IOP Publishing

Computational Intelligence Based Solutions for Vision Systems

Varun Bajaj and Irshad Ahmad Ansari

Chapter 1

Drone-based vision system: surveillance during calamities

Ankit Charan Janbandhu, Sachin Sharma, Irshad Ahmad Ansari and Varun Bajaj

This chapter describes how drones and computer vision can work together to enhance surveillance. A drone-based surveillance system is proposed that can be used in disaster situations for monitoring the location and the number of people present, and provide information about the area so that rescue teams can use this information to be more effective in their work. Moreover, this chapter demonstrates that the BlazeFace model is more efficient than the existing face detection algorithm in terms of computation. It also discusses the use of odometry to track location in an indoor environment and a description of centroid tracking in terms of tracking faces. The drone used in this chapter is DJI Tello, and the tools used are the DJI Tellopy library, OpenCV Python and the MediaPipe library.

1.1 Introduction

In the last few decades, surveillance and human detection through monitoring have become part of current technology. In terms of human rescue during calamities and disasters, surveillance has a wide scope for further advancements in terms of human face detection, recognization and observation in extreme situations such as floods, earthquakes, fires, dust storms or any other disasters. A drone-based rescue system needs to be effective and adequate for providing information about the ground level conditions. In the present scenario, the drone-based system is equipped with new technological advancements such as the incorporation of computer vision [1], convolution neural networks [2] and deep learning [3]. Previously, surveillance was performed using a limited approach using cameras during natural or artificial calamities. Now, the worst affected areas can be reached easily using drones, even in adverse conditions.

doi:10.1088/978-0-7503-4821-8ch1

Previously surveillance was possible only through using hot-air balloons with cameras, airplanes with a camera attached to the outer surface [4] and, a later point of time when cameras became smaller in size, a camera could be attached to a bird's body to observe a location for the objective of surveillance and rescue [5]. These conventional methods are expensive, time-consuming and can provide inaccurate information, which causes a lot of ambiguity and a waste of resources. Now drones equipped with computer vision provide an efficient drone model to tackle these challenging disaster situations in an effective manner [6–9]. By using a face detection technique and moving the drone over the target location, authentic information can be provided about the number of people present and also their location, as the drone can use its computer vision to observe difficult and impenetrable areas that are impossible for humans to reach. This system of surveillance has the power to tackle calamities and disasters in a systematic manner help people be rescued as fast as possible.

During a period of disaster, this drone-based computer vision could make a significant difference in terms of the allocation of human resources. For example, during the 2013 the Uttrakhand floods in India drones provided a clear picture of the natural disaster, and thousands of people were found using drones for the first time in Indian search and rescue operations [10]. The rescue of people was the major goal of the government during the mountain tsunami in Uttrakhand. Using a drone-based vision system, surveillance and recovery were achieved by providing immediate help to the people facing disaster. As the drone detects faces, it can provide exact numbers. The Kerala floods in 2018 also demonstrated the effective use of drones [11] in hidden places where helicopters and humans could not reach. By using drone-based computer vision during a flood, tsunami, earthquake, fire or other calamity one can also obtain information about the number of people present at the location.

In this chapter the focus will be on the two primary components: the drone and computer vision. This chapter will also elaborate on the objectives of face detection and counting the number of affected people using drones.

1.2 Surveillance system

Surveillance systems have evolved in the last two decades, in which time video surveillance has, in general, transitioned towards being a larger part of society. Although there are still many debates regarding the ethics of video surveillance technology, recent advancements in technology have led to the use of this technology in many different parts of society. Companies use video surveillance to provide security and monitor employees, monitor customer activities, and handle large crowds to reduce crowding and power consumption. Video surveillance systems have been around for a long time and they have evolved from black and white cameras to color cameras, to even the ability to be fired from a drone.

