Maximizing object detection using sUAS

Manore, Curtis, Manjunath, Pratheek, Larkin, Dominic

Curtis Manore, Pratheek Manjunath, Dominic Larkin, "Maximizing object detection using sUAS," Proc. SPIE 11605, Thirteenth International Conference on Machine Vision, 1160528 (4 January 2021); doi: 10.1117/12.2586989

SPIE.

Event: Thirteenth International Conference on Machine Vision, 2020, Rome, Italy

Maximizing Object Detection Using sUAS

Curtis Manore[a], Pratheek Manjunath[a], and Dominic Larkin[a]

[a]United States Military Academy, West Point, New York, USA

ABSTRACT

This paper examines optimal look-angles for a camera which is mounted on a small unmanned aerial system (sUAS), that provides for maximized object detection on the ground. Using a generic convolutional neural network (CNN), this research identifies the best angle for detecting a ground target from an aerial perspective. The study involves altering camera angles on an sUAS that is flown along a fixed trajectory and then determining the angle which provides the highest detection rate of predefined objects, which are emplaced at known locations on the ground. The experiment is conducted in simulation and validated on a physical quadcopter. The results of this paper directly influence the U.S. Army's research efforts on training neural networks and developing object detection algorithms.

Keywords: Camera Angle, Pixel Density, Object Detection, Neural Network, UAS, Remote Surveillance

1. INTRODUCTION

Many Department of Defense (DoD) agencies seek to implement object detection on UAS to gain a better vantage point during surveillance and reconnaissance missions. Another application of aerial object detection is in search and rescue operations conducted by military and law enforcement. Advances in sensor technology make it possible to employ sensor systems on unmanned vehicles to identify occluded soldiers during rescue operations [1], detecting enemy combatants, or for monitoring at border crossings [2]. DoD agencies have an interest in detecting more than just people. Especially in combat zones, it is of interest to identify or detect objects, buildings, and terrain features. Hence, developing effective image classification and object detection techniques is essential to enhance the U.S. military's understanding of the battlefield and remain technologically ahead of near-peer adversaries.

Specifically, sUASs have unique advantages over larger unmanned aerial vehicles (UAVs) and satellites for Intelligence Surveillance and Reconnaissance (ISR) purposes. Apart from the operational and logistical benefits that multirotor sUASs provide, they also reduce the post-processing requirement on the data recorded. This computational advantage arises from the fact that images captured by them are often from a low altitude, which limits the clutter, image frame, and file size. This advantage allows for faster and easier object identification when compared to their high-altitude counterparts. Furthermore, detecting an object from an sUAS can now be done with little to no human involvement, thanks to advances in Artificial Intelligence for object classification. From a military fielding perspective, sUAS can be commanded at lower echelon levels and integrated into smaller units (platoons and companies), whereas larger UAVs demand more resources and serve as assets to higher levels of command (battalion or regiment).

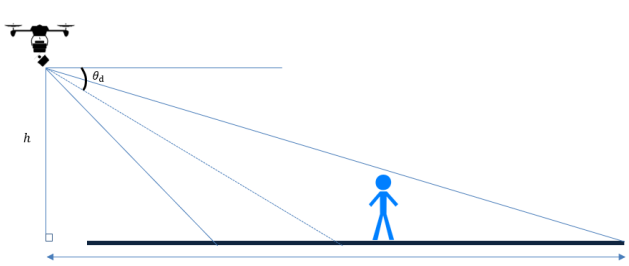
1.1 Background

The primary objective of the study is to test different camera look-angles for object detection. This experiment also begins to explore the limits of neural network data sets, as well as the detection algorithm and software itself. The camera angle is measured with respect to the horizontal axis of the sUAS, and attempts are made to experimentally find the camera angle that provides the most instances of object detection from a slow-moving sUAS with the intent of detecting small, stationary objects on the ground.

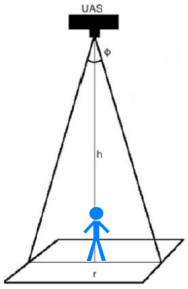
The quadcopter experiences design trade-offs at different camera angles. From an angled field of view (FOV), as shown in Fig. 1a, the users have the advantage of detecting an object at a greater horizontal range or distance from the vehicle. This distance provides a tactical advantage to the operator of the sUAS, but the distance decreases the number of pixels covering the object, which lowers the image quality of the object being detected. If the neural network is not provided enough object pixels, it will fail to detect it. In Fig. 1b, with a 90-degree top-down view, the neural network has sufficient pixels (where the object is in view) available to detect the object, but has a minimal horizontal range for detection; i.e., the object must be right underneath the vehicle to be in perspective. The orientation of the object also plays a role in detection. If the sUAS is looking straight down on the object, the neural network must be trained to recognize the top-down orientation of the object, not just its profile view.

