

# The International Journal of Robotics Research

<http://ijr.sagepub.com/>

---

## **A cascaded method to detect aircraft in video imagery**

Debadeepta Dey, Christopher Geyer, Sanjiv Singh and Matthew Digiioia  
*The International Journal of Robotics Research* published online 10 August 2011  
DOI: 10.1177/0278364911412807

The online version of this article can be found at:  
<http://ijr.sagepub.com/content/early/2011/08/09/0278364911412807>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



Multimedia Archives

**Additional services and information for *The International Journal of Robotics Research* can be found at:**

**Email Alerts:** <http://ijr.sagepub.com/cgi/alerts>

**Subscriptions:** <http://ijr.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

# A cascaded method to detect aircraft in video imagery

Debadepta Dey<sup>1</sup>, Christopher Geyer<sup>2</sup>, Sanjiv Singh<sup>1</sup> and Matthew Digiioia<sup>3</sup>

## Abstract

*Unmanned Aerial Vehicles (UAVs) have played vital roles recently in both military and non-military applications. One of the reasons UAVs today are unable to routinely fly in US National Airspace (NAS) is because they lack the sense and ability to avoid other aircraft. Although certificates of authorization can be obtained for short-term use, it entails significant delays and bureaucratic hurdles. Therefore, there is a great need to develop a sensing system that is equivalent to or has greater performance than a human pilot operating under Visual Flight Rules (VFR). This is challenging because of the need to detect aircraft out to at least 3 statute miles, while doing so on field-of-regard as large as 30° (vertical) × 220° (horizontal) and within the payload constraints of a medium-sized UAV. In this paper we report on recent progress towards the development of a field deployable sense-and-avoid system and concentrate on the detection and tracking aspect of the system. We tested a number of approaches and chose a cascaded approach that resulted in 100% detection rate (over about 40 approaches) and 98% tracking rate out to 5 statute miles and a false positive rate of 1 every 50 frames. Within a range of 3.75 miles we can achieve nearly 100% tracking rate.*

## Keywords

Sense-and-avoid, UAV, field robotics, perception

## 1. Introduction

Medium and small Unmanned Aerial Vehicles (UAVs) (such as Raven<sup>1</sup>, Predator<sup>2</sup>, Tigershark<sup>3</sup>, and Shadow<sup>4</sup>) are typically commanded via waypoints with the operators at remote locations. Such UAVs generally do not have the payload to carry radar systems or a traffic collision avoidance system (TCAS) or transponders. Collision avoidance is currently done by flight planning, use of ground- or air-based human observers and segregated air spaces. Lack of an autonomous sense-and-avoid system along with absence of regulations is preventing commercial UAVs from flying in the US National Airspace (NAS). The Radio Technical Commission for Aeronautics, Special Technical Committee 203 has been given the task to prepare Sense and Avoid Minimum Performance Standards for Unmanned Aircraft Systems (Radio Technical Commission for Aeronautics 2010) by December 2013. These proposals will be used by the Federal Aviation Administration (FAA) to determine final regulations. UAVs must not degrade the existing safety of the NAS, but the metrics that determine this are yet to be fully determined. It is still possible to state functional requirements and determine some performance minimums. For both manned and unmanned aircraft to fly safely in the

same airspace UAVs will need to detect other aircraft and follow the same rules as human pilots.

Key specifications of the international committee F38 on UAS systems standard F2411-04 (ASTM International 2004) proposed requirements which include a field of regard of 220° (horizontal) × 30° (vertical), minimum detection range of 3 statute miles under visual flight rules and a required miss distance of 500 feet. Without this capability, widespread utilization of UAVs will not be possible.

In this paper we focus on the sensing of aircraft with passive vision. Small size, low weight and power requirements make cameras attractive for this application. Multiple cameras can be used to cover the wide field-of-regard. A typical image of an aircraft at ranges of the order of a few miles is a few pixels in diameter. Figure 1 shows a 11 × 11 window

<sup>1</sup>School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

<sup>2</sup>Robot Corporation, Bedford, Boston, USA

<sup>3</sup>Penn State Electro-Optics Center, Freeport, PA, USA

## Corresponding author:

Debadepta Dey, School of Computer Science, Carnegie Mellon University, Newell Simon Hall 1111, Pittsburgh, PA 15213, USA  
Email: debadeep@cs.cmu.edu



**Fig. 1.** Selection of  $11 \times 11$  sub-windows showing the image of the Piper Archer II which was used as the approaching aircraft for collecting imagery, at a range of 1.5 miles. The camera and lens used had 0.41 milliradian/pixel resolution and a field of view of  $30^\circ(H) \times 21^\circ(V)$ .

around the image of the approaching aircraft at various ranges. Part of the challenge in detecting aircraft in such a wide field of regard reliably is the low signal to background ratio. Active sensors such as radar are not feasible because of their prohibitive power and size requirements (Bernier et al. 2005) for UAVs. Passive vision provides a low-cost, low-power solution albeit at the cost of a relatively high false-positive rate.

The contributions of our work include the formulation of a novel, cascaded method for long-range detection of aircraft for UAV sense-and-avoid purposes, and the evaluation of the approach on a large corpus of real imagery of flying aircraft versus range. Our approach uses a novel descriptor for low signal-to-noise ratio targets. This descriptor is then used to train a supervised classifier using a closed-loop training method to distinguish between true and false detections. The closed-loop training method uses the false positives from the commonly used morphological filtering stage as the negative examples. This suppresses a large percentage of the false positives even before the tracking stage of our approach. We achieve 100% detection (over about 40

approaches) and 98% tracking rate out to 5 statute miles and a false-positive rate of 1 every 50 frames. This is beyond the 3 statute miles limit imposed by proposed regulations.

In Section 2 the related work in vision-based sense-and-avoid systems for UAVs is discussed. In Section 3 we detail the requirements imposed on a complete sense-and-avoid system by regulations. In Section 4 we discuss the details of the vision-based aircraft detection algorithm. In Section 5 we outline our efforts to collect imagery of flying aircraft. In Section 6 we present details about the result of our algorithm on the corpus of real ground-truth imagery of aircraft. Section 7 details the relative performance of different cameras with which we collected data. Finally in Section 8 we discuss the path forward towards a field-deployable sense-and-avoid system.

## 2. Related work

Utt et al. (2005) described a fielded vision-based sensory and perception system with potential for implementation on small UAVs. McCandless (1999) proposed an optical flow

method for detecting aircraft. This is suitable only for moving objects and therefore is not useful for a target on a collision course, which will not appear to be moving. The use of morphological filtering is popular in computer vision-based sense-and-avoid systems (Gandhi et al. 2000; Carnie et al. 2006). However, this approach generates a significant number of false positives and requires tracking of the features over a large number of frames. Petridis et al. (2008) used AdaBoost to detect aircraft in low-resolution imagery. Track-Before-Detect (TBD) is an approach used especially with infrared imagery (Fernandez et al. 1990; Arnold and Pasternack 1990). Defence Research Associates have implemented a vision-based sense-and-avoid system on a Predator UAV system, which can track targets using three cameras and custom hardware McCalmont et al. (2002).

A field-deployable sense-and-avoid system must be able to operate in a variety of atmospheric conditions including fog, haze and even directly against the glare of the Sun. The operation of the system must not degrade beyond an acceptable level under all of these conditions. We developed an image formation model which accounts for the various atmospheric conditions. We used this model to predict the signal-to-background ratio of the image of the aircraft. The signal-to-background ratio is a measure of the detectability of the aircraft in the image. The image formation model is described in detail by Geyer et al. (2009). The image formation model also allows us to determine the suitability of any sensor combination before using the sensor and also to determine the minimum derived resolution for achieving a specified performance. The performance of the image formation model has been validated by the vast corpus of real imagery of flying aircraft that we collected during the course of this project.

