Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2023.034488
Article

Tech Science Press

Drone for Dynamic Monitoring and Tracking with Intelligent Image Analysis

Ching-Bang Yao[1], Chang-Yi Kao[2,*] and Jiong-Ting Lin[3]

[1]Chinese Culture University, Taipei, Taiwan
[2]Soochow University, Taipei, Taiwan
[3]Chinese Culture University, Taipei, Taiwan
*Corresponding Author: Chang-Yi Kao. Email: edenkao@scu.edu.tw

Received: 18 July 2022; Accepted: 22 September 2022

Abstract: Traditional monitoring systems that are used in shopping malls or community management, mostly use a remote control to monitor and track specific objects; therefore, it is often impossible to effectively monitor the entire environment. When finding a suspicious person, the tracked object cannot be locked in time for tracking. This research replaces the traditional fixed-point monitor with the intelligent drone and combines the image processing technology and automatic judgment for the movements of the monitored person. This intelligent system can effectively improve the shortcomings of low efficiency and high cost of the traditional monitor system. In this article, we proposed a TIMT (The Intelligent Monitoring and Tracking) algorithm which can make the drone have smart surveillance and tracking capabilities. It combined with Artificial Intelligent (AI) face recognition technology and the OpenPose which is able to monitor the physical movements of multiple people in real time to analyze the meaning of human body movements and to track the monitored intelligently through the remote control interface of the drone. This system is highly agile and could be adjusted immediately to any angle and screen that we monitor. Therefore, the system could find abnormal conditions immediately and track and monitor them automatically. That is the system can immediately detect when someone invades the home or community, and the drone can automatically track the intruder to achieve that the two significant shortcomings of the traditional monitor will be improved. Experimental results show that the intelligent monitoring and tracking drone system has an excellent performance, which not only dramatically reduces the number of monitors and the required equipment but also achieves perfect monitoring and tracking.

Keywords: Drone; deep learning; face detection; human pose intention; equidistant track; remote monitoring

1 Introduction

Monitoring equipment has been widely used in the public environment, such as in supermarkets, stations, banks, or even residential areas can find the presence of monitors. With the continuous growth of monitoring systems, the drawbacks of traditional monitoring systems are becoming more and more prominent. First of all, it is challenging for humans to concentrate for a long time in the face of boring surveillance pictures. Even if more professional staff is adopted, the labor cost is quite expensive. Therefore, the traditional monitor only plays the role of post-inquiries. In recent years, AI has become an important direction for the development of modern society where more and more applications are intelligent. With the rapid development of AI-related technology and improvement of computer computing capabilities, the application of drones has dramatically extended in many fields such as disaster rescue, transportation, police security, and plant protection [1].

However, the traditional monitoring system requires manual monitoring in the control room at any time to check whether there are any abnormal phenomena. Therefore, although remote monitoring saves much trouble in monitoring, in essence, this monitoring method still requires a certain degree of human resources requirements. Furthermore, when an abnormal situation occurs, it is also limited that the monitoring camera cannot move with the object, resulting in loopholes in monitoring operations, and even essential pictures cannot be recorded.

With the rapid development of control and AI-related technologies [2], unmanned aerial vehicles (UAV) are quite common in modern society and have become one of the prevalent aerospace industries in recent years [3,4]. Compared with officially manned aircraft, UAVs are often used in aerial photography, detection, military, and other fields due to the advantages of lightness, convenience, and speed [5]. In recent years, UAVs have been used not only in military applications but also in daily life, especially in the monitor field. Therefore, the most significant advantage of using UAVs in the remote monitoring environment [6] is that it can significantly improve the main disadvantage that the monitoring equipment cannot move with the object in the traditional monitoring environment. On the other hand, in recent years, although automated image processing and analysis have also been applied to surveillance photography and even combined with AI technology [7–9], there is still a vast amount of image traffic, so the speed of image analysis often fails to catch up with real-time analysis requirements. Therefore, to satisfy the needs for image processing and identification is very important, especially when the amount of data to be processed is enormous.

In order to improve the two main shortcomings of the above-mentioned traditional monitoring systems and the automated image processing, this study combines face recognition technology in AI image recognition and analysis technology [10], and gesture recognition technology to conduct real-time environmental monitoring and analysis [11], and determine abnormal security conditions. Then, with the instant message transmission technology, the latest monitoring scene images are sent back to the remote monitoring room, and the AI will find out whether there are intruders with suspicious behavior at any time and then send them back in order to obtain the best and real-time information. The monitoring room of the terminal and the AI will find out whether there are intruders with suspicious behavior at any time and then send it back to obtain the best real-time information. This study proposes an intelligent drone monitoring system that uses deep learning methods. First, OpenCV obtains the images taken by the drone, then uses Dlib HOG algorithm to detect human faces [12,13], and then uses OpenPose to identify and mark the critical points of the human body and analysis of posture while realizing timely human-drone interaction based on the posture of the detected human body. Meanwhile, the UAV will keep track and follow equidistantly anytime and anywhere during the monitoring process [14,15].

2 Related Works

This research is based on AI image processing and analysis technology in the remote control of UAV, so that UAV can automatically track and monitor the intruding person after judging the abnormal condition of the monitored place according to the image analysis, and the tracking and monitoring image Continuous and immediate return to the remote monitor. The following is an analysis of related technologies and references used in this research.

2.1 Neural Network

Neural Network is a computational model designed to imitate the biological nervous system. There are usually several layers in a neural network, and each layer normally has numerous neurons. Those neurons sum up the inputs of the previous layer and then convert the activation function as the output of the next layer. Each neuron has a special connection relationship that the output value of the previous layer is passed to the neuron of the next layer after weight calculation. Normally, the activation function is usually a non-linear transformation. Most of the activation functions are Sigmoid functions or hyperbolic tangent functions. This study uses OpenCV modules to implement the most typical multi-layer perceptron (MLP) model of Artificial Neural Networks (ANN) to improve the accuracy of face recognition.

2.2 Machine Learning

Machine learning (ML) is a system that builds algorithms to enable computers to learn from data or to improve performance by accessing data. Machine learning focuses on training computers to learn from data and to automatically make continuous improvements based on experience gained from training, rather than running jobs according to explicit code. Therefore, machine learning is a branch of Artificial Intelligence, which uses historical data for training through algorithms, and generates a model after training is completed. When adding new data, it is easy to make predictions using a trained model. Machine learning can be divided into four types: Supervised Learning, Unsupervised Learning, Semi-supervised Learning, and Reinforcement Learning, according to the characteristics of the learning method and input data.

2.3 Deep Learning

Deep learning [16] is a branch of machine learning. The biggest difference is that machine learning uses ready-made human knowledge to extract features from a large amount of training data then uses these features to train the model so that the model can judge. However, deep learning discards human knowledge and directly uses the neural network architecture that mimics the transmission of human neurons to learn the characteristics of training data from a large amount of data. In other words, this kind of neural network method will greatly reduce the time before processing the data and sufficiently improve the model's accuracy [17]. Additionally, deep learning is a type of technology for performing machine learning. Early use of the Central Processing Unit (CPU) would cause the situation that it was unable to perform heavy calculations. Later, the application of Graphics Processing Unit (GPU) and the introduction of CUDA architecture launched by NVIDIA made deep learning develop rapidly.

2.4 Image Processing

Dlib uses the Histogram of Oriented Gradient (HOG) to detect the position of the human face. HOG uses the calculated gradient to find the relevant information (appearance, features) of the face in the photo, then uses the bar graph to make statistics. The method process is as follows: (1) First, the input image should be grayscale and then performed Gamma Normalize correction. The advantage is that it uses the histogram of directional gradients (HOG) to obtain the features of the frontal face. The biggest advantage is that the processing and calculation speed is fast, and it is unnecessary to adjust the parameters to improve recognition [18]. This is the basis for developing the intelligent detection module in this study using this technology.

It will rule out the influence of light-generating factors because the actual human body may appear on different occasions, and the collected light will be different. (2) Next, images will be divided into cells, calculating the Direction and Magnitude of the gradient in each cell. (3) Take 8*8 cells as an example. The formula that is used to calculate the horizontal and vertical pixels, to calculate the gradient in two directions, to calculate the direction of the gradient as shown in Eqs. (1), (2), (3), (4), respectively. (4) Count the direction and size of the calculated gradients on a histogram. The histogram is the vector of 9 bins and the corresponding angles are 0, 20, 40, … 160. Each bin will be divided into 0 to 20, 21 to 40, 41 to 60, 61 to 80, 81 to 100, 101 to 120, 121 to 140 and 141 to 160, a total of 8 intervals. The horizontal axis of the histogram is the angle, and the vertical axis is the size, which integrates all the cells. (5) Finally, a HOG feature map with size and directionality will be obtained.

$$G_x(x,\ y) = I(x+1,\ y)\ - I(x-1,\ y) \tag{1}$$

$$G_y(x,\ y) = I(x,\ y+1)\ - I(x,\ y-1) \tag{2}$$

$$G_x(x,\ y) = \sqrt{G_x(x,\ y)^2 + G_y(x,\ y)^2} \tag{3}$$

$$\theta(x,\ y) = \tan^{-1}\frac{G_x}{G_y} \tag{4}$$

2.5 OpenPose

Proposed by researchers at Carnegie Mellon University (CMU), OpenPose is now regarded as the most advanced method of real-time human pose estimation. Compared with other methods used in human pose estimation [19], OpenPose has the advantage of having excellent stability and accuracy of body prediction in a different environment. On the other hand, traditional recognition adopts the “top-to-down” method, which will cause the problem of inaccurate recognition when the portraits cross or overlap.