Send correspondence to pratheek.manjunath@westpoint.edu

Thirteenth International Conference on Machine Vision, edited by Wolfgang Osten, Dmitry Nikolaev, Jianhong Zhou
Proc. of SPIE Vol. 11605, 1160528 · © 2021 SPIE · CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2586989



(a) Detection of Object at an Angle, θ_d



(b) Detection of Object at $\theta_d = 90°$

Figure 1. Pixel separation on the ground

1.2 Literature Survey

There have been various studies conducted on camera angles and the placement of imaging devices centered around detecting humans. Some of the techniques and algorithms that appear widely used today for facial recognition or for detecting humans assume the perspective is that of surveillance cameras on the ground [3]. With the entry of Artificial Neural Networks into Image Processing, medical imaging saw the entry of advanced techniques such as edge detection [4]. Although work has been done on the detection from different views and using multi-modal sensors [5], it was rarely tested on a physical sUAS. There remains much research to be done between the capture of imagery and the algorithms used in image processing. This research addresses this gap in research by examining camera angles for remote sensing using unmanned aerial vehicles.

2. RELATED WORK

Some of the current studies involve the use of an articulated gimbal to mount the sensors or cameras. This allows the operator to control the orientation of the payload on-the-fly. However, gimbals increase aircraft weight, system complexity and cost. It also reduces flight endurance due to the added power consumption [6]. In the research presented in this paper, the camera angle is fixed, eliminating the need for a gimbal. This minimizes system complexity, cost and energy draw. An additional benefit of using a fixed-angle mount is that it makes the results of our study independent of a gimbal's kinematic design and dynamic model. Inconsistencies arising from actuator limits and vibrations at singularity configurations, which could be introduced without the operator's knowledge, are eliminated while recording sensor data.

Researchers have studied the accuracy of object localization for ground robot systems. Hsu mounted a charge-couple device (CCD) camera on a ground robot and placed objects within the robot's field of view. Using the CCD camera, they approximated the distance to the target and performed an error analysis between the actual and perceived distance between the object and the robot. They found that a larger camera incline angle has a more significant measurement error [7]. While this experiment was performed on a ground robot system, the measurement accuracy could be used to indicate how well an sUAS can accurately detect objects.

In 2018, Du et al. compared the object detection performance of sUAS from different viewpoints. Du et al. used an sUAS to detect cars using a neural network, and they monitored detection from three perspectives - front, side, and top-down (nadir). The research found that having a front or side view of the car provided better detection rates and increased confidence than a nadir look-angle. This is to be expected since most training data sets are dominated by the front (profile) and side views of an object, thereby providing the classifier with greater detail to process these views of the car [8]. While Du's research depends on the quality of the neural network, it provides insight into an angled view allowing better detection as opposed to a top-down view.

Petrides et al.'s research investigated the bird's eye view approach further by providing equations for the effective height of detection from a right angle. Fig. 2 shows the trigonometric breakdown for such a top-down perspective.

From Fig. 1b, Petrides et al. took the following trigonometric expression

$$h = \frac{r}{2tan(\frac{\varphi}{2})} \tag{1}$$

and rearranged it to relate to pixel density and object size [9],

$$h_d \leq \frac{\hat{p}}{2tan(\frac{\hat{\varphi}}{2})} \times \sqrt{\frac{obj}{rec}} \tag{2}$$

where h_d is the effective altitude range for detection, $\hat{\varphi}$ is the camera field of view angle according to the select number of pixels, $\hat{p}$ is the camera resolution, horizontally or vertically, in pixels, obj is the top-down object size in square meters, and rec is the number of pixels trained on the object of interest in square pixels.

This research uses Petrides's work to develop trigonometric equations for the effective altitude range for detection. An equation for the angled camera view was developed by implementing a directional vector to add the angle to the equation. This directional vector is multiplied by the original height of detection and taking the magnitude results in the optimal detection distance for an angled camera. These experiments test this relationship.

