The 4th International Conference on Digital Arts, Media and Technology
and 2nd ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering

Tracking People and Objects with an Autonomous Unmanned Aerial Vehicle using Face and Color Detection

Olarik Surinta
Multi-agent Intelligent Simulation Laboratory (MISL)
Faculty of Informatics, Mahasarakham University
Maha Sarakham, Thailand
olarik.s@msu.ac.th

Sanya Khruahong
Department of Computer Science and Information Technology
Faculty of Science, Naresuan University
Phitsanulok, Thailand
sanyak@nu.ac.th

***Abstract*—We propose a people and object tracking algorithm for an autonomous unmanned aerial vehicle (UAV). It uses as a surveillance camera and can move anywhere. The camera from UAV is not fixed l ocation a s c losed-circuit t elevision. T he face detection and objection detection are applied to support our proposed. In this research, the UAV model for this paper was AR-Drone 2.0. It has a constraint on the front camera because it has fixed the position of view and cannot change the view during flight. W e d esigned t wo e xperiments. F irst, t he f ace detection using images and applied to the popularity of the face detection, is a Haar-cascade classifier and max-margin object detection with convolutional neural network based features because they have high precision in analysis. Second, color detection system, which only focuses on the color of objects which can developed as an obstacle detection system. The results of the experiment can be accepted to adapt to tracking people and objects in the smart-city.**

***Index Terms*—unmanned aerial vehicle, drone, autonomous flight c ontrol, f ace d etection, c olor d etection, t racking algorithm**

I. Introduction

Recently, the tracking of people and objects detection are seen as the crucial factors for surveillance in the smart-city, and can be used to increase safety for the digital society. These methods can applied to support the closed-circuit television (CCTV) systems. Blind-spots in CCTV coverage can compromise coverage and lead to failure. There is a need to find other techniques providing better surveillance.

An unmanned aerial vehicle (UAV), called a *drone* [1], has been developed for many research areas [2]. It can controlled by a user or a program designed for an auto-flying the mission. Such as the military, disaster rescue facility, agriculture and transportation [2, 3]. It can lead to improving the smart device of UAV which is easy to control. Although, the remote control can control the drone with the frequency band (Hertz: Hz) which the most famous of the frequency band for the drone is 900 MHz and 2.4 GHz is suitable for Wi-Fi devices [4]. The low-frequency band can send the signal to be far more than the high-frequency band, but the signal is not stable.

Many companies produce commercial UAVs, but not functional for controlling the UAVs by coding any programs. However, a successful company called *Parrot* produces a UAV model called AR.Drone as shown in Fig. 1. For this UAV model, the developer can code for controlling the UAV by auto-flying and many researchers use it for research development [1, 5, 6]. The developer can communicate by using a smartphone or computer via Wi-Fi signal that are broadcasts from the AR.Drone. Also, it will fix IP address, which can support to being stable flight. The camera on the UAV can be recording video with a resolution of 720p which is enough for adapting to the image processing scheme.

II. Related Work

For the task of face detection, the Viola-Jones face detector [7] is well-known method that first proposed for object and then for pedestrian detection [8] and nowadays, this technique, called *Haar-cascade classifier* , has become a standard technique for face detection. The Viola-Jones face detector computes feature vector based on the Haar feature. It calculates from the rectangle detector or sub-window. The detector scans through the image with the size of 24×24 pixel resolution. The size of the detector will increase by 1.2 times. Then, the set of the feature vector is given to the AdaBoost classifier

Fig. 1: An example of the AR.Drone 2.0 used in this research.

which is the weak classifier. This approach can process in real-time and get high precision.

Dalal and Triggs [9] proposed a feature extraction method called *histograms of oriented gradients (HOG)* and first designed for human detection. This method can use the shape for calculation of the intensity of gradients. Each sub-sampling of images will change the edge images as a cell, in which the orientation of the edge image will be calculated for each cell. Then the orientation can lead to create the histogram and send to learn with the support vector machine (SVM) using the linear kernel. The SVM is used as a binary classifier to be trained on positive and negative images, accordingly it is used to classify each of the sub-window images. Moreover, many research uses the HOG-SVM method [10, 11] for face detection as well.