However, the advantage of OpenPose lies in its use of a convolution pose machine (CPM) and the “down-to-top” method for image recognition. First, it will predict the lower body with more straightforward features, and the whole body. This will significantly reduce the calculation time and leave more time to process and identify the overlapping or crossing areas.

When an RGB image is an input, the Openpose will first perform image analysis through the first ten layers of the VGG-19 model to obtain a Feature Map, and then the image features will be input into two-branch multi-level CNN, multi-stage means that the network will be stacked one by one at each stage, increasing the depth of the neural network, so that the subsequent stages can have more accurate output. Two branches indicate that CNN will have two different outputs. The first branch predicts the 2D confidence map of the key joint points of the human body, and the second branch obtains the affinity areas (Part Affinity Fields, PAFs), which are used to predict the limbs. Therefore, the intelligent follower module of this study [20] benefits from the Openpose gesture detection and uses the advantages of PAF technology to quickly identify the body posture of many people in order to further develop the function of controlling the drone with a specific posture, as well as the application functions of detecting and recognizing the movements of the target person, and maintaining an equidistant following and photography.

This research combines the advantages of Dlib and Openpose, and uses the advantages of fast classification of HOG and SVM technology to achieve fast detection and recognition. At the same time, using PAF for fast recognition of body posture, an intelligent, fast recognition, and follow-up module can be made [21], and Its operation mechanism has the most significant advantage of real-time calculation and fast and high accuracy.

3 System Architecture and Module

This research mainly integrates the face recognition function, human posture judgment function, and our automatic monitoring and tracking algorithm to provide a community environment security system with automatic recognition and tracking function. The system modules and main functions of the automatic monitoring and analysis system using UAVs [22] developed in this research are as follows:

3.1 The Intelligent Monitoring and Tracking Algorithm

We take the advantages of OpenPose, an open source library of Carnegie Mellon University (CMU), in the accuracy of face, hand, and human pose judgment, combined with facial recognition technology, then integrated with drone control, equidistant judgment and tracking algorithm to propose-The Intelligent Dynamic Monitoring and Tracking system [23,24]. Furthermore, we utilize the advantages of drone own agility and deep learning algorithm to improve the intelligent dynamic monitoring and tracking algorithm of this article, which has the function of automatic identification, automatic judgment, dynamic adjustment of monitor angle and position and automatic tracking and shooting.

The intelligent drone dynamic monitoring and tracking system proposed in this article adopt TIMT algorithm (The Intelligent Monitoring and Tracking algorithm) and the judgment modules [25] is shown in Fig. 1 below, which is divided into seven steps: First, the system connects the drone remotely and automatically controls takeoff from beginning to the completion of preparatory action. Then, the system module will utilize OpenCV to obtain the drone's picture and send them to the remote pre-set computer and mobile device in time. Third, the system uses Open-Pose by deep learning to identify the critical points of the human body and constructs the shape of the human body [26]. Fourth, the system determines the relevant actions and classifies actions based on the detected pose. Fifth, the "Monitoring and Tracking Module" will allow drones to analyze human movements and determine the movement's meaning while the drone maintains a proper distance from the human [27,28]. Sixth, the system uses Dlib to perform face detection with the HOG algorithm and then judges the specific movement of the monitored person and changes the position. Finally, the distance between the monitored person, the drone can track and follow equidistantly anytime and anywhere.

3.2 System Modules and Functions

This "Intelligent Dynamic Monitoring and Tracking System" contains five modules, including Drone remote control module, a face detection module, a pose estimation, and analysis module and Equidistant tracking and following modules. The function under each module mainly includes image processing, face detection, pose classification and so on, as shown in Fig. 2 below.

- ***The Intelligent Monitoring and Tracking Algorithm***

We take the advantages of OpenPose, an open source library of Carnegie Mellon University (CMU), in the accuracy of face, hand and human pose judgment, combined with facial recognition technology, then integrated with drone control, equidistant judgment and tracking algorithm to propose-The Intelligent Dynamic Monitoring and Tracking system [29]. Furthermore, we utilize the advantages of drone own agility and deep learning algorithm to improve the intelligent dynamic monitoring and tracking algorithm of this article, which has the function of automatic identification, automatic judgement, dynamic adjustment of monitor angle and position and automatic tracking and shooting [30,31].

3.3 Drone Remote Control Module

In the beginning, the system interface will establish a connection with the drone through Wi-Fi wireless network and confirm the drone status and flight data before searching the target intelligently. At the same time, the remote computer or the preset device will display the returned picture taken by the drone with the appropriate frame size. Moreover, the drone module is an instant connection between the host and the drone by the socket.

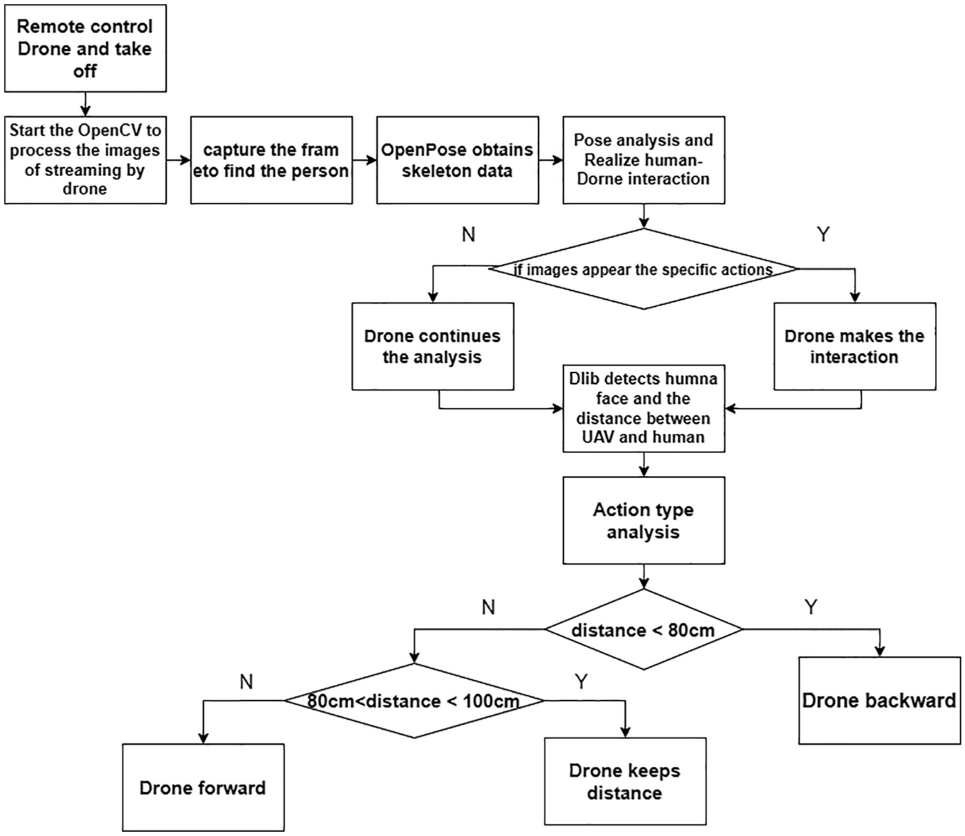

Remote control Drone and take off
Start the OpenCV to process the images of streaming by drone
capture the fram eto find the person
OpenPose obtains skeleton data
Pose analysis and Realize human-Dorne interaction
N
Y
if images appear the specific actions
Drone continues the analysis
Drone makes the interaction
Dlib detects humna face and the distance between UAV and human
Action type analysis
N
Y
distance < 80cm
N
Y
80cm<distance < 100cm
Drone backward
Drone forward
Drone keeps distance