Fellow researchers at the Army Research Laboratory (ARL) arrived at the following equation relating angled view with the height of detection:

$$h = \frac{GR_{swath}}{[cot(\theta_d - \frac{\theta_{FOV}}{2}) - cot(\theta_d + \frac{\theta_{FOV}}{2})]} \tag{3}$$

Where GR_{swath} is the width of the ground covered from by the camera, θ_d is the angle the camera is tilted, and θ_{FOV} is the angle of the camera's field of view. Using the Army Research Lab's equation for an angled view, we calculated the required height of detection for each camera angle tested.

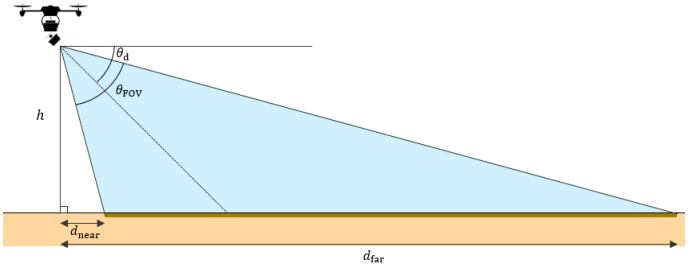


Figure 2. sUAS Angled View

3. THEORY

This research focuses on identifying an appropriate camera angle which yields the optimal pixel density required by a neural network to detect the object of interest located on the ground, with a degree of high confidence. The following are definitions or terminology used in this paper:

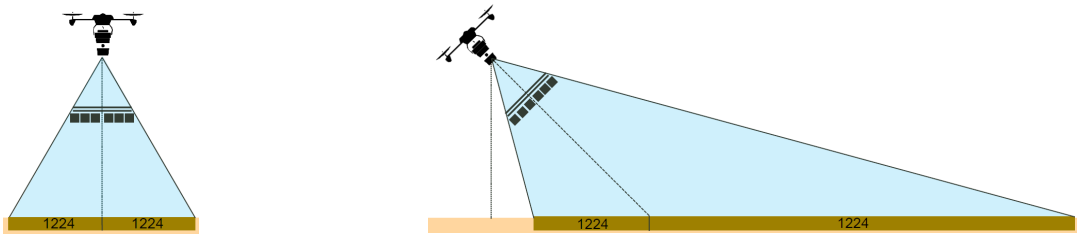
1. Distance to target: the horizontal separation between UAS and object of interest on the ground.
2. Camera angle: Angle between the image sensor's center and the horizontal plane, parallel to the ground.
3. Ground Swath: area on the ground, which lies within the camera's horizontal field of view. This can also be modeled as a linear track on the ground upon which the camera's pixels are projected. By rearranging Eqn. 3, one may arrive at the ground swath distance.
4. Field Of View: The camera's horizontal field of view (HFOV) is contingent on the lens configuration. It is always perpendicular to the direction of flight and is a crucial constraint. The flight path must ensure the object falls within the HFOV. The vertical field of view is of less concern as it is parallel to the direction of flight, is usually less than HFOV and only limited by the distance flown.
5. Pixel Separation: The distance between adjacent pixels incident along the ground swath. In Eqn. 4 H_{FOV} is the number of horizontal pixels for the camera, and θ_p is the angle per pixel.

$$\theta_p = \frac{\theta_{FOV}}{H_{FOV}} \tag{4}$$

6. Pixel Density: Number of pixels present in a unit measure of the ground swath. Although pixel density is conventionally represented as pixels in a unit area, we are linearizing this for analytical simplicity. The units of measure are $pixels/mm$.

3.1 Pixel Separation Analysis

From a nadir (top-down or 90°) camera view, a uniform pixel distribution can be expected. As depicted in Fig. 3a, the pixels equally illuminate the ground swath.



(a) Pixel Separation in Nadir View (b) Pixel Separation in Slanted View

Figure 3. Pixel separation on the ground

Eqn. 4 lists the range of angles at which each pixel, within the camera lens's field of view, will be projected.

$$angles = [-\frac{\theta_{FOV}}{2}, -\frac{\theta_{FOV}}{2} + \theta_p, \ \dots, 0, \ \dots, \frac{\theta_{FOV}}{2} - \theta_p, \frac{\theta_{FOV}}{2}] \tag{5}$$

Expanding this relationship to pixel separation from a non-nadir view to a slanted view, we can express pixel separation on the ground as:

$$GR_{pixels} = \frac{h}{\tan(\theta_d - angles)} \tag{6}$$

In Fig. 3b, one can notice that pixels scatter more to the far side of the UAS than those nearer to the UAS.

