicles, and cars) from aerial platforms (May and Krouglicof, 2013; Cho et al., 2013; Nussberger et al., 2014, 2015). A notable example of a moving ground object detection system being modified to detect potential collision threats below the horizon is presented in Nussberger et al. (2014). The system of (Nussberger et al., 2014) [borrowed from (Saleemi and Shah, 2013)] involves performing image registration (i.e. computing and applying an homography from tracked image features) between pairs of consecutive frames, followed by

two-frame differencing to determine image regions that have changed during the inter-frame period. A bank of extended Kalman filters is then used to track candidate detections from the frame differences, and those found to move with respect to the ground based on a digital terrain map are declared as target detections (Nussberger et al., 2014).

As an alternative to frame differencing, May and Krouglicof (2013) propose a Fourier domain technique based on regional phase correlation for detecting moving aircraft in a scene. This regional phase correlation approach has characteristics that make it attractive for real-time implementation on unmanned aircraft with field programmable gate arrays (FPGAs). However, the approach only performs a search for motion between consecutive frames, which limits the range at which aircraft can be detected.

Importantly, Zarandy et al. (2015) recently noted that true collision threats can appear stationary with respect to ground clutter (as illustrated in Figure 2d). These collision threats are unlikely to be detected by systems that search for aircraft with large apparent motion between two consecutive frames. Recently, progress has also been made in detecting aircraft from their visual appearance (rather than motion) by training convolutional neural networks (CNNs) to recognise aircraft in unstabilised image sequences (Rozantsev et al., 2015). Although this appearance-based approach is promising, its performance is fundamentally limited by the availability of training data (which is currently extremely limited and difficult to collect for all classes of aircraft). Detection approaches that search for aircraft based on their visual motion characteristics therefore remain an important cornerstone of vision-based aircraft detection, although there is a strong case for examining approaches that detect small motions over multiple consecutive frames.

In order to detect small slow moving ground objects from aerial platforms, Reilly et al. (2010) and Schubert and Mikolajczyk (2014) have recently proposed multi-stage multi-frame image processing pipelines. The approaches of Reilly et al. (2010) and Schubert and Mikolajczyk (2014) involve processing multiple frames with image registration, background subtraction, and motion candidate tracking stages. In (Reilly et al., 2010), image registration is performed by estimating homographies between frames using tracked image features (e.g. Harris corners and SIFT features). In contrast, Schubert and Mikolajczyk (2014) propose a Fourier phase correlation method of homography estimation that avoids the problem of selecting features to track by operating globally over full frames. The phase correlation approach of (Schubert and Mikolajczyk, 2014) is also well suited for implementation on parallel computing hardware such as graphics processing units (GPUs) and FPGAs.

By performing image registration with homographies, Reilly et al. (2010) and Schubert and Mikolajczyk (2014) assume a planar scene. As noted by Reilly et al. (2010), this assumption is often violated, and is particularly problematic when the scene contains out of plane objects such as buildings. These homography-based image registration approaches therefore perform poorly near the edges of out of plane objects. As a result, the background subtraction stages of (Reilly et al., 2010) and (Schubert and Mikolajczyk, 2014) produce motion frames with false motion candidates near the edges of out of plane objects. Reilly et al. (2010) propose a simple approach to remove many false motion candidates by suppressing regions of the background with strong image gradient magnitudes. In contrast, Schubert and Mikolajczyk (2014) rely on temporal filtering with optical flow and ad-hoc dynamic programming to reject false motion candidates and perform final moving object detection.

In this paper, we propose a system for detecting aircraft in ground clutter below the horizon. Our system is inspired by the ground object detection approaches of Reilly et al. (2010) and Schubert and Mikolajczyk (2014), but uses a new method for overcoming key challenges in rejecting false targets in motion frames. In particular, recent HMM temporal filtering developments in (Lai et al., 2013) for sky-region vision-based aircraft detection have led to HMM filters outperforming ad-hoc dynamic programming approaches similar to those used by Schubert and Mikolajczyk (2014). We propose exploiting these recent HMM temporal filtering developments to detect below-horizon aircraft by combining them with a modified version of the gradient suppression approach proposed in (Reilly et al., 2010).

3 Vision-Based Detection System

In this section, we present our system for detecting potential collision threats in a sequence of 8-bit grayscale camera frames g_k where $k \geq 1$ is the frame number (or time index). After each new frame g_k is received, our system processes the window of frames $\mathcal{W}_k \triangleq \{I_{(k,1)}, I_{(k,2)}, \dots, I_{(k,w)}\}$ where $I_{(k,\ell)}$ denotes the frame $g_{k-w+\ell}$ with intra-window index $1 \leq \ell \leq w$, and w is the length of the window (for ease of presentation we assume that w is odd). The main stages of our system are shown in Figure 3, and include: (i) image registration; (ii) background and gradient subtraction; and (iii) temporal filtering.

As illustrated in Figure 3, at each time step k , the frames in the window $\mathcal{W}_k$ are registered to a common reference frame to remove camera ego-motion (which is present because the camera is mounted on a moving aircraft). This image registration stage produces the window of

Figure 3: Stages of our vision-based detection system. The frames are converted to 8-bit grayscale prior to image registration.

registered frames $\mathcal{R}_k \triangleq \{r_{(k,1)}, r_{(k,2)}, \dots, r_{(k,w)}\}$ where $r_{(k,\ell)}$ is the registered version of frame $I_{(k,\ell)}$ for $1 \leq \ell \leq w$. We use the registered frames $\mathcal{R}_k$ to estimate a background image b_k of the static ground clutter in the scene. Motion frames $\mathcal{M}_k \triangleq \{m_{(k,1)}, m_{(k,2)}, \dots, m_{(k,w)}\}$ showing regions of candidate moving targets are then produced by subtracting the background image b_k from the registered frames (and by also suppressing regions of strong gradients in the background image). Finally, temporal filtering is used to reject false targets in the motion frames $\mathcal{M}_k$ due to noise and other artefacts being present after image registration and background subtraction. We now describe the details of each stage of our system.

3.1 Image Registration

In the image registration stage of our system, we estimate a homography between each frame $I_{(k,\ell)}$ in the window $\mathcal{W}_k$ and the centre frame $I_{(k,[w+1]/2)}$. We apply these homographies to the frames in the window $\mathcal{W}_k$ to construct the window of registered frames $\mathcal{R}_k$. We perform image registration to the centre frame of the window $\mathcal{W}_k$ since it is the frame closest in time to all other frames in $\mathcal{W}_k$. We also compensate for potential errors in image registration by downsizing the registered frames in $\mathcal{R}_k$ from 1280×960 pixels to 1024×768 pixels using linear interpolation (which results in no reduction in the field of view). Here, we shall present two existing methods for estimating the required homographies. The first method is a standard approach using image features (Szeliski, 2006), whilst the second method is the Fourier-based phase correlation approach previously proposed in (Schubert and Mikolajczyk, 2014). We

Figure 4: Stages in Image Registration with Phase Correlation.

shall later experimentally compare these two image registration approaches in our system.