Figure 1: Flow chart of intelligent dynamic monitoring and tracking system

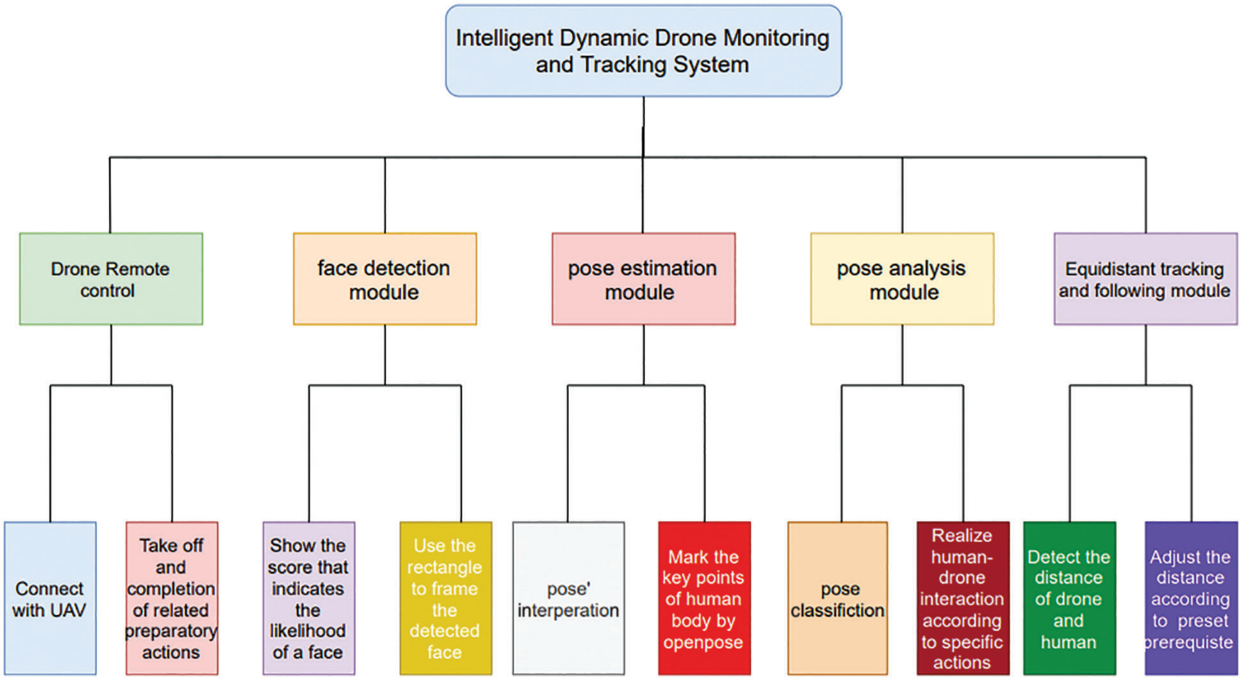

Intelligent Dynamic Drone Monitoring and Tracking System
Drone Remote control
face detection module
pose estimation module
pose analysis module
Equidistant tracking and following module
Connect with UAV
Take off and completion of related preparatory actions
Show the score that indicates the likelihood of a face
Use the rectangle to frame the detected face
pose' interperation
Mark the key points of human body by openpose
pose classifiction
Realize human-drone interaction according to specific actions
Detect the distance of drone and human
Adjust the distance according to preset prerequiste

Figure 2: Function decomposition diagram

3.4 *Face Detection Module*

The intelligent dynamic drone system integrates Dlib into this module and uses Dlib which provides a well-trained module to identify 68 feature points of the face, including nose, eyes, eyebrows, and mouth. The system will detect and recognize the human faces before judging the action by deep learning. The traditional deep learning method is to compare, classify, or predict each data's feature value [26,29]. In this study, we adopt an HOG direction gradient bar graph that it will collect contour features, preprocess the image to grayscale, calculate the gradient strength and direction of each pixel to make statistics, and finally forms the HOG of the image after a series of integrations, which to achieve the purpose of better face detection. After successful face detection, a green frame will be used to frame the face and follow the face according to its movement of the face. A score will appear on the top of the frame. The higher the score, the higher the possibility of a human face. The timely frame rate (FPS) also appears on frame, as shown in Fig. 3 below.

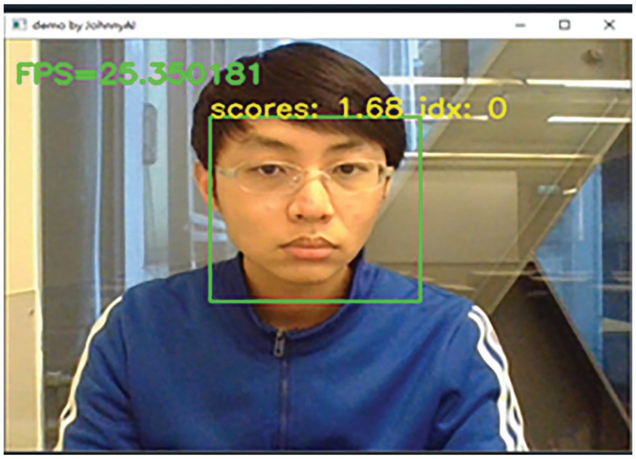
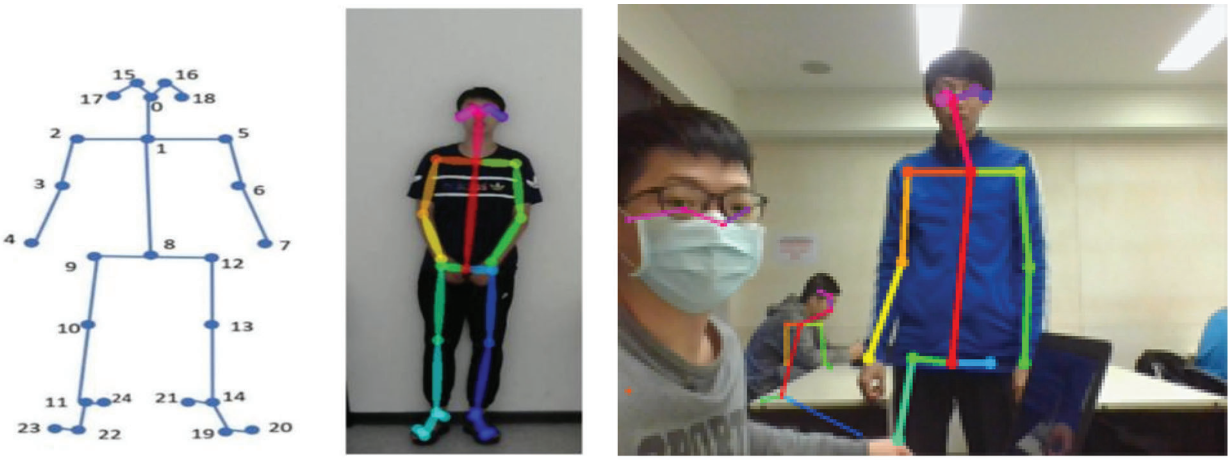


Figure 3: Dlib face detection graph

3.5 *Human Pose Estimation Module*

This module writes posture analysis and segment algorithm programs for different postures of people for automatic analysis and judgment, so that the drone can understand and interpret various postures of people. At the same time, through this module, users can control the drone's flight with different actions.

This module utilizes the OpenPose library to identify human poses and adopts the CNN deep learning architecture to find the Confidence Map of each joint position and the newly defined Part Affinity Fields (PAF) of OpenPose to find the corresponding Posture vector. After the model integrates the above two features, it can further predict each limb segment and successfully mark the outline of the entire body skeleton, which contains a total of 25 feature values, as shown in Fig. 4a below. This study uses the posture judgment of openpose as the basis and the Intelligent Dynamic Monitoring and Tracking algorithm proposed to identify the user's specific actions, and then let the drone make the corresponding flight actions [32,33]. At the same time, this function is used to monitor the environment automatically and track specific targets, as shown in Fig. 4b below.



(a) (b)

Figure 4: (a) Openpose human pose estimation graph. (b) Human pose judgment graph of Openpose in a Multi-person environment

3.6 Specific Pose Analysis Module

The intelligent UAV system of this study will automatically recognize the specific posture made by the monitored person through this module, and make a timely response, to achieve a better monitoring effect [32]. In this study, the drone could recognize five specific postures, respectively "the right arm open ", " the right arm close ", " the left arm open ", " the left arm close ", hands crossed on the chest, and the" hands on the neck ". The intelligent drone system will interpret these specific postures and then realize a human-drone Interaction. This study uses the key points and line segments of Openpose, and by calculating the angle between the line segment and the adjacent line segment, you can determine the specific posture you set up and use this specific posture as a further application for operating the drone. There are the four specific gestures, as shown in the Figs. 5a–g below:

3.7 Equidistant Tracking and Following Module

This equidistant tracking and the following mechanism is the core module of this study's intelligent monitoring and tracking UAV. The steps for automatic monitoring and tracking algorithm are as follows:

3.7.1 Calculate the Distance Between the Drone and the Target

According to the imaging principle formula, the module will calculate the distance between the drone and the person. While using the imaging principle, the ratio between the height or size of the object in the photo and the actual height or size of the object is equal to the ratio of the focal length to the actual distance between the camera and the object.