### 3. Summary of regulations

Both civil and defense authorities have been involved in drafting regulations for UAV integration into NAS. This is in light of the fact that a large number of UAV operations in NAS will be for security and defense purposes. Hence, a summary of the proposed regulations from both civil and defense organizations is presented in this section.

In July 2003, manufacturers, members of the Association for Unmanned Aerial Vehicle Systems International (AUVSI) and other interested parties voted to create through American Society for Testing and Materials (ASTM International) the International Committee F38 on Unmanned Aerial Systems (UAS). In November, 2004 the Office of the Secretary of Defense (OSD) (Office of the Secretary of Defense 2004) released a plan for airspace integration for unmanned aviation. This document established top-level timelines and program milestones in relation to the FAA's strategic plans. This is in accordance with the mandate by Congress for the defense and civil authorities to work out the timelines and standards for full integration of UAVs into manned airspace. In November 2006 the

FAA (Federal Aviation Administration 2006) updated rules to establish protocols and guidelines for obtaining Certificates of Authorization (CoA) for temporary UAV operation in manned airspace. The definition of Remotely Operated Aircraft (ROA) was also incorporated to account for the fact that most UAVs are operated remotely.

In 2004 the F38 committee released standard F2411-04 which proposed requirements for sense-and-avoid systems (ASTM International 2004). These proposed standards serve as a baseline for current and future design of sense-and-avoid systems and the most relevant points are presented here. F2411 defined classes of sense-and-avoid systems, as well as functional and non-functional requirements for collision detection. The F2411 standard also defines three classes of sense-and-avoid systems based on their sense-and-avoid capabilities and where and how they apply them:

*Class 1 (Pilot-in-the-loop):* A system that warns a remote operator of a potential collision with another vehicle. The remote operator is responsible for evasive maneuvers.

*Class 2 (Automated Air):* In addition to detecting threats, a Class 2 system initiates a maneuver to avoid a potential mid-air collision or near mid-air collision autonomously.

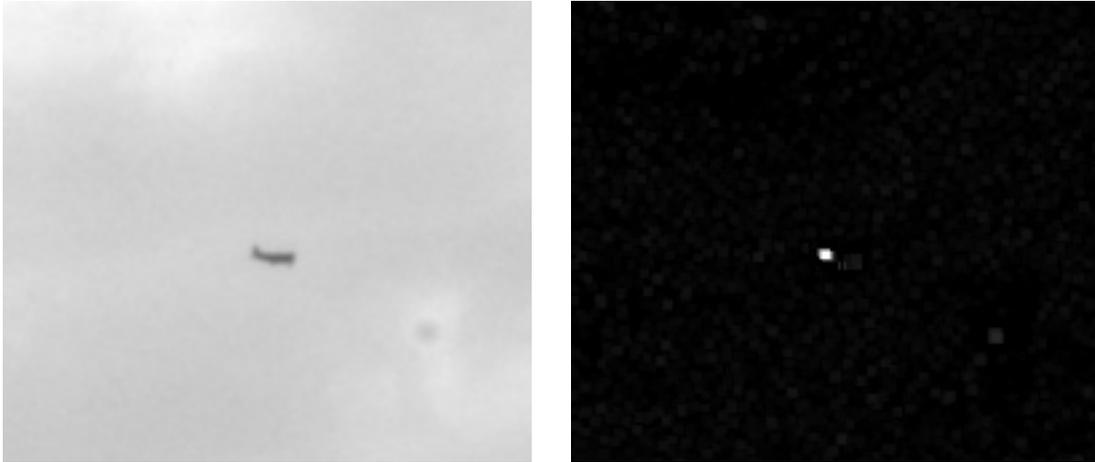
*Class 3 (Automated Air and Surface):* Class 3 systems have the additional capability of detecting and avoiding collisions with vehicles while taxiing on the runway.

Table 1 lists requirements as specified in the ASTM F2411 standard, as well as the Highly Capable UAVs (HCU) Payloads Planning Document. Where they differ, we note the two different requirements. These are only a subset of the requirements that are most likely to affect sensing requirements.

### 4. Aircraft detection and tracking

We experimented with a number of different approaches to detecting small targets with low signal-to-background ratios with an emphasis on methods that have both high detection rates and low computational complexity.

We have developed a multi-stage method that starts with a large number of candidates and winnows these down. The approach is a combination of existing approaches demonstrated in the literature. We start with a morphological filter (Gandhi et al. 2000) that looks for high-contrast regions in the image that are most likely to be aircraft. Next we use a classifier that has been trained on positive and negative examples and finally we track the candidates over time to remove false positives. We chose the tracking algorithm based on Shafique and Shah (2005) at the recommendation of the object tracking survey of Yilmaz et al. (2006). Neither (extended) Kalman filters nor particle filters were appropriate based on the number of false positives present. Other candidates were Multiple Hypothesis Tracking (MHT) and Joint Probability Data Association (JPDA).



**Fig. 2.** The image on the left shows part of the image of the Piper Archer II at a range of 2.87 miles. The image on the right shows the result of the morphological operation of the left image in Stage 1 of the processing pipeline. The dark aircraft image shows up as bright white spot.

**Table 1.** Basic requirements from the ASTM F2411 standard definition and Highly Capable UAV (HCU) Payloads Planning Document.

F2411 Class	Class 3: Autonomous air and ground sense and avoid (HCU 7.1.4.2.4, HCU 7.3.5)
Required miss distance	500 feet (F2411 4.2.1)
Field of regard	270°(H) × 40°(V) (HCU 7.3.1.1.1) 220°(H) × 30°(V) (F2411 4.2.2)
Minimum detection range	3 statute miles (HCU 7.3.1.1.2); ‘at a range to allow a resolution maneuver that results in a required miss distance of 500 feet or greater.’ (F2411 4.2.1)
Angular resolution	0.2 mrad or 0.011° (HCU 7.3.1.1.2.1)
Environmental	Day, night, bright light, and any weather as long as there is 3 statute miles visibility (HCU 7.3.1.1.2.1, HCU 7.3.1.2.4)
Accuracy and reliability	False alarm rates, false positive rates to be decided (HCU 7.3.1.2.3)

It was the simplest method to implement and best adapted to situations of the kind we ran into: large number of distracting false positives; non-Gaussian. In the following we discuss each ‘stage’ of detection in detail.

