IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 05-10

UAV Drone for Object Detection and Identification with Flight Stability

Shashank Singh[*1], Renua Tekam[2], Vikramaditya Thakur[3], Prof. Vikrant Chole[4]
[1,2,3,4]*Department of Computer Science and Engineering, G.H. Raisoni Academy of Engineering and Technology, Nagpur, Maharashtra, India*
Corresponding Author: Prof. Vikrant Chole

Abstract: *This paper describes the improvements in surveillance, using an Unmanned Aerial Vehicle (UAV), when integrated with the "capabilities" of Artificial Intelligence (AI). The system relies on Convolution Neural Network based deep learning model to detect and identify entities such as humans, animals, and vehicles in a noisy environment. The system also uses multiple instances of a closed loop feedback algorithm called the Proportional Integral and Derivative (PID) to achieve a high degree of flight stability and thus capable of tracking objects, up to certain degrees of automaticity and to balance the flight dynamics of the system*
Keywords: *UAV, Autonomous, Surveillance, Machine Learning, Artificial Neural Network*.

I. Introduction

Drones have become increasingly popular lately and have found applications in a wide array of disciplines. Despite their accessibility and availability to public and organizations, they still require the operator to provide manual operation codes to control and interpret the actions of the drone. In the field of surveillance particularly, though it is very feasible to survey an area with the help of a drone, being able to interpret its output data manually in real time and that too quickly is a difficult task. The task becomes even more difficult when the environment has unpredictable and strong noise such as wind, visual impairment, etc.

Our aim was to develop an UAV with the capability of providing real time, highly accurate surveillance reports independent of manual interpretation. We have used machine learning model to detect and identify objects in real time which gives us advantage in noisy environments such as during a Tsunami, avalanche or in a forest fire. Our UAV provides a video stream of the area it is currently surveying to the ML model frame by frame, which then reads every pixel of the stream to determine the objects present in the frame that it is analyzing.

Upon detection, it tries to identify the object, based on the weights and biases that it has adjusted during the training phase. Meanwhile, the PID algorithm maintains the Yaw, Pitch and Roll of the UAV, generating values to counter the presence of the external noises such as wind, thermal current, rain, etc. The PID algorithm reads the value of different sensors such as Gyroscope, and Voltmeter to determine the voltage level of the UAV battery and tries to match it to the desired values. Upon encountering variations, the algorithm adjusts motors until the current values match the desired values. We have applied separate PID loops for each Yaw, Pitch and Roll to achieve maximum stability.

The UAV also provides a limited autonomous tracking facility where it uses the object detection capability of the ML model to decide current position of the object of interest. Depending on the object's location, the UAV is able to change its position to keep that object of interest in the center of the view.

All these facilities bundled together in our system makes it highly efficient to be used by novice pilots in an unfamiliar or unstable environment thus making it a viable tool to be used in search and rescue operations, military operations, surveillance, traffic analysis, and much more.