(1) Calculate the distance of the object:

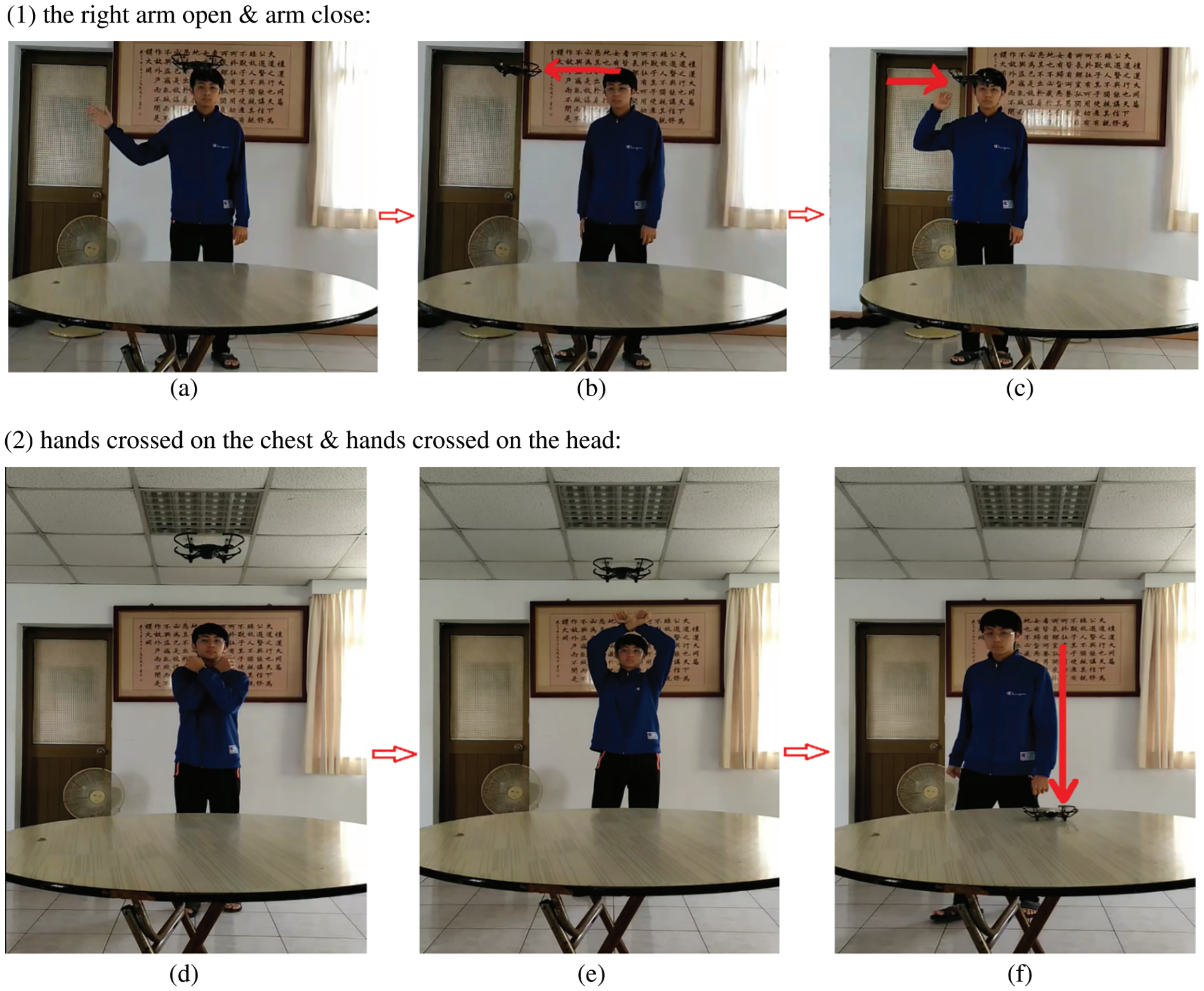
Use the formula:

Distance (cm) = focal length (mm) * subject length (cm)/image size (pixels)

it first deduces the focal length of the drone, and then inverts the actual distance. Therefore, it can get

Focal length (mm) f = distance (M) * sensor size (mm)/length of the object to be shot (L)

Therefore, the calculated focal length can be used to calculate the distance between the drone and the object being photographed and monitored.



(1) the right arm open & arm close:

(a) (b) (c)

(2) hands crossed on the chest & hands crossed on the head:

(d) (e) (f)

Figure 5: (a) Gesture guidance. (b) Drone flying to the right. (c) Gesture guide the drone to fly back. (d) Gesture guide to take a picture. (e) Hands gesture to land the drone. (f) Drone landing

(2) Calculate the distance between people:

After the drone first recognizes the face, it will immediately frame the recognized face, and you can use the marked face width (pixel) in the photo to reverse it, and the actual face width, to find The focal length value of the camera can be obtained. Then, the focal length value can be used to obtain the actual distance value between the drone and the person being photographed. This study uses the UAV's automatic monitoring module to show that when the UAV automatically performs monitoring, it will detect while taking pictures. Once a face is detected and recognized, the face will be marked, and the distance to the subject will be calculated, as shown in Fig. 6.

After that, the drone maintains the same equidistant distance from the monitored person [31,33,34]. First, this module adopts the "camera imaging principle". When estimating the distance of the person in front of the Lens, cameras use similar triangles to calculate the distance from the lens to a known object or target. Next, this article preset that the drone keeps a safe monitoring distance of 80 to 120 cm from the person being monitored. If the distance between the drone and the person is less than 80 cm, the drone will automatically move forward. If the distance exceeds 120 cm, the drone will automatically move back to a reasonable distance.

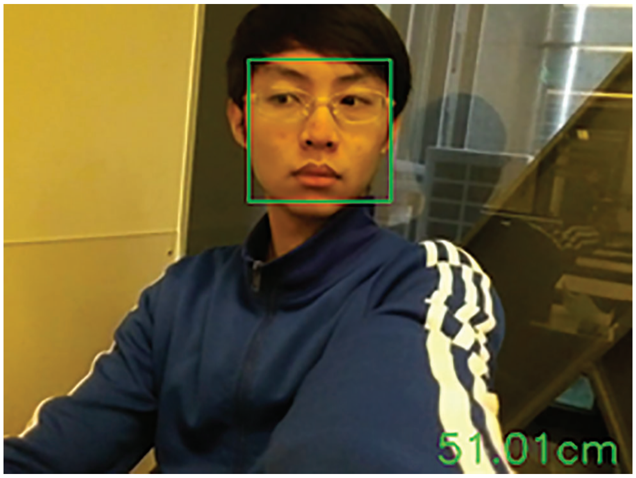


Figure 6: The drone instantly calculates the distance to the target being tracked

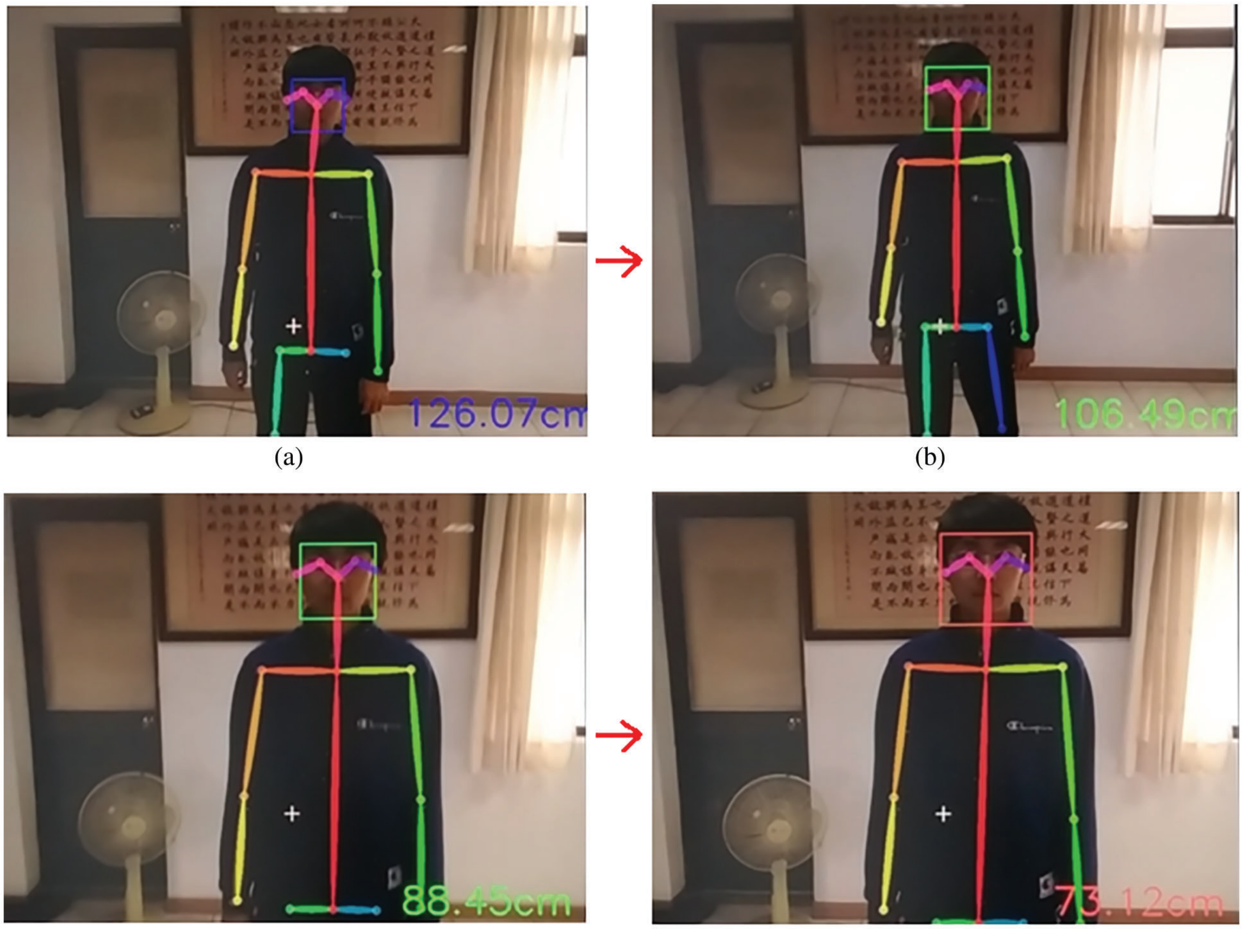
3.7.2 Calculate the Difference Between the Positions of the Monitoring Targets in the Continuous Images