The 1990s was a decade of many societal and technological changes. One of these was the emergence of widespread video surveillance systems. During World War II, Great Britain and France faced a desperate shortage of weapons after Germany had seized control of much of Eastern Europe. Britain's only means of response was to make some changes to its military, to make sure that it was ready to defend itself. One of those changes was to put cameras on the back of tanks. The tanks were fitted with cameras to make sure that the British troops would be able to see where they were going. The cameras were also used to help British troops avoid being shot by the Germans. The Germans in turn used closed-circuit television (CCTV) technology to monitor the movements of soldiers, tanks and other military equipment to ensure that German troops were following the correct procedures and to prevent any harm being done to civilians due to false information [12, 13]. A further development occurred after the war, in the 1960s, when the number of police officers participating in law enforcement was falling rapidly. One of the ways to create a safe environment for police officers was by using CCTV surveillance equipment. In the 1980s, generally large factory buildings or quiet residential areas were chosen as possible sites for multi-story CCTV surveillance systems.

Today there are many different types of CCTV cameras in homes and businesses. As cameras become more sophisticated, they prevent the threat of burglary. Facial recognition surveillance utilizes data obtained from video surveillance cameras to automate the identification of individuals. It is often used in crime-fighting applications, but it can be used by people in everyday life as well. Typically, the systems use image-recognition algorithms to identify people in video footage. CCTV cameras can collect images of intruders without having to shoot them or have other devices involved. Video surveillance is a valuable technology for anyone who needs to create a secure shop. The main benefits of video surveillance include the ability to monitor the entrance and exit points of your shop. It also gives you new options when it comes to securing sensitive areas and the online presence of a shop. Video surveillance is very easy to set up and can be used in different ways. You can install a full-body-style camera and use the image as a sales tool. There is nothing new about video surveillance systems and still we continue to see advancements in the field every year.

1.2.1 The importance of surveillance systems

The purpose of computer vision is to provide an understanding of the physical world in order to perform necessary tasks in the environment. Computer vision is involved in surveillance, video surveillance, surveillance cameras and other applications. The video surveillance industry has grown dramatically over the past few years. It is not uncommon for companies to install video surveillance systems to protect their businesses from theft of confidential data, robberies and vandalism of their property. However, the technology of video surveillance is becoming more and more sophisticated and can now identify images and audio of unusual movements of people and vehicles. These include people on foot, cars on the street, and people moving into and out of buildings. With the advancing technology of these systems it has become possible to identify individuals or vehicles on video as well as identify crime. There are various fields in which surveillance systems play an important role. Some of them are discussed below.

1.2.1.1 Retail

A shop owner can use video surveillance equipment to help him/her keep an eye on the customers coming into his/her shop, as shown in figure 1.1. The video surveillance system can be installed in different ways depending on the needs of the store owner. There are many different types of surveillance systems that are used in retail. From security cameras positioned on store walls, to scanners that can be used to check the bags of customers, many different types of surveillance systems are used to help retailers and individual employees safeguard their businesses.

1.2.1.2 Agriculture

In the ever-expanding world of agriculture, surveillance cameras have become an essential tool, as shown in figure 1.2. The free-market economy has given rise to the idea that agricultural production is possible without limits or limits on who can do it. If you are looking to start a farm, the benefits of implementing surveillance cameras are many. With the great number of cameras available on the market today, it is easy to install them with little to no up-front cost. A surveillance camera is an automated or robotic device that records images of events occurring on or near the farm. It is used to monitor farm animals, the movement of livestock, disease infestation, agricultural activities, accidents and other farm activities. The use of surveillance cameras also allows the quality of air and water to be monitored, and the environment to be preserved.

1.2.1.3 Private households

Surveillance cameras are now installed in many homes. Not everyone has access to them, but many people can access their cameras remotely. They are widely used for security purposes, as shown in figure 1.3. These cameras are designed to capture unwanted events, such as theft, vandalism, noise, accidents and so on.

Figure 1.1. A shop under video surveillance.

Figure 1.2. Drone surveillance of a farm.