3.1.1 Homography Estimation with Image Features

Homography estimation with image features is arguably the standard approach to image registration (Szeliski, 2006). In the feature-based approach, we first detect local features in each frame of the window $\mathcal{W}_k$. We then compute local descriptors for each feature with either the speeded-up robust features (SURF) algorithm or the scale-invariant feature transform (SIFT) algorithm. We finally estimate the homography between each frame $I_{(k,\ell)}$ in $\mathcal{W}_k$ and the centre frame $I_{(k,[w+1]/2)}$ using matched pairs of descriptors and random sample consensus (RANSAC) based robust least squares regression.

3.1.2 Homography Estimation with Phase Correlation

The second method of homography estimation that we consider exploits the phase correlation properties of translated images. As illustrated in Figure 4, the phase correlation approach is a multi-stage process. We first partition each frame $I_{(k,\ell)}$ in the window $\mathcal{W}_k$ into a collection of N tiles (or regions) $\mathcal{I}_{(k,\ell)} \triangleq \{t_{(k,\ell)}^1, t_{(k,\ell)}^2, \dots, t_{(k,\ell)}^N\}$. We then compute the 2D fast Fourier transform (FFT) $T_{(k,\ell)}^j$ of each tile $t_{(k,\ell)}^j$ from the collection $\mathcal{I}_{(k,\ell)}$ (i.e. we compute $T_{(k,\ell)}^j$ for all $1 \leq \ell \leq w$ and all $1 \leq j \leq N$). We apply a Blackman window function to each tile before computing its FFT in order to prevent edge effects caused by the circular nature of the FFT [see (Harris, 1978) for details].

After computing the FFT of each tile in each frame of the window $\mathcal{W}_k$, we compute the cross-power spectrums (Szeliski, 2010, p. 390)

$$F_{(k,\ell)}^j = \frac{T_{(k,\ell)}^j \cdot \bar{T}_{(k,[w+1]/2)}^j}{|T_{(k,\ell)}^j \cdot \bar{T}_{(k,[w+1]/2)}^j|}$$

between the j th tile in frame $I_{(k,\ell)}$ and the j th tile in the centre frame $I_{(k,[w+1]/2)}$ for all $1 \leq j \leq N$ and all $1 \leq \ell \leq w$ with $\ell \neq (w+1)/2$. Here, the multiplication and division operations are element-wise, and $\bar{T}_{(k,[w+1]/2)}^j$ denotes the complex conjugate of $T_{(k,[w+1]/2)}^j$. We then calculate the normalised cross-correlations $f_{(k,\ell)}^j$ by performing the inverse FFT of the cross-power spectrums $F_{(k,\ell)}^j$ for all $1 \leq j \leq N$ and all $1 \leq \ell \leq w$ with $\ell \neq (w+1)/2$.

The normalised cross-correlations $f_{(k,\ell)}^j$ provide information useful for estimating the horizontal and vertical translations $(\Delta x_{(k,\ell)}^j, \Delta y_{(k,\ell)}^j)$ of the j th tile in frame $I_{(k,\ell)}$ from the j th tile in the centre frame $I_{(w+1)/2}$ (Szeliski, 2010, p. 390). Let $f_{(k,\ell)}^j(x, y)$ denote the value of the normalised cross-correlation $f_{(k,\ell)}^j$ at the image Cartesian location (x, y) . Due to the shift properties of the Fourier transform, an estimate of the translation $(\Delta x_{(k,\ell)}^j, \Delta y_{(k,\ell)}^j)$ is given by (Szeliski, 2010, p. 390)

$$(\Delta x_{(k,\ell)}^j, \Delta y_{(k,\ell)}^j) = \arg \max_{(x,y)} f_{(k,\ell)}^j(x, y)$$

for all tiles $1 \leq j \leq N$ and all $1 \leq \ell \leq w$ with $\ell \neq (w+1)/2$. We then estimate the homography between each frame $I_{(k,\ell)}$ in $\mathcal{W}_k$ and the centre frame $I_{(k,[w+1]/2)}$ from the tile translations $(\Delta x_{(k,\ell)}^j, \Delta y_{(k,\ell)}^j)$ for all $1 \leq j \leq N$ using RANSAC based robust least squares regression.

3.2 Background and Gradient Subtraction

In the second stage of our system, we use the registered 8-bit grayscale frames $\mathcal{R}_k$ to estimate the static background of the scene together with its gradient magnitude. We first estimate the static background of the scene, b_k , by calculating the pixel-wise median of the set of registered frames $\{r_{(k,1)}, r_{(k,2)}, \dots, r_{(k,[w+1]/2)}\}$ (i.e., the first half of $\mathcal{R}_k$). By computing the pixel-wise median of the first half of the window, we aim to avoid including slow moving targets that appear later in the window $\mathcal{R}_k$ in the estimate of the background.

Following background estimation, we next calculate a contrast-enhanced image of gradients in the background frame b_k . As in (Saleemi and Shah, 2013), we define the contrast-enhanced image of gradients as

$$e_k \triangleq \exp(\|\nabla b_k\|)^\gamma$$

where $\gamma > 0$ is a design constant, and $\|\nabla b_k\|$ is the spatial gradient magnitude of the background frame b_k . Here, the gradient magnitude $\|\nabla b_k\|$ is normalised before computing the exponential, and the exponential is normalised before raising it to the power of γ . As described in (Saleemi and Shah, 2013), the design constant $\gamma > 0$ is useful for highlighting regions of the background frame b_k with gradient information in e_k . Due to the normalisation of both $\|\nabla b_k\|$ and e_k to 8-bit grayscale images, regions of b_k with small spatial gradient magnitudes are assigned low intensity values in the image e_k for large values of γ (e.g., $\gamma > 1.5$). Conversely, for small values of γ (e.g., $\gamma < 1.5$), regions of b_k with both small and large spatial gradient magnitudes are assigned high intensity values in the image e_k . We

shall later report that in our experiments $\gamma = 1.5$ offered suitable system performance.

Given the background b_k and gradient e_k images, we identify regions of potential motion by subtracting them from each of the registered frames $r_{(k,\ell)}$ in the (full) window of registered frames $\mathcal{R}_k$. In particular, we compute the set of motion frames $\mathcal{M}_k \triangleq \{m_{(k,1)}, m_{(k,2)}, \dots, m_{(k,\ell)}\}$ where

$$m_{(k,\ell)} = |r_{(k,\ell)} - b_k| - (e_k \oplus S)$$

for all $1 \leq \ell \leq w$, and where $\oplus$ denotes morphological dilation by the structuring element S (we select S as a 7×7 pixel square). Similar to (Reilly et al., 2010), here we propose subtracting the morphologically dilated version of the gradient image e_k in order to suppress large intensity differences between the registered frames and the background image that arise due to imperfect image registration (rather than true moving objects).