3.2 Pixel Density Analysis

If we analyze pixel density as the number of pixels covering an object, we can determine this density requirement based on a neural network's ability to detect that object. As pixel scatter increases, the number of pixels incident on an object decreases. Hence the resolution and definition needed for a classifier to confidently recognize that object is also reduced. [10] For example, consider the below specifications:

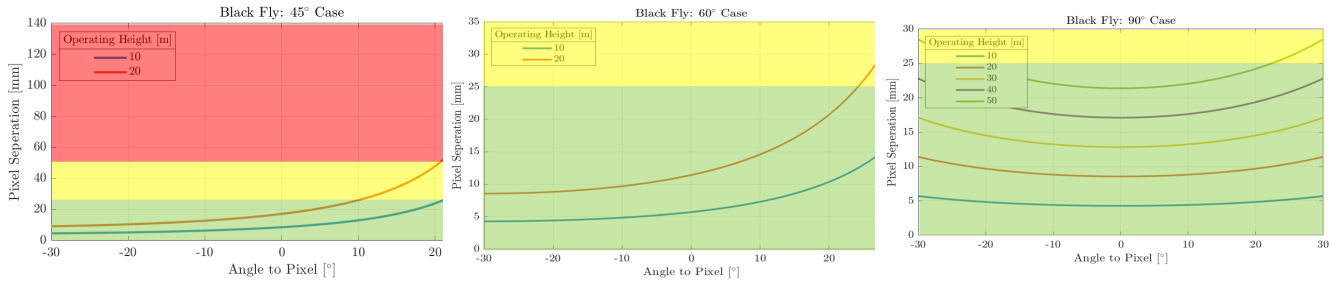
(i) Camera resolution = 2448 x 2048
(ii) HFOV = 60°
(iii) Object dimension = 6 inches
(iv) Minimum number of pixels needed for object detection = 6

Based on these requirements, we plot the pixel separation distance versus the scatter angle. If the pixel separation is less than 1 pixel per inch or 1 pixel per 25.4mm, it meets the minimum threshold for detection and is shaded green. As the scatter angle increases, the pixels are separated by distances greater than 25mm and are shaded red. These angles may not provide the resolution needed for the neural network to accurately and confidently classify the object.

4. SYSTEM SETUP

4.1 Simulation

Preliminary tests were conducted using a simulated 3DR Iris quadcopter in a software-in-the-loop (SITL) framework, comprising of Gazebo, MAVROS, and QGroundControl. *Gazebo* allowed the user to control and access the sUAS model through a well-developed architecture of Robot Operating System (ROS) nodes. *MAVROS* is a ROS node that translates ROS messages to MAVLink messages, thereby bridging the divide between high-level application languages used within the ROS ecosystem and low-level MAVLink protocol employed by the flight controller. The simulated 3DR Iris SDF model is controlled by a PX4 Flight Management Unit (FMU) and is equipped with a PX4Flow camera. The PX4Flow camera's technical specifications were modified to match those of the actual color camera used on the physical quadcopter. *QGroundControl* is a popular Ground Control Station (GCS) software used for configuring FMUs and planning missions.



(a) 45 Degrees (b) 60 Degrees (c) 90 Degrees

Figure 4. Pixel density requirement for detection at different angles

The detection algorithm used in simulation is YOLOv3 and DarkNet. To avoid getting a false detect, i.e. detecting other objects instead of the one emplaced in the image frame, YOLOv3 was limited to only register detects on the desired object. The system measures the number of detects on the object per flyover, and the flyovers are iterated over for the same flight pattern until repeated results are received.

For this experiment, a fire hydrant is chosen as the object to detect. A fire hydrant is small and asymmetric and meets the detection goals of current U.S. Army research in this field. It is also within Gazebo's object database as well as within the YOLOv3 predefined set of objects.

4.2 Hardware Infrastructure

A quadcopter was constructed using the Tarot 650 Carbon Fiber frame to serve as the physical testbed for hardware validation. This sUAS was custom-built by the Robotics Research Center and included a Pixhawk 2.1 flight controller, Hex Here+ RTK GNSS receiver, FLIR Blackfly-S USB3 5.0 MP camera, and Intel NUC.