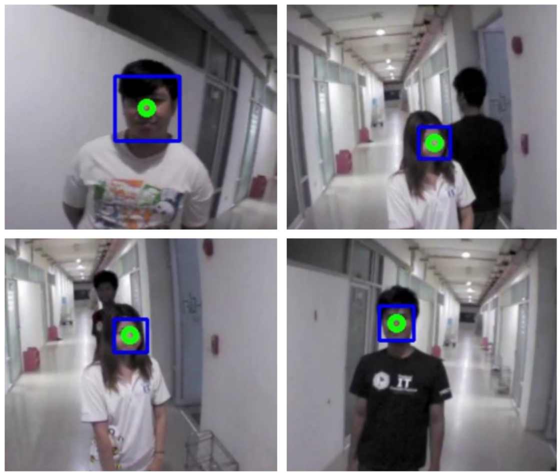
Furthermore, in [12], max-margin object detection (MMOD) is proposed for finding the object. In this approach, the non-overlapping sub-window image slide through image and uses the window scoring function to find the fewest possible detection mistakes. As for the feature extraction method, the spatial pyramid bag-of visual-words model combined with the HOG descriptor method are computed from each pixel location. The MMOD method is performed well on the face detection dataset and benchmark (FDDB).

In addition, in [13], a color image is converted into the HSV color space and an extracted area of the traffic light color such as red, yellow and green colors as the region of interest (ROI). The pixels of traffic light colors are represented as the set of feature vectors. Then the feature vector is given to the machine learning technique to create a model that uses it to classify if the ROI is that of the traffic light or not. In [14], the Haar-like feature and AdaBoost classifier was applied, and reported to have high accuracy.

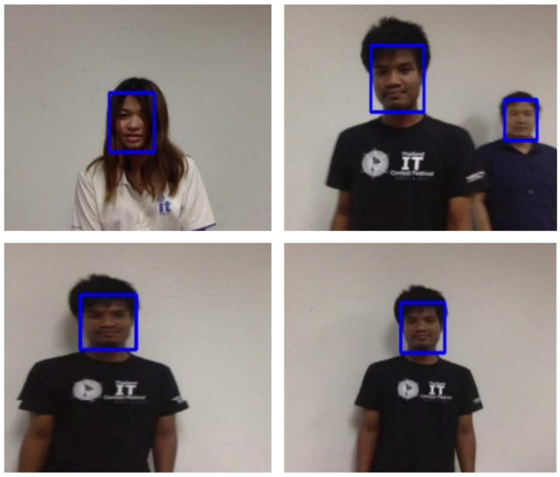
In [15, 16], the UAV was applied to marker recognition and using augmented reality (AR) for flying and tracking to the target by automatic flight and auto-landing.

Nevertheless, Davis [12] presents the max-margin object detection (MMOD) for finding the object. It is different to Dalal and Triggs because all area of the image will be used to analyze with HOG and then applied to the linear function which can develop to finding the objects. Their cases study has three samples, consist of TU Darmstadt cows, INRIA pedestrains and face detection dataset and benchmark (FDDB). Binagkit and Widyantoro [13] propose how to detect to the color of a traffic light, which the image is changed to being HSV color space. After that it is set the period of color such as red, yellow and green for extraction region of interests (ROI). The ROI is selected to learning with SVM by using RBF kernel.

Contribution. In this research, the aim of the study is a detection development of UAV for object detection and tracking people. The primary contribution has two approaches. First, the two well-known techniques of face detection including Haar-cascade classifier [7, 8] and max-margin object detection [12] with convolutional neural network based features (MMOD-CNN). Second, our research focuses on the color detection of the objects [13]. These two approaches can lead to tracking the people and objects based on face and color, respectively.



(a)



(b)

Fig. 2: Some examples of face detection using (a) Haar-cascade classifier and (b) MMOD-CNN methods.