After identifying the following face, the UAV immediately compares and calculates the positions of the tracking targets in the continuous images. The drone will continue to detect the following face in the surveillance. When the position of the face in the photo is found to be offset, it will calculate and analyze the relative position of the tracking target in the front and rear images.

The operation of the UAV detection and automatic following target modules, the operation between the core modules can be divided into the following four steps: (1) First, the "UAV" will check the environmental image being shot to detect whether someone. Therefore, the people in the film will be identified and framed for display, as shown in Fig. 7a below. (2) Next, calculate the position and distance of the identified characters, as shown in Fig. 7b. (3) Then, the posture recognition module of the UAV system recognizes the human body posture of the person in the image. (4) Finally, continue to track the people in the continuous images, determine the direction the target person is moving, and then keep a certain distance from the person being followed, automatically monitor and take pictures [34], and send the data back to the monitor, as follows Fig. 7d. The actual demonstration process of the whole UAV identification and automatic following target calculation steps, as shown in Figs. 7a–d below, shows that the detected person continues to approach.

3.7.3 Instantly Judge the Most Likely Direction of Movement of the Target

Next, according to the position difference of the tracking target, we will calculate the relative coordinate change of the targeted person's position between the two frames, then it is further judged what the relative moving direction of the target person. Finally, the drone will continue to analyze the image path of the subsequent pictures to determine the movement method of the tracking target. In fact, this study uses Openpose as the module for human posture recognition, the main reason which it can take into account are the characteristics of high accuracy and fast calculation. Therefore, this same principle is also why Dlib is used for face recognition in this study. Dlib extracts features, and obtains target's face features through a HOG, which is characterized by fast computing speed and high accuracy.



(a) (b) (c) (d)

Figure 7: The drone detects the target and judges that the target person continues to approach

4 Experiment and Discussion Result

This research focuses on the experiment of UAV's face recognition for abnormal intruders, as well as the experimental design of automatic monitoring and tracking. It simulates the results of the experimental recognition rate in various environments, which are described as follows:

4.1 Experimental Environment

The experimental environment is divided into two parts as follows:

(1) The environment of face detection and frame of the target:

Among them, the function of the UAV detection and monitoring module is based on the training model of OpenCV plus Dlib, which is used to capture and identify facial features, and then use Linear SVM (Support Vector Machine) to distinguish the belonging classification to be able to perform facial detection and identification. In the testing part, the experimental method of this study is to simulate various scenarios where someone appears, and no one appears and let the drone take a total of 40 photos for testing to determine whether the face can be correctly detected and framed in the photo correctly.

(2) The environment in which human body movements are recognized and followed:

This research system uses the human limb key point marking, PAF (A Part Affinity Field), a model built by Openpose as a training model, and based on this, develops the recognition of specific human poses and applies it to the control function of UAVs. The experimental method of this study is also to simulate various scenes with people appearing and no people appearing, let the drone take a total of 40 photos, and then test whether this module can correctly identify the posture of the human body and whether it can be kept correctly at a fixed distance to trace the target person.

4.2 Experimental Design

The experiments of this study are proceeded according to different distances and different indoor brightness environments to measure the automatic identification, ability to track targets, and flight accuracy of the intelligent UAV in this study. Therefore, there are the four main module functions of the "Intelligent Monitoring and Tracking System" proposed by this research which contains: UAV image processing and face recognition functions, UAV body posture recognition and judgment, system monitoring and tracking functions, and equidistant follow the function of UAV, we designed and conducted the following four sets of experiments, which are described as follows:

$$CRR(Correction\ Rate) = \frac{Number\ of\ Correction}{Total\ Cases}$$

4.2.1 The Recognition Rate of Face Detection with and Without Masks and the Different Angles

This experiment is divided into two groups without masks and wearing masks, and each conducts face detection and recognition experiments. When the drone correctly frames the detected face, it is considered a successful recognition. According to the different angles facing the UAV, this experiment is divided into four different detection angle experiments, and 40 experiment times are carried out for each angle. The experimental results are shown in Table 1 below:

Table 1: Experimental results of face detection and recognition with and without masks

	Front	Turn 30 degrees	Turn 60 degrees	Turn 90 degrees
Without mask	97.5%	77.5%	35%	0%
With mask	87.5%	60%	22.5%	0%
Number of experiments	40	40	40	40

In this experiment, when the drone faces the target person directly, the facial recognition rate is the highest, which are 97.5% (without a mask) and 87.5% (with a mask). Then, as the angle at which the drone captures the face of the target person gradually increases, it only captures the side of the face, so the recognition rate gradually decreases with and without a mask. When the face turns to a 60-degree skew, the recognition rate drops to 35% (without a mask) and 22.5% (with a mask). Finally, until the face of the target person is turned to a full side view of 90 degrees, the recognition rate of the face becomes 0. Because when the face is turned to 90 degrees, there is absolutely no way for any feature values to be extracted, resulting in unrecognizable results.

4.2.2 The Accuracy of Face Detection Under the Overall Different Brightness Environment

In order to test the accuracy of face detection under the brightness of different light sources in the overall environment. The experiment chooses to adjust the bright-ness of 25%, 50%, 75%, and 100% in a closed classroom for testing, as shown in Fig. 8 below. According to the experiment, the accuracy of the UAV's face detection by the different brightness is obtained. In the environment of different brightness, each brightness is tested for 20 times, and the accuracy of face detection is 95% and 95%, respectively. 97%, 98%, as shown in Fig. 9 below.

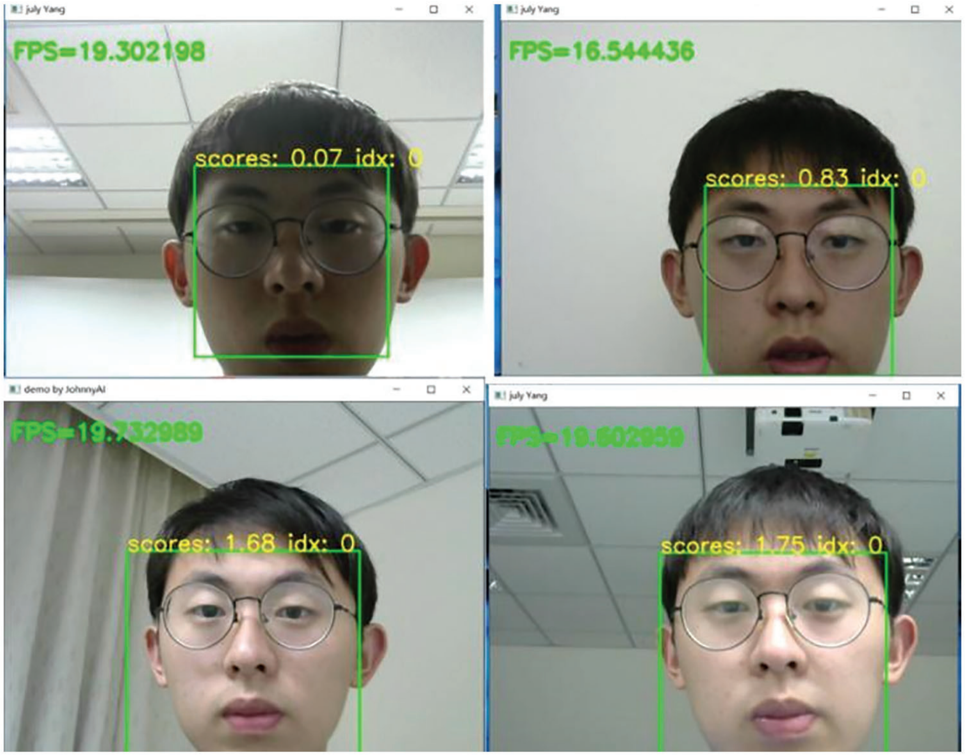


Figure 8: The test situations under different brightness environments

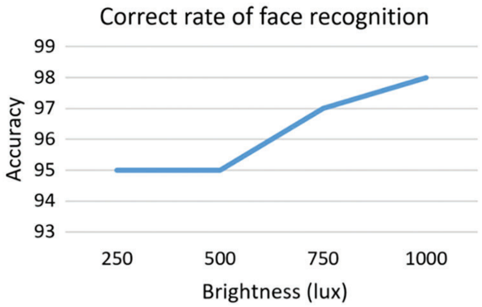


Figure 9: Face detection accuracy rate under different brightness

4.2.3 The Accuracy of Face Detection Under the Changing Background and Distance Under Environment

In this experiment, by changing different backgrounds, we examine the degree of influence on the face recognition rate of this research system and Google's Teachable Machine system. In addition, this experiment also changed the distance between the target and the UAV, and then tested the accuracy of the intelligent monitoring UAV's face detection and recognition, compared with the accuracy of the Teachable Machine platform system. The experimental results are shown in Table 2 below.

