Detection versus False Alarm Characterisation of a Vision-Based Airborne Dim-Target Collision Detection System

John Lai, Jason J. Ford, Luis Mejias, Peter O’Shea, and Rodney Walker
Australian Research Centre for Aerospace Automation (ARCAA)
Queensland University of Technology
GPO Box 2434, Brisbane Queensland 4001
Email: {js.lai, j2.ford, luis.mejias, pj.oshea, ra.walker}@qut.edu.au

Abstract—This paper presents a preliminary flight test based detection range versus false alarm performance characterisation of a morphological-hidden Markov model detection approach to vision-based airborne dim-target collision detection. On the basis of compelling in-flight collision scenario data, we calculate system performance operating characteristic (SOC) curves that concisely illustrate the detection range versus false alarm rate performance design trade-offs. These preliminary SOC curves provide a more complete dim-target detection performance description than previous studies (due to the experimental difficulties involved, previous studies have been limited to very short flight data sample sets and hence have not been able to quantify false alarm behaviour). The preliminary investigation here is based on data collected from 4 controlled collision encounters and supporting non-target flight data. This study suggests head-on detection ranges of approximately 2.22 km under blue sky background conditions (1.26 km in cluttered background conditions), whilst experiencing false alarms at a rate less than 1.7 false alarms/hour (ie. less than once every 36 minutes). Further data collection is currently in progress.

Keywords—collision; detection; false alarm; hidden Markov model; morphology; vision;

I. INTRODUCTION

Machine vision and dim-target image processing techniques have been previously suggested by many authors as potential tools for addressing detection aspects of the unmanned aerial vehicle (UAV) airborne collision avoidance problem (which is also known as the sense-and-avoid problem) [1], [2]. Compared to other sensing options (such as radar), machine vision sensors are relatively low cost and have minimal size, weight, and power (SWaP) footprints, allowing them to remain feasible even in small to medium class UAVs. In addition to their SWaP advantages, machine vision offers a non-cooperative sensing approach that does not rely on transponders or other special equipment to be installed on other aircraft (cooperative approaches like the Traffic Alert and Collision Avoidance System (TCAS) require cooperative radio communications via onboard transponders [3]). Furthermore, the parallels between machine vision and human vision can be an advantage in gaining future regulatory approval, given the focus of current standards on equivalence with human see-and-avoid capabilities [4], [5].

Over the last three decades, a two-stage processing paradigm has emerged for detection of dim, sub-pixel sized targets [6]–[10]. These two stages are: 1) an image pre-processing stage that highlights potential targets with spatial attributes of interest (within a single frame); and 2) a temporal filtering stage that exploits target dynamics across a sequence of image frames.

The principle objective of the image pre-processing stage is to suppress non-target features within the image such as background noise and clutter. Image morphology has become one of the most prevalent pre-processing approaches in the dim-target detection problem [11]–[13] (specific implementations include the Hit-or-Miss filter [14], the Close-Minus-Open filter [15], and the Top-Hat filter [16]). Admittedly, early image morphology applications were focused on the detection of small target in infrared (IR) images [17]–[19]. However, much of the recent image morphology research activity has related to the problem of dim-target detection in the context of visual spectrum systems [6]–[8], [20], [21].

Following the image pre-processing stage, the temporal filtering stage attempts to highlight or extract image features exhibiting target-like temporal evolution across successive image frames (using track-before-detect filtering approaches). Two basic temporal filtering approaches have been well studied within the dim-target tracking community: Viterbi based approaches [7], [9], [10], [22], [23], and Bayesian based approaches [6], [21], [22], [24]–[26]. An explanation of the difference between the two paradigms is provided in [26], with performance comparisons provided in [21], [22] (on the basis of synthetic computer generated data).

Efforts to characterise the performance of dim-target detection techniques on the basis of real target data have continued over the last few years [6], [20], [27]–[30]. In [20], [27], [28], dim-target detection techniques were applied to data sets collected from ground based sensors; these data sets illustrated the basic feasibility of image based dim-target detection. In [6], [29], [30], practical in-flight airborne dim-target detection (that is, detection on the basis of an imaging sensor onboard an aircraft) was demonstrated in
a limited number of test cases. These test cases illustrated the importance of imaging sensor stability in dim-target detection, which is a significant challenge in the airborne environment due to platform vibrations and unpredictable aerodynamic disturbances. However, long-term false alarm analysis (over data sets more than just a few hundred image frames in length) and performance trade-offs were not addressed in this work.