#### 4.1. Stage 1: Morphological filtering

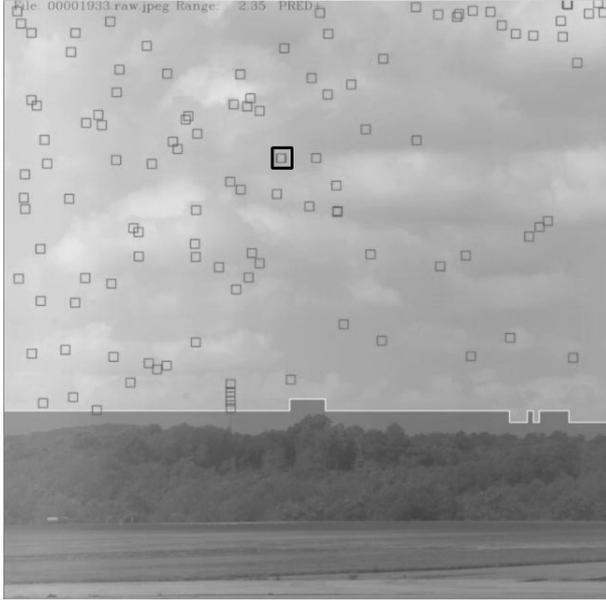
In the first stage, we apply a morphological filter that detects deviations from the background intensity. We use two types,

one favors dark targets against lighter backgrounds (positive), and the other favors light targets against darker backgrounds (negative). The positive morphological filter takes the form:

$$\mathcal{M}^+(x, y) = \mathcal{I}(x, y) - \max_{|i| \leq w} \min_{|j| \leq w} \mathcal{I}(x + i, y + j), \quad (1)$$

$$\max_{|i| \leq w} \min_{|j| \leq w} \mathcal{I}(x, y + i + j)$$

Here  $\mathcal{M}^+(x, y)$  refers to the output of the morphological operator at pixel location  $x, y$  in the image  $\mathcal{I}$  and  $w$  is half the length of the window side being considered around the pixel location  $x, y$ . As long as no  $2w + 1$  sub-window (we used  $w = 2$ ) contains all target pixels (higher intensity) and no background pixels (lower intensity), then all sub-windows will contain at least one (darker) background pixel. Since the background could be noisy, the maximums have the effect of finding a conservative greatest lower-bound for the background intensity. The difference, then, yields an estimate of the difference between signal and background for the pixel. The negative morphological filter,  $\mathcal{M}^-$ , swaps min for max, and negates the expression. From  $\mathcal{M}^+$  we choose the top  $n_+$  pixels above a threshold  $T_+$ , while suppressing local non-maxima, and construct a list of detections. We do the same for  $\mathcal{M}^-$ . Figure 2 shows an example aircraft image and the result of the morphological filtering on the example image. Figure 3 shows all the points of interest (indicated by black boxes) detected by the morphological operator in an image. Although the true location (indicated by the bold black box) of the aircraft is detected by the morphological operator, a large number of false positives can be seen in the image. The subsequent classification and tracking stages reduce the number of false positives significantly.



**Fig. 3.** Stage 1 output on example image. Example output of the morphological filter on an image. The black boxes are the points of interest detected by the detector. The bold black box represents the manually located aircraft. The below the horizon part of the image is automatically detected and not processed further. Although the position in the image of the approaching aircraft has been correctly detected by the morphological filter a large number of false positives are present in the frame.

## 4.2. Stage 2: Construction of a shape descriptor and SVM-based classification of potential targets

**4.2.1. Stage 2a: Construction of a shape descriptor** In order to discriminate between the actual image of the aircraft and the large number of false positives that are generated by the morphological filter in Stage 1 it is necessary to find a representation of the target that distinguishes it from false positives. In this section we describe the 19-attribute shape descriptor that we calculate for each sub-window in an image frame which contains the aircraft image. Using such a representation allows us to train a classifier (see Section 4.2.2) to detect false positives and remove them from further consideration.

For each detection we fit a Gaussian function to its  $(2r + 1) \times (2r + 1)$  sub-window (we settled on  $r = 7$ ) and construct a shape descriptor for the detection. Through trial and error we found a descriptor that was a good discriminator in test sequences. The descriptor encodes the parameters of the fitted Gaussian, as well as statistics computed from the residual image. We use an axis-aligned Gaussian, parameterized as follows:

$$\mathcal{G}(x, y; \sigma_x, \sigma_y, b, s) = b + \frac{s}{2\pi\sigma_x\sigma_y} e^{-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}}. \quad (2)$$

Here  $\sigma_x, \sigma_y$  are the standard deviations of the Gaussian along the  $x, y$  directions in the image,  $b$  is the background intensity and  $s$  is the intensity of the image of the aircraft or target. We center the Gaussian at the pixel with the largest absolute deviation from the window's mean intensity. We use gradient descent to minimize the sum of square errors between the input sub-window and  $\mathcal{G}(\cdot; \xi)$ , minimizing over  $\xi = (\sigma_x, \sigma_y, b, s)$ . To do this efficiently, we avoid repeated calls to the exponential function by pre-computing both a set of template  $\mathcal{G}$  over a range of  $\sigma_x$  and  $\sigma_y$  pairs, with  $(b, s) = (0, 1)$ , and a set of finite difference approximations to the partial derivatives of  $\mathcal{G}$  with respect to  $\sigma_x$  and  $\sigma_y$ . Figure 4 shows an example aircraft image's intensity profile and the fitted two-dimensional Gaussian window centered on the image of the aircraft.

Using the best fitting Gaussian  $\mathcal{G}^*$ , we compute a shape descriptor from the residual difference between the input image and  $\mathcal{G}^*$  in upper-left (UL), upper-right (UR), lower-left (LL), lower-right (LR), and center (C) regions. We construct both positive and negative half-sign sums. For example,

$$S_{UL}^+ = \sum_{\substack{1 \leq x \leq w \\ 1 \leq y \leq w}} \max [0, \mathcal{G}^*(x, y) - \mathcal{I}(x, y)],$$

...

$$S_C^- = -\sum_{\substack{w/2 < x < 3w/2 \\ w/2 < y < 3w/2}} \min [0, \mathcal{G}^*(x, y) - \mathcal{I}(x, y)].$$

Then, we construct min and max of positive and negative half-sign sums, e.g.  $S_{max}^+ = \max(S_{UL}^+, \dots, S_C^+)$ , and for each statistic we take its log normalized by the background intensity  $b$ , e.g.  $s_{max}^+ = \log(S_{max}^+/b)$ . We also compute the estimated signal-to-background ratio:

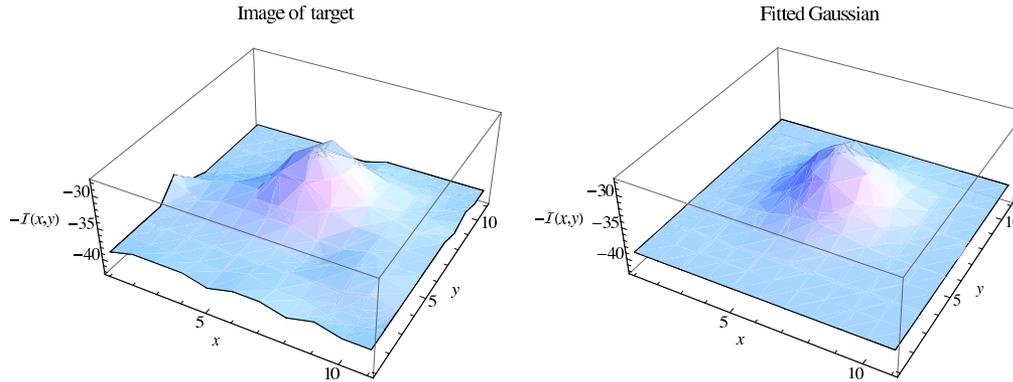
$$SBR = \frac{|b| + |s|/2\pi\sigma_x\sigma_y}{|b|}. \quad (3)$$

Finally, the shape descriptor we use is

$$\mathbf{d} = (b, s, \sigma_x, \sigma_y, SBR, s_{min}^+, s_{max}^+, s_{min}^-, s_{max}^-, s_{UL}^+, s_{UR}^+, s_{LL}^+, s_{LR}^+, s_C^+, s_{UL}^-, s_{UR}^-, s_{LL}^-, s_{LR}^-, s_C^-)$$

We associate this 19-dimensional vector with each detection.