II. Methodology

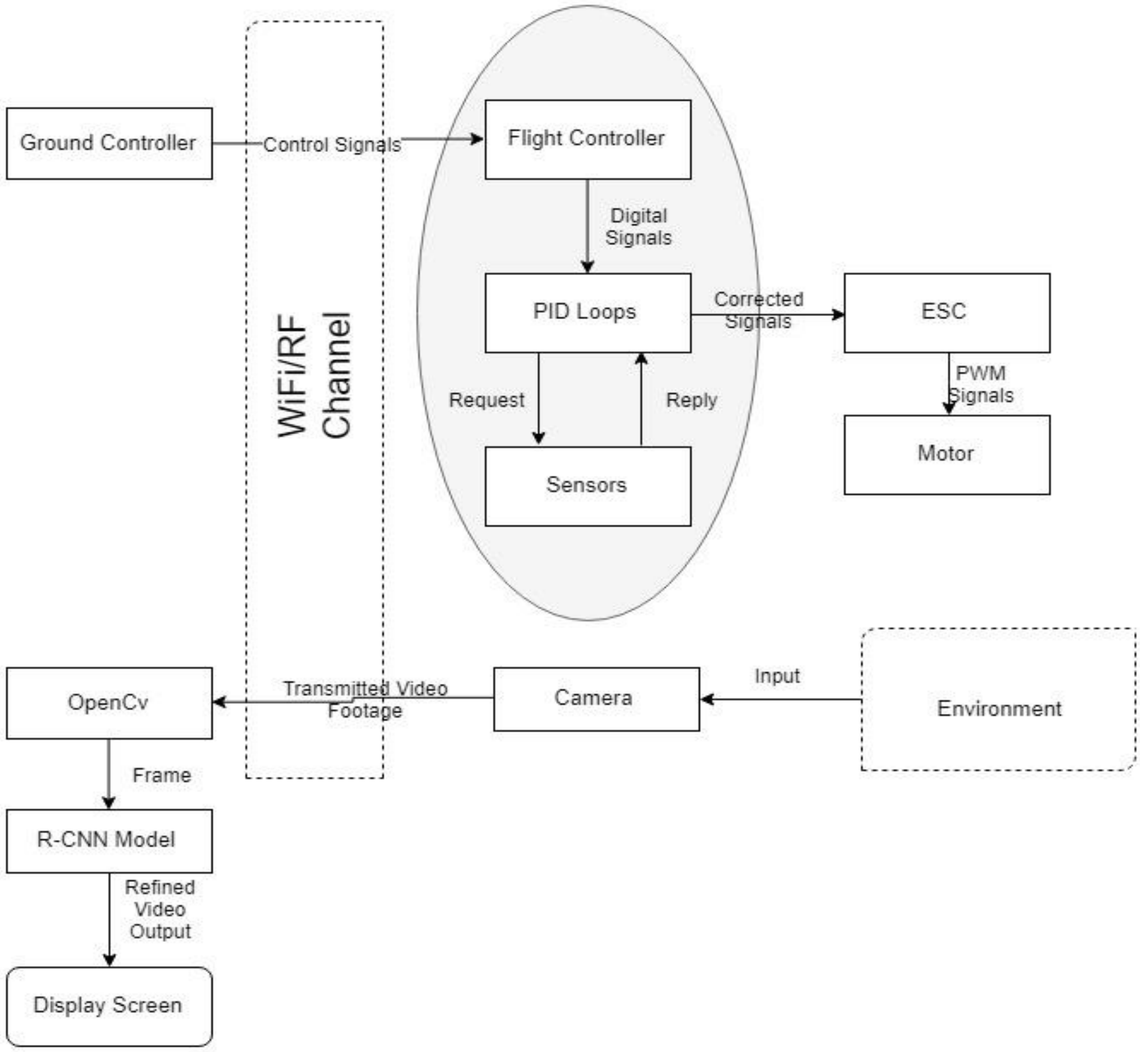

Figure 1: *Block* diagram *showing the complete operational environment of this project*

A. Image Processing

The expectation of the grouping procedure is to classify all pixels in a computerized picture into one of a few "land spread classes or subjects". This classified information is utilized to deliver topical maps of the land spread present in a picture. Regularly, multispectral information is utilized to play out the grouping and, in fact, the phantom example present inside the information for every pixel is utilized as the numerical reason for order (Lillesand and Kiefer, 1994). The target of picture characterization is to recognize and depict, as a remarkable dim dimension (or shading), the highlights happening in a picture as far as the item or sort of land spread these highlights really speak to on the ground.

B. CNN

A classical CNN can just disclose to you the class of the items, not where they are found. It is really conceivable to relapse jumping boxes straightforwardly from a CNN yet that can occur for one item at any given moment. In the event that various items are in the visual field, at that point the CNN jumping box relapse can't function admirably because of impedance. In R-CNN the CNN is compelled to concentrate on a solitary district at any given moment since that way impedance is limited since it is normal that just a solitary object of intrigue will rule in a given locale. The districts in the R-CNN are recognized by specific inquiry calculation pursued by resizing with the goal that the locales are of equivalent size before they are bolstered to a CNN for order and jumping box relapse. Quicker R-CNN is the quicker and more improved usage of R-CNN that we have utilized.
The main advantage in using Faster R-CNN is that it replaces the slow selective search algorithm with a fast neural net and introduces the Region Proposal Network.

Working of Region Proposal Network (RPN):

- At last layer of an initial CNN, a 3x3 sliding window moves over the component guide and maps it to a lower measurement.
- It generates multiple possible regions for each sliding-window location based on *k* fixed-ratio Anchor Boxes.
- Each region proposal consists of an Objectness score for that region and a tuple of four coordinates representing the bounding box of the region.

In other words, we look at each location in our last feature map then consider *k* different boxes centered on it. We output the coordinates for a box if it contains any object in it.