More recently, a direct comparison study between the Viterbi and Bayesian based temporal filtering approaches was made on the basis of image sequences collected during staged UAV-on-UAV mid-air collision course experiments [6]. This UAV collision course study suggested that both dim-target detection approaches offered similar detection range performance; however, the collected data sets were not extensive enough to allow proper evaluation of false alarm performance statistics (although some partial evidence about false alarm susceptibility was provided).

Together, these previous studies have demonstrated the basic feasibility of using image based dim-target detection techniques for the airborne collision avoidance problem, but these studies have also highlighted that false alarm performance is an important issue that is yet to be appropriately characterised. It is also worth highlighting that the collection of collision engagement data is not straightforward and typically requires specialised equipment and careful planning [6], [29], [30].

The key contribution of this paper is to provide the first credible assessment of vision based dim-target airborne collision detection systems through the evaluation of two key system operating characteristics: detection range and false alarm rate. This is performed using airborne flight test data, and we present our results in terms of system operating characteristic (SOC) curves that illustrate important trade-off relationships between detection range and false alarm rate. The calculated SOC curves provide the first systematic characterisation of false alarm performance of an airborne dim-target detection system.

This paper is structured as follows: In Section II we will present a description of the vision-based dim-target airborne collision detection approach examined in this paper (the morphological-hidden Markov model (morphological-HMM) approach). In Section III, we describe our data collection and analysis approach and present our system operating characteristic (SOC) curves. Some brief concluding remarks are provided in Section IV.

II. DIM TARGET DETECTION APPROACH

The purpose of this paper is to evaluate the detection and false alarm performance of a candidate dim-target vision based airborne collision detection system. The candidate system under evaluation is the morphological-HMM dim-target detection approach (a Bayesian approach) [6], [21], [24], [31]. This dim-target detection approach involves two processing stages: 1) a morphological image pre-processing stage, followed by 2) a track-before-detect hidden Markov model temporal filtering stage. The calculations involved in each stage are highly parallelizable, and previous studies have exploited graphics processing unit (GPU) based hardware to realise real-time execution performance (in [6], unoptimised implementations of the morphological-HMM algorithm can process 1024-by-768 pixel images at up to 30 frames per second). More detail on the two processing stages is given in the following sections.

A. Morphological Image Pre-Processing

In [6], a Close-Minus-Open (CMO) morphological filtering stage was used to simultaneously detect both positively contrasting (brighter than background) and negatively contrasting (darker than background) features within an image. A directional decomposition technique [15] is exploited in a manner similar to [21], [24], whereby filtering in the horizontal and vertical directions are performed independently using 1D row and column structuring elements (structuring elements 5 pixels in length were used in our implementation). The minimum of the horizontal and vertical response on a pixel by pixel basis is taken as the final morphological filter output.

B. Track-before-detect hidden Markov model Temporal Filtering

The HMM temporal filtering stage of the candidate dim-target detection approach assumes that, if present, the target resides within (and evolves over) a set of discrete 2D grid points or pixel locations \( \{(i, j) | 1 \leq i \leq N_v, 1 \leq j \leq N_h \} \), where there are \( N_v \) vertical and \( N_h \) horizontal locations in the image. Each pixel location \((i, j)\) is interpreted as a unique state of the HMM, which can be indexed by the single index value \( m = [(j - 1)N_v + i] \in \{1, 2, \ldots, N\} \) (corresponding to the column stack operation applied to the image grid).