3.3 Temporal Filtering

The final stage of our detection system involves searching the window of motion frames $\mathcal{M}_k$ for small features that move slowly across the image-plane. In this paper, we consider two temporal filtering approaches: a hidden Markov model (HMM) temporal filtering approach similar to those used in the sky-region work of (Lai et al., 2008, 2011, 2013); and, an ad-hoc Viterbi temporal filtering approach similar to those used in the sky-region work of (Carnie et al., 2006) and (Gandhi et al., 2006). Both temporal filtering approaches allow us to adopt a track-before-detect approach in which we declare target detections based on expected target

motion.

3.3.1 HMM Temporal Filtering

In order to introduce our HMM temporal filter, we require an HMM representation of aircraft image motion. Following (Lai et al., 2013), we assign each pixel (m, n) in the $N_h \times N_v$ pixel image to a unique *state* of the HMM representation. We assume that the aircraft occupies pixel $X_{(k,\ell)}$ in motion frame $m_{(k,\ell)} \in \mathcal{M}_k$ at time $k \geq 1$ where $X_{(k,\ell)}$ takes values in the set $\mathcal{X} \triangleq \{1, 2, \dots, N_v N_h\}$, and where $1 \leq \ell \leq w$. The aircraft motion is then naturally observed as transitions between pixels, and the key parameters of our HMM representation are:

1. *Intra-Window Transition Probabilities* $A^{ij} \triangleq P(X_{(k,\ell+1)} = i | X_{(k,\ell)} = j)$ that describe the probability of the aircraft transitioning from pixel j to pixel i between motion frames $m_{(k,\ell)}$ and $m_{(k,\ell+1)}$ within the window $\mathcal{M}_k$ (here, $1 \leq \ell < w$);
2. *Inter-Window Transition Probabilities* $A_{\mathcal{M}}^{ij} \triangleq P(X_{(k+1,1)} = i | X_{(k,w)} = j)$ that describe the probability of the aircraft transitioning from pixel j to pixel i between the last motion frame $m_{(k,w)}$ of the window $\mathcal{M}_k$, and first motion frame $m_{(k+1,1)}$ of next window $\mathcal{M}_{k+1}$; and
3. *Measurement Probabilities* $B^i(m_{(k,\ell)}) = P(m_{(k,\ell)} | X_{(k,\ell)} = i)$ that describe the likelihood of observing the motion frame $m_{(k,\ell)}$ given that the aircraft is located in the i th pixel in $m_{(k,\ell)}$ of the window $\mathcal{M}_k$ (here, $1 \leq \ell \leq w$)

where $1 \leq i, j \leq N_v N_h$, and $k \geq w$.

Figure 5: Example of the 3×3 pixel neighbourhood (centred on each pixel) of permitted transitions under the intra-window transition matrix A and intra-window set Q .

We shall collect the *intra-window* transition probabilities in the matrix $A \in \mathbb{R}^{N_v N_h \times N_v N_h}$, and the *inter-window* transition probabilities in the matrix $A_{\mathcal{M}} \in \mathbb{R}^{N_v N_h \times N_v N_h}$. We highlight that both $\sum_{i=1}^{N_v N_h} A^{ij} = 1$ and $\sum_{i=1}^{N_v N_h} A_{\mathcal{M}}^{ij} = 1$ for all j . We shall model aircraft as slowly moving by limiting the possible transitions in the intra-window matrix A to the 3×3 pixel neighbourhood surrounding each pixel (as shown in Figure 5). We expand this neighbourhood of non-zero transition probabilities to 23×23 pixels in the inter-window transition matrix $A_{\mathcal{M}}$ in order to account for the time difference between the end of $\mathcal{M}_k$ and the start of $\mathcal{M}_{k+1}$. The transition matrices A and $A_{\mathcal{M}}$ are therefore both sparse with allowed transitions assigned probabilities of $1/9$ and $1/529$, respectively. To ensure that an HMM filter can be applied, we consider transitions that would otherwise describe the aircraft as leaving the image, as (artificially) describing transitions to the opposite side of the image.

In principle, the measurement probabilities $B^i(m_{(k,\ell)})$ can be estimated from observation data. However, similar to Lai et al. (2013), in this paper we directly interpret the intensities of the motion frames $m_{(k,\ell)}$ as approximate likelihoods in the sense that

$$B^i(m_{(k,\ell)}) = m_{(k,\ell)}^i + 1$$

for all $1 \leq i \leq N_v N_h$. We will use $B(m_{(k,\ell)})$ to denote the $N_v N_h \times N_v N_h$ diagonal matrix with i th element on the diagonal $B^i(m_{(k,\ell)})$.

We are now in a position to present our HMM temporal filter. The conditional mean estimate $\hat{X}_{(k,\ell)}$ of the aircraft pixel location $X_{(k,\ell)}$ in motion frame $m_{(k,\ell)}$ of $\mathcal{M}_k$ is given by the HMM filter recursion (Elliott et al., 1995)

$$\hat{X}_{(k,\ell)} = \begin{cases} N_{(k,\ell)} B(m_{(k,\ell)}) A \hat{X}_{(k,\ell-1)} & \text{if } 1 < \ell \leq w \\ N_{(k,\ell)} B(m_{(k,\ell)}) A_{\mathcal{M}} \hat{X}_{(k-1,w)} & \text{if } \ell = 1 \end{cases}$$

for $k \geq w$ where $N_{(k,\ell)}$ are scalar normalisation factors. For $k < w$, we initialise the HMM filter with the uniform distribution in the sense that $\hat{X}_{(k,\ell)}^i \triangleq 1/(N_v N_h)$ for all $1 \leq i \leq N_v N_h$, all $1 \leq \ell \leq w$, and all $k < w$.

The scalar normalisation factors $N_{(k,\ell)}$ produced by the HMM filter provide information about the likelihood of observing motion frames $m_{(k,\ell)}$ given that an aircraft is present (Elliott et al., 1995). We therefore detect aircraft on the basis of the moving average test statistic:

$$\eta_{(k,\ell)} \triangleq \begin{cases} (1 - \tau^{-1}) \eta_{(k,\ell-1)} + \tau^{-1} \log \frac{1}{N_{(k,\ell)}} & \text{if } 1 < \ell \leq w \\ (1 - \tau^{-1}) \eta_{(k-1,w)} + \tau^{-1} \log \frac{1}{N_{(k,\ell)}} & \text{if } \ell = 1 \end{cases}$$

for $k \geq w$. Here, we define $\eta_{(k,\ell)} \triangleq 0$ for all $\ell < w$ when $k < w$. Similar to the sky-region work of (Lai et al., 2011, 2013), $\tau > 0$ is an HMM forgetting parameter which we set to $\tau = 10w$.