Paper Outline. This paper is organized as follows: We survey related work in section II and describe the proposed people and object tracking algorithm in section III. Section IV, experimental settings and the results are presented and section V finalizes the paper with summary and future work.

III. Proposed People and Object Tracking Algorithm

A. Face Detection

We compare two face detection techniques namely Haar-cascade classifier [7] and max-margin object detection [12] with convolutional neural network based features (MMOD-CNN). As for the experiment, these two approaches applied to the UAV. The results found that the precision of detection was directly related to the constrained illumination environment. Examples of the face detection using Haar-cascade classifier and MMOD-CNN methods are shown in Fig. 2(a) and Fig. 2(b), respectively.

B. Color Detection

To detect candidate regions of interest (ROI) from the image. The color image changes the color space from the RGB to HSV color space [13, 14] because the RGB color space represents the color of the R, G, and B channels. When the illumination slightly changes, it directly impacts the color image. The proposed method used the HSV color space to detect the ROI. This research focuses on three different colors including blue, green, and orange. The process of finding the ROI from the blue image is shown in Fig. 3.

C. Tracking Algorithm

The conception of our approach is using image processing and machine learning to control the UAV automatically by using face and object detection methods. In our method, the centroid of ROI $C_{i,j}$ that was found in the face and color detection steps are calculated by Equation 1.

$$
\begin{aligned}
i &= \tfrac{(x_2 - x_1)}{2} \\
j &= \tfrac{(y_2 - y_1)}{2}
\end{aligned}
\tag{1}
$$

where x_1 and x_2 are pixel position on x-axis and y_1 and y_2 are pixel position on y-axis.

The centroid of the window $C_{h',w'}$ is computed by $h^{'} = h/2$ and $w^{'} = w/2$, which h and w are the height and width of the window, respectively.

The algorithm is called *turnDrone* use $C_{i,j}$ and $C_{h',w'}$ for turning the UAV as shown in Algorithm 1.

Algorithm 1 turnDrone

Input: Set $f(h,w)$ to initial the window size, $C_{i,j}$ is the centroid of the ROI and $C_{h',w'}$ is centroid of the window

```
while C_{h',w'} ≠ C_{i,j} do
  if c_i < c_{h'} then
    move drone down, update C_{i,j}
  else
    move drone up, update C_{i,j}
  end if
  if c_i < c_{w'} then
    turn drone right, update C_{i,j}
  else
    turn drone left, update C_{i,j}
  end if
end while
```

The Algorithm 2 is called *movingDrone*, this algorithm used to support flying which compares the size of the ROI. In this step, the size of the ROI S is compared to the threshold value T. If S is higher than T, the UAV will move backward. On the other, it will move forward. Moreover, this algorithm will calculate the distance between the UAV and object for suitable distance while flying and protection to crash the object.

Algorithm 2 movingDrone

Input: Set T to initial acceptance ratio, Sum pixel of object: $Z_0 = \sum o(i,j)$

```
while Z_o ≠ T do
  if Z_o > T then
    move drone backward, update Z_o
  else
    move drone forward, update Z_o
  end if
end while
```

IV. Experimental and Results

In the experiment, we used a web camera on a laptop computer to experiment with algorithms in the indoor environment. These two face detections techniques performed the high precision of the face detection. Subsequently, we performed face detection techniques using the frontal camera on the UAV. The camera on UAV takes video and send to a computer which is the flight center control. It processes the images then apply to our algorithms for auto-flight control. The precision of the frontal camera on the UAV was slightly less when compared to the web camera on the laptop computer.

The result of validation with images on the web camera, obtained the accuracy to detect although an illustration may affect to calculate. Secondly, using images from UAV, we found that the result was not accurate detection, because of having some problem such as the quality of images that obtained from the UAV and sometimes lost the Wi-Fi signal.

The experiment for measuring the precision of UAV in the Wi-Fi signal area is designed. The camera on UAV takes video and send to computer server which being the center control flight. It processes the images then apply to our tracking algorithms for auto-flight of UAV. Each method is tested with 10 times.