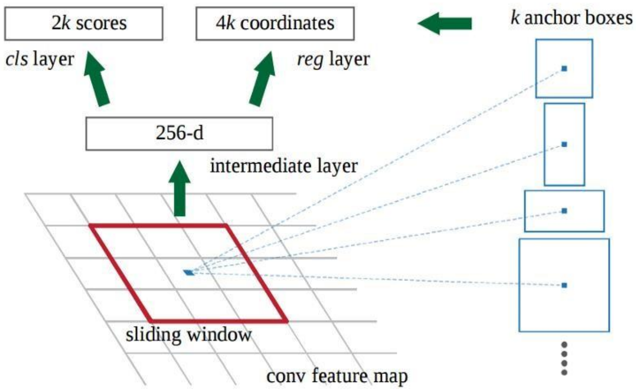

Figure 2.*Image representing scan of pixels in the image using a sliding window*[8]

The N bounding boxes' Softmax Probability is represented by the 2k scores. Although the RPN produces (output, generates) bounding box coordinates, it does not try to classify any potential objects; its sole job is still proposing an object's region. If an anchor box has an objectness score above a certain threshold, the box's coordinates are then passed as a region proposal.

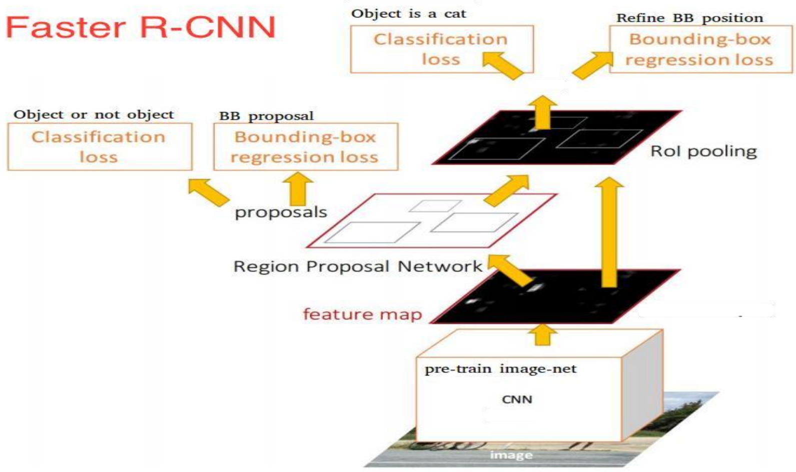

Figure 3.*Faster R-CNN working model*[8]

Altogether, Faster R-CNN achieves a much better speed with state-of-the-art accuracy. It is worth noting that although future models did a lot to increase detection speed, only few models could manage to outperform Faster R-CNN by a significant margin. In other words, Faster R-CNN may not be the simplest or fastest method to detect objects, but it has one of the best performances in CNN.

C. Object Tracking/ Path Planning:

To truly track the individual in space, we required a profundity expectation of the environment which required various other sensors thus reducing the payload capacity of the system, since that was difficult to overcome, we attempted to utilize the features of our OpenCV highlighters. When the system has identified an object of interest, it fundamentally generates 4 numbers portraying a container to highlight the object. These numbers are: its upper left corner (x1, y1) and its base right corner (x2, y2).

Given those, we were able to figure out the focal point of the rectangle highlighter and furthermore its region. To register the territory, we just needed to figure out the width as (x2-x1) and height as (y2-y1) and

duplicate them. With respect to the inside, the center is figured essentially as z1 = (x2+x1)/2 and z2 = (y2+y1)/2 i.e., (z1, z2) denotes the center of the rectangle. The focal point of the square shape lets us know whether the individual is focused in the image or on the right or left side. In other words, we tried to match the center of the highlighting rectangle to the center of the camera view or screen. With this data, we were able to send control signals to our system to decide whether to move to one side or to the other. Concerning its vertical hub, so as to get the individual the focal point of the shot. Similarly, we can order the system to go up or down if the individual is identified to be in the upper piece of the picture.

The area of the square shape can give us a rough data of how close the individual is. A greater area of the shape implies that the object of interest is close, while a smaller area demonstrates that the object of interest is far away. In view of this data, we can propel the system forward or backward so as to maintain the ideal area of the highlighter square thus maintaining a constant and steady distance from the object.