4.3 Software Infrastructure

The Intel NUC companion computer on the UAS ran Ubuntu 18.04 as the operating system with ROS Melodic. The flight control unit uses Ardupilot v4.1.0 firmware, and the camera is controlled using the FLIR Spinnaker SDK. The code repository is located at https://github.com/cmanore25/suas_px4sim2.

5. TESTING AND METHODOLOGY

This paper focuses on the outcomes of simulation testing. Implementation on physical hardware was delayed due to COVID-19 and is still underway. The flexibility of the Gazebo simulation allowed tests to be ran remotely.

5.1 Initial Testing

The general procedure adopted for testing in simulation is outlined in Fig. 5

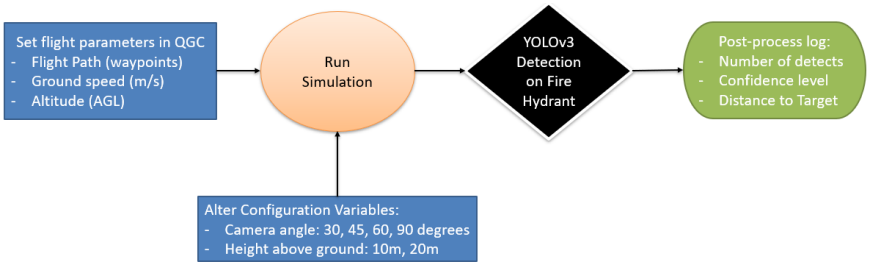


Figure 5. Flowchart of Testing Methodology

A mission plan was created consisting of a rectangular corridor with dimensions 100m x 20m, a constant ground speed of 1 *m/s*, and the test altitude. The 1 *m/s* was chosen after studying the forward pitch of sUAS at various speeds. The slow speed helps to minimize the impact of sUAS pitch on camera tilt.

The configuration variables include the camera look angle (measured by the depression from the quadcopter's horizontal) and the quadcopter's altitude above ground. The angles tested are 30, 45, 60, and 90 degrees at heights of 10 and 20 meters above the ground.

After running the Gazebo simulation, the image stream from the camera is fed through YOLOv3's detection algorithm. The highest resolution setting of *608 x 608* was selected as this model has the most accuracy. [11].

The post-process log for each iteration is analyzed. Of interest to us are:

(i) Number of detects - indicates that a known object has been identified in the image frame

(ii) Number of positive detects - frames in which the fire hydrant was detected when present in the camera's field of view. The number of detects and frames with object in FOV formed the ratio as shown in Eq. 7.

(iii) Confidence level - a number expressed as a percentage. The confidence level measures how confidently YOLO can recognize the fire hydrant.

(iv) Distance to target - measured as the horizontal displacement between sUAS and location of fire hydrant. Distance to target is used to measure the relationship between confidence levels and the distance at which the confidence levels are recorded.

$$\%\ of\ detections = \frac{\#\ of\ detects}{total\ frames\ with\ object} \tag{7}$$

5.2 Improved Testing

After conducting the first round of tests and flyovers, we noticed that the initial test procedure outline above had a flaw: it was heavily dependent on the neural network's ability to detect objects only when presented with a straight-down view. Since most object classifiers are predominantly fed with the profile views of objects, the slant or side views led to no detection. YOLOv3 could not detect the fire hydrant from 30° because the neural network was not trained to detect a fire hydrant from looking at the bottom or side of the object. This led to the 90° view performing the best, which is not very intuitive and also not the research objective. To fix this issue, the dependency on the neural network's lack of perspective had to be eliminated. Tilting the object to match the 3DR Iris's camera angle made this possible. The improved method allows the camera to see the fire hydrant's silhouette from all angles, making pixel density the primary determinant of object detection within the simulation.