The image from the web camera can be used for face detection. However, one important factor for testing of the face detection; it is the UAV model. We applied our methods to AR.Drone 2.0 model which the camera cannot change the angle view. Thus, the appropriate height to use AR.Drone for face detection is between 150 and 190 centimeters.

As for the color detection, the result found that green and blue color obtained the better result than the orange color. Because the orange color (see Fig. 4(a)) is closely to skin of human also processing appears some skin of arms on the image as shown in Fig. 4(b).

In addition, the face detection needs more computation time than the color detection and so needs longer time to track people. Because the UAV has to move and turn during

(a) (b) (c)

Fig. 3: An illustration of finding the region of interest (ROI). (a) The original image presented with blue color. (b) The algorithm filtered the image containing the blue color and represented as white color on the black background. (c) The ROI of the blue color image.

TABLE I: The range of HSV color space used in this research.

Color	RGB value	HSV Lower Range	HSV Upper Range
Orange	40, 99, 225	0, 100, 100	20, 255, 255
Blue	142, 52, 18	102, 100, 100	122, 255, 255
Green	65, 178, 141	30, 100, 100	50, 255, 255

tracking. Furthermore, the range of HSV values used in this research is shown in Table I.

Our results showed that using the color detection obtains the better detection performance than the face detection. This is because of color detection being robust to changes in illumination.

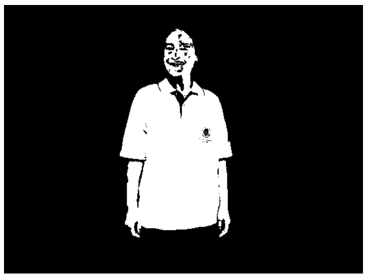
Also, the experiment showed the error occurs when finding the ROI from the orange color image as shown in Fig. 4. In addition, we perform color detection to detect and track a small green ball as an object. The UAV was moved following the object and face as shown in Fig. 5 and Fig. 6, respectively. keeping the same distance between the UAV and the object.

V. Conclusion

In this paper, we proposed the approaches to tracking people and object by using the auto-flying control of the unmanned aerial vehicle (UAV). We applied to Haar-cascade classifier and max-margin object detection with convolutional neural network based features (MMOD-CNN) for face detection, and color detection to find the region of interest from the image. The experiment was designed to use an image from the web camera of a computer and a camera for the frontal of the UAV. The result showed that the color detection outperforms the face detection due to computation time. Color detection experimented with three colors including blue, green, and orange. The color detection could even detect and track the small objects. We found that color detection may fail, for example with orange color because this color is closed to Asian skin color. Finally, the experiment showed that these approaches can adapted to tracking people and objects, and they can be applied to be the monitoring system in the smart-city.

In the future, we will develop with another UAV model, which can change the angle of the camera. Nevertheless, some approaches would be applied to our algorithms, which can lead to being highly stable and highly efficient auto-flight. Our approach may refer to ontology and machine learning. Moreover, the deep learning technique will be designed which can lead to achieving high precision of face detection and good tracking systems.

(a)

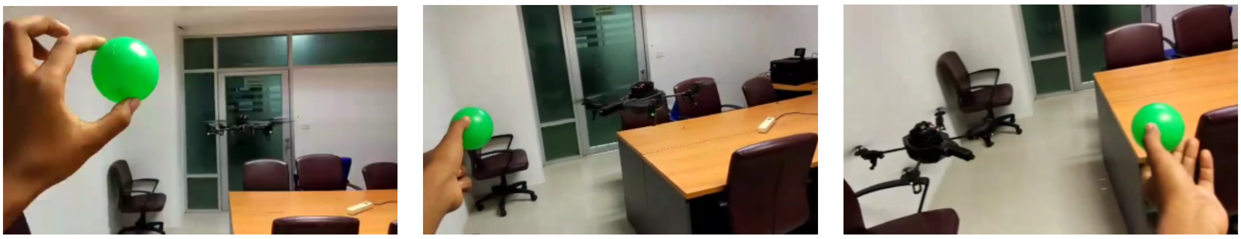


(b)

Fig. 4: An error occurs when finding the ROI from the orange color. (a) The original image presented with orange color. (b) The binary image that containing both the orange color and skin, because of the skin of the people was in a range of the orange color.