**4.2.2. Stage 2b: SVM-based classification of potential targets** The computed shape descriptors,  $\mathbf{f}$ , are then used to train a supervised classifier. We chose a *support vector machine* (SVM) (Cristianini and Shawe-Taylor 2000) as the supervised classifier. The SVM was trained using descriptors from positive and negative examples taken from a sequence of hand-labeled images. For negative examples we used the false negatives produced by the morphological filter. We used radial basis functions as the kernel and obtained a probability of detection from a true target using the distance of an example from the classification boundary. During training we construct empirical densities of  $x$  for positive ( $p_x^+$ ) and negative ( $p_x^-$ ) classes using a mixture of Gaussians, and store a log-likelihood ratio function



**Fig. 4.** (Left) An example of an image of a target shown with inverted intensity. The aircraft image is the lump in the middle of the height field. This is an  $11 \times 11$  sub-window centered around the location of the image of the aircraft. (Right) The fitted Gaussian shape to the intensity profile of the aircraft image in the left figure.

$\ell(x) = \log p_x^+ / p_x^-$  in a look-up table keyed on  $x$ . We choose the kernel bandwidth just large enough to make the odds monotonic in  $x$ .

We keep only those detections whose odds exceed a minimum value of  $p_{\min}$ . Figure 5 shows an example image where as compared against Figure 3 a large percentage of the false positives have been eliminated. The few remaining false positives are eliminated in the tracking stage.

### 4.3. Stage 3: Tracking

The purpose of this stage is to track detections over time, associating detections to a list of tracked targets. Since many of the false positives are intermittent, we also use tracking to reduce the false-positive rate. We arrived at a simple procedure for target tracking that provides a full screen tracking system for high-definition imagery.

First, we always maintain a list of targets, and in steady state, it is the job of the tracker to associate to every existing target a detection. With any remaining detections, it also decides whether to create new targets.

For each existing target we consider a set of candidate detections, which are chosen from a wide search area around the predicted position of the target. Detections outside this area are not considered for pairing with that particular target. This gating (Kolawole 2002) technique prevents targets from being associated with unlikely candidate detections. For each potential matching detection, we evaluate the likelihood that the target and detection are associated given their respective descriptors.

Then, given a list of the likelihoods for the possible pairings we construct a graph with a node for each target and each detection, and edges between possible pairings, where the weights are the log likelihoods of pairings.

We construct a cost matrix, whose rows correspond to targets, columns to detections, and entries are the log likelihoods of the potential pairing, with  $-\infty$  given to non-candidate pairings. The goal is to choose entries from the matrix, no more than one from every row and no more than one from every column such that the sum of probabilities is a maximum. We use the Hungarian algorithm to find this matching (Papadimitriou and Steiglitz 1998). For the number of targets we typically have, usually less than 200, this computation can be computed in less than 8 ms.

Figure 6 shows the result of tracking consecutive frames on an example image. The track history is overlaid as a black trail on the image. The bold black box represents the actual location of the aircraft and the black box inside the bigger box the correctly tracked location. The other black boxes represent detected dead pixels and dust particles on the camera lens. Dead pixels and dust particles are very persistent and we detect such objects and remove them from further consideration. A list of known locations of dead pixels in an imager can also be supplied.

## 5. Data collection

We collected ground to air imagery of aircraft with ten different camera/lens combinations at Penn State Electro-Optics facilities at Jimmy Stewart Airport, Indiana, PA, USA. The data was collected on four separate days of 2, 16, 22 July and 5 September 2008. Four different infrared cameras were also used to acquire imagery. The infrared imagery has not been fused with the visual spectrum imagery yet. The total list of combinations of visual spectrum cameras we tested is listed below in Table 2.

For the ground-to-air data collection we used a Directed Perception D-100 pan tilt unit (PTU) with a geo-pointing



**Fig. 5.** Stage 2b output on an example image. The points of interest detected by the morphological filter in Stage 1 are assigned probabilities by a trained support vector machine (SVM) of being true positives. Only those detections which exceed a minimum value are kept. This step eliminates a large percentage of the false positives. As compared with Figure 3 the above image contains only a few false positives. The true location of the aircraft in the image is represented by the bold black box and the remaining points of interest after SVM-based false positive elimination are shown by black boxes. In spite of the elimination of a large percentage of the false positives, the true positive (inside the bold black box) is correctly preserved.



**Fig. 6.** Stage 3 output on an example image. The detected track history of the actual aircraft in the image has been overlaid as a black trail on the image. The actual aircraft position is represented by the bold black box and the detected and tracked position by the smaller black box which is shown positioned on the ground-truth location. The other black boxes represent detected dead pixels and detected dust particles on the camera lens. There are no false positives remaining in Stage 3 in contrast to Stages 1 and 2. See Figures 3 and 5. The inset image on the bottom left of the figure shows the aircraft zoomed in on the example image as detected by the algorithm.

**Table 2.** List of the camera/lens pairs that we tried during testing. In order, this table lists in each column: (i) the camera; (ii) the lens; (iii) the resolution of the camera; (iv) the number of megapixels (megapixel); (v) the field-of-view of the lens.

Camera	Lens	Res.	mpix	FOV
IPX-4M15	Nk105mm	2048×2048	4.2	11° × 11°
IPX-16M3	Nk105mm	4872×3248	15.8	25° × 17°
IPX-4M15	Zs85mm	2048×2048	4.2	13° × 13°
IPX-16M3	Zs85mm	4872×3248	15.8	31° × 21°
IPX-4M15	Zs50mm	2048×2048	4.2	22° × 22°
IPX-16M3	Zs50mm	4872×3248	15.8	50° × 34°
Lu125	CmpVF@36mm	1280×1024	1.3	18° × 14°
IPX-2M30	CmpVF@36mm	1600×1200	1.9	24° × 18°
Lu125	Cmp25mm	1280×1024	1.3	25° × 20°
IPX-2M30	Fujinon 25mm	1600×1200	1.9	33° × 25°

module (GPM). We mounted the cameras on the PTU and used a Piccolo autopilot to send the GPS stream from the approaching aircraft to the GPM. The GPM automatically computes the angles the PTU needs to be pointing at in order to always have the aircraft in the field of view of the

cameras. A Piper Archer II was used as the approaching aircraft during the data collection. Figure 7 shows the PTU and camera setup used at the airport.

In Figure 8 we show the pattern the approaching aircraft flew as we gathered imagery on 22 July 2008. The approaching aircraft flew out to 5–6 miles each pass. To date we have collected 2.5 TB of imagery of which in 2 TB the position of the aircraft has been picked out by hand for ground-truth purposes. This corpus of real imagery has been used to analyze the performance of our algorithm.

