A Novel Technique for Rejecting Non-Aircraft Artefacts in Above Horizon Vision-Based Aircraft Detection

Jasmin James¹, Jason J. Ford¹ and Timothy L. Molloy²

Abstract—Unmanned aerial vehicle (UAV) operations are steadily expanding into many important applications. A key technology for better enabling their commercial use is an onboard sense and avoid (SAA) technology which can detect potential mid-air collision threats in the same manner expected from a human pilot. Ideally, aircraft should be detected as early as possible whilst maintaining a low false alarm rate, however, textured clouds and other unstructured terrain make this trade-off a challenge. In this paper we present a new technique for the modelling and detection of aircraft above the horizon that is able to penalise non-aircraft artefacts (such as textured clouds and other unstructured terrain). We evaluate the performance of our proposed system on flight data of a Cessna 172 on a near collision course encounter with a ScanEagle UAV data collection aircraft. By penalising non-aircraft artefacts we are able to demonstrate, for a zero false alarm rate, a mean detection range of 2445m corresponding to an improvement in detection ranges by 9.8% (218m).

I. INTRODUCTION

The unmanned aerial vehicle (UAV) market worldwide is projected to grow by US $47.8 Billion by 2025, with a compounded growth of 18.8%. This growth is driven by the increasing use of UAVs in various commercial applications, such as monitoring, surveying and mapping, precision agriculture, aerial remote sensing, product delivery and many more [1]–[3]. The consistent increase in the global market has propelled efforts to ensure that routine, standard and flexible UAV operations are integrated into the national airspace such that they do not compromise the existing safety levels [4]. The risk of mid-air collision is an important safety concern that is both faced and posed by UAVs. The capability to avoid mid-air collisions would allow UAVs to more routinely operate in common airspace [5], [6].

In order to reduce the risk of a mid-air collision, the national airspace is strictly regulated with several safety layers [4], [5], [7]. The first few layers involve operational procedures, air traffic management and cooperative collision avoidance systems. The final layer is for potential mid-air collision threats that are not caught by the other layers e.g. aircraft that are not communicating their presence. In these situations non-cooperative collision avoidance is necessary. For a manned aircraft this final safety layer involves a pilot visually seeing and then avoiding a collision threat. For UAVs this final safety layer involves sensing and avoiding potential non-cooperative, mid-air collision threats with a proficiency matching or exceeding that of human pilots. A general guideline for how well human pilots detect potential collision threats is given in [8] where pilots, who were alerted to the presence of potential collision threat, were able to detect them with an 86% success rate at a median range of 2593m.

For sense and avoid (SAA), in small to medium sized UAVs, machine vision has been established as a potential technology as vision sensors have power, cost, size and weight benefits over other sensing approaches [9]. To meet the approximate human guidelines (as well as providing sufficient time for avoidance manoeuvres) it is desirable to detect aircraft as early as possible in an image sequence. At these ranges aircraft appear in vision sensors as a very small number (approx 1-10) of locally dim pixels that poorly contrast with the background (see Figure 1 for an example of aircraft size). In this paper we will focus on the “sense” aspect of SAA, specifically vision-based aircraft detection. There have been a variety of collision avoidance strategies to address the “avoid” aspect suitable for use with vision-based aircraft detection approaches (see [9], [10] and references therein), however that is out of scope for this paper.

The most promising approaches for vision-based aircraft detection that have been presented in the literature exploit the use of a multi-stage detection pipeline [7], [11]–[15]. The first stage typically exploits image pre-processing where spatial and visual aircraft features are highlighted and background clutter is suppressed. Popular approaches that

Fig. 1. An example of the above horizon conditions and aircraft target size which we are trying to detect. Note the texture in the clouds and the similarities between the aircraft and cloud features.
have been proposed in the literature to achieve this include morphology [7], [11], [12], and image frame differencing [7], [13]. Various machine learning [16] and deep learning [17], [18] approaches have been investigated to exploit the visual appearance of aircraft. Recently [15] used a deep CNN fused with morphological processing in the image pre-processing stage to detect aircraft above the horizon with a mean detection range of 2527m and no false alarms. However there are several drawbacks in using learnt approaches in this application as data is expensive and challenging to collect.

