Object Tracking With Drone

Ying Ying Lai
School of Information Technology
Monash University Malaysia
ylai0018@student.monash.edu

I. Introduction

Recently, object detection and tracking with unmanned aerial vehicle (UAV) has been increasingly popular in the field of computer vision and remote sensing. It is widely used in multitude of real-life applications due to its high practical value. Some of the practical use cases include information gathering for security surveillance purposes and autonomous drone system for search and rescue. In this paper, a real-time UAV-based object detection and tracking system is presented. Several image processing techniques are performed on the image captured from the drone's video stream to uniquely identify a tennis ball. The drone will constantly keep the location of the ball at the middle of its view and maintain its distance within a certain range.

II. System Implementation

An overview of the colour-based object detection and tracking system is illustrated in Figure1.

Once a frame is acquired from the drone's camera, the image will be pre-processed before applying an object detection and tracking algorithm. The location of the tennis ball will be calculated and is used to determine the velocity of the drone. A set of instruction will then be sent to the drone for execution. A detailed explanation for the methodology of each step is presented at the sections below.

A. Image Pre-processing

A computer program is run to obtain a video frame from the drone's camera. Each frame will be transformed and manipulated through a series of mathematical operations before a object detection algorithm can be applied. The image pre-processing operations is implemented using the OpenCV open-source Python Library.

Firstly, the frame is resized to have a dimension of 640×480 pixels. A smaller frame will require lesser time to process, therefore achieving a higher FPS (frame per second) and smoother movement of the drone.