We now repeat the model construction that is provided in [6] (an alternative model description is also provided in [31]). Let \( x_k \) denote the HMM state that is active at time \( k \). Between consecutive image frames, the target may move between different pixel locations (or equivalently, transition between different HMM state values). The probability of transition between different HMM states will be described by the transition probabilities matrix \( A \), where the \( m \)th element \( A_{mn} = P(x_{k+1} = \text{state } m | x_k = \text{state } n) \) describes the probability of moving from a pixel position (state) \( n \) to any other pixel position (state) \( m \). In this airborne collision...
detection problem, transition probabilities will be used to describe the expected mean target motion in the image plane (and we remind that motion in the image plane is not directly equivalent to motion in the real world). We will let initial probabilities be denote by the vector $\pi$, where the $m$th element $\pi_m = P(x_1 = \text{state } m)$ for $1 \leq m \leq N$ describes the target’s probability of initially being located in state $m$. Finally, we will let measurement probabilities $B^m(Y_k) = P(Y_k | x_k = \text{state } m)$ for $1 \leq m \leq N$ define the probability of the observed image measurement $Y_k \in R^{N_x \times N_y}$, given that the target is at pixel location (or HMM state) $m$ (see [33] for more details about the parameterisation of HMMs).

In this airborne detection application, as argued in [24], we can calculate $B^m$ using the following relationship between target location $x_k$ and the pre-processed measurements $Y_k$:

$$B^m(Y_k) = \frac{P(Y_k | x_k = \text{state } m)}{P(Y_k | x_k \neq \text{state } m)} \quad (1)$$

for $1 \leq m \leq N$, where $Y_k^m$ denotes the pixel in $Y_k$ in the same location as state $m$. The relationship (1) allows $B^m$ to be calculated on a per-pixel basis (rather than the whole image), and only requires access to the per-pixel probabilities $P(Y_k^m | x_k = \text{state } m)$ and $P(Y_k^m | x_k \neq \text{state } m)$. Furthermore, using the approach described in [6], the per-pixel measurement probabilities $P(Y_k^m | x_k = \text{state } m)$ and $P(Y_k^m | x_k \neq \text{state } m)$ can be empirically estimated by using samples of the morphologically processed data. Specifically, the per-pixel probability $P(Y_k^m | x_k = \text{state } m)$ can be estimated on the basis of the statistics of the pixel value at non-target locations in the samples of morphologically processed data. Similarly, $P(Y_k^m | x_k \neq \text{state } m)$ can be estimated on the basis of the statistics of the pixel value at target locations in the samples of morphologically processed data. In our experience, per-pixel probabilities estimated from as little as 100 morphologically processed image frames can ensure reasonable dim-target detection performance (and we have observed that algorithm performance is not overly sensitive to this part of the filter design).

We will delay discussion of $A$ design until later in Section II.D when we discuss HMM filter bank design. We next introduce our HMM filter.

**C. HMM Temporal filtering approach**

The HMM temporal filtering step can be mechanised using the forward part of the forward-backward HMM procedure [34] as follows. For $1 \leq m \leq N$:

1. Initialisation: Let $\alpha_k^m$ denote the probability $P(Y_1, Y_2, \ldots, Y_k, x_k = \text{state } m)$. Then $\alpha_k^m = \pi_m B^m(Y_1)$.

2. Recursion: At time $k > 1$, set $\alpha_k^m = \sum_{n=1}^{N} \alpha_{k-1}^m A^{mn} B^m(Y_k)$.

so that the conditional mean filtered estimate of the state $\hat{x}_k^m := E[x_k = \text{state } m | Y_1, Y_2, \ldots, Y_k]$ is given by

$$\hat{x}_k^m = \frac{\alpha_k^m}{\sum_{m=1}^{N} \alpha_k^m}. \quad (2)$$

Alternatively, the conditional mean estimate can be calculated as [33]:

$$\hat{x}_k = N_k B_k(Y_k) A \hat{x}_{k-1}. \quad (3)$$

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; and $\hat{x}_k$ is a $N \times 1$ vector consisting of elements $\hat{x}_k^m$ for $1 \leq m \leq N$.