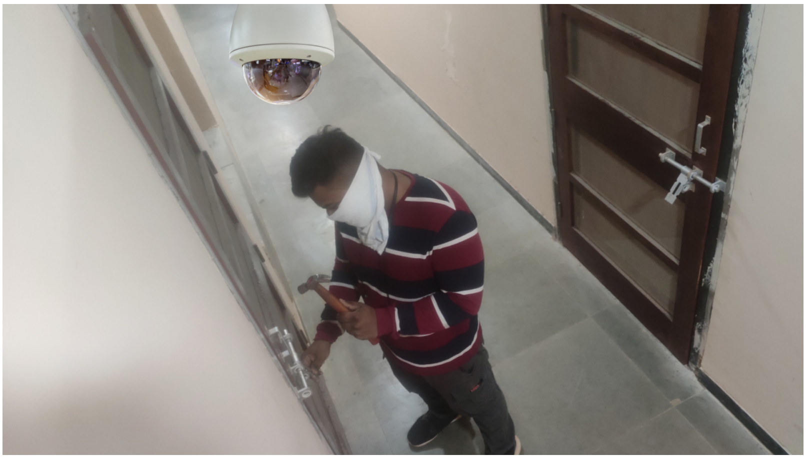


Figure 1.3. A house under CCTV surveillance.

1.2.1.4 Public places

When it comes to public safety, surveillance cameras are very useful for security and law enforcement, as shown in figure 1.4. It is becoming increasingly common for us to see those cameras on our street corners, on our local shopping centers, or even on our local bus. Surveillance cameras are used to catch robbers prowling the streets, criminals committing crimes and speeding vehicles.

Figure 1.4. A street under CCTV surveillance.

1.2.2 The use of drones in surveillance system

The recent buzz around artificial intelligence (AI) and the rise of the Internet of Things (IoT) have inspired many to start experimenting with drones. Drone technology has advanced tremendously over the last few years. Today, robots are being developed to assist humans with everyday tasks. For example, an AI-powered drone could be used to harvest crops or deliver products to consumers. Surveillance is a major concern today and drones can be used to fulfill that role. There are many different types of surveillance systems you can use with drones [14], but the three most common are umbrella coverage, perimeter coverage and ground-penetrating radar (GPR).

During the past two years, the booming drone industry has grown enormously. Drones are becoming increasingly popular in dangerous situations. They are now being used to save lives and provide humanitarian assistance to people in need, as shown in figure 1.5. The military uses drones to get information about people who are kidnapped by terrorists. When it comes to aerial photography, drones are often considered an excellent alternative to fixed cameras. The advantages of using drones include their wide perspective, the possibility for high-resolution photographs and their low cost of operation. With the use of AI and computer vision technology, many drones are now used in hazardous areas where people may need help. With the use of intelligent unmanned aerial vehicles, the number of people trapped in a critical situation can be detected and counted and suitable aid can be provided in those areas.

1.3 Proposed method

This section has the primary objectives of developing an understanding of the various computer vision methods for detecting humans, and to create a system that automatically counts the people in a chaotic environment such as a natural disaster. A DJI Tello drone has been used to test the code in this work.

Figure 1.5. The use of a drone in surveillance systems.

The topics listed below will be addressed in the following sections of this chapter:

- Section 1.3.1. Detecting human faces—An overview of the various facial recognition methods available to date is given in this section, as well as the methods used to accomplish this task and their advantages.
- Section 1.3.2. Tracking human faces—This section explains how to trace a face in an image and why this is so important.
- Section 1.3.3. Locating and capturing human faces—The purpose of this section is to demonstrate how to locate a drone in an indoor environment and how to store images of different faces in a database while performing the drone's location.
- Section 1.3.4. Counting the number of people—The goal of this section is to demonstrate how to count the number of people in a frame as well as how tracking can assist us in counting people and tracking their locations.
- Section 1.3.5. Drone deployment and testing—This section shows how to deploy our code in DJI Tello, which will be compared with some test results to check the effectiveness.

1.3.1 Detecting human faces

Face discovery is a computer technology that identifies the face and dimensions of a human face in an electronic image. The face attributes are found and also any other items such as trees, buildings and bodies are disregarded from the electronic image. It can be regarded as a specific instance of object-class detection, where the work is finding the location as well as sizes of all items in an image that are related to that class. Face detection can be seen as a more general situation of face localization. In face localization the task is to identify the locations and sizes of a recognized number of faces (typically one). Primarily, there are two types of approaches to discover faces in a given electronic image—attribute based and photograph based techniques.