We declare a candidate aircraft detection when the test statistic $\eta_{(k,\ell)}$ exceeds a given threshold $h > 0$. Selecting the detection threshold h involves making trade-offs between detection

and false alarm performance. In our later experiments, we shall select and report detection thresholds h that raise no false alarms on our collected data (i.e., zero false alarm thresholds).

3.4 Ad-Hoc Viterbi Temporal Filtering

Ad-hoc Viterbi temporal filtering is a long-time competitor to HMM temporal filtering in sky-region collision target detection (Lai et al., 2011, 2013). Motivated by these studies, later we shall examine it alongside HMM temporal filtering for below-horizon collision target detection.

Following Lai et al. (2013), let $Q \triangleq \{(1, 1), (0, 1), \dots, (-1, -1)\}$ denote the 3×3 neighbourhood of allowed intra-window transitions shown in Figure 5. Similarly, let $Q_{\mathcal{M}} \triangleq \{(11, 11), (10, 11), \dots, (-11, -11)\}$ denote the 23×23 neighbourhood of allowed inter-window transitions. We highlight that Q and $Q_{\mathcal{M}}$ specify the same 3×3 and 23×23 neighbourhood of valid intra- and inter- window transitions as the transition matrices A and $A_{\mathcal{M}}$ of the HMM temporal filter.

Now, let $D_{(k,\ell)}$ denote an $N_h \times N_v$ pixel image such that $D_{(k,\ell)}^{ij}$ is the intensity output of the ad-hoc Viterbi filter at pixel (i, j) . For $k \geq w$, the output of the ad-hoc Viterbi filter is given by the recursion

$$D_{(k,\ell)}^{ij} = \begin{cases} (1 - \beta)m_{(k,\ell)}^{ij} + \beta \max_{(m,n) \in Q} D_{(k,\ell-1)}^{i-m,j-n} & \text{if } 1 < \ell \leq w \\ (1 - \beta)m_{(k,\ell)}^{ij} + \beta \max_{(m,n) \in Q_{\mathcal{M}}} D_{(k-1,w)}^{i-m,j-n} & \text{if } \ell = 1 \end{cases}$$

for all $1 \leq i \leq N_v$ and $1 \leq j \leq N_h$ where $0 < \beta < 1$ is a Viterbi forgetting parameter (Lai

et al., 2013). Here, following the studies of (Carnie et al., 2006; Lai et al., 2011, 2013), we selected the forgetting parameter $\beta = 0.75$, and we initialised $D_{(k,\ell)}^{ij} = 0$ for all $1 \leq k \leq w$, all $1 \leq \ell \leq w$, all $1 \leq i \leq N_v$, and all $1 \leq j \leq N_h$. Intuitively, the output of the ad-hoc Viterbi filter output $D_{(k,\ell)}$ can be used to detect the presence of an aircraft in the sequence of windows $\mathcal{M}_k$. Here, we shall declare an aircraft detection when the test statistic,

$$\lambda_{(k,\ell)} \triangleq \max_{ij} D_{(k,\ell)}^{ij},$$

exceeds a given threshold $h_V > 0$ for any $k \geq w$, and $1 \leq \ell \leq w$.

Before we examine the performance of our detection system, we will describe several important implementation details.

4 Implementation Details

In this section, we describe the implementation of our detection system. We implemented all stages in C/C++ using the OpenCV library and the Nvidia Compute Unified Device Architecture (CUDA) application programming interface (API). OpenCV was used to perform general image manipulation and processing (for example, morphological dilation and gradient magnitude calculations). The CUDA API enabled us to implement the greediest computational stages of our system (e.g., image registration, temporal filtering, and background subtraction) on a GPU.

4.1 Image Registration Implementation

We implemented feature-based image registration as described in Section 3.1.1 using the SIFT and CUDA SURF routines from OpenCV 3 with the GPU brute-force matcher and the OpenCV 3 implementation of RANSAC in the function `findHomography`. We similarly implemented image registration with phase correlation as described in Section 3.1.2 using the CUDA fast Fourier transform (CUFFT) library, custom CUDA kernels, and the OpenCV 3 implementation of RANSAC in the function `findHomography`. In selecting the size and shape of the tiles for phase correlation, we gave consideration to the amount of inter-frame motion we needed to handle (since larger tiles enable larger motions to be estimated), and to computational burden (more tiles implies more FFTs and the CUFFT routines are optimised for inputs with dimensions that factor into powers of two). We aimed to select square tiles with side lengths L that factor into powers of 2 (e.g., a side length of 256 pixels). Furthermore, although overlapping tiles has previously been found to improve registration and motion estimation performance (Schubert and Mikolajczyk, 2014), for ease of implementation we only introduced tile overlap when tiles would otherwise exceed the dimensions of the frame. Similarly, we did not perform sub-pixel interpolation of the normalised cross-correlation f_i^j to determine translations with sub-pixel accuracy.

4.2 Temporal Filtering Implementation

The HMM and ad-hoc Viterbi temporal filtering stages were implemented in CUDA using predominantly custom CUDA kernels. Our implementations exploit the sparse structure of the transition matrices, and are similar to those presented previously by Lai et al. (2011)

and Mejias et al. (2010). Before we describe the complexity of our implementation, we highlight that our implementation is not overly optimised (e.g., it does not exploit host computer threading, and some routines like RANSAC are performed sequentially on the host computer rather than on the GPU).

4.3 Computational Effort

The computational effort of our detection system with phase correlation image registration and HMM temporal filtering is dependent on the choice of system parameters. These parameters (and their baseline values) are summarised in Table 1. We shall later discuss the computational effort associated with the feature-based methods of image registration.

At each time step k , our implementation of image registration with phase correlation involves: N forward FFTs at $O(L^2 \log L)$ each; $O(NL^2w)$ multiplications and divisions to compute cross-power spectrums; Nw inverse FFTs at $O(L^2 \log L)$ each to compute cross-correlations; Nw peak searches at $O(N)$ each to compute translations; and w RANSAC operations each with complexity less than $O(N^2)$. At each time step k , our background and gradient subtraction implementation involves: finding the pixel-wise median of w frames using the quick select algorithm with worst case complexity $O(w^2)$ (frames are each $N_h \times N_v$ pixels); and computing and subtracting the edge image e_k at $O(N_v N_h)$. It follows that at time step k , phase correlation image registration has total complexity $O(N^2 L^2 w)$, background and gradient subtraction has total complexity $O(N_v N_h w^2)$, and our HMM temporal filter has total complexity less than $O(N_v^2 N_h^2 w)$.