The following important relationships hold between filter outputs and the measurement likelihood:

$$P(Y_1, Y_2, \ldots, Y_k) = \sum_{m=1}^{N} \alpha_k^m = \prod_{\ell=1}^{k} \frac{1}{N_{\ell}}. \quad (4)$$

In [6], a measurement likelihood based test statistic is introduced for declaring the presence of a target. The window averaged test statistics $\gamma_k$ can be defined via the moving average filter of window length of $L$ as follows:

$$\eta_k = \left(\frac{L-1}{L}\right) \eta_{k-1} + \left(\frac{1}{L}\right) \log \left(\frac{1}{N_k}\right). \quad (5)$$

At time $k$, if $\eta_k$ exceeds a predefined threshold (a design parameter that controls the sensitivity of the detection system), then a target is declared to be present, and the target is considered to be located at state

$$\gamma_k = \arg \max_m (\hat{x}_k^m). \quad (6)$$

In [6], a window length of $L = 10$ was found to produce good detection results.

**D. HMM Filter Bank**

Whilst a single HMM filter can provide some detection capability, previous studies have demonstrated that the simultaneous use of multiple HMM filters organised into a filter bank can lead to substantially improved detection performance [24]. The performance advantage occurs because a bank of HMM filters can provide more appropriate approximation of the range of possible target temporal behaviors that need to be detected (in comparison to a single HMM model that tries to provide a “lumped” compromised description of all possible behaviors). In this paper, as proposed in [6], [24], we evaluate a HMM filter bank consisting of four HMM filters. Each HMM filter has the same morphologically pre-processed image data as its input (and there is no coupling between HMM filter intermediate quantities).

The transition probability matrix $A$ of each filter in the HMM filter bank are designed based on expected target
motion. Previous studies have shown that collision-course targets tend to manifest as relatively stationary features on the image plane [35]. Hence, we can exploit slow-motion transition designs for each member HMM in the filter bank that limit target interframe motion to 1 pixel; that is, between consecutive image frames the target either remains stationary or moves to an adjacent neighboring pixel. For example, consider the slow-motion transition designs illustrated by the patches in Figure 1. For each patch, if we consider the target to be located in the centre highlighted pixel at frame $k$, then the shaded pixels represent the possible target locations at frame $k + 1$.

In practice, however, a combination of factors that occur in the airborne detection problem (such as image jitter and frame rate limitations) degrade the effectiveness of slow-motion transition designs. Drawing on past experiences and insights [6], the proposed system implements the four transition designs/patches illustrated in Figure 2 for our HMM filter bank. These patches describe motion on the image plane with a maximum displacement of 2 pixels per frame, and collectively capture the range of target dynamics likely to be encountered in our airborne data collection experiments (a bias towards vertically upward motion in 3 of the patches is due to the target aircraft being flown at a higher altitude than the camera aircraft (see Section III.A)).

When a HMM filtering bank is used for collision detection, a detection is declared whenever the test statistic of any of the individual filters passes the predefined threshold. If multiple filters simultaneously declare detection, the filter with the largest test statistic is used to estimate target location.

### III. Flight Test Results

#### A. Data Collection

Our approach to characterising the detection range and false alarm performance of the morphological-HMM detection technique requires the collection of two types of image data: 1) images containing a collision-course target/intruder aircraft (target data); and 2) images of the airborne sky background without any targets present (non-target data). Our primary data collection platform is a specially modified Cessna 172 aircraft [36] that is capable of recording image and aircraft state data in-flight.

We used the Cessna’s externally mounted camera to collect grayscale, 8-bit per pixel, target and non-target image data at 15Hz with a resolution of 1024-by-768 pixels. The camera field-of-view (FOV) was approximately $20^\circ$-by-$15^\circ$ horizontally and vertically, respectively.

The collection of target data required the deployment of a second aircraft to act as an intruder/target. We have conducted numerous flight experiments that have, in a deliberate but controlled manner, brought target and camera aircraft in close proximity to each other in order to recreate realistic collision-course scenarios. This has allowed us to capture image data highly representative of what would be perceived in actual airborne collision situations.