The feature based method tries to extract features from the image and match them with knowledge of facial features. In contrast, the picture-based strategy attempts to obtain the best match between the training and testing photos. Various methods are available for identifying faces from a still photo or video.

As illustrated by figure 1.6, various methods are available for facial recognition [15], but the image-based method is employed in this study. Drones are used for this task and they have low computational power, so we need a method that is computationally friendly.

The new BlazeFace framework provided by Google [16] allows one to detect faces with greater accuracy. BlazeFace is a lightweight model for seeing faces in images and is an adaptation of a single-shot detection (SSD) method [17].

The BlazeFace method detects faces in an image and creates a bounding box around each face to indicate its location in the frame. It also produces six facial keypoint coordinates (for the eye centers, ear tragions, mouth center and nose tip) that allow us to estimate face rotation.

The BlazeFace model uses the depthwise separable convolution method [18]. The benefit of this method over the standard convolution method [19] can be explained by the following example. If an input image is 32 × 32 × 3 (the three values here refer to channels, as a color image has three channels: red, green and blue). Thus, to detect edges or some other feature, if the size of the kernel is taken as 5 × 5 × 3 and it is convolved with the image and slid over the image, it gives an output that has a size of 28 × 28 × 1. The total number of multiplications performed to obtain this result is $5 \times 5 \times 3 \times 28 \times 28 = 98\,000$. For example, to detect 64 features of the image, a 64, 5 × 5 × 3 kernel has been used, so now the outcome will have 64, 28 × 28 × 1 image stacks back-to-back, as shown in figure 1.7. Now the total number of multiplications