Table 2: Experimental results of face detection and recognition with the changing background and distances

	System comparison	Similar background	Changing background	Extend the distance
Without mask	OpenCV + Dlib	97.5%	95%	95%
	Teachable Machine	98%	95%	62.5%
With mask	OpenCV + Dlib	87.5%	82.5%	80%
	Teachable Machine	92.5%	87.5%	52.5%
Total number of each of experiments		40	40	40

Without wearing a mask, the recognition accuracy of the face recognition function developed based on OpenCV and Dlib is similar to that of Google's Teachable Machine, above 97%. Both are equally good. However, when the test background becomes complicated, especially when the tester is kept away from the drone so that the target person and the background are confused, the face recognition accuracy rate of Teachable Machine will drop rapidly by about 62%, but the real-time recognition rate of the drone in this study is 80%, which is almost only a little lower. In the above situation, if you change to wearing a mask, the situation is almost the same as not wearing a mask.

4.2.4 The Degree of Influence on Human Body Posture Recognition Under the Overall Environment of Different Brightness

In order to test the accuracy of the system's recognition of human posture in the overall environment, the experiment selects the influence of the system on the judgment of human posture in an environment with a light brightness of 50% and a brightness of 100%. There are four possible situations for the determination of posture: (1) the posture has been posed, and the system has correctly determined the correct posture, (2) the posture has not been posed, but the system has detected the posture incorrectly, (3) the posture has been posed and the system has not detected it, (4) the posture has not been posed, and the system has not detected it either. The above four situations are represented by TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative), respectively. The experimental results are expressed in terms of Sensitivity, Specificity, and Accuracy.

In order to find out whether the drone will affect the real-time recognition ability of the video shot by the drone under the environment of different lighting levels, the impact on the real-time recognition rate of the captured human face under different lighting levels is carried out. Among them, the experimental illuminance, commonly known as lux, represents the luminous flux the subject surface receives per unit area. 1 lux is equivalent to 1 lumen/square meter, that is, the luminous flux irradiated vertically by a light source with a distance of one meter and a luminous intensity of 1 candle per square meter of the subject. Illumination is an important indicator to measure the shooting environment. The illuminance is suitable for reading, sewing. is about 500 lux.

Finally, through experiments, it was found that under different environments, accuracy, sensitivity, and specificity, under the conditions of a 50% brightness environment, the UAV's posture discrimination experiment is shown in Fig. 10 below. The results of the three indexes are respectively 98%, 96.1%, and 95%. In the environment condition of 100% brightness, the UAV's posture discrimination experiment is shown in Fig. 11 below. Again, the three indicators have improved which are respectively 98.9%, 96.5%, and 94.1%. Finally, in this experiment, in two different brightness environment systems, the UAV's posture discrimination experiment, and the experimental data on the three indicators of Accuracy, Sensitivity, and Specificity, are organized as shown in Table 3 below.

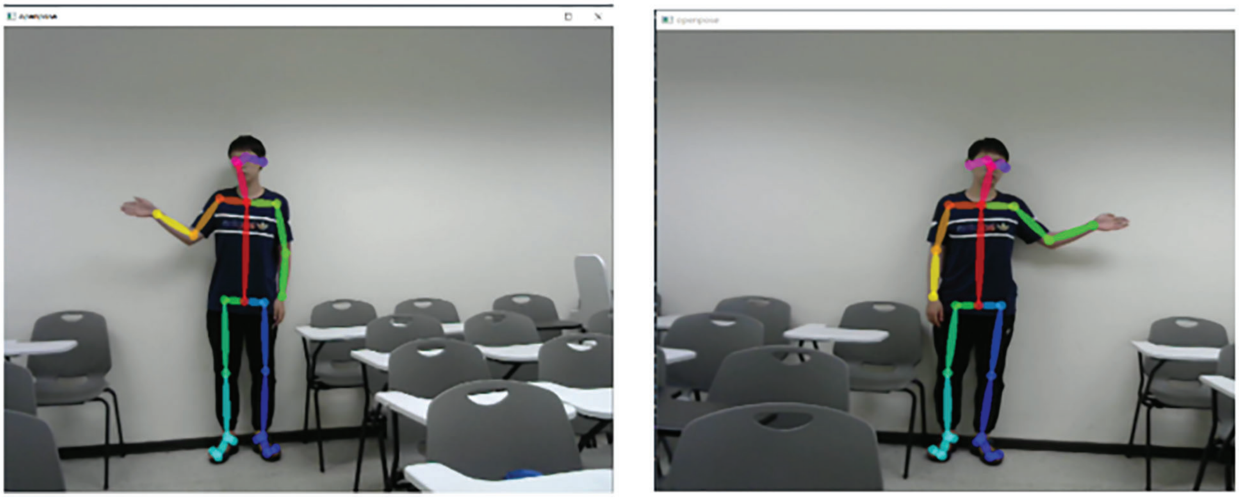


Figure 10: Under 50% brightness environment, the drone's judgment of posture

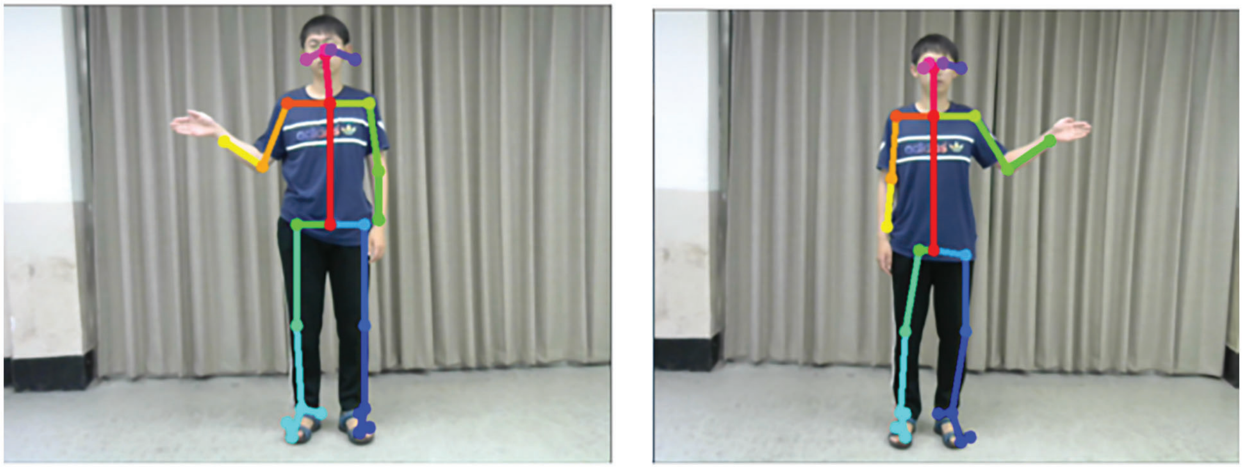


Figure 11: Under 100% brightness environment, the drone's judgment of posture

Table 3: Experimental results of posture judgment in different environments

Index\light	50%	100%
Accuracy	98%	98.9%
Sensitivity	96.1%	96.5%
Specificity	95%	94.1%

4.2.5 The Effect of the Reaction Time Required for Drone Tracking at Different Distances from People

In order to test the reaction time of the UAV during monitoring, equal distances are required for the test. The distance between the monitored person and the UAV is 100, 150, and 200 cm was selected in the experiment, as shown in Fig. 12 below.

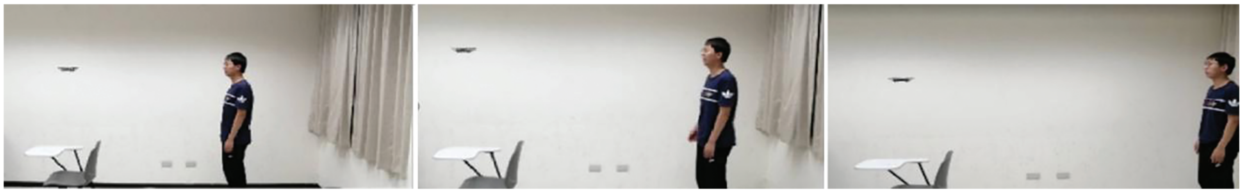


Figure 12: Testing of drones at different distances from people (near, medium, and far)

Each of the experiments was tested 20 times, and it was found that the system's reaction time (t) at the distance of 100 cm was about 0.5 s. On the other hand, the reaction time at the distance of 150 cm is 1 s, and the reaction time at the distance of 200 cm is about 1.2 s, as shown in Fig. 13 below.