Author Manuscript

We benchmarked our implementation (with baseline parameters given in Table 1) on a Ubuntu 14.04 Linux desktop computer with an Intel i5-2400 CPU @ 3.10 GHz, 16 GB of memory, an Nvidia GeForce GTX 590 GPU, a 7200 RPM hard drive, CUDA 7.0, and OpenCV 3. The computation times for the different stages of our detection system are presented in Table 2. From Table 2, we see that the compute times of the registration, and temporal filtering stages are indeed linear in window length w (which agrees with our complexity claims). Importantly, our implementation achieves an overall frame rate that enables real-time performance on this desktop computer for window lengths less than 7 (15 Hz is desirable for real-time performance). Similar to previous observations made by Lai et al. (2012) and Mejias et al. (2010), the performance of our CUDA implementations will scale with future improvements in GPU hardware.

5 Flight Experiments and Results

In this section, we report the design and results of a series of flight experiments that we conducted to evaluate our vision-based detection system.

5.1 Flight Experiment Design and Data Collection

In order to collect data suitable for characterising the performance of our detection system, we conducted several flight experiment campaigns during the period April 2015 to July 2015 in the Watts Bridge area of Queensland. All of our flight experiments were near head-on encounters between two manned aircraft: a specially equipped Cessna 172R sensor aircraft

(a)

(b)

Figure 6: Sensor pod mounted on wing strut of sensor aircraft. (a) Pod with cover on for flight. (b) Pod with cover removed to show downward tilted sensor (and an IMU that is not used in this work).

[see Greer et al. (2010) for details]; and a Socata Trinidad target aircraft (with a wingspan of 9.97m).

We designed the data capture system and near collision geometry so that the target was 45 pixels wide before leaving the field of view. This target size corresponds approximately to the best detection results reported in (Nussberger et al., 2014). To achieve this target size, we equipped the sensor aircraft with an uncalibrated Basler Scout machine vision camera with a Navitar NMV-12m-23 lens. The sensor lens provided a field of view of 17.1×22.1 degrees, with an angular resolution of 0.0175 degrees per pixel. The camera was configured to capture 1280×960 pixel frames in 8-bit Bayer mode at approximately 15 Hz (we later converted all frames to grayscale). The sensor was mounted in a custom built pod attached to the Cessna’s starboard wing strut (see Figure 6a). The pod enabled the sensor to be tilted downwards to maximise the ground field of view coverage (see Figure 6b). For the purpose of collecting truth data, the sensor aircraft carried an FSAS Novatell SPAN GPS/INS system, and the target aircraft carried a GPS receiver.

Our encounter design followed the approach described in (Lai et al., 2013) with several alterations to accommodate a below-horizon target. In particular, the sensor aircraft was flown at a desired altitude of between 823m and 914m, and at an airspeed of between 46.3 m/s and 51.4 m/s. The target aircraft operated at a vertical separation greater than 60m

(a) (b)
(c) (d)

Figure 7: Example frames showing typical ground clutter in data captured with the sensor aircraft heading (a) North-East, (b) South-West, (c) South-East, and (d) North-West.

below the sensor aircraft.

Our experiments were conducted over four days between the hours of 10 am and 4 pm in scattered cloud lighting conditions with wind less than 10.3 m/s. Overall, we collected a total of 63 encounters which are summarised in Table 3. Within each encounter, there are frames where the target aircraft is visible (i.e., target frames), and frames where the target aircraft is not visible (i.e., non-target frames). We captured the target in front of a variety of ground clutter by flying the same encounter geometry with four different sensor aircraft headings – North-East, South-West, North-West, and South-East.

Figure 7 provides a visual overview of the ground clutter present in encounters with different sensor aircraft headings. Buildings, roads, fields, scattered vegetation, and distant hills are present in the ground clutter of all of our collected encounters. However, during encounters filmed with a North-East heading, hills dominate the ground clutter with a forested range present in the centre of the field of view. In contrast, a lake (Lake Wivenhoe) and its surroundings form the ground clutter during encounters filmed with a South-East heading. These South-East encounters also contain visible smoke from fires. Smaller bodies of water (rural dams) are present in the encounters filmed with a North-West heading.

5.2 Results

We now examine the performance of our system on the collected data. We shall perform several studies of our system that illustrate its practical performance with a variety of parameters, and how it compares to alternative approaches. In each of our studies, we shall report detection ranges and the number of missed detections for the minimum detection threshold h that achieves no false alarms on the non-target frames (i.e. zero false alarm thresholds). In the future, collection of additional non-target data will enable further examination of system performance at other thresholds and false alarm rates.

5.2.1 Temporal Filtering Study

In our first study, we compare the performance of our system with HMM temporal filtering with a version that uses an ad-hoc Viterbi temporal filtering stage. The other stages are as described in Section 3 (with phase correlation image registration), and we used the baseline parameter values given in Table 1. We also examine the sensitivity of our system's performance to choice of the threshold h in the HMM temporal filtering stage.

Figure 8 shows the detection ranges achieved on each case of our data set by our system. Importantly, our system detects all targets in our data set with no missed detections or false alarms. The detection ranges do not vary significantly with the type of ground clutter (cf. Figure 7). The minimum detection range that our system achieves is 1.26 km (on case number 63), and the maximum detection range is 2.6 km (on case number 3).

Figure 9 shows that the performance of the system with an ad-hoc Viterbi temporal filtering

Figure 8: Detection ranges at zero false alarms for each case in the data set using our detection system with an HMM temporal filtering stage. The ground clutter varies with the pass heading of the sensor equipped aircraft (see Figure 7).

Figure 9: Detection ranges at zero false alarms for each case in the data set using our detection system with an ad-hoc Viterbi temporal filtering stage. The aircraft went undetected in 44 cases (missed detection cases are absent from this figure). The ground clutter varies with the pass heading of the sensor equipped aircraft (see Figure 7).

stage is poor. Indeed, we were unable to find a threshold for the ad-hoc Viterbi approach that enabled detection of all targets with no false alarms (hence many cases with missed detections are missing from Figure 9). Furthermore, the aircraft that are detected with the ad-hoc Viterbi filtering stage are much closer than they are when detected with the HMM temporal filtering stage (cf. Figure 8).

The number of missed detections, mean detection ranges, and standard errors of the mean (SEOM) detection ranges are summarised in Table 4. The mean and SEOM were computed over all cases in the data set (excluding cases where there were missed detections). From Table 4, we see that our system detects all targets at a mean detection range of over 1.8 km with no false alarms. In contrast, the system with an ad-hoc Viterbi temporal filtering stage only detects the target with zero false alarms in 19 cases (at a mean range of 416 m). We therefore conclude that HMM temporal filtering outperforms ad-hoc Viterbi filtering, and we shall only consider HMM temporal filtering in the remainder of this paper.