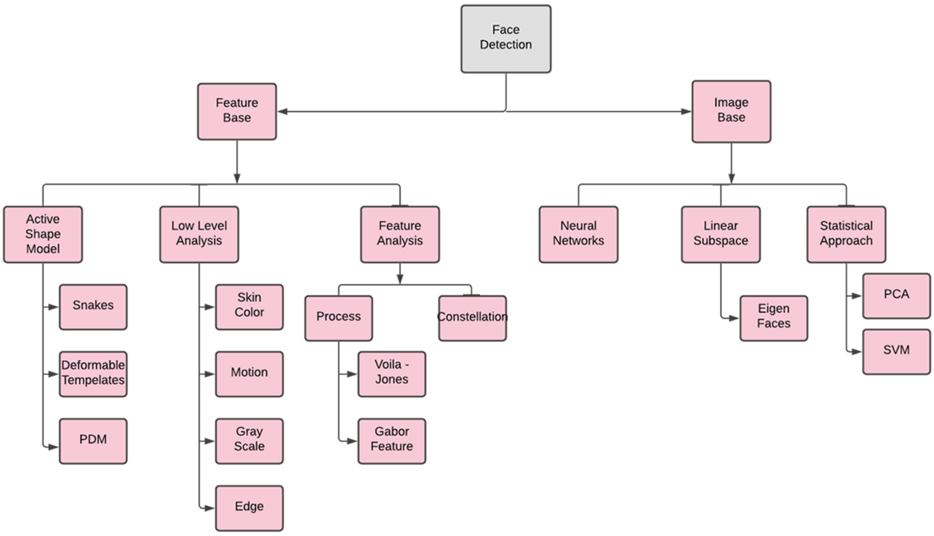


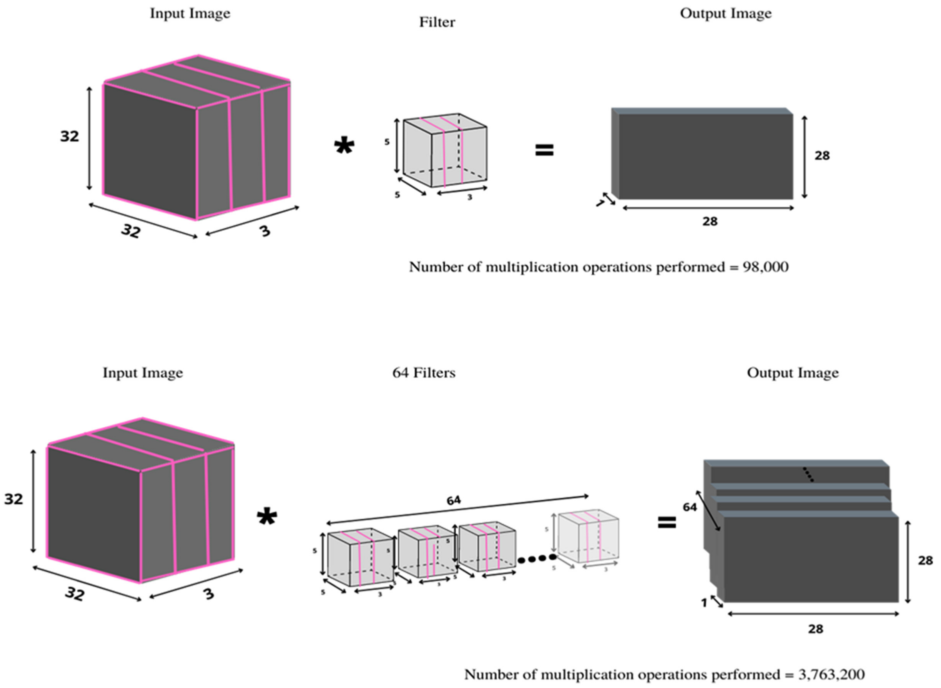
Figure 1.6. Face detection methods.

performed is 64 × 5 × 5 × 3 × 28 × 28 = 3 763 200 operations, so it can be seen that traditional convolution calculations need high computational power.

Usually, the filters for all input channels are applied in one step and then combined at the same time, while depthwise separable convolution [20] is applied in two steps:

1. *Depthwise convolution (at the filtering stage).*

   Unlike standard convolution, depthwise convolution applies convolution to a single input channel at a time. For example, if an input image of size 32 × 32 × 1 is used for depthwise convolution, a 5 × 5 × 1 shape kernel has been applied. That is why it is necessary to have three such kernels. After convolution its output image will be of size 28 × 28 × 3. The number of multiplications will be 5 × 5 × 3 × 28 × 28 = 58 800 operations. This concludes the first phase. Now this will be succeeded by pointwise convolution, as shown in figure 1.8.
2. *Pointwise convolution (at the combination stage).* Pointwise convolution involves performing a linear combination of each of these layers. Here, the input is an image of size 28 × 28 × 1. The filter has the shape 1 × 1 × 3. This is a 1 × 1 operation over all three layers. If there is a need to detect 64 features, then 64 such filters are needed to see the different features, just as was done in standard convolution. It gives an output of an image size of 28 × 28 × 1, and there are 64 such images stacked together. Now the number of multiplications performed is 64 × 1 × 1 × 3 × 28 × 28 × 1 = 150 528.



Number of multiplication operations performed = 98,000

Number of multiplication operations performed = 3,763,200

Figure 1.7. Convolution using traditional methods.