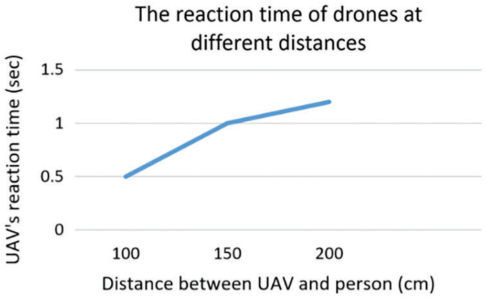


Figure 13: Testing of drones at different distances from people (far, medium, and near)

4.2.6 At Different Distances from People (far, Medium, and Near), the Accuracy of the UAV System in Judging the Posture

In order to obtain the correct rate of the system's posture judgment, in this experiment, the distance between the human and the drone is divided into three types: short (100 cm), medium (150 cm), and long-distance (200 cm). The right arm is opened, and the right arm is closed. The two actions are tested 20 times each, and the results are recorded to check whether the drone can correctly follow the command of the gesture and complete the corresponding flight action. Among them, the right arm is extended to guide the drone to fly to the right side, and the right arm is closed to guide the drone to fly to the left. Using the gesture recognition algorithm, different gestures are used for real-time judgment, to facilitate the control of the UAV. The control sensitivity and control accuracy experiments are carried out in Figs. 14 to 16 under different indoor brightness environments.

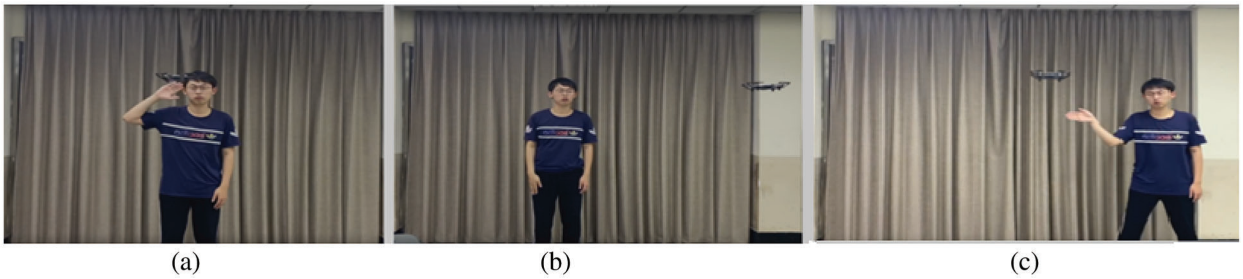


Figure 14: In the case of a short distance, close the arm with the right hand to control the drone to fly 30 cm to the left, and open the arm with the right hand to control the drone to fly 30 cm to the right

As shown in Fig. 14a, the right-hand arm is close to guide the drone to fly 30 cm to the left, which is used to judge the specific postures to apply in the human-computer interaction. It is found through experiments that when the distance is 100 cm, the correct rate can be as high as 100%; when the distance is 150 cm, the correct rate is as high as 90%, as shown in Figs. 15a to c below. Finally, when the distance is 200 cm, the accuracy rate is as high as 90%, as shown in Figs. 16a to d below.

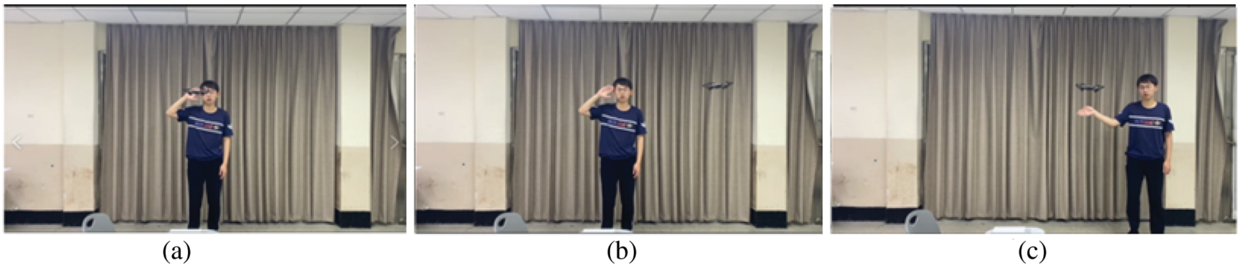


Figure 15: In the case of a medium distance, close the arm with the right hand to control the drone fly 30 cm to the left, and open the arm with the right hand to control the drone to fly 30 cm to the right

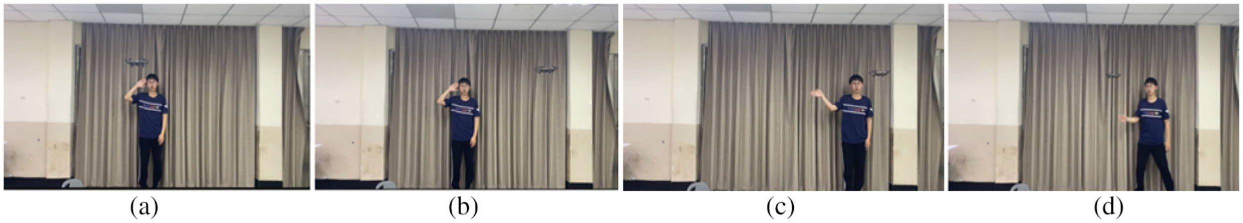


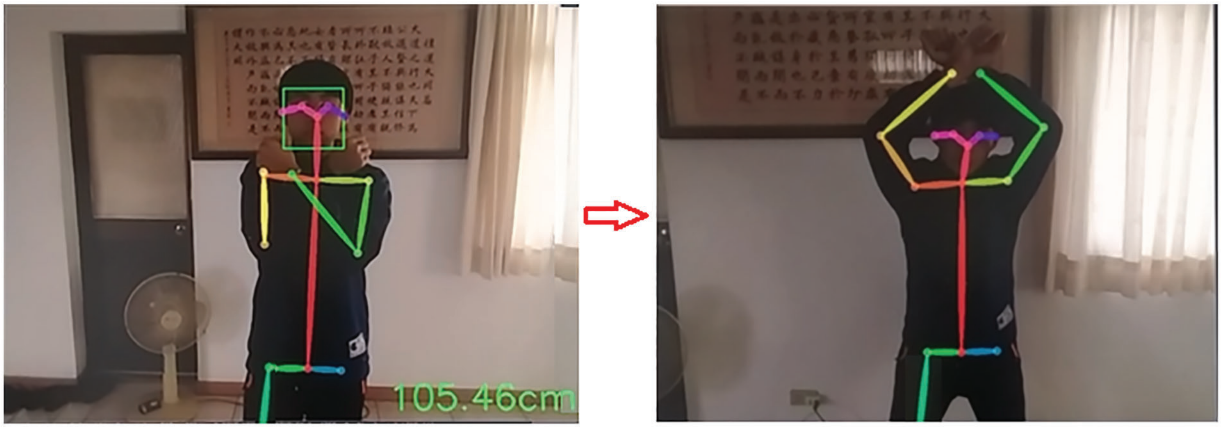
Figure 16: In the case of a long distance, close the arm with the right hand to control the drone fly 30 cm to the left, and open the arm with the right hand to control the drone to fly 30 cm to the right

4.2.7 The Experiment of Commanding the Drone to Take Pictures and Land in Time with Specific Gestures

Finally, the experiment is to control the camera and landing functions of the drone with specific gestures. Similarly, this experiment was repeated 20 times for each type of gesture, and the results were recorded. The hands are crossed in front of the chest to let the drone take pictures, as shown in Fig. 17a. In addition, when the hands are crossed on the head, it means to let the drone land and complete the flight, as shown in Fig. 17b. The experimental results show that gestures make the drone take pictures with a success rate of 95%, and the average response time is 0.9 s. At the same time, another gesture experiment is to cross the hands on the head to let the drone land to complete the flight, and the success rate is 95%, and the average reaction time is 1.4 s. This result confirms that the UAV intelligent monitoring module can accurately detect and rapidly analyze the movements of the target person.

4.3 Experimental Analysis and Discussion

The experiment is mainly divided into three parts, the accuracy of the system's face detection under different luminosity, the accuracy of human posture recognition and reaction time, and the accuracy of the system's judgment at different distances. Experiments have proved that the intelligent UAV system has an accuracy of nearly 95% of face detection and human posture accuracy, a short response time (less than 2 s), and the landing of the UAV is very stable.