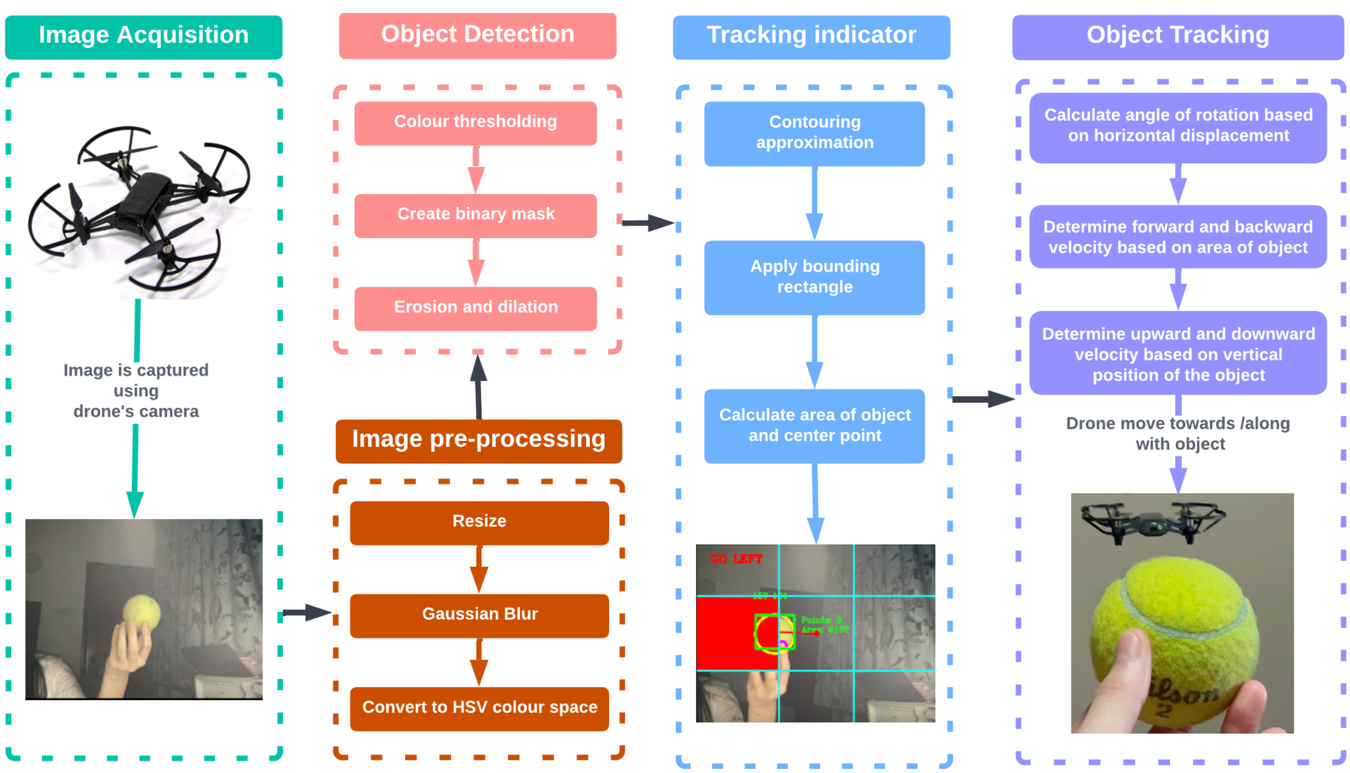


Figure 1. Pipeline of the proposed framework

Then, Gaussian blurring is applied to the frame to remove noise and other high frequency content[1]. Finally, the image is converted to the HSV colour space. Based on a study conducted by Cucchiara et. al [2], "HSV corresponds closely to the human perception of colour" and it has been proven to achieve better detection accuracy. It is preferred to the RGB colour space in localisation and approximating the measurements of moving objects due to its effectiveness in distinguishing and suppressing shadows, which prevent shadows from being be misclassified as part of the object.

B. Object Detection

To localise the object, a possible range of HSV colours that a tennis ball may have is determined before creating a binary mask. The lower and upper boundaries for the colour is selected based on the HSV colour map in Figure 2. Experiment to find out the best colour range for the yellowish-green colour of a tennis ball is carried out in multiple indoor settings with different lighting. The optimal range for each H, S and V component is:

1) Hue: [24, 64]
2) Saturation: [85, 255]
3) Value: [6, 255]

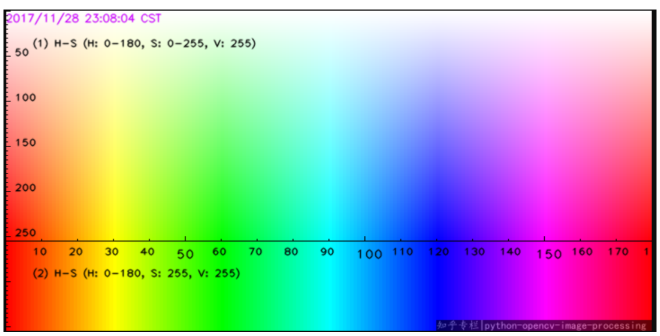


Figure 2. HSV Colour Map [3]

The binary mask consists of white and black pixels. The former represents the pixels with HSV values within the lower and upper boundaries and the latter represents the pixels with out of range values. Bitwise conjunction between the frame and binary mask is performed for visualising the results of colour space segmentation of the tennis ball (Figure 3).

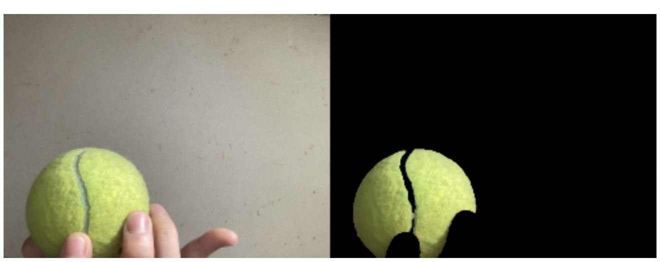
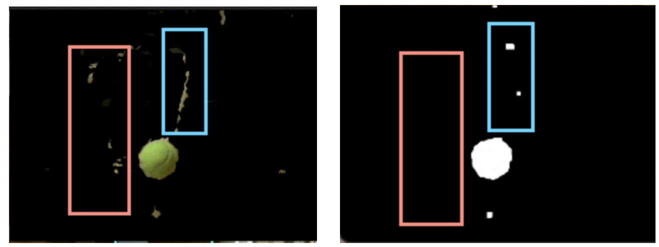


Figure 3. **Left:** Raw frame acquired from the drone's camera. **Right:** Result of colour space segmentation.