In the second stage temporal filtering approaches are often utilised to emphasise and extract features that possess aircraft-like dynamics such as: Viterbi-based filtering [11], [13], Kalman filtering [7], and hidden Markov model (HMM) filtering [12]-[14], [19].

Finally the detection logic stage aims to utilise the information available from the image pre-processing and temporal filtering stages in order to declare whether an aircraft is present or not. Some more simple approaches include looking for a detection in a $5 \times 5$ neighbourhood in the previous frame [20] and checking if the aircraft is increasing in area [21]. In [14] they proposed a new way of modelling aircraft emergence and were able to obtain a theoretically optimal detection logic. In this paper we aim to build off the theoretically optimal detection logic presented in [14] however we also aim to exploit the visual appearance of aircraft emergence and common false alarms.

The key contributions of this paper are casting the vision-based aircraft detection problem as an optimal stopping problem where we are able to introduce a penalty on detecting non-aircraft artefacts such as cloud features. We are able to establish that the optimal solution occurs on first entry of the change posterior into a stopping region characterised by the union of convex sets. Using these properties we propose a new detection logic stage that is able to improve detection ranges and false alarm rates.

The rest of this paper is structured as follows. In Section II we set up our model of aircraft dynamics and observations, this lets us cast our problem as an optimal stopping problem and propose two candidate rules for aircraft detection. In Section III we evaluate the performance of our proposed rules on flight data of near collision course encounters. We provide concluding remarks in IV.

II. PROPOSED SYSTEM

In this section we describe our proposed vision-based aircraft detection system. Similar to several state of the art approaches we exploit a multi-stage detection pipeline as seen in Figure 2. An image sensor captures an image which is then stabilised (the data we will test on was stabilised by a GPS-INS sensor see [22] for full details). We will now describe the other stages of our proposed detection system.

A. Image Pre-Processing

As noted in the introduction morphological processing is commonly exploited in the image pre-processing stage. For a greyscale image $I$, the dilation by a morphological structuring element $S$ is denoted by $I \oplus S$ and erosion is denoted by $I \ominus S$. Similar to several state of the art approaches [12], [14] we use bottom hat $[I \ominus S] \ominus I$ morphological processing which emphasises dark targets In [23] they observed that targets are generally darker than the background (supporting the notion that bottom hat filters are appropriate).

B. Proposed Temporal filtering

For our temporal filtering stage we use a HMM approach. Let us consider an aircraft in an image sequence which we are trying to detect as soon as possible. The aircraft emerges over time and is (potentially) first visually apparent in an image frame (that is, it has not visually emerged yet, or there is no collision threat). Let us denote this Markov chain as $X_k \in \{e_1, e_2, \ldots, e_N, e_{N+1}\}$ where $e_i \in \mathbb{R}^{N+1}$ are indicator vectors with 1 as the $i$th element and 0 elsewhere. For $i \in \{1, \ldots, N\}$, $e_i$ corresponds to the aircraft being visually apparent at the $i$th pixel and $e_{N+1}$ corresponds to the aircraft not being visually apparent. We denote this the out of image state.

The following are the three key parameters of a HMM description of target image motion for $1 \leq i, j \leq N+1$

1) State transition probabilities: $A^{i,j} = P(X_{k+1} = e_i | X_k = e_j)$ gives the probability of moving between different states. This can either be from the out of image state to a pixel in the image, or moving between different pixels. An aircraft not current located in the image is able to transition from the out of image state to any pixel in the image allowing for the possibility that an aircraft can visually emerge anywhere as it approaches from a distance.
Fig. 3. The possible transition of an aircraft at pixel (state) X. An aircraft is able to transition to its current pixel (state) or any of the neighbouring pixels beside or above indicated by the darker grey regions.

Possible aircraft inter-frame motion can be modelled by a transition patch (see [12] for detailed explanation of patches). In this paper we will use the patch that allows transitions to side and above pixels as seen in Figure 3.

2) Initial probabilities:
\( \pi^i = P(X_0 = e_i) \) denotes the probability that the target is initially located in state \( e_i \). We initialise our filter with an equal probability of \( \pi^i = \frac{1}{N+1} \) for all states.