By reducing the threshold h used to declare aircraft detections in the HMM temporal filtering stage from its zero false alarm value of $h = 1.483$, our system is able to detect aircraft at greater ranges but at the cost of incurring false alarms. Figure 10 provides a summary of the

(a)

(b)

Figure 10: Sensitivity of (a) Mean Detection Range, and (b) Number of False Alarms to choice of threshold h in HMM Temporal Filtering Stage.

trade-off between choice of threshold and the detection ranges and false alarms our system achieves on our data set. Importantly, the mean detection range appears to increase linearly with decreasing threshold value, whilst the number of false alarms increases marginally as the threshold is reduced to $h = 1.2$ from its zero false alarm value of $h = 1.483$. These results suggest that our system’s performance is not overly sensitive to the choice of threshold.

5.2.2 Window Length Study

We now investigate the sensitivity of our system to the choice of window length w . This study is practically significant because the computational effort and compute time of our system grows linearly with the window length w (cf. Table 2). In order to conduct our study, we used the baseline system with parameters given in Table 1, and varied the window length between $w = 3$ and $w = 11$. The recorded number of false alarms, mean detection ranges, and SEOMs on each of our 63 cases are shown in Table 6.

The results shown in Table 6 illustrate that our system is able to detect all of the aircraft in our data set with no false alarms for window lengths greater than 3. For small window lengths (i.e., when $w = 3$ and the background frame b_k is only computed with 2 frames), our system is unable to detect one aircraft without incurring false alarms on our non-target data. Importantly, increasing the window length appears to offer diminishing returns in terms of improvements in mean detection range.

The diminishing increases in mean detection range as the window length grows can be attributed to the limitations of our image registration and background subtraction stages. In particular, use of the pixel-wise median for background estimation implicitly requires that there is no optical flow within the window of registered frames $\mathcal{R}_k$, and the use of homography image registration implicitly requires that the optical flow is uniform over the entire field of view in the window $\mathcal{W}_k$. However, since the scene is non-planar and registration is being performed with homographies, there will be some optical flow within the window of registered frames $\mathcal{R}_k$, and it will grow with the window length w . Hence, we expect the mean detection ranges achieved by our system with zero false alarms to plateau (and eventually decrease) after $w = 11$. Similarly, the optical flow will only be approximately uniform across the field of view when the sensor aircraft is sufficiently high in altitude (since the camera is oblique to the scene resulting in the optical flow being greatest at the bottom of the field of view and least at the top). Again, this effect will have greatest impact on the system's performance for larger window lengths w but we have insufficient data captured at different sensor aircraft altitudes to further characterise its performance impact.

5.2.3 Image Registration Study

Our third study examines the impact of image registration on the detection performance of our system. In particular, we investigate the detection performance of our system when the homographies are estimated using one of the three methods described in Section 3.1, namely, homography estimation with phase correlation; homography estimation with SURF image features; or homography estimation with SIFT image features. We also investigate

how the number of tiles N in the phase correlation approach affects detection performance. To conduct our study, we varied the image registration stage of our baseline system with parameters given in Table 1. The results of our study are presented in Table 6.

The results in Table 6 suggest that our system performs best with feature-based image registration (with SIFT outperforming SURF at the cost of additional compute time). Table 6 also suggests that the detection performance of our system with phase correlation image registration is very sensitive to the number of tiles N . For $N = 4$ and $N = 6$ tiles, the phase correlation image registration is very poor resulting in blurred background images b_k and non-target artefacts in the motion frames m_ℓ . As a result of these artefacts, the HMM filter test statistics $\eta_{(k,\ell)}$ are larger on the non-target frames, and smaller on the target frames compared to when phase correlation image registration is performed with $N = 12$ tiles. Although more tiles and more sophisticated tiling schemes could be used to improve the performance of phase correlation image registration [cf. (Schubert and Mikolajczyk, 2014)], these modifications could compromise its ability to operate in real-time on our current hardware.

5.2.4 Gradient Subtraction Study

In our final study, we demonstrate the important role that the gradient subtraction stage plays in our system. To conduct our study, we again used our system with phase correlation image registration, HMM temporal filtering, and the baseline parameters given in Table 1. We varied the exponent, γ , used to compute the contrast-enhanced image of gradients e_k . The results are presented in Table 7.

The results in Table 7 suggest that subtracting background gradient information can sig-

nificantly improve detection performance. We see from Table 7 that selecting γ around 1.5 leads to the best detection performance on our data set. When $\gamma < 1.25$ and $\gamma > 1.75$ so that the gradient image e_k contains more or less gradient information than when $\gamma = 1.5$, the mean detection range is reduced whilst the number of missed detections is increased. In effect, gradient subtraction is masking sources of false alarms on the nontarget data, which enables us to set lower HMM thresholds h as γ decreases (and still achieve zero false alarms). However, subtracting too much gradient information (by selecting $\gamma < 1.25$) leads to missed detections and reduced detection ranges. In contrast, subtracting too little gradient information (by selecting $\gamma > 1.75$) enables many sources of false alarms to appear in the motion frames, which forces us to set higher HMM thresholds h in order to achieve no false alarms. Since subtracting too little gradient information ($\gamma > 1.75$) is akin to simply removing the gradient subtraction stage, without gradient subtraction, our system is only able detect all of the aircraft in the data set by incurring false alarms.

6 Discussion, Lessons Learnt, and Future Work

Through our experimental studies, we have demonstrated the feasibility of reliable aircraft detection below the horizon solely on the basis of vision using our system. In particular, the results of our temporal filtering and window length studies demonstrate that our multi-frame HMM filtering system facilitates visual detection of aircraft at ranges beyond those achieved by consecutive frame differencing (see for example, the $w = 3$ case in our window length study). Indeed, the detection performance of our system improves with increasing window length (at least until $w = 11$). Furthermore, our image registration study suggests that

feature-based image registration approaches currently offer the best detection performance, but have longer compute times than phase correlation approaches with simple tile arrangements. Our gradient subtraction study highlights that many false alarms due to imperfect image registration can be avoided by suppressing regions of the background that have strong gradient magnitudes. Finally, all of our studies demonstrate that by introducing HMM temporal filtering and adapting a track-before-detection filtering approach, we are able to avoid false alarms by selecting appropriate detection thresholds.

Although we investigated the impact of different HMM thresholds on the performance of our system, we have not examined the impact of different intra- and inter-window pixel transition neighbourhoods. Studies of different transition neighbourhoods have however been conducted for HMM filters used to detect aircraft above the horizon in order to characterise the sensitivity of the filters to jitter (or superior motions) (Lai et al., 2011). These studies have found that the tracking performance of HMM filters with transition neighbourhoods similar to those shown in Figure 5 degrades when jitter is present. A number of strategies for improving the performance of HMM filters in the presence of jitter have therefore been proposed including the use of multiple filters with different pixel transition neighbourhoods (Lai et al., 2011; Wainwright et al., 2011; Lai et al., 2013). Importantly, these strategies could be used to improve the performance of the HMM temporal filtering stage in our below horizon aircraft detection system.