Then, morphological operations such as erosion and dilation are used to remove addition noise and refine the edges of the object. Erosion is first apply to remove additional noise. The effect of erosion can be seen in Figure4. The bright region of the the image is minimised, meanwhile the dark region get bigger[4]. Since Gaussian blurring actually removes high frequency content from the image, erosion may further blurred the boundary of the tennis ball. Therefore, dilation is applied to refine the edges of the object by adding pixels to the boundaries.



(a) Before (b) After

Figure 4. Before and after erosion and dilation

C. Contouring and Applying Tracking Indicator

Using OpenCv's built-in function, only the contour with the maximum area is drawn on the frame. This prevents leftover noise to be detected as the object. Next, we find the center point of the object from the circumscribed circle which encloses the object with minimum area. Besides that, a bounding rectangle is also drawn using OpenCv's $boundingRect()$ function to get the coordinates of the top-left corner and dimensions of the rectangle [5]. The center point, coordinates and area of the object will be used to track the displacement of the tennis ball with respect to the middle point of the frame. Tracking indicators are displayed on the frame as shown in Figure 5.

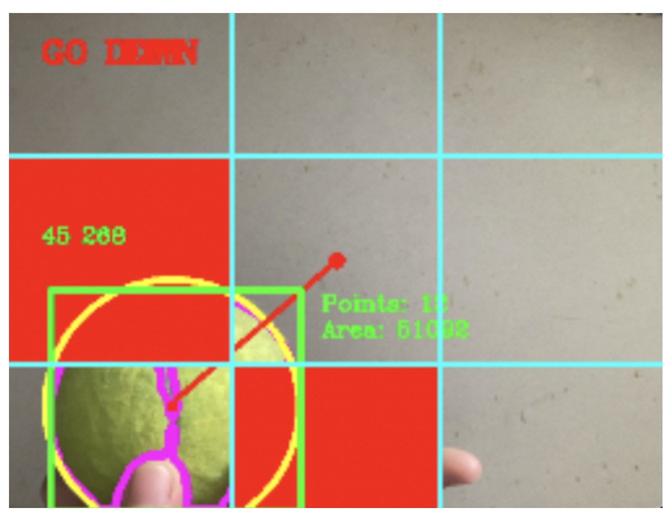


Figure 5. The area and coordinates of the top-left corner of the bounding rectangle is displayed on the frame at all times. Warning signs will pop up if and only if the object is close to the edge of the frame.

D. Object Tracking

The movement of the drone is controlled by three variables, eg. the yaw velocity, the up-down velocity and the forward-backward velocity.

1) Yaw velocity: The yaw velocity is defined as the angular velocity of the rotation. A PID controller is used to determine the rotational speed of the drone. The horizontal displacement of the center of the object from the middle point of the frame is calculated. This value is used as the error. The goal is to minimize error by adjusting the speed of the drone's motor.

2) Forward and Backward Velocity: The forward and backward velocity is determined based on the area of the object. The acceptable distance between the drone and the object is set to a range between 4000 and 8000. Area larger than 8000 indicates that the drone is too close to the object, hence it will have a backward velocity of 20. Similarly, area smaller than 4000 implies that the drone is too far away from the object, therefore it will be assigned a forward velocity of 20.

3) Up Down Velocity: The logic for determining the up and down velocity is rather straightforward. The drone will either move up or down depending on the vertical displacement between the center point of the object and the frame.