3) Measurement probabilities:
\( b^i(y_k) = P(y_k|X_k = e_i) \) specify the probability of obtaining the observed image measurement \( y_k \) given that the target is actually in pixel location \( e_i \). At each time \( k > 0 \) we obtain a noise corrupted, greyscale image \( y_k \) that has been morphologically processed using bottom hat. We denote the measurement of the \( i \)th pixel at time \( k \) as \( y_k^i \) and following [12], [14] for \( i,j \in \{1, \ldots, N+1\} \) our diagonal matrix of (unnormalised) output densities is then given by the approximation
\[
B^{ij}(y_k) = \begin{cases} b^i(y_k) & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}
\]
where \( b^i(y_k) = y_k^i + 1 \) for \( i \in \{1, \ldots, N\} \) and \( b^{N+1}(y_k) = 1 \) (see [14] for justification). Essentially, the likelihood of an aircraft is proportional to the strength of the morphology output.

For \( k > 0 \) we can now calculate \( \tilde{X}_k \in R^{N+1} \) via the HMM filter [24], which can be regarded as an indicator of likely target locations
\[
\tilde{X}_k = N_k B(y_k) A \tilde{X}_{k-1}.
\]
With the initial condition \( \pi \) and where \( N_k \) are scalar normalisation factors defined by
\[
N_k^{-1} \triangleq \langle 1, B(y_k) A \tilde{X}_{k-1} \rangle.
\]
Here, \( \langle \cdot \rangle \) denotes an inner product.

C. Proposed Detection Logic

We now present our proposed detection logic. We first cast our problem as an optimal stopping rule and establish an optimal policy. We then propose a general pragmatic greedy decision rule for use as our detection logic.

1) Problem Formulation: Our desired detection logic can be characterised by the following continuing and stopping costs,
\[
\begin{align*}
C(\tilde{X}) & \triangleq \bar{c}_1 \tilde{X} \\
S_i(\tilde{X}) & \triangleq c_2 \tilde{X}^{N+1} + c_i \tilde{X}.
\end{align*}
\]
where \( \bar{c}_i \in \mathbb{R}^{N+1} \), \( c_i^{N+1} = 0 \), \( c_i > 0 \) for \( i \in \{1, \ldots, N\} \) denotes the delay penalty (when an aircraft is present but a detection has not been declared), \( c_2 > 0 \) denotes the false alarm penalty (when an aircraft is not present but a detection has been declared) and \( \bar{c}_i \in \mathbb{R}^{N+1} \) for \( i \in \{1, \ldots, N\} \) denotes a penalty for incorrectly declaring a detection on a non-aircraft artefact. Given the location decision \( i \) such that \( S(\tilde{X}) \triangleq \min_{i \in \{1, \ldots, N\}} S_i(\tilde{X}_\tau) \), we seek to design a stopping time \( \tau \geq 0 \) that minimises the following cost criterion
\[
J(\tau, \tilde{X}) \triangleq \mathbb{E} \left[ \sum_{k=0}^{\tau-1} C(\tilde{X}_k) + S(\tilde{X}_\tau) \mid \tilde{X} \right].
\]
Where \( \mathbb{E}[\cdot | \tilde{X}] \) denotes the expectation operation corresponding to the probability measure where the initial state has distribution \( \tilde{X} \).

2) Optimal Policy: Let us consider a stopping action \( u_k \in \{1 \text{ (continue), 2 (stop)} \} \). Then, there is an optimal policy \( \mu^*(\tilde{X}_k) \) to minimise our cost criterion \( J \) given by the value function \( V(\tilde{X}_k) \triangleq \min_{\tau} \{ J(\tau, \tilde{X}_k) \} \). This value function can be described by the following recursion (similar to [25, pg. 258])
\[
V(\tilde{X}_k) = \min \left\{ C(\tilde{X}_k) + \mathbb{E} \left[ V \left( \tilde{X}^+(\tilde{X}_k, y) \right) \mid \tilde{X}_k \right] + \bar{S}(\tilde{X}_k) \right\}
\]
where \( \tilde{X}^+(\tilde{X}_k, y) = (1, B(y) A \tilde{X}_k)^{-1} B(y) A \tilde{X}_k \), and \( B(y) = \text{diag}(b^1(y), \ldots, b^{N+1}(y)) \). If we let \( Q(\tilde{X}) \triangleq C(\tilde{X}) + \mathbb{E}[V(\tilde{X}^+(\tilde{X}, y))|\tilde{X}] \) denote the total cost incurred if continuing, then the optimal policy is given by
\[
\mu^*(\tilde{X}_k) = \begin{cases} 1 \text{ (continue) if } Q(\tilde{X}_k) < \bar{S}(\tilde{X}_k) \\
2 \text{ (stop) if } Q(\tilde{X}_k) \geq \bar{S}(\tilde{X}_k). \end{cases}
\]
In general, the value recursion \( V \) is difficult to compute making it unsuitable for use in the vision-based aircraft detection application. We can however establish some properties of the optimal solution that allow us to propose some practical rules.