Despite our experimental studies being amongst the most extensive to date, further data collection is still required to fully characterise the detection range, missed detection, and false alarm performance of our system. Indeed, due to the difficulty of collecting appropriate

data, we have been unable to examine the performance of our system in encounters where the target displays almost no motion relative to the background (as illustrated in Figure 2d). Our system will struggle to detect these targets since they will be estimated as part of the background image b_k , and largely removed in the background subtraction stage. However, these targets may be detected when they are close enough to exhibit significant looming behaviour (i.e., changes in apparent size). Some support for this conjecture is provided in (Molloy et al., 2014) where target looming information in the sky-region is extracted from HMM filter outputs. As an alternative solution, Zarandy et al. (2015) recently suggested that the sensor aircraft could execute regular heading changes to improve observability of targets with minimal motion with respect to the ground clutter. Future work will investigate the suitability of our system for detecting aircraft with minimal motion relative to the background ground clutter.

Another issue that we have been unable to resolve is that of rejecting moving ground vehicles. Although we are able to detect all aircraft in our data set without detecting any moving ground vehicles, rejecting these moving ground vehicles was done solely on the basis of HMM detection threshold selection. We therefore anticipate situations with our system where trade-offs need to be made between detecting moving ground vehicles and detecting aircraft. Assessing alternative (vision-based) methods for rejecting moving ground objects appears important since at a distance, moving ground objects and aircraft can both appear as locally bright (or dim) pixels (as illustrated in Figures 1 and 11). Again, differences in visual looming behaviour appear worth investigating for the purpose of reliable aircraft detection in the presence of moving ground vehicles.

Figure 11: Example frame showing a ground vehicle (small white dot circled and bottom right insert) that is tracked but not detected by our system.

Finally, our system has a number of important features worth highlighting. In particular, in contrast to the system in (Nussberger et al., 2014), our system operates purely on the output from a single monocular camera sensor, and does not require access to navigation sensor data or digital terrain maps (which significantly reduces system cost). Furthermore, our system is not overly sensitive to ground clutter present in a semi-rural Australian environment. Although the performance of our system in other environments (e.g., mountain regions and suburbia) requires further examination, we anticipate that it should remain suitable in environments with less clutter. Indeed, the detection ranges achieved by our system on our data set with structured ground clutter are comparable to those achieved by the state-of-the-art system of (Lai et al., 2013) in the less cluttered sky-region. Future development of our system will include extending it to support detection across a wider field of view by using either multiple cameras or wide field of view lenses.

7 Conclusion

In this paper, we have considered the detection of potential mid-air collision threats below the horizon solely on the basis of vision. We described a novel vision-based system for detecting potential collision threats that appear below the horizon in structured ground clutter. We characterised the detection performance of our system (and assessed the feasibility of vision-based aircraft detection below the horizon) by collecting a data set of 63 near collision head-on encounters between a below-horizon target aircraft and a camera equipped manned aircraft

acting as a UAS surrogate. Our system (with suitable parameters) detects all 63 targets in our data set at a mean range of 1890m with no false alarms on 1.1 hours of non-target data.

Acknowledgments

This work was supported by Boeing Research and Technology Australia. The authors gratefully acknowledge the contribution of the pilots and engineers involved in conducting the reported flight experiments (Dmitry Bratanov, Dirk Lessner, Gavin Broadbent, and Dean Gilligan). The authors are also grateful to Jasmin James for verifying the results, and Andrew English and Adam Jacobson for the many fruitful discussions.

References

- Carnie, R., Walker, R., and Corke, P. (2006). Image processing algorithms for UAV "sense and avoid". In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 2848–2853.
- Cho, S., Huh, S., Shim, D., and Choi, H. (2013). Vision-based detection and tracking of airborne obstacles in a cluttered environment. *Journal of Intelligent & Robotic Systems*, 69(1-4):475–488.
- Clothier, R. A., Williams, B. P., and Fulton, N. L. (2015). Structuring the safety case for unmanned aircraft system operations in non-segregated airspace. *Safety Science*, 79:213–228.
- Dey, D., Geyer, C., Singh, S., and Digioia, M. (2011). A cascaded method to detect aircraft

in video imagery. *International Journal of Robotics Research*, 30(12):1527–1540.

Elliott, R., Aggoun, L., and Moore, J. (1995). *Hidden Markov Models: Estimation and Control*. Springer, New York, NY.

Fasano, G., Accardo, D., Tirri, A., Moccia, A., and De Lellis, E. (2014). Morphological filtering and target tracking for vision-based UAS sense and avoid. In *Unmanned Aircraft Systems (ICUAS), 2014 International Conference on*, pages 430–440.

Gandhi, T., Yang, M.-T., Kasturi, R., Camps, O., Coraor, L., and McCandless, J. (2006). Performance characterization of the dynamic programming obstacle detection algorithm. *Image Processing, IEEE Transactions on*, 15(5):1202–1214.

Greer, D. G., Mudford, R., Dusha, D., and Walker, R. A. (2010). Airborne systems laboratory for automation research. In *27th International Congress of the Aeronautical Sciences*, Acropolis Conference Centre, Nice. ICAS.

Harris, F. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51–83.

Lai, J., Ford, J., O’Shea, P., and Walker, R. (2008). Hidden Markov Model Filter Banks for Dim Target Detection from Image Sequences. In *Computing: Techniques and Applications, 2008. DICTA ’08. Digital Image*, pages 312–319.

Lai, J., Ford, J. J., Mejias, L., and O’Shea, P. (2013). Characterization of Sky-region Morphological-temporal Airborne Collision Detection. *Journal of Field Robotics*, 30(2):171–193.

- Lai, J., Ford, J. J., Mejias, L., O'Shea, P., and Walker, R. (2012). See and Avoid Using Onboard Computer Vision. In Plamen, A., editor, *Sense and Avoid in UAS*, pages 265–294. John Wiley & Sons, Ltd, Hoboken, NJ.
- Lai, J., Mejias, L., and Ford, J. J. (2011). Airborne vision-based collision-detection system. *Journal of Field Robotics*, 28(2):137–157.
- May, K. and Krouglicof, N. (2013). Moving target detection for sense and avoid using regional phase correlation. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 4767–4772.
- Mcfadyen, A. and Mejias, L. (2016). A survey of autonomous vision-based see and avoid for unmanned aircraft systems. *Progress in Aerospace Sciences*, 80:1 – 17.
- Mejias, L., McNamara, S., Lai, J., and Ford, J. (2010). Vision-based detection and tracking of aerial targets for UAV collision avoidance. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 87 –92.
- Molloy, T. L., Ford, J. J., and Mejias, L. (2014). Looming aircraft threats : shape-based passive ranging of aircraft from monocular vision. In *Australian Conference on Robotics and Automation 2014*, The University of Melbourne, Melbourne, VIC.
- Morris, C. (2005). Midair collisions: Limitations of the see-and-avoid concept in civil aviation. *Aviation, space, and environmental medicine*, 76(4):357–365.
- Nussberger, A., Grabner, H., and Van Gool, L. (2014). Aerial object tracking from an airborne platform. In *Unmanned Aircraft Systems (ICUAS), 2014 International Conference on*, pages 1284–1293.