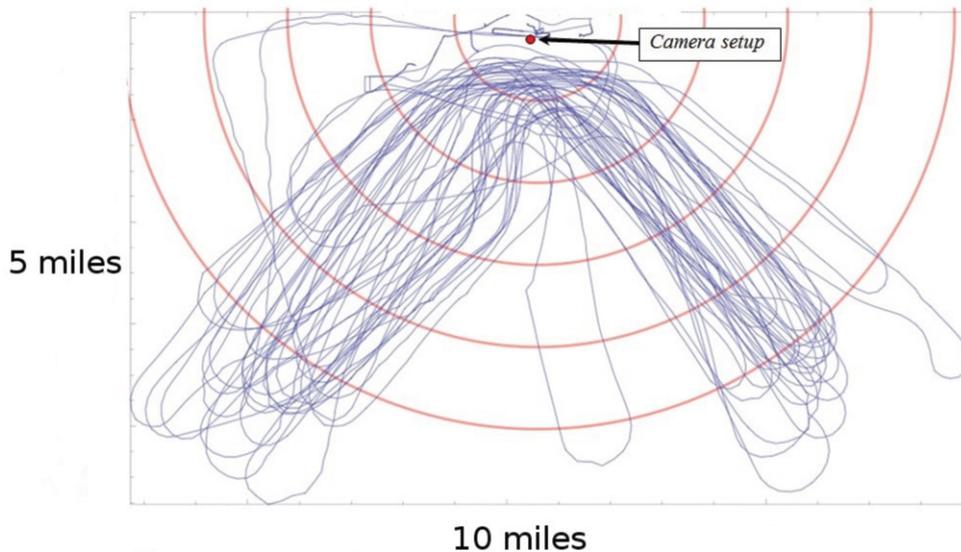
## 6. Detection and tracking results

We evaluated the performance of each stage of the algorithm using receiver operator characteristic (ROC) curves, which measure specificity (ability to reject outliers) and sensitivity (ability to detect true target) of a detector on about 2 TB of imagery of above the horizon flying aircraft. Here we present the results using the imagery from the Imperx 4 megapixel camera with a Zeiss 85 mm lens which includes data from 4 days of data collection.

Stage 2 improves the false-positive rate by a factor between 6 and 17 depending on the detection rate over Stage



**Fig. 7.** A Directed Perception D-100 pan tilt unit was used along with a geo-pointing module to automatically keep the approaching aircraft in the field of view of the mounted cameras. The pan tilt unit tracked the location of the approaching aircraft using the GPS stream over a radio modem from the aircraft.



**Fig. 8.** A manned aircraft equipped with a GPS was flown in a series of flights such that it was in the field of view of the ground-based cameras on 22 July 2008. The circles show the distance to the cameras in miles.

1. Refer to Table 3. We get a vast improvement with tracking in Stage 3. In the case of both Stage 1 and Stage 2 the variable affecting rates is a threshold. For Stage 1, the threshold is the value returned by the morphological filter at the detection. For Stage 2, the threshold is the probability according to the SVM classifier, that the detection is a target.

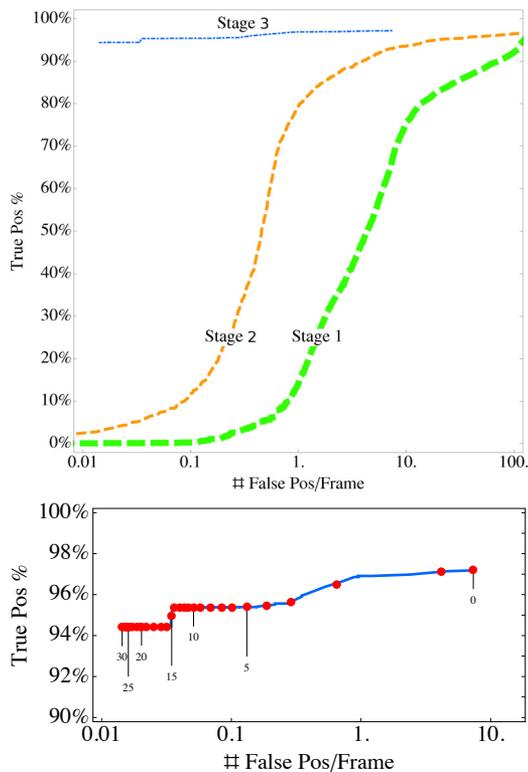
Figure 9 (top) shows the ROC curve for Stage 1, Stage 2 and Stage 3 of the algorithm. Whereas before the value affecting performance was a threshold on the output of a filter or classifier, in this case the threshold is the number of frames for which a target has been tracked. It is to be noted that the best overall detection rate of Stage 3 is higher

**Table 3.** Shows the number of false positives (FP) per frame for Stage 1 and Stage 2 as a function of the true positive percentage. Stage 2 reduces the false positive rate by a factor between 6 and 17 times as Stage 1.

TP%	Stage 1 FP/frame	Stage 2 FP/frame	FP Reduction Factor
95%	120	20	5.9×
90%	66	3.9	17×
80%	14	1.0	14×
70%	8	0.66	12×
60%	6.2	0.56	11×

**Table 4.** Performance of various stages of the algorithm. Stage 3 achieves a false positive (FP) reduction rate of 571 times over Stage 1

TP%	Stage 1 FP/frame	Stage 2 FP/frame	Stage 3 FP/frame	FP Reduction Factor
97%	—	—	7.3	—
95%	120	20	0.035	571×
90%	66	3.9	—	—
80%	14	1.0	—	—
70%	8	0.66	—	—
60%	6.2	0.56	—	—



**Fig. 9.** The top shows the receiver operator characteristic curve for true positive and false positive for the three main stages of the algorithm on 4 megapixel imagery. The curve for Stage 3 shows almost perfect tracking rate with a false-positive rate of as low as 0.02 per frame. The bottom shows the effect of varying the minimum number of frames that a potential target has to be tracked for before being declared as a target in Stage 3. This curve is very flat as most true positives have long tracks and false positives have short tracks.

than the best overall detection rate of Stage 2, even though it is based on the output of Stage 2. We believe that this is a temporal effect, in that detections that are intermittently below threshold are picked up by the tracker. The detection rate decreases slightly at closer ranges. This is due to

the fact that the algorithms were not optimized for close ranges.

Figure 9 (bottom) shows the effects of the variance of the minimum number of frames that a potential target has to be tracked for before being declared as a target. The points on the curve are the number of frames that a target has to have been tracked for in order for it to be declared a possible target. In our experiments we let this threshold go up to 30 frames, at which point the false-positive rate was 0.014 FP/frame and detection rate was 95%. It is to be noted that this curve is very flat. Most of the true positives have long tracks and almost all outliers have short tracks.

Overall there is a significant decrease in the number of false positives per frame. We add to Table 3 the results of Stage 3 and present them in Table 4. The entries for tracking rates below 95% are not filled in because we chose not to evaluate the threshold frames beyond 30. If we had, the detection rate would have eventually fallen. We find that tracking in Stage 3 improves the false-positive rate by a factor of over 500 times over Stage 2.

We found a reasonable compromise in false-positive and true-positive rate when we insisted that targets be tracked for at least 10 frames. Then the overall tracking rate was 95%, the false-positive rate was 0.05 false positives per frame. Between 2.5 and 3.75 miles the tracking rate is nearly 100%. Since the signal-to-background ratio increases as the range to the intruder aircraft decreases it can be assumed that our detection and tracking algorithm will continue to have nearly 100% tracking rate at ranges lower than 2.5 miles. If we let the targets be tracked for 25 frames then the false-positive rate is reduced to about 0.02 false positives per frame (1 false positive in every 50 frames). Refer to Extension 1 for a video of the 3 stage tracking procedure on a single approach of the aircraft.

About 80% of the false positives that made it through the tracking of at least 10 frames were items that are of interest to collision avoidance. Most of the false positives were birds or landmarks on the ground that were not segmented out by the horizon detector (e.g. an antenna in the distance). These targets are of interest and could be considered useful (Figure 10). If we account for the useful targets detected then the false-positive rate at the 10 frames tracking threshold drops

from 0.05 to 0.01 false positives per frame (1 false positive in every 100 frames). At the 25 frames tracking threshold the false positive rate drops to 0.004 false positives per frame (1 false positive every 250 frames).