D. PID:

PID controllers are found in a wide scope of utilizations for mechanical procedure control. PID or Proportional Integral Derivative is a set of three controllers are consolidated so that it delivers a control flag. As an input controller, it conveys the control yield at wanted dimensions. PID controller keeps up the yield with the end goal that there is zero blunder between procedure variable and set point/wanted yield by shut circle activities. The typical control function of PID is:

$$u(t) = k_p e(t) + k_i \int_0^t e(t')dt' + k_d \frac{de(t)}{dt}$$

Here k_p,k_i and k_d are the coefficients of the proportional, integral and derivative parameters, respectively. Calculating these arbitrary contents required us to use the Ziegler–Nichols tuning method.

The k_p attribute provides the power required to move from the current state to the desired state. The k_i attribute helps in keeping the system in the desired position once it has reached there. And the k_d value dampens or smooths out the curve of force as soon as the system reaches closer to its desired position.

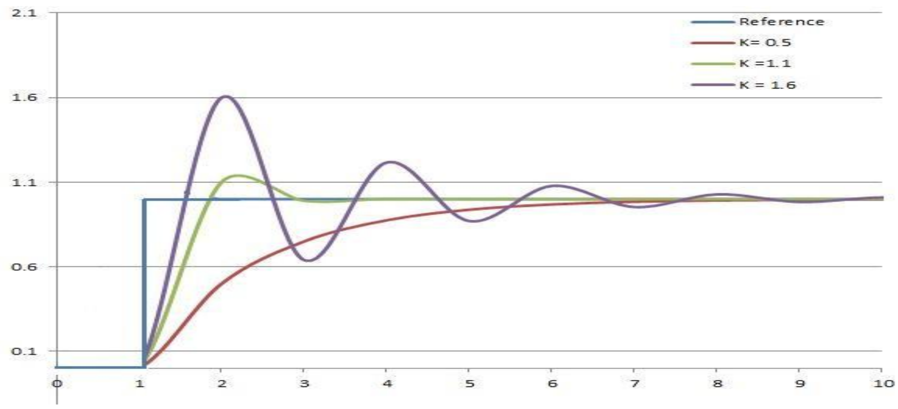

Figure 4. Graph showing the effect of the kp, ki, and kd value on the stability of the system. Blue line indicates overshoot, red indicates undershoot while the green is the ideal case. [7]

In Ziegler–Nichols tuning method, initially all the constants are set to zero. The proportional constant is increased until the system reaches a state of steady oscillation. Once that value is achieved, which is called the ultimate gain, the other constants are calculated using the table below.

Control Type	k_p	T_i	T_d	k_i	k_d
PID	$0.6k_u$	$\frac{T_u}{2}$	$\frac{T\_u}{8}$	$\frac{2k_u}{T_u}$ 1.	$\frac{3k_uT_u}{40}$

Table 1. Table indicating the formulas to calculate k_p, k_i, and k_d value based on Ziegler–Nichols tuning method.

III. Results And Discussion

This empirical study tries to prove the importance of Artificial Intelligence in human vision assistance. It does not act as a substitute for human interaction for monitoring operations, but rather as an aid to the same thing. One of the main limitations of human eye is that it can only focus on what is directly in front of it. The peripheral vision, though fully able to view the environment, does so very poorly by neglecting most of the details. This proves to be a major problem during an event of disaster such as flood. Being able to detect people in a flood like situation while ignoring the thrash that's flowing with it, in time proves to be the deciding factor in saving someone's life.

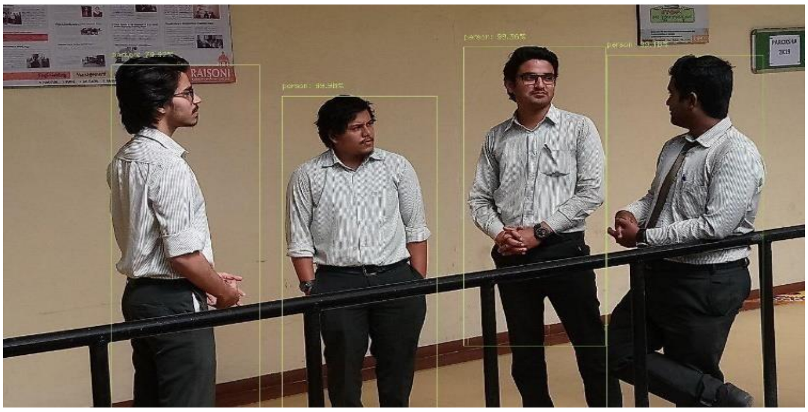

Figure 5. *Detection and recognition of objects (Human) in a single video frame captured by drone camera. Each detected object is given a confidence value indicating how confident the system is to identify it.*

Based on the patterns that it finds in each frame, it tries to decide the kind of object it is looking at. The system gives a certain confidence value to each kind of object that the given detected object may look like and selects the one with the highest confidence.

