Cambridge Small (2020), **37**, 1–3

ARTICLE TYPE

Low-cost Computation Method for Monocular Distance Estimation in Unmanned Aerial Vehicles

Matthew Wang*

Duke University, Durham, NC USA
*Corresponding author. Email: matthew.wang187@duke.edu

Abstract

Monocular object distance estimation methods for Unmanned Aerial Vehicles (UAV's) allow UAV's to measure the distance of objects with just the drone camera. Many present monocular distance estimation models use deep learning techniques, which can be computationally expensive. This research uses You Only Look Once (YOLO) and a proposed distance estimation equation to accomplish monocular distance estimation at low cost.

1. Introduction

Unmanned Aerial Vehicles (UAVs) are commonly used as a photogrammatry surveying tools. With cheaper drones becoming more accessible and popular, monocular depth estimation techniques (depth estimation with only an image) are being developed since most cheaper drones do not have range sensors such as LIDAR. Developing techniques that use low computational power should be considered, as many drones are controlled by mobile devices with low computation power. The current most popular methods of monocular distance estimation use deep learning techniques such as Convolutional Neural Networks and Generative Adversarial Networks. However, deep learning methods are computationally expensive and cannot be ran efficiently on smaller devices. Their performance accuracy will also highly vary on the quality of the model's training. A low cost method for monocular depth estimation for real-time UAV footage using trigonometric techniques is proposed.

Given the UAV's altitude, camera angle, camera specification, and image pixel coordinates of the object of interest, we can calculate the distance from the UAV to the object. This requires projecting a coordinate in the image to a coordinate in the real world. The model makes the assumption the ground is flat and objects are touching the flat ground. Our technique uses a geometric abstraction of a cone projecting onto a flat plane, where the tip of our cone is the camera location, the cone base is the image plane, and the flat plane is the real world ground. The given pixel coordinate on the image plane is projected on the flat plane to calculate distance between the camera and object on the ground.

2. Object Distance Estimation

2.1 Input Variables

The model requires information about the drone and the object to calculate distance. It requires the drone's flying altitude, drone camera angle, coordinates of the object's bounding box in the image, and camera specifications such as focal length and resolution. The values can be obtained from sensors on the drones and input into the model.
H : Drone Altitude in Meters
α : Drone camera angle in radians, where the 0 radians is where the camera points horizontally to horizon.
$(x_{min}, y_{min}), (x_{max}, y_{max})$: Object bounding box upper left and bottom right corners
f_x : Camera focal length (in mm)

2.2 Foot Coordinates

The foot coordinate is the center-bottom point of an object's bounding box on the image plane.

It is assumed that the earth is flat, objects of interest are standing on the ground, and bounding boxes drawn around the objects are accurate. The foot coordinates of the objects bounding box (x_f, y_f) is defined as follows:

$$x_f = (x_{min} + x_{max})/2$$

$$y_f = y_{max}$$

2.3 Principle Point

Define the center point of the image be (x_c, y_c). Principle point is the point on the ground plane that is pointed to by the center point of the image plane. That is, the location on the ground that the camera is pointing directly center to.

Given the drone camera's angle α and the altitude H, the spacial distance h between the camera and the principle point is as follows:

$$h = H * sec(\alpha)$$

2.4 Foot Angle

The foot angle ϕ is the angle in radians between the vectors formed by the principle point, drone, and the object's foot point considering the camera's focal length f_x.

$$\phi = \tan^{-1}(\sqrt{(x_f - x_c)^2 + (y_f - y_c)^2}/f_x$$

2.5 Image Plane Angle

The image plane angle θ is the angle describing the object's location on the image plane using polar coordinates with respect to the y-axis. It can be defined as the angle on the image plane between the three points $(x_f, y_f), (x_c, y_c), (x_c, y_{image-height})$.

$$\theta = \tan^{-1}(y_f - y_c)/(x_c - x_f)$$

2.6 Distance Equation Derivation

It is assumed that in the 3D space, the principle point is at $(0,0,0)$, the drone is hovering at point $(0,0,h)$, and the object of interest is at (x', y', z'). The object's coordinates can be described in the polar form, $x' = r\cos\theta$, $y = r\sin\theta$, $z = z$. The distance between the drone and object is

$$d = dist((x', y', z'), (0,0,h)) = \sqrt{x'^2 + y'^2 + (z' - h)^2} = \sqrt{r'^2 + (z' - h)^2}$$

When the camera angle changes (drone looks up and down), the ground plane is tilting along the y-axis in respect the the drone. Therefore, the equation of the ground plane can be expressed as $z' = \alpha' * y' = \alpha' * r\sin\theta$, where $\alpha' = \tan^{-1}\alpha$

We know $\tan(\phi) = r'/(h - z')$. Solving for r' and replacing z', we get

$$r' = h * \tan(\phi)/(1 + \alpha' * \cos\theta * \tan\phi)$$

Finally, distance estimation equation is as follows:

$$d = \sqrt{r'^2 + (z' - h)^2}$$

where

$$r' = h * \tan(\phi)/(1 + \alpha' * \cos\theta * \tan\phi)$$

$$z' = \alpha' * r' \sin\theta$$

$$\alpha' = \tan^{-1}\alpha$$