It is to be emphasized that the tracking rate is the accuracy of detecting the aircraft in each of the image frames containing the aircraft for a given approach of the aircraft averaged over all of the approaches of the intruder aircraft towards the ground station (see Section 5). The data we have used contains about 40 such approaches. Our detection algorithms were able to successfully detect an incoming aircraft on every approach starting at 5 miles. Before this system can be fielded extensive testing on several thousand hours of data taken under different atmospheric conditions and geographic locations will have to be carried out.

We have developed and demonstrated a vision-based algorithm that achieves a reasonable true-positive rate of approximately 98% out to 5 statute miles and a false-positive rate of 1 in every 50 frames which exceeds the minimum range requirement of 3 statute miles imposed by regulatory requirement.

## 7. Comparison of camera performance

The effects on the performance of an aircraft detection system can be decoupled into controllable system factors and uncontrollable environmental factors. Environmental factors include atmospheric properties, in particular visibility due to weather conditions and attenuation in the atmosphere; lighting conditions; and intruder aircraft properties. For all of the uncontrollable factors, we have to be able to accommodate any range of these factors that are within the necessary or desired operating range for the system. For example, if the visibility do not meet visual meteorological conditions (VMC), then we cannot plan to fly under visual flight rules (VFR). However, if the conditions are just above VMC, i.e. the visibility is 3.1 miles, then the sense-and-avoid system still must be able to detect aircraft out to at least 3 miles. Similarly if there is sufficient light such that the UAV is flying VFR, then no matter how dark it is, we still ought to be able to detect out to 3 miles.

Our first observation is that since the threshold objective is to enable a system to be equivalent in reliability and safety to a pilot operating under VFR, we can assume that the environmental factors are no worse than the limits of VMC. In other words, though these factors may be uncontrollable, at least to meet a threshold objective, we can assume that these effects meet visual meteorological conditions.

As for the controllable system factors, for an aircraft detection approach based on imagery (although we expect later to possibly integrate other sensing modalities), the primary factors that affect detection performance in a single image are: signal and image processing, and the detection algorithms; the optics including lens focal length, aperture, etc.; and, the imaging device, including its pixel pitch, the

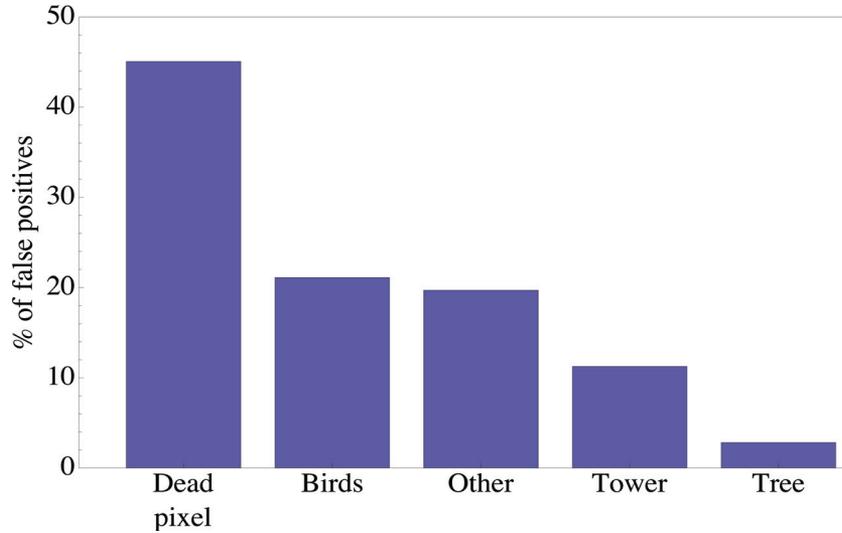
exposure used, noise characteristics, and dynamic range, to name a few. In a complete system, numerous other factors affect performance some of which are dependent on the platform, the aircraft dynamics, for example, and where the resulting performance is then determined by the capabilities of tracking and state estimation algorithms.

Our second observation is that the controllable factors are either correlated or are decoupled. In particular, if the modulation transfer function (MTF) of the lens is appropriate for the resolution of the imager, then the resolution of the imager and optics determine a number of radians per pixel, which can then be varied. We further expect the performance of the detector to be independent of the imaging device. The purpose is to reduce a combinatorial explosion of possibilities, and to characterize the detectors performance in quantities that are independent of the imager. Hence, we examine the performance of the top-performing detector as a function of the signal to background ratio which is an empirical measure of the contrast ratio of the image of the aircraft to the background as introduced in Equation (3). Then, the signal-to-background ratio depends on the imaging device (optics and sensor), as well as the environmental factors. As long as the signal-to-background ratio generated under the worst conditions by the chosen imager is above threshold, the combination of the detector with the imaging device is expected to perform above threshold. We take this argument at least as far as informing testing. It is certainly possible that coupling effects, atypical noise characteristics of an imager having affects on detection for example, may cause deviations from predicted performance; if there are significant deviations from expected performance, then we ought to be able to understand what effects that have not been modeled could be contributing to inconsistencies in predictions.

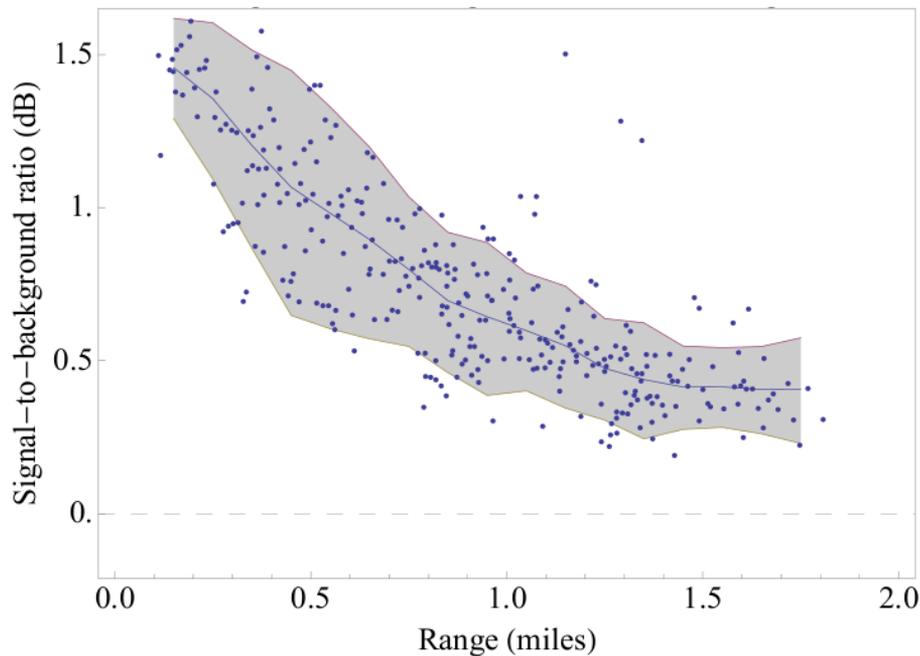
Figure 11 shows the signal-to-background ratio, of the aircraft, versus the range of the aircraft being imaged using the Lumenera 125 camera with a 25 mm lens and with data collected on 22 April at Jimmy Stewart Airport, PA. We also fit a curve to the data, and the gray area shows three times the median deviation from the curve in each direction. We observe that the signal-to-background ratio falls steadily as the range increases. Thus, there is a high correlation between the signal-to-background ratio and range.