- Nussberger, A., Grabner, H., and Van Gool, L. (2015). Robust aerial object tracking in images with lens flare. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 6380–6387.
- Pellebergs, J. (2010). The MIDCAS project. In *27th International Conference of the Aeronautical Sciences*.
- Reilly, V., Idrees, H., and Shah, M. (2010). Detection and Tracking of Large Number of Targets in Wide Area Surveillance. In Daniilidis, K., Maragos, P., and Paragios, N., editors, *Computer Vision - ECCV 2010*, volume 6313 of *Lecture Notes in Computer Science*, pages 186–199. Springer Berlin Heidelberg.
- Rozantsev, A., Lepetit, V., and Fua, P. (2015). Flying objects detection from a single moving camera. In *Conference on Computer Vision and Pattern Recognition*.
- Saleemi, I. and Shah, M. (2013). Multiframe many-many point correspondence for vehicle tracking in high density wide area aerial videos. *International Journal of Computer Vision*, 104(2):198–219.
- Schubert, F. and Mikolajczyk, K. (2014). Robust registration and filtering for moving object detection in aerial videos. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 2808–2813.
- Szeliski, R. (2006). Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision in Computer Graphics and Vision*, 2(1):1–104.
- Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.

- Wainwright, A. L., Ford, J. J., and Lai, J. S. (2011). Fat and thin adaptive HMM filters for vision based detection of moving targets. In Drummond, T., editor, *Australasian Conference on Robotics and Automation (ACRA 2011)*, Monash University, Melbourne, VIC. Australian Robotics & Automation Association.
- Zarandy, A., Nagy, Z., Vanek, B., Zsedrovits, T., Kiss, A., and Nemeth, M. (2013). A five-camera vision system for uav visual attitude calculation and collision warning. In Chen, M., Leibe, B., and Neumann, B., editors, *Computer Vision Systems*, volume 7963 of *Lecture Notes in Computer Science*, pages 11–20. Springer Berlin Heidelberg.
- Zarandy, A., Nemeth, M., Nagy, Z., Kiss, A., Santha, L., and Zsedrovits, T. (2014). A real-time multi-camera vision system for uav collision warning and navigation. *Journal of Real-Time Image Processing*, pages 1–16.
- Zarandy, A., Zsedrovits, T., Pencz, B., Nameth, M., and Vanek, B. (2015). A novel algorithm for distant aircraft detection. In *Unmanned Aircraft Systems (ICUAS), 2015 International Conference on*, pages 774–783.

Table 1: System Parameters and Baseline Values.

Parameter	Symbol	Baseline Value
Input Frame Width	N/A	1280
Input Frame Height	N/A	960
Registered Frame Width	N_h	1024
Registered Frame Height	N_v	768
No. Tiles	N	12
Tile Side	L	256
Window Length	w	11
Gradient Scalar	γ	1.5
HMM Forgetting Parameter	τ	$10w$
HMM Threshold	h	1.483

Table 2: Measured compute times per time-step (in milliseconds) and measured overall frame rate (in frames per second) of baseline system with phase correlation image registration, background and gradient subtraction, and HMM temporal filtering stages. System parameters are given in Table 1.

w	Compute Time Per Time Step (ms)			Frame Rate
	Registration	Background	Temporal	(Hz)
3	10.5	6.0	24.6	24
5	18.6	6.8	26.5	19
7	24.5	7.4	28.3	16
9	30.1	9.1	30.2	14
11	36.7	10.2	32.0	12

Table 3: Summary of collected data set. The total target data is 0.35 hours in duration and the total non-target data is 1.1 hours in duration.

Date	No. Cases with Sensor Aircraft Heading:				No. Frames	
	North-East	South-West	North-West	South-East	Target	Non-Target
April 8	4	7	0	0	3211	8053
July 14	7	6	0	0	3482	15216
July 15	10	9	0	0	6158	17170
July 16	5	4	7	4	6177	18165
Total	26	26	7	4	19028	58604

Table 4: Results of temporal filtering study. Missed detections, mean detection ranges, and standard errors of mean (SEOM) detection ranges are reported at a fixed zero false alarm threshold calculated from our non-target data.

Algorithm	No. Missed Detections	Detection Range	
		Mean (m)	SEOM (m)
HMM	0	1890	43.19
Ad-hoc Viterbi	44	416	40.21

Table 5: Results of window length study. Missed detections, mean detection ranges, and standard errors of mean (SEOM) detection ranges are reported at a fixed zero false alarm threshold calculated from our non-target data.

w	No. Missed Detections	Detection Range	
		Mean (m)	SEOM (m)
3	1	1404	34.94
5	0	1546	34.91
7	0	1857	39.54
9	0	1859	40.53
11	0	1890	43.19

Table 6: Results of image registration study. Missed detections, mean detection ranges, and standard errors of mean (SEOM) detection ranges are reported at a fixed zero false alarm threshold calculated from our non-target data. The compute time is per time step k .

Registration	No.	Tile Side	No. Missed	Detection Range		Compute
	Tiles	(pixels)	Detections	Mean (m)	SEOM (m)	Time (ms)
Phase Corr.	4	512	56	1294	93.25	26.5
Phase Corr.	6	384	36	1333	47.25	35.3
Phase Corr.	12	256	0	1890	43.19	36.7
SURF	N/A	N/A	0	1893	43.15	309.6
SIFT	N/A	N/A	0	1915	47.09	> 500

Table 7: Results of gradient subtraction study. Missed detections, mean detection ranges, and standard errors of mean (SEOM) detection ranges are reported at a fixed zero false alarm threshold calculated from our non-target data.

γ	No. Missed Detections	Detection Range	
		Mean (m)	SEOM (m)
0.5	8	1177	41.64
1.0	3	1675	40.98
1.25	1	1784	42.90
1.5	0	1890	43.19
1.75	6	1843	43.08
2.0	13	1847	53.07
2.5	40	1783	61.29

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Molloy, TL;Ford, JJ;Mejias, L

Title:

Detection of aircraft below the horizon for vision-based detect and avoid in unmanned aircraft systems

Date:

2017-10-01

Citation:

Molloy, T. L., Ford, J. J. & Mejias, L. (2017). Detection of aircraft below the horizon for vision-based detect and avoid in unmanned aircraft systems. *Journal of Field Robotics*, 34 (7), pp.1378-1391. <https://doi.org/10.1002/rob.21719>.

Persistent Link:

<http://hdl.handle.net/11343/292907>