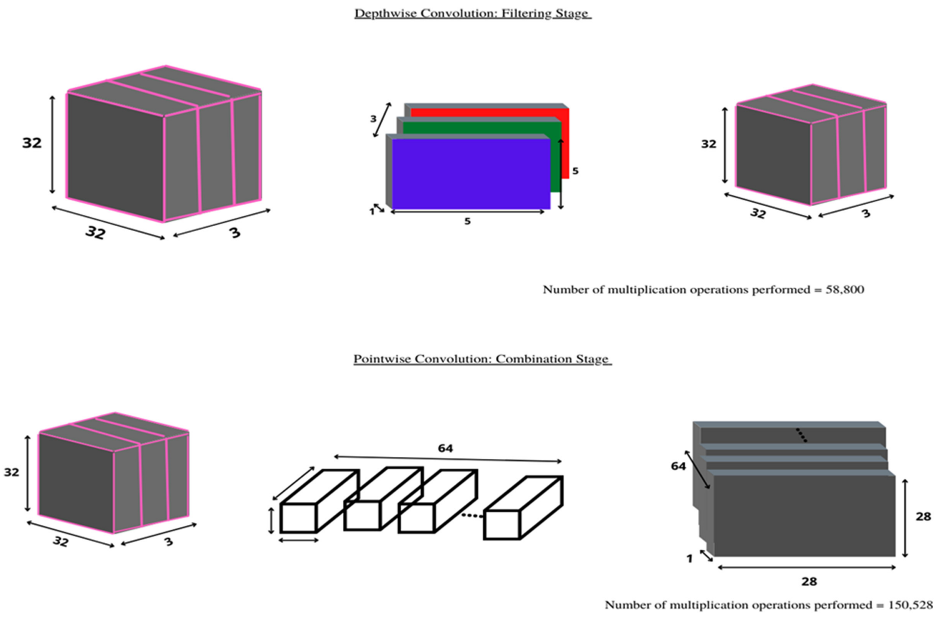

Depthwise Convolution: Filtering Stage
32
32
3
3
5
1
5
32
32
3
Number of multiplication operations performed = 58,800
Pointwise Convolution: Combination Stage
32
32
3
64
64
28
1
28
Number of multiplication operations performed = 150,528

Figure 1.8. Convolution using depthwise separable convolution.

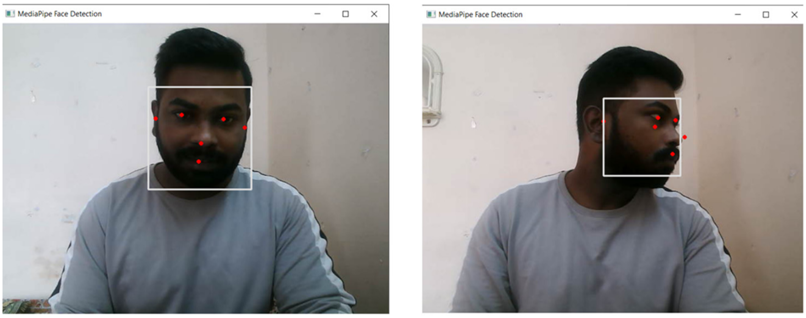
Now the total number of multiplications performed is (depthwise convolution stage + pointwise convolution stage) 58 800 + 150 528 = 209 328 multiplications, which is lower compared to the standard convolution, in which 3 763 200 operations were required. Thus the BlazeFace model uses depthwise separable convolution so that it will perform fewer calculations and thus perform well in low-energy computational devices.

The BlazeFace model is adapted from SSD but there is a difference. The BlazeFace model is specially designed to detect faces and the SSD is used to detect objects having a lot of variance. The makers of the BlazeFace model claims that there is limited variance in human faces, i.e. every face has eyes, one nose and one mouth, so the number of anchors is reduced. Instead of using 4 × 4 and 2 × 2 feature map sizes, they have reduced the architecture which stops at an 8 × 8 feature map without further downsampling and use 2 + 2 + 2 = 6 anchors of 8 × 8, which also provides extra speed to this model.

It can see in figure 1.9 how the author’s face is detected by the BlazeFace model and also how the six landmarks are projected. This was done using the Google MediaPipe library which uses the BlazeFace model for face detection.

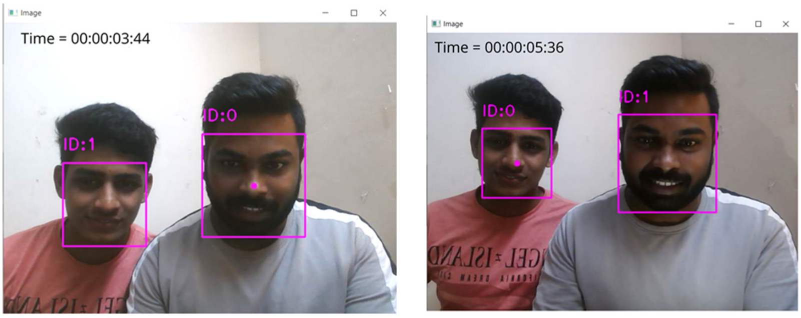
1.3.2 Tracking human faces

Although the tracking of faces is an essential part of the application, the MediaPipe library does not support active tracking, as can be seen in figure 1.10. For two people in the same image, it assigned two different IDs at two different time intervals. For this work, the faces’ IDs must remain unique in the camera’s frame. This is the first



Location : Power & Control Lab (PDPM IIIT Jabalpur)

Figure 1.9. Face detection using the MediaPipe library.