Two types of target data flight experiments were carried out: 1) head-on collision experiments where aircraft are converging from opposite directions; and 2) tail-chase experiments aimed at replicating an overtaking collision scenario where a faster aircraft approaches a slower aircraft from behind. A hired Cessna 182 aircraft equipped with an onboard GPS played the role of the target aircraft in our experiments. State data logged on both target and camera aircraft enabled image ego-motion compensation and calculation of detection ranges post-flight. In total, 2 head-on and 2 tail-chase image sequences were collected. Figures 3 and 4 are contrast enhanced samples of target images (cropped to a size of 300-by-300 pixels) captured against blue sky and cloudy backgrounds, respectively. The images are roughly centered on the target, and both targets are at a distance of approximately 1.5 km. The performance of the morphological-HMM detection approach over the 4 target image sets is considered in the next section.
B. Preliminary Performance Characterisation

In the context of airborne collision detection, detection range is an intuitive measure of performance. A long detection range is desirable as it allows for more time to plan and execute avoidance manoeuvres. However, increasing the system sensitivity in order to achieve longer detection ranges tends to result in more false alarms. Studies into automated systems have shown the adverse impact of excessive false alarms [37]. The specific concern in the airborne collision avoidance problem is that false alarms would trigger unnecessary avoidance manoeuvres. Hence, a proposed detection approach is not fit-for-purpose until a low false alarm rate can be achieved.

The purpose of the study presented here is to characterise the important trade-off between detection range and false alarm rate for the morphological-HMM target detection algorithm. We consider the effect of different background conditions (clear ‘blue sky’ versus cluttered/cloudy) and different collision geometries (tail-chase versus head-on).

To quantify the two sky background types used in our study, we exploited some spatial gradient statistics that have found application in cloud texture classification [38]. In particular, we calculated the mean gray-level difference statistics for sample images drawn from our blue sky and cloud cluttered image sequences, where larger difference values indicate the presence of more structured cloud fields in the image. The image in Figure 5 is representative of our blue sky data and yielded a mean gray-level difference of $3.4 \times 10^{-3}$. The same statistic calculated for Figure 6, which is representative of our cloud cluttered data, resulted in a value of $5.9 \times 10^{-3}$. Alternatively, the gray-level pixel variance for Figures 5 and 6 are 288 and 988, respectively, highlighting the more inhomogeneous nature of our cloudy data set (here, pixel variance is the variance of a row or column vector containing the gray-level values of each pixel in the image).

False alarm performance was characterised based on detections induced by non-target image sequences approximately 0.6 hours in length. This data length limits our characterisation to false alarm rates no lower than 1.7 false alarms/hour (even so, this study represents a significant improvement over previous investigations [6], [29], [30] which
useful signal-to-noise ratio (SNR) quantity, at least 200 consecutive frames). Detection ranges were evaluated based on detections declared on target image sequences correlated temporally with recorded aircraft state data. Detection range and false alarm statistics were combined to construct the SOC curves shown in Figures 7 and 8. Each curve is based on data from one target image sequence and one non-target image sequence (in each target sequence, the intruder aircraft was in the camera FOV for at least 200 consecutive frames).

Finally, for each target data set we have also evaluated a useful signal-to-noise ratio (SNR) quantity, ∆DSNR, which has previously been exploited in [6] to provide a measure of detection confidence. Based on the filtering output of the morphological-HMM detection algorithm, ∆DSNR is given by:

\[
\Delta \text{DSNR} = 20 \log_{10} \left( \frac{P_T}{P_F} \right) \text{dB},
\]

(7)

where \(P_T\) is the maximum target pixel intensity at the filtering output and \(P_F\) is the highest non-target pixel response at the filtering output. The ∆DSNR value (i.e. detection confidence) is suppressed by strong responses away from the true target location.

Results: Figure 7 illustrates two SOC curves of the morphological-HMM target detection approach under blue sky background conditions. The top curve corresponds to a tail-chase collision scenario whereas the bottom curve is for a head-on collision scenario. Each marker on the curve represents a specific threshold at which the target aircraft was detected. We highlight that as the detection threshold was increased (i.e. as we move towards the left-hand end of the curves), the morphological-HMM algorithm eventually reported zero false alarms. For the tail-chase scenario, this first occurs at a range of approximately 2.62 km, and for the head-on scenario the range was approximately 2.22 km.