6. RESULTS

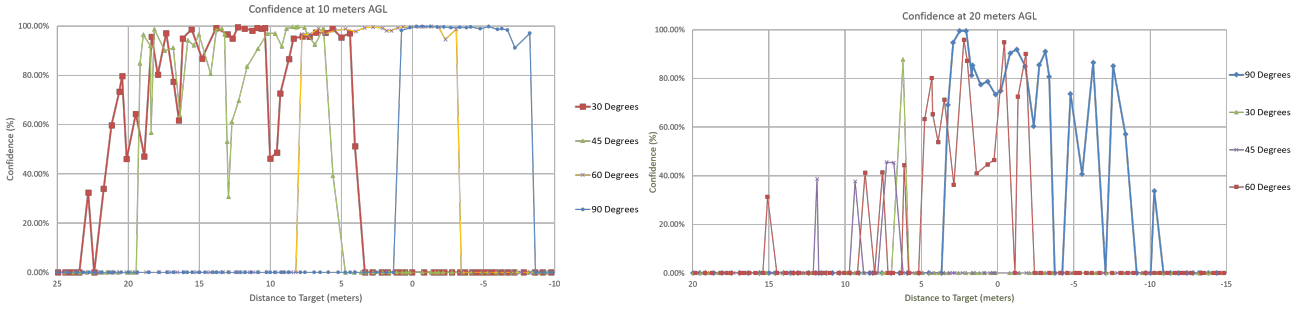
The analysis of the post-process log of each simulation iteration are discussed here.

Table 1. Detection Percentage

Camera Angle	***10 meters AGL***	***20 meters AGL***
30 Degrees	49.4 %	1.08 %
45 Degrees	100 %	5.97 %
60 Degrees	100 %	39.1 %
90 Degrees	100 %	65.7 %

When looking at the detection percentage from Table 1, 30° had the worst performance, detecting only 49.4% of the time from 10 meters and 1.08% from 20 meters. 45, 60, and 90° camera angles detected the fire hydrant in all frames where the object was visible when flying at 10 meters Above Ground Level (AGL). At 20 meters AGL, 45° did not have many detects, and even 90° struggled with 65.7% of detects. Overall, the 90° camera angle still had the highest performance at these heights.

The graphs in Figures 6a and 6b show how confident YOLOv3 was in determining that the object was a fire hydrant, given by a percentage with respect to the horizontal distance to the object. The data points at 0% confidence indicate no detect for that horizontal distance. At 10 meters AGL, 30° struggled to have high confidence levels; however, it was able to first detect the fire hydrant over 20 meters away from it at low confidence. 45° displayed higher confidence at a distance,



(a) 10 meters AGL

(b) 20 meters AGL

Figure 6. Confidence vs Distance-to-target

reaching 90% confidence or better at 15 meters away. 60° consistently held high confidence, but had to be closer to the object – starting to detect at 8 meters away. 90° had high confidence, but the fire hydrant needed to be right underneath the quadcopter or a little past it.

At 20 meters AGL, confidence levels were much lower for all camera angles. 30° only had 1 detect at 7 meters, and 45° also struggled to produce high confidence in its detects. The detection for the 60° camera angle was better, but again had to be fairly close to the object. 90° was the only camera angle that produced almost 100% confidence at some points, but had to be underneath or past the target.

These findings were consolidated into a *'Look-up table'* (Table 2). This makes it more intuitive to determine the appropriate camera angle to use for different applications. *Range* is the horizontal distance away from the object where the user would like to detect objects, and *Confidence* is the typical confidence level possible for each detection. Range and Confidence were separated into three different categories: low, medium, and high.

Table 2. Summary of Camera Angle Selection

	Confidence (%)		
Range (m)[a]	***Low (< 40)***	***Medium (40 - 80)***	***High (> 80)***
Low (< 10)	–	60°, **90°**	60°, **90°**
Medium (10 - 20)	30°, **45°**, 60°	30°, **45°**, 60°	**30°**
High (> 20)	**60°**	–	–

[a]Stand-off Distance from Object (in meters).

7. CONCLUSIONS

Unmanned systems provide excellent aerial surveillance and detection capabilities, and have uses in many multi-domain environments. The camera angle that maximizes the detection on an object depends on what the user needs based on the application. Some of the areas for future work discussed below will help ascertain the appropriate camera look-angle based on the scenario where the UAS is being fielded. This research will continue to scale in complexity and realism. Below are some of the areas in which there is interest to evolve this project in the near future:

1. Hardware Implementation - Test and validate results using a physical quadcopter. Hardware implementation was initially planned within the scope of the project, but complications due to COVID-19 limited the feasibility.
2. Position of object in image frame - Study the relevance of position of an object within an image frame. Overfly the target object such that the course over ground is offset to one side, so as to obtain a slant and off-center view. Tests can be run wherein the fire hydrant is positioned to the left or right side of the quadcopter, instead of directly beneath it, to see how the relative position of the object within the image frame affects the neural network's detection.
3. UAS Attitude - Study the effect of UAS attitude on object perspective - while in flight, how significant is the effect of roll or pitch on the camera's look-angle? Testing on a physical system would allow the researchers to consider real-world factors such as wind and airspeed at which flight attitude begins to have an impact on the look-angle.