References

[1] K. Surakul, S. Sodsee, and S. Smanchat, “A control of multiple drones for automatic collision avoidance,”



(a) (b) (c)

Fig. 5: Illustration the AR.Drone 2.0 while tracking and following the green object.

Fig. 6: Illustration the AR.Drone 2.0 flies in indoor environments using the front camera to detect face from the image.

Information Technology Journal, vol. 11, pp. 91–101, 2015.

[2] A. J. A. Rivera, A. D. C. Villalobos, J. C. N. Monje, J. A. G. M. nas, and C. M. Oppus, “Post-disaster rescue facility: Human detection and geolocation using aerial drones,” in *IEEE Region 10 Conference (TENCON)*, Nov 2016, pp. 384–386.

[3] W. G. Aguilar, V. P. Casaliglla, and J. L. Polit, “Obstacle avoidance based-visual navigation for micro aerial vehicles,” *Electronics*, vol. 6, pp. 1–23, 2017.

[4] M. LaFay, *Drones for Dummies*. Wiley, Jul 2015.

[5] L. Orozco, F. Coloma, C. Oyos, D. Alulema, D. Morocho, and F. Encalada, “Design and implementation of a control of position system for the quadricopter AR.Drone 2.0 by Parrot,” in *IEEE International Conference on Automatica (ICA-ACCA)*, Oct 2016, pp. 1–6.

[6] P. Suwansrikham and P. Singkhamfu, “Indoor vision based guidance system for autonomous drone and control application,” in *International Conference on Digital Arts, Media and Technology (ICDAMT)*, Mar 2017, pp. 110–114.

[7] P. Viola and M. J. Jones, “Robust real-time face detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[8] P. Viola, M. J. Jones, and D. Snow, “Detecting pedestrians using patterns of motion and appearance,” in *9th IEEE International Conference on Computer Vision*, vol. 2, Oct 2003, pp. 734–741.

[9] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, Jun 2005, pp. 886–893.

[10] H. Bristow and S. Lucey, “Why do linear SVMs trained on HOG features perform so well?” *CoRR*, vol. abs/1406.2419, pp. 1–8, 2014. [Online]. Available: http://arxiv.org/abs/1406.2419

[11] H. S. Dadi and G. K. M. Pillutla, “Improved face recognition rate using HOG features and SVM classifier,” *IOSR Journal of Electronics and Communication Engineerin*, vol. 11, pp. 34–44, 2014.

[12] D. E. King, “Max-margin object detection,” *CoRR*, vol. abs/1502.00046, pp. 1–8, 2015. [Online]. Available: http://arxiv.org/abs/1502.00046

[13] J. L. Binangkit and D. H. Widyantoro, “Increasing accuracy of traffic light color detection and recognition using machine learning,” in *10th International Conference on Telecommunication Systems Services and Applications (TSSA)*, Oct 2016, pp. 1–5.

[14] H. Park, N. Kim, H. Hernsoo, and Y. Han, “Traffic light detection and recognition using AdaBoost classifier based on haar-like feature,” in *ISER 59th International Conference*, Jun 2017, pp. 6–9.

[15] J. Kim, Y. S. Lee, S. S. Han, S. H. Kim, G. H. Lee, H. J. Ji, H. J. Choi, and K. N. Choi, “Autonomous flight system using marker recognition on drone,” in *21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*, Jan 2015, pp. 1–4.

[16] T. Zhao and H. Jiang, “Landing system for AR.Drone 2.0 using onboard camera and ROS,” in *IEEE Chinese Guidance, Navigation and Control Conference (CGNCC)*, Aug 2016, pp. 1098–1102.