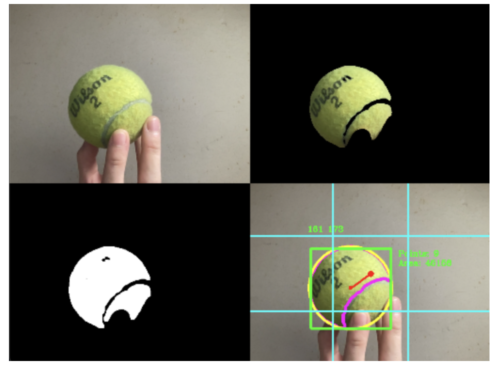
Let (x_{object}, y_{object}) and (x_{frame}, y_{frame}) be the coordinates of the center of the object and the frame respectively. The acceptable displacement is set to a value of 100. Hence, the value of $y_{object})$ must be within the range of $[y_{frame} - 100$, $y_{frame}+100)]$. If y_{object} is out of the acceptable displacement range, it will be assigned an up-down velocity of either positive or negative 25. Otherwise, it will remain at its current height.

III. Results

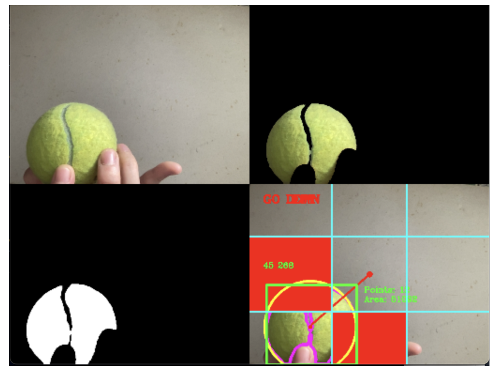
Figure 6 shows sample visual output after processing the frame. Figure 6a and 6b shows the sample output when the test is carried out in settings with limited noise, whereas Figure 6c and 6d are carried out in noisy settings. Results show that the proposed framework is able to accurately detect the tennis ball in settings with and without noise. Besides that, warning signs are accurately displayed when the tennis ball is not located at the middle of the frame. Overall, the object detection and tracking system is working as expected.

IV. Discussions

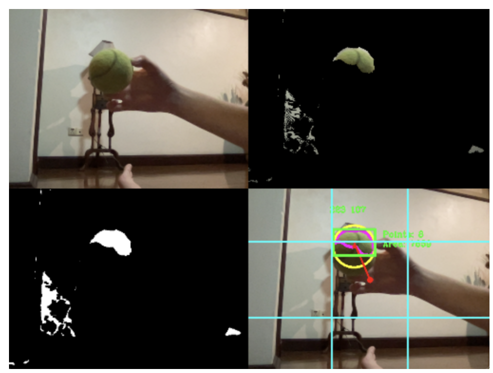
The framework has the ability to detect multiple objects at the same time. Since the proposed framework is implemented with a colour-based segmentation technique using the HSV colour space, it is able to detect more than one object, as long as it falls within the predefined colour range. In this experiment, the system is programmed to only detect object with the largest surface area. Hence, only one object is tracked even in noisy settings, which aligned with the objective of this research. However, this implementation imposes some limitations on the framework nevertheless. In situations where the noise cannot be completely removed, other object with similar colour tone might be detected as the object if the area captured is larger than that of the tennis ball (as shown in Figure 7).



(a)



(b)



(c)

(d)

Figure 6. Visual samples of the annotated frame

Figure 7. The highlighted region shows the actual position of the tennis ball.

REFERENCES

[1] *Smoothing images*. URL: https://docs.opencv.org/4.x/d4/d13/tutorial_py_filtering.html.

[2] R. Cucchiara et al. "Improving shadow suppression in moving object detection with HSV color information". In: *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585)*. 2001, pp. 334–339. DOI: 10.1109/ITSC.2001.948679.

[3] Raghunath Dayala. *Color filtering/segmentation/detection – HSV*. Aug. 2020. URL: https://cvexplained.wordpress.com/2020/04/28/color-detection-hsv/.

[4] *Eroding and dilating*. URL: https://docs.opencv.org/3.4/db/df6/tutorial_erosion_dilatation.html.

[5] *Contour features*. URL: https://docs.opencv.org/4.x/dd/d49/tutorial_py_contour_features.html.