In an effort to explain the characteristics observed in the previous section, we developed a model of image formation for aircraft imaged above the horizon, and have compared the model's predicted characteristics to the observed image characteristics. We constructed a head-on silhouette of a Piper Archer III, which has the same geometry as the Piper Archer II, from a head-on image. We then used a model of image formation that takes into account atmospheric effects that was proposed by Nayar and Narasimhan (1999). In this model we have aimed to take into account the effects of haze and atmospheric scattering and potential defocus due to poor or mismatched optics.

Using this model we can predict the signal-to-background ratio of the image of an aircraft as a function



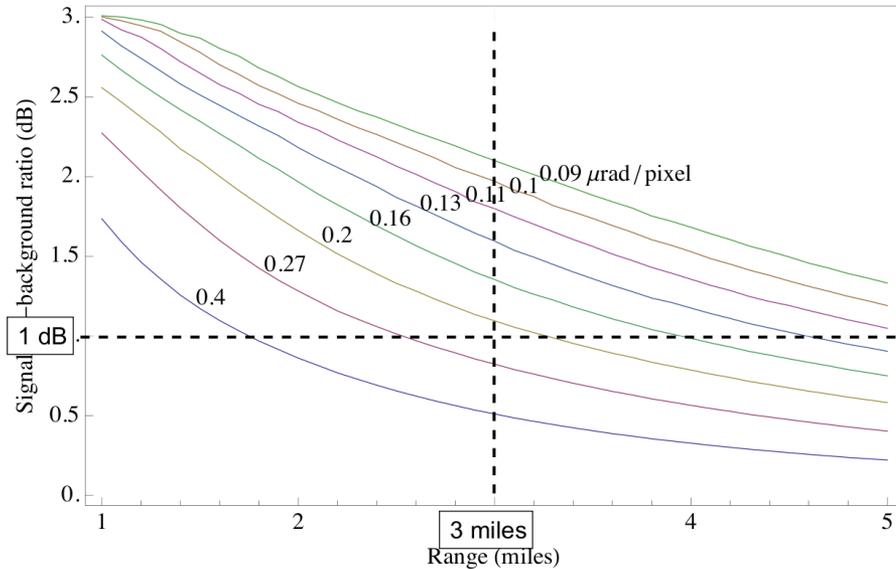
**Fig. 10.** About 80% of detected false positives which are not aircraft are not really false positives and can be of advantage to the overall system. The false positives under the ‘other’ category in the figure were due to image regions that looked like aircraft at a distance and were not discriminative enough to be rejected by the classifier. This group can be considered as real false positives as they are hazards.



**Fig. 11.** Signal-to-background ratio (SBR) versus range to the aircraft in miles. The SBR decays as the range to the aircraft increases. The black dots are the individual SBR of the aircraft image in collected imagery. The center line is a curve fitted to the data and the gray area shows three times the standard deviation from the curve in each direction.

of range, with different curves for different resolution. We measure resolution in milliradians per pixel (mrad/pixel). The signal-to-background ratio is a key measure of identifiability in the image, therefore such a prediction can give some information about what resolution would be sufficient for target detection out to the desired range. Figure 12 shows a plot of the predicted signal-to-background ratio

as a function of range with different curves for different resolutions. We note from collected data that a signal to noise ratio of at least 1 dB is required for detecting an aircraft out to 3 miles with a rate of 95%. Using this information and the model of image formation the system designer can choose the resolution of the imaging system required to achieve a minimum performance under any specified worst-case



**Fig. 12.** Predicted signal-to-background ratio (SBR) versus range to the aircraft. Different resolution camera/lens pairs are represented by different curves. We note that in order to have 95% detection rate at a range of 3 miles a SBR of at least 1 dB is necessary. This imposes a constraint on which camera/lens pairs are suitable for achieving this minimum performance under all ranges of atmospheric conditions defined under visual flight rules (VFR). The set of camera/lens pairs whose predicted curves pass below the boundary imposed by 1 dB SBR at 3 miles are unsuitable for this scenario.

atmospheric conditions independent of the specific capabilities of any detection algorithm. The details of the image formation model and validation using the data collected using the camera/lens pairs listed in Table 2 can be found in Geyer et al. (2009).

Figure 13 shows the plot of the predicted resolutions of each camera/lens pair listed in Table 2 versus the measured resolution for each pair. The predicted resolutions ( $x$ -axis) are obtained by the image model whereas the measured resolutions ( $y$ -axis) are obtained by the curve fitted to the observed signal-to-noise ratio versus range from the collected data. If any point is above the line  $y = x$  then the camera performed worse than predicted; if it is below the line then it performed better than expected. We notice several trends that indicate the possibility of systematic biases, although their causes are not known:

- In general, the IPX-4M15 seems to perform worse than the IPX-16M3. Because these sensors have the same pixel pitch, for each lens tried, the points for the IPX-4M15 and IPX-16M3 lie on a vertical line. For each lens, the IPX-4M15 is consistently higher (therefore worse) than the IPX-16M3.
- The Zeiss 50 mm lens outperforms both the Zeiss 85 mm and the Nikon 105 mm lenses. For both the IPX-4M15 and the IPX-16M3, the measured resolution is better than predicted only for the Zeiss 50 mm lens; for the Zeiss 85 mm and Nikon 105 mm the measured resolutions are worse than predicted for both cameras.
- The Lu125 performs worse than the IPX-2M30. The Lu125 consistently performed worse than predicted,

and the IPX-2M30 consistently performed better than expected.

Differences in lens can be caused due to uncalibrated lenses but differences in imagers are clearly present as they are being compared with the same lenses. That these trends seem to be consistent across cameras and lenses seem to indicate that the differences between measured and predicted SBR versus range curves is due to inherent performance differences in the cameras and lenses, instead of problems with the model.

Table 5 shows the list of camera/lens pairs and their estimated detection ranges assuming a 1 dB minimum signal-to-noise ratio.

We note that such an analysis allows the system designer to extrapolate performance on any specific resolution camera/lens pair and predict performance using any other camera/lens pair without having to obtain actual data.

## 8. Future work

Detecting aircraft below the horizon presents different challenges. It is not possible to eliminate the part of the scene below the horizon and search for aircraft in the sky in scenes where the aircraft is against the ground and hence can have a much more confusing background. Collecting such below the horizon imagery of aircraft and evaluating detection performance is essential for a complete system.

Currently, our system detects bearing to targets that must be avoided. An important extension will be to estimate the range to the target so that precise maneuvers can be

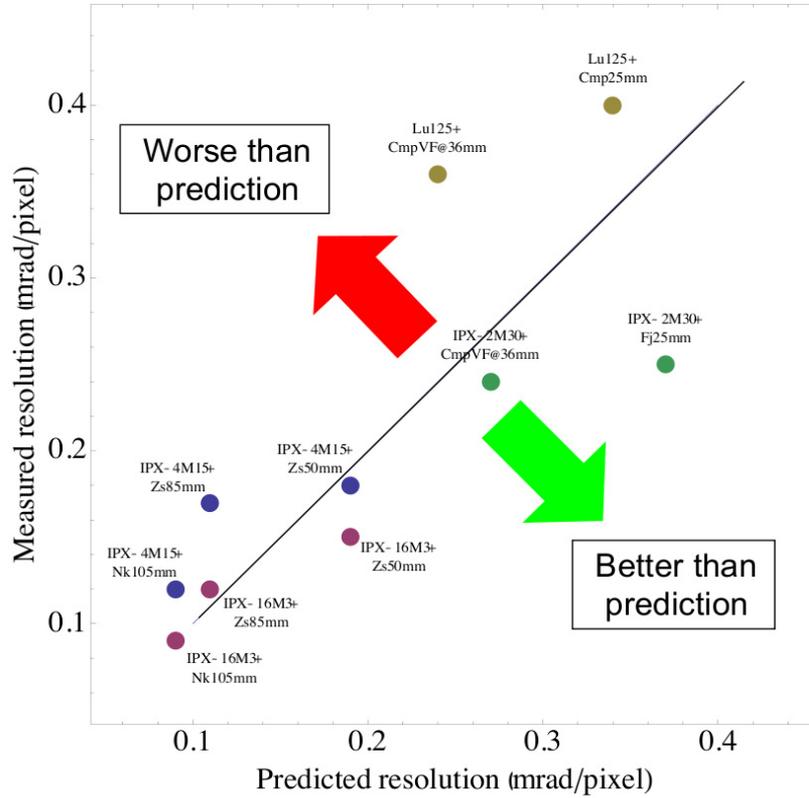


Fig. 13. Predicted resolutions of each camera/lens pair versus the measured resolution for each pair.