We also examined detection performance under cluttered/cloudy sky background conditions for both tail-chase and head-on collision scenarios, as illustrated in Figure 8. Here, the morphological-HMM algorithm was able to detect the target aircraft at a distance between 1.27 km and 1.24 km without incurring any false alarms. We note that under the same background conditions, there was not a significant difference in detection ranges between head-on and tail-chase scenarios. However, a comparison of results across different backgrounds reveals that cloud clutter backgrounds unfortunately yielded noticeably shorter detection ranges for similar false alarm rates. This suggests that cloud clutter has a camouflaging effect on the target, hence requiring the target to become more distinct (i.e. closer to the observer) before it can be reliably detected.

We have also observed some interesting false alarm performance characteristics related to image properties. The bottom half of Figure 9 illustrates image features that tend to generate false alarms in the morphological-HMM detection approach; that is, cloud structures that exhibit small locally-dark regions. In contrast, the more uniform structured clouds in the top half of Figure 9 represent conditions far less conducive to producing false alarms.

To put the detection distance results that we have reported in the context of a real-life scenario, the Federal Aviation Administration (FAA) advisory circular on pilots’ role in collision avoidance suggests that a conflicting aircraft must be detected at least 12.5 seconds prior to the time of impact for safe collision avoidance [39]. Considering aircraft closing speeds in our data collection experiments of approximately 100 m/s, our results show the morphological-HMM approach providing adequate warning in blue-sky conditions (> 20 seconds) and borderline protection in cloudy conditions (≈ 12 seconds).

We also highlight that Figures 5 and 6 illustrate the trade-off relationship between detection distance and false alarm performance. In particular, the system false alarm rate is not fixed and varies depending on the desired detection distance (improving detection distance can be done at the expense of
Figure 9. Image features conducive to false alarms. The locally dark cloud features in the bottom half of the image tend to cause false alarms, whereas the more uniform clouds structures above are less conducive to false alarms.

Table I

<table>
<thead>
<tr>
<th>Background Type</th>
<th>Collision Geometry</th>
<th>Detection Range (km)</th>
<th>Detection Confidence ΔDSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue Sky</td>
<td>Head-on</td>
<td>2.22</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Tail-chase</td>
<td>2.62</td>
<td>85</td>
</tr>
<tr>
<td>Cloudy</td>
<td>Head-on</td>
<td>1.27</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>Tail-chase</td>
<td>1.24</td>
<td>131</td>
</tr>
</tbody>
</table>

a higher incidence of false alarms).

Finally, Table I shows the detection ranges and corresponding detection confidence levels (average ΔDSNR over 10 consecutive frames) achieved by the morphological-HMM algorithm when no false alarms were reported. We observe that the detection algorithm must reach a higher confidence level to detect targets in cloud cluttered data than in blue sky background data. Furthermore, within the same background type, tail-chase targets tend to require lower detection confidence levels than head-on targets.

Admittedly, more non-target data is required if we want stronger confidence that these early, but promising, results are truly representative of the detection algorithm’s operating characteristics. We are currently working towards the collection of more than 20 hours of non-target data which will eventually allow us to report evidence of system performance with a resolution approaching 0.05 false alarms/hour (compared to the 1.7 false alarms/hour resolution possible with the current data set).

IV. CONCLUSION

This paper presented the preliminary analysis of the performance of a vision-based morphological-HMM airborne collision avoidance system. This analysis suggest that, in a moderately ‘blue sky’ environment, reliable vision-based detection of potential head-on airborne collision with a general aviation aircraft (ie. a Cessna aircraft) can be made at distances out to 2.22 km (1.27 km in cluttered background conditions), whilst experiencing false-alarms at a rate less than 1.7 false alarms/hour (ie. less than once every 36 minutes).

ACKNOWLEDGMENT

This research was supported under Australian Research Council’s Linkage Projects funding scheme (project number LP100100302) and the Smart Skies Project, which is funded, in part, by the Queensland State Government Smart State Funding Scheme.

The authors would like to thank all flight pilots (Rhys Mudford, Duncan Greer, and Felipe Gonzalez), and test engineers (Scott McNamara, Chris Turner, Richard Glassock, and Reece Clothier) involved in conduct of the reported flight tests and the construction of the specialised equipment. We also thank Brendan Williams and Geoff Casey for their assistance with risk management and experimental planning stages.

REFERENCES