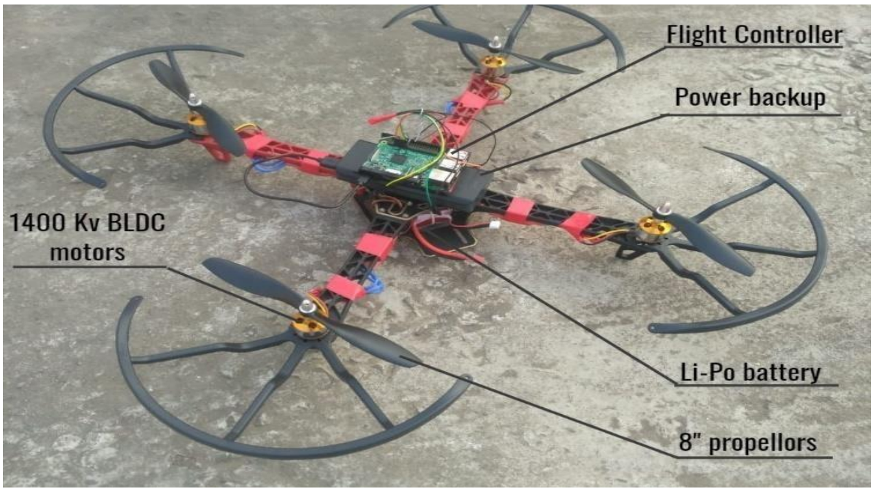

Figure 6. *Raspberri Pi 3B based quadcopter model which runs the PID loops and the control signal in the RPI flight controller.*

Our UAV is powered by a 11.1V LiPo Battery consisting of four 1400Kv BLDC motors with an 8" prop on each motor. The Flight Controller is designed using a Raspberry Pi 3B which handles WiFi communication and PID loops in it. The UAV has a thrust to weight ration of around 1:4 which makes it capable of carrying large payloads if needed. Our empirical study can be looked as an effort to develop a system that can be an improvement on current surveillance methods by providing accurate and real time assistance to the user for the during critical events and thus, maximizing the productivity and saving time.

IV. Conclusion

UAV Drone with object recognition is applicable in variety of situations ranging from civilian rescue to traffic analysis. Due to the constant requirement of control input by an operator, real time analysis of data becomes even more difficult with increasing environmental noises. Integrating the UAV with the characteristics of Artificial intelligence helps to overcome these problems. The drone uses convolution neural network for image recognition, and Proportional – Integral – Derivative to balance and counter the drone against environmental noises during flight. Blending the capabilities of AI and UAV helps faster processing as majority of calculations are done by the AI thus, reducing the efforts made into data processing and drone control, simultaneously (or synchronously).

References

[1]. LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey (2015), "Deep learning". Nature. 521 (7553): 436–444. Bibcode:2015
[2]. Natur.521..436L.doi:10.1038/nature14539. PMID 26017442.
[3]. JinghuaZhong (Spring 2006"PID Controller Tuning: A Short Tutorial" (PDF). Retrieved 2011-04-04
[4]. Leishman, J.G. (2000). Principles of Helicopter
[5]. Aerodynamics. New York, NY: Cambridge University Press.
[6]. Stafford, Jesse Spring. 2014."How a Quadcopter works | Clay Allen". University of Alaska, Fairbanks. Retrieved 2015-01-20.
[7]. Hubel, DH; Wiesel, TN. 1959. "Receptive fields of single neurons in the cat's striate cortex". J. Physiol. 148 (3): 574–91. doi:10.1113/jphysiol.1959.sp 006308. PMC 1363130. PMID 14403679.-
[8]. LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey 2015. "Deep learning". Nature. 521 (7553): 436–444. Bibcode:2015 Natur.521..436L doi:10.1038/nature14539 PMID 26017442.
[9]. https://en.wikipedia.org/wiki/File:PID_varyingP.jpg, March 8th 2019
[10]. Shoqing Ren, Kaiming He, Ross Girshick, Jian Sun 2016. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks".: Cornel University arXiv:1506.01497.