(a) Gesture make drones take pictures (b) Experiments with landing drones

Figure 17: (a) In the case of a long distance, hands crossed chest. (b) Hands crossed over head

4.3.1 The Impact of Different Detection Techniques When the Background has Complex Changes

The intelligent surveillance module of this study is also compared with Google's Teachable Machine AI platform in face detection and recognition, as shown in Table 2: when the background changes to more complex, the intelligent surveillance module For face detection and recognition, the performances of both are outstanding, and they are similar, but when the distance becomes farther, and the target person is kept away from the distance from the drone, the accuracy of the research system only decreases. , but the recognition rate of the Google platform has dropped rapidly. The reason is that the system in this study uses Dlib HOG plus Linear SVM. Since face recognition is performed quickly by lifting the features, it is less affected by the distance of the background environment. Hence, larger background differences, but Google's AI platform will be affected by changes in the background when the distance changes.

4.3.2 Instant Analysis of Some Main Factors Affecting Face Detection and Body Gesture Recognition

In addition, under different brightness environments, the TIMT algorithm of the UAV in this study can achieve a correct recognition rate of more than 95% and correctly calculate the distance to the target person. Therefore, the advantages and goals of intelligent environmental monitoring and dynamic tracking can be achieved. Furthermore, the average response time from the completion of face recognition and body posture recognition to the command to let the drone start to perform actions is within 0.7–1.6 s, which proves that the response time is quite fast. Nevertheless, occasionally, there will be situations where the posture cannot be judged. There are two possible reasons: (1) The drone delays in transmitting the image, resulting in a situation where the posture judgment cannot be judged, and (2) the drone is in face detection. Furthermore, during the measurement process, because the drone is too skewed in the shooting angle, the reticle of the face recognition may sometimes be offset and not stable enough, which will affect the measured distance and become inaccurate. In the future, image processing techniques that correct for skew effects can be used to improve the accuracy of the Dlib face judgment module in marking the face frame.

5 Conclusions

The intelligent monitoring and tracking drone system proposed in this article combines with deep learning algorithm of Dlib and OpenPose and adopts the python library to develop the TIMT algorithm, which enables the system to have dynamic monitoring and automatic tracking functions.it will significantly improve the disadvantages of fixed-point camera. The system will detect and analyze the human face and the body gestures immediately and judge the distance between the drone and the monitored person, and then track and follow the monitored person at an equal distance according to the movement through the intelligent recognition and tracking module. Experiments have confirmed that this research system can accurately and instantly identify and effectively provide continuous monitoring and tracking functions with mobile capabilities. Therefore, it can effectively improve traditional monitoring systems' effective monitoring range and efficiency. On the other hand, in the process of flight, the drone will also automatically and timely make movements according to the relevant postures and actions of the monitored person to achieve better human-computer interaction, greatly improve the shortcomings of traditional surveillance camera blind spots, and effectively improve the quality of better Monitoring and monitoring screen.

Acknowledgement: No support given which is not covered by the author's contribution.

Funding Statement: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] U. Z. Uddin, J. J. Lee and T. S. Kim, "An enhanced independent component-based human facial expression recognition from video," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 4, pp. 2216–2224, 2009.

[2] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *IEEE Transactions on Affective Computing*, vol. 6, no. 1, pp. 99–111, 2015.

[3] R. A. Khan, A. Meyer, H. Konik and S. Bouakaz, "Frame work for reliable, real-time facial expression recognition for low resolution images," *Pattern Recognition Letters*, vol. 34, no. 10, pp. 1159–1168, 2013.

[4] S. S. Sarmah, "Concept of artificial intelligence, Its impact and emerging trends," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 11, pp. 2164–2168, 2019.

[5] H. M. Jayaweera and S. Hanoun, "A dynamic artificial potential field (D-APF) UAV path planning technique for following ground moving targets," *IEEE Access*, vol. 8, pp. 192760–192776, 2020.

[6] A. Mollahosseini, B. Hasani and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2017.

[7] S. Tu, S. U. Rehman, M. Waqas, Z. Shah, Z. Yang *et al.,* "ModPSO-CNN: An evolutionary convolution neural network with application to visual recognition," *Soft Computing*, vol. 25, no. 3, pp. 2165–2176, 2021.

[8] Y. Lecun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[9] N. Dalal and B. Trigs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, San Diego, CA, USA, pp. 886–893, 2005.

[10] Y. Li, J. Zeng, S. Shan and X. Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, 2018.

[11] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei and Y. Sheikh, "Openpose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–182, 2019.

[12] C. Shan, S. Gong and P. W. Mcowan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[13] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[14] M. P. J. Ashby, "The value of cctv surveillance cameras as an investigative tool: An empirical analysis," *European Journal on Criminal Policy and Research*, vol. 23, pp. 441–459, 2017.

[15] W. Hahne, "AI security cameras: What are the advantages of outdoor security cameras? June," 2022. [Online]. Available: https://www.a1securitycameras.com/.

[16] M. C. Hwang, L. T. Ha, N. H. Kim, C. S. Park and S. J. Ko, “Person identification system for future digital tv with intelligence,” *IEEE Transactions on Consumer Electronics*, vol. 53, no. 1, pp. 218–226, 2007.

[17] C. C. Chang and C. J. Lin, “Libsvm: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology (Tist)*, vol. 2, no. 3, pp. 27, 2011.

[18] L. Y. Lo, C. H. You, Y. Tang, A. S. Yang, B. Y. Li *et al.,* “Dynamic object tracking on autonomous UAV system for surveillance applications,” *Sensors*, vol. 21, no. 23, pp. 7888, 2021.

[19] C. H. Huang, Y. T. Wu, J. H. Kao, M. -Y. Shih and C. -C. Chou, “A hybrid moving object detection method for aerial images,” in *Proc. of the Pacific-Rim Conf. on Multimedia*, Shanghai, China, pp. 357–368, 2010.

[20] J. P. Škrinjar, P. Škorput and M. Furdi´c, “Application of unmanned aerial vehicles in logistic processes,” in *Proc. of the Int. Conf. New Technologies, Development and Applications*, Sarajevo, Bosnia and Herzegovina, pp. 359–366, 2018.

[21] B. Kim, H. Min, J. Heo and J. Jung, “Dynamic computation offloading scheme for drone-based surveillance systems,” *Sensors*, vol. 18, pp. 2982, 2018.

[22] Y. Liu, Q. Wang, H. Hu and Y. He, “A novel real-time moving target tracking and path planning system for a quadrotor UAV in unknown unstructured outdoor scenes,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, pp. 2362–2372, 2018.

[23] S. Wang, F. Jiang, B. Zhang, R. Ma and Q. Hao, “Development of UAV-based target tracking and recognition systems,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, pp. 3409–3422, 2019.

[24] E. Lygouras, N. Santavas, A. Taitzoglou, K. Tarchanidis, A. Mitropoulos *et al.,* “Unsupervised human detection with an embedded vision system on a fully autonomous UAV for search and rescue operations,” *Sensors*, vol. 19, pp. 3542, 2019.

[25] Y. Feng, K. Tse, S. Chen, C. -Y. Wen and B. Li, “Learning-based autonomous UAV system for electrical and mechanical (E&M) device inspection,” *Sensors*, vol. 21, pp. 1385, 2021.

[26] A. Bochkovskiy, C. -Y. Wang and H. -Y. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” *Semantic Scholar*, Arxiv 2020, Arxiv:2004.10934, vol. 5, pp. 10934–10951, 2020.

[27] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier and L. V. Gool, “Robust tracking-by-detection using a detector confidence particle filter,” in *Proc. of the IEEE 12th Int. Conf. on Computer Vision (ICCV)*, Kyoto, Japan, pp. 1515–1522, 2009.

[28] G. Ning, Z. Zhang, C. Huang, X. Ren, H. Wang *et al.,* “Spatially supervised recurrent convolutional neural networks for visual object tracking,” in *Proc. of the IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Baltimore, MD, USA, 2017.

[29] M. Siam and M. ElHelw, “Robust autonomous visual detection and tracking of moving targets in UAV imagery,” in *Proc. of the IEEE 11th Int. Conf. on Signal Processing (ICSP)*, Beijing, China, pp. 1060–1066, 2012.

[30] R. V. Aragon, C. R. Castaño and A. C. Correa, “Impact and technological innovation of uas/drones in the world economy,” in *2020 Int. Conf. on Innovation and Trends in Engineering (IEEE CONIITI 2020)*, Bogota, Colombia, 2020.

[31] J. Haugen and L. Imsland, “Monitoring moving objects using aerial mobile sensors,” *IEEE Transactions on Control Systems Technology*, vol. 24, pp. 475–486, 2015.

[32] M. Blösch, S. Weiss, D. Scaramuzza and R. Siegwart, “Vision based MAV navigation in unknown and unstructured environments,” in *Proc. IEEE Int. Conf. on Robotics and Automation*, Anchorage, AK, USA, pp. 21–28, 2010.

[33] S. A. P. Quintero and J. P. Hespanha, “Vision-based target tracking with a small UAV: Optimization-based control strategies,” *Control Engineering Practice*, vol. 32, pp. 28–42, 2014.

[34] S. Ren, K. He, R. Girshick and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.