Table 5. Predicted and estimated detection ranges for all tested camera pairs assuming a 1 dB minimum signal-to-noise ratio is required for 95% detection. We also list the number of cameras required for each camera/lens combination tested to cover a 270°(H) × 40°(V) field of view.

Camera	Lens	Res.	mpix	FOV	#cams	mrad	Predicted range	Estimated range
IPX-4MI5	Nk105mm	2048×2048	4.2	11° × 11°	67	0.09	> 5.0	5.0
IPX-16M3	Nk105mm	4872×3248	15.8	25° × 17°	20	0.09	>5.0	>5.0
IPX-4MI5	Zs85mm	2048×2048	4.2	13° × 13°	48	0.11	>5.0	3.9
IPX-16M3	Zs85mm	4872×3248	15.8	31° × 21°	13	0.11	> 5.0	>5.0
IPX-4MI5	Zs50mm	2048×2048	4.2	22° × 22°	17	0.19	3.4	3.6
IPX-16M3	Zs50mm	4872×3248	15.8	50° × 34°	5	0.19	3.5	4.1
Lu125	CmpVP @36mm	1280×1024	1.3	18° × 14°	33	0.24	2.8	2.0
IPX-2M30	CmpVF @36mm	1600×1200	1.9	24° × 18°	19	0.27	2.6	2.9
Lu125	Cmp25mm	1280×1024	1.3	25° × 20°	17	0.34	2.1	1.8
IPX-2M30	Fujinon 25mm	1600×1200	1.9	33° × 25°	10	0.37	1.9	2.7

planned. We are currently investigating active ranging systems that can be pointed at potential targets, to estimate range and further reduce false positives.

Fusing infrared imagery with visible spectrum imagery is another area that we need to address that can help further reduce false positives.

Although the current algorithm takes about 0.8 seconds per 4 megapixel frame on an AMD Athlon X2 3800+ processor, most of the computation is image processing and hence amenable to parallelization. Specialized hardware such as digital signal processors are promising.

All of the above issues affect how a collision detection and warning system should be designed so as to cover the desired field-of-regard.

**Note**

1. See [http://www.avinc.com/uas/small\\_uas/](http://www.avinc.com/uas/small_uas/)
2. See [http://www.avinc.com/uas/small\\_uas/](http://www.avinc.com/uas/small_uas/)
3. See <http://www.microav.com/tigershark.html>
4. See [http://www.aaicorp.com/products/uas/uas\\_main.html](http://www.aaicorp.com/products/uas/uas_main.html)

## Funding

This material is based upon work supported by the Unique Missions Division, Department of the Army, United States of America under Contract No. W91CRB-04-C-0046. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of Army or Penn State EOC.

## References

- Arnold J and Pasternack H (1990) Detection and tracking of low-observable targets through dynamic programming. *Proceedings of SPIE* 1305: 207.
- ASTM International (2004) Standard specification for design and performance of an airborne sense-and-avoid system, F2411-04. West Conshohocken, PA: ASTM International.
- Bernier R, Bissonnette M and Poitevin P (2005) DSA radar: development report. In *Proceedings of UAVSI 2005*.
- Carnie R, Walker R and Corke P (2006) Image processing algorithms for UAV 'sense and avoid'. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*, pp. 2848–2853.
- Cristianini N and Shawe-Taylor J (2000) *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press.
- Federal Aviation Administration (2006). Special Operations Order 7610.4. Available at: [http://www.faa.gov/regulations\\_policies/orders\\_notices/index.cfm/go/document.information/documentID/99640](http://www.faa.gov/regulations_policies/orders_notices/index.cfm/go/document.information/documentID/99640).
- Fernandez M, Aridgides T and Bray D (1990) Detecting and tracking low-observable targets using IR. *Proceedings of SPIE* 1305: 193.
- Gandhi T, Yang M-T, Kasturi R, Camps OI, Coraor LD and McCandless J (2000) Detection of obstacles in the flight path of an aircraft. In *Proceedings of CVPR*. Los Alamitos, CA: IEEE Computer Society Press, pp. 2304–2311.
- Geyer C, Dey D and Singh S (2009) *Prototype Sense-and-Avoid System for UAVs*. Technical Report CMU-RI-TR-09-09, Robotics Institute, Carnegie Mellon University.
- Kolawole M (2002) *Radar Systems, Peak Detection and Tracking*. Amsterdam: Elsevier.
- McCalmont J, Utt J and Deschenes M (2002) Detect and avoid technology demonstration. In *Proceedings of the American Institute of Aeronautics and Astronautics Infotech*.
- McCandless J (1999) Detection of aircraft in video sequences using a predictive optical flow algorithm. *Optical Engineering* 38: 523.
- Nayar S and Narasimhan S (1999) Vision in bad weather. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*, Vol. 2.
- Office of the Secretary of Defense (2004). *Airspace Integration Plan for Unmanned Aviation*. Available at: <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA431348>.
- Papadimitriou C and Steiglitz K (1998) *Combinatorial Optimization: Algorithms and Complexity*. Courier Dover Publications.
- Petridis S, Geyer C and Singh S (2008) Learning to detect aircraft at low resolutions. In Gasteratos A, Vincze M and Tsotsos JK (eds), *Proceedings of ICVS (Lecture Notes in Computer Science*, Vol. 5008). Berlin: Springer, pp. 474–483.
- Radio Technical Commission for Aeronautics (2010) *Special Committee (SC) 203. Minimum Performance Standards for Unmanned Aircraft Systems*. RTCA Paper No. 065-10/PMC-790, Radio Technical Commission for Aeronautics, Inc. Available at: [http://www.rtca.org/CMS\\_DOC/SC-203-TOR-PMC%20Approved-04-26-2010.pdf](http://www.rtca.org/CMS_DOC/SC-203-TOR-PMC%20Approved-04-26-2010.pdf).
- Shafique K and Shah M (2005) A noniterative greedy algorithm for multiframe point correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27: 51–65.
- Utt J, McCalmont J and Deschenes M (2005) Development of a sense and avoid system. In *AIAA Infotech at Aerospace*.
- Yilmaz A, Javed O and Shah M (2006) Object tracking: a survey. *ACM Computing Surveys* 38: 13.

## Appendix: Index to Multimedia Extensions

The multimedia extension page is found at <http://www.ijrr.org>

### Table of Multimedia Extensions

Extension	Type	Description
1	Video	Illustration of the multi-stage aircraft detection algorithm