3) Structure of the Optimal Policy: We now establish that our optimal stopping problem can be solved by finding \( N \) convex stopping sets (rather than solving the dynamic programming recursion equations directly). We define the stopping region for state \( e_i \) as follows \( R^k_{\delta} \triangleq \{ \tilde{X} : S_i(\tilde{X}) \leq Q(\tilde{X}) \} \).
Theorem 1. Consider the value recursion \( \Phi \) and let \( \mathcal{R}_S \triangleq \bigcup_{i \in \{1, \ldots, N\}} \mathcal{R}^i_S \) be the union of the \( N \) regions \( \mathcal{R}^i_S \). Then, the optimal stopping time given by

\[
\tau^* \triangleq \inf \{ k : \hat{X}_k \in \mathcal{R}_S \}
\]

Furthermore, the regions \( \mathcal{R}^i_S \) forming the optimal stopping region \( \mathcal{R}_S \) are convex and contain \( c_i \).

Proof. See Appendix for proof.

Theorem 1 establishes that the solution to our optimal stopping problem can be solved by finding the \( N \) convex stopping sets \( \mathcal{R}^i_S \). Given the parallels with our problem setup and quickest change detection and identification, we expect the stopping region \( \mathcal{R}_S \) to not be a connected region in general [26].

4) A Greedy Decision Rule: We now propose a practical greedy rule which can be applied in our application and establish some performance bounds.

Let us define \( \mathcal{R}^g_S \triangleq \{ \hat{X}_k \in \mathcal{R}_S \} \) and the union of sets \( \mathcal{R}_g = \bigcup_{i \in \{1, \ldots, N\}} \mathcal{R}^i_g \). We also define the probability of a false alarm (PFA) as \( \text{PFA} \triangleq P(X_k = c_{N+1}) \).

Lemma 1. For \( i \in \{1, \ldots, N\} \), we consider the union of sets \( \mathcal{R}^i_g \subset \mathcal{R}^i_S \) and the value recursion \( \Phi \). Then, the greedy decision rule given by

\[
\tau^g \triangleq \inf \{ k : \hat{X}_k \in \mathcal{R}_g \}
\]

achieves the performance bound

\[
\text{PFA} \leq \frac{c_m}{c_m + c_2},
\]

where the maximum delay constant \( c_m \triangleq \max_i \bar{c}_i \).

Proof. See Appendix for proof.

Lemma 1 suggests that a pragmatic solution to our optimal stopping problem is the greedy (sub-optimal) stopping rule \( \tau^g \). Moreover, if this rule is used then the cost function parameters can be related to a false alarm performance trade-off. The lemma also provides some insight into the role of cost parameters and false alarm performance.

Now, rather than solving the dynamic programming recursion equations \( \Phi \) directly or trying to calculate the convex stopping sets \( \mathcal{R}^i_S \) (both of which are computationally expensive), we are able to just run the HMM filter \( \Phi \). We highlight that you do not need to know the value of the costs constants \( \bar{c}_1, \bar{c}_2, \bar{c}_3 \) to run the filter.

5) Proposed Greedy Detection Logic: We now present two proposed greedy rules for use in the detection logic stage in our multi stage detection pipeline. Due to Lemma 1 we know that these stopping rules are a greedy (sub-optimal) solution to our cost criterion \( \Phi \).