4. Sensor Degradation - Classifier's immunity to sensor degradation and object occlusion - what pixel density is required to overcome the poor visibility due to weather conditions or presence of obstacles? Since the pixel density directly relates to pixel scattering and thereby to camera angle, decoding this relationship would lead to a better understanding of the object classifiers limitations. This will help in the development of training data sets where pixel separation can be accounted for.
5. Detection Range - The authors would like to explore the relationship between pixel density and range - the horizontal displacement between UAS and object on the ground - in greater detail. Based on our preliminary findings presented above, we know that the confidence level of detection must be sacrificed to gain greater range. For certain applications, stand-off distance from a target maybe more desirable than mere pixel density [12]. What camera angles would be best suited for such purposes? Along the same line of investigation, the relationship between the slant range - the hypotenuse distance between camera and object on the ground - and pixel density can also be established.

ACKNOWLEDGMENTS

This work is supported by the Defense Threat Reduction Agency. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Department of Defense or the U.S. Government.

The authors would like to thank Christopher Korpela and the Robotics Research Center at West Point for providing resources and equipment, Brian Phelan from Army Research Laboratory's Sensors and Electron Devices Directorate for providing guidance on the relationship between camera angle and projected pixel density, and Erik Kiser from the Department of Mathematical Sciences at the United States Military Academy for assistance in analyzing camera geometries.

REFERENCES

[1] Levin, E., Zarnowski, A., Mccarty, J. L., Bialas, J., and Banaszek, A., "Feasibility study of inexpensive thermal sensors and small uas deployment for living human detection in rescue missions,"
[2] Jalled, F., "Object Detection Using Image Processing," tech. rep.
[3] Yohannan, L., Baby, M., Monson, M., Yadhukrishnan, P. D., and Raju, C. K., "Optimal camera positions for human identification," in [*2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*], 1218–1221 (2019).
[4] Ozturk, O. and Kutucu, H., "Detection of bone fractures using image processing techniques and artificial neural networks," in [*2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*], 1–5 (2017).
[5] Patterson, M. C. L. and Brescia, A., "Integrated sensor system for uas," *Unmanned Air Vehicle Systems: 23rd International Conference, 7-9 April 2008, Bristol, United Kingdom* **1**, 476–489 (Apr 2008).
[6] Egnal, G., "Wide area persistent surveillance with no gimbal," in [*Defense Transformation and Net-Centric Systems 2012*], Suresh, R., ed., **8405**, 84050J, SPIE (may 2012).
[7] Hsu, C.-C. J., Lu, M.-C., and Lu, Y.-Y., "Distance and Angle Measurement of Objects on an Oblique Plane Based on Pixel Number Variation of CCD Images," *IEEE Transactions on Instrumentation and Measurement* **60**, 1779–1794 (May 2011).
[8] Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., Zhang, W., Huang, Q., and Tian, Q., "The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking," in [*Computer Vision – ECCV 2018*], Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, Y., eds., **11214**, 375–391, Springer International Publishing, Cham (2018).
[9] Petrides, P., Kyrkou, C., Kolios, P., Theocharides, T., and Panayiotou, C., "Towards a holistic performance evaluation framework for drone-based object detection," in [*2017 International Conference on Unmanned Aircraft Systems (ICUAS)*], 1785–1793 (June 2017). ISSN: null.
[10] Choi, D. Y., Choi, J. H., Choi, J. W., and Song, B. C., "Cnn-based pre-processing and multi-frame-based view transformation for fisheye camera-based avm system," in [*2017 IEEE International Conference on Image Processing (ICIP)*], 4073–4077 (2017).
[11] Redmon, J. and Farhadi, A., "Yolov3: An incremental improvement," *arXiv* (2018).
[12] Colorado, J., Devia, C., Perez, M., Mondragon, I., Mendez, D., and Parra, C., "Low-altitude autonomous drone navigation for landmine detection purposes," in [*2017 International Conference on Unmanned Aircraft Systems (ICUAS)*], 540–546 (June 2017).