Greedy Rule 1: We first consider a a simple stopping rule of the form

\[
\tau^{g_1} = \inf \left\{ k > 0 : \max_{i \in \{1, \ldots, N\}} \hat{X}^i_k \geq h^{g_1} \right\}, \quad (8)
\]

where \( \hat{X}^i_k \) is the \( i \)-th element of \( \hat{X}_k \). Intuitively, this rule declares a detection when the probability of being in one of the first \( N \) states is higher than a threshold.

Greedy Rule 2: We introduce a mapping \( M(\cdot) \) which reshapes the pixel elements of the vector \( \hat{X}_k \) to an image matrix and also introduce the inverse mapping \( M^{-1}(\cdot) \). Our second stopping rule is of the form

\[
\tau^{g_2} = \inf \left\{ k > 0 : \max_{i \in \{1, \ldots, N\}} \zeta^i(\hat{X}_k) \geq h^{g_2} \right\}, \quad (9)
\]

where \( \zeta^i(\cdot) \) is the \( i \)-th element of \( \zeta(\cdot) \) and \( \zeta(\cdot) \triangleq M^{-1}(M(\cdot) \ast \omega) \). Our second stopping rule is of the form

\[
\tau^{g_2} = \inf \left\{ k > 0 : \max_{i \in \{1, \ldots, N\}} \zeta^i(\hat{X}_k) \geq h^{g_2} \right\}, \quad (9)
\]

where \( \hat{X}^i_k \) is the \( i \)-th element of \( \hat{X}_k \). Intuitively, this rule declares a detection when the probability of being in one of the first \( N \) states is higher than a threshold.

Greedy Rule 2: We introduce a mapping \( M(\cdot) \) which reshapes the pixel elements of the vector \( \hat{X}_k \) to an image matrix and also introduce the inverse mapping \( M^{-1}(\cdot) \). Our second stopping rule is of the form

\[
\tau^{g_2} = \inf \left\{ k > 0 : \max_{i \in \{1, \ldots, N\}} \zeta^i(\hat{X}_k) \geq h^{g_2} \right\}, \quad (9)
\]

where \( \zeta^i(\cdot) \) is the \( i \)-th element of \( \zeta(\cdot) \) and \( \zeta(\cdot) \triangleq M^{-1}(M(\cdot) \ast \omega) \). The convolution operation with the kernel

\[
\omega = \begin{bmatrix}
-1 & -1 & -1 \\
-1 & 1 & -1 \\
-1 & -1 & -1
\end{bmatrix}.
\]

Intuitively we know that aircraft emerge in an image at approximately 1 pixel in size and hence only occupy one state. As our aim is to detect this aircraft emergence, we propose a kernel that penalizes the states around the aircraft as a type of penalty. Additionally, we found that this kernel was an effective way to reject cloud artifacts that are generally larger in size (more pixels) than an aircraft emerging in an image sequence.

III. RESULTS

In this section we examine the performance of our two proposed detection logic stages in the vision-based aircraft detection application. As a baseline we compare to the theoretically optimal intermittent signal detection (ISD) rule presented in [14], we denote this the baseline ISD rule.

We compare the performance of our proposed detection systems on 15 near mid-air collision course encounters between two fixed wing aircraft; the data collection aircraft was a ScanEagle UAV and the other aircraft was a Cessna 172. An example of the flight path of a head on near collision course encounter is presented in [15]. Figure 4 is an example of the size of aircraft we are detecting. For full details of the flight experiments see [22].
for each rule that achieve zero false alarms (ZFAs) in this data set. In practice, detection thresholds could be adaptively selected on the basis of scene difficulty such as proposed in [27].

The resulting ZFA detection ranges are presented in Figure 6. The mean detection distance and standard error was 2295 m and 22 m for Greedy rule 1, 2445 m and 61 m for Greedy rule 2 and 2227 m and 52 m for the baseline ISD rule.

Both our Greedy rules improved performance relative to the baseline ISD rule but importantly Greedy rule 2 improved detection ranges by a mean distance of 218 m (9.8%).

Intuitively, the performance benefits from Greedy rule 2 are from penalising the non-aircraft artefacts and as a result more effectively rejecting false alarms. This allows for a lower ZFA threshold to be set and therefore an earlier detection can be declared.

IV. Conclusion

In this paper we presented a new technique for the modelling and detection of aircraft above the horizon that is able to penalise non-aircraft artefacts (such as textured clouds and other unstructured terrain). We first posed the vision-based aircraft detection problem in a Bayesian setting and established some optimal properties. We then proposed a practical greedy rule and developed some bounds for characterising its performance. Finally, we investigated the performance of two greedy rules on real flight data where we were able to improve detection ranges by a mean distance of 218 m (9.8%) relative to a current state of the art vision-based aircraft detection technique.

REFERENCES

[1] Researchandmarkets.com, “Global UAV Drones Market Analysis, Trends, and Forecasts 2019-2025,” Tech. Rep., 2019. [Online]. Available: https://www.globenewswire.com/news-release/2019/11/22/195244/0/en/Global-UAV-Drones-Market-Analysis-Trends-and-Forecasts-2019-2025.html
A. Proof of Theorem 1

For \( i \in \{1, \ldots, N\} \), we define the cost for the location decision \( i \) as

\[
J_i(\tau, \hat{X}) \triangleq \mathbb{E} \left[ \sum_{k=0}^{\tau-1} C(\hat{X}_k) + S_i(\hat{X}_\tau) \right] \hat{X}
\]

(10)

and consider the function \( V_i(\hat{X}_k) \triangleq \min_{i \in \{1, \ldots, N\} \{J_i(\tau, \hat{X}_k)\}} \) for the stopping time recursion described by

\[
V_i(\hat{X}_k) = \min \left\{ C(\hat{X}_k) + \mathbb{E} \left[ V_i \left( \hat{X}^+(\hat{X}_k, y) \right) \right] \bigg| \hat{X}_k \right\}, S_i(\hat{X}_k).
\]

(11)

Noting that the cost is linear here, then according to [25, Theorem 7.4.2], \( V_i(\hat{X}_k) \) are concave in \( \hat{X}_k \). Moreover, our location decision \( \hat{S}(\hat{X}) \triangleq \min_{i \in \{1, \ldots, N\}} S_i(\hat{X}_\tau) \) for each \( \hat{X}_k \), \( J_i(\tau, \hat{X}) = \min_{i \in \{1, \ldots, N\}} V_i(\hat{X}_k) \), and hence \( V(\hat{X}_k) = \min_{i \in \{1, \ldots, N\}} V_i(\hat{X}_k) \) (swapping the order of the minimisation operations) implies that \( V(\hat{X}_k) \) is concave in \( \hat{X} \) (concavity is preserved under minimum operations). Therefore, \( V(\hat{X}_k) \) concavity gives that \( R_\hat{S} \) is convex sets (see similar proof steps in [25, Theorem 12.2.1]).

We proceed by considering the \( N \) location decisions as individual optimal stopping problems. For \( i \in \{1, \ldots, N\} \) if \( S_i(e_i) = 0 \) then \( e_i \in R_\hat{S}^i \). When \( \hat{X}_k \in R_\hat{S}^i \) for any \( i \) implies \( S_i(\hat{X}) \leq Q(\hat{X}) \) and under optimal rule (7) implies that the optimal action is to stop. Contrarily, \( \hat{X} \notin R_\hat{S}^i \) for any \( i \) implies \( S_i(\hat{X}) > Q(\hat{X}) \) and under optimal rule (7) implies the optimal action is to continue. Hence the union of the \( R_\hat{S}^i \) regions defines the optimal stop region as given in the theorem statement.

B. Proof of Lemma 1

For \( i \in \{1, \ldots, N\} \), \( C(\hat{X}_k) \geq S_i(\hat{X}_k) \) implies that \( Q(\hat{X}_k) \geq S_i(\hat{X}_k) \) and therefore, according to the optimal policy (7), we should stop when \( \hat{X}_k \in R_\hat{S}^i \). Further, if \( S_i(e_i) = 0 \) then \( e_i \in R_\hat{S}^i \) gives our first lemma result.

Simple algebra shows that \( C(\hat{X}_k) \geq S_i(\hat{X}_k) \) implies \( (c_m + c_2)\hat{X}_k^{N+1} + c_i \hat{X}_k < c_m \) (given that \( c_m (1 - \hat{X}_k^{N+1}) \geq c_i \hat{X}_k \)). Taking the expectation operation on both sides and using the idempotent property gives our lemma’s PFA performance bound result.