Location : Power & Control Lab (PDPM IIIT Jabalpur)

Figure 1.10. Proposed system without the centroid tracking algorithm.

challenge, and the second is that the camera might assign a different ID to the same face when it comes back into the frame after the drone moves. When the drone moves it is desirable that an ID be assigned to the face if it is in the frame, and if it disappears from the frame the drone it will wait for a specific number of frames. If the face reappears in this period it will again be assigned the same ID.

For solving this problem, a centroid tracking algorithm is used. It is a multi-step process, as shown in figure 1.11. The following steps are used to perform the detection and tracking of human faces using the centroid tracking algorithm.

Step 1. Calculating the centroid of the bounding box.

By using the BlazeFace model, the information of the bounding box can be traced and by applying this information the calculation of the centroid is completed, as shown in figure 1.12.

Step 2. Computing the Euclidean distances between the new bounding box and existing faces.

First, the Euclidean distance is calculated for every subsequent frame of the video stream. However, instead of assigning a unique ID to the object, first the centroid is calculated from the boundary box. This is followed by calculating the Euclidean

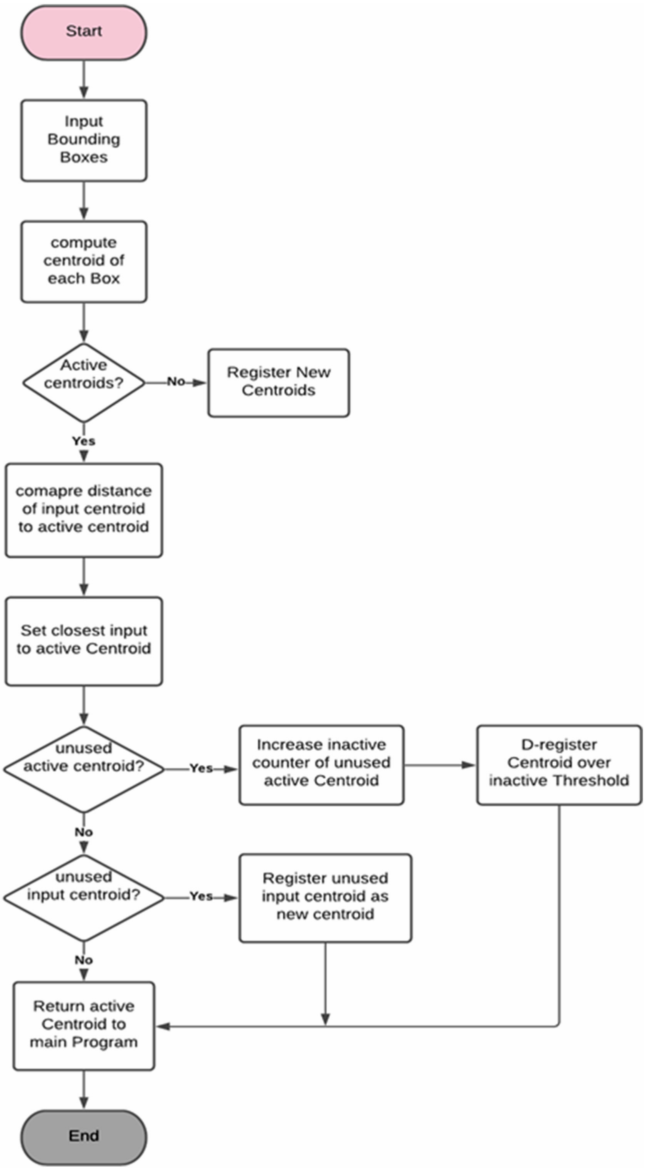

Start
Input Bounding Boxes
compute centroid of each Box
Active centroids?
No
Register New Centroids
Yes
comapre distance of input centroid to active centroid
Set closest input to active Centroid
unused active centroid?
Yes
Increase inactive counter of unused active Centroid
D-register Centroid over inactive Threshold
No
unused input centroid?
Yes
Register unused input centroid as new centroid
No
Return active Centroid to main Program
End

Figure 1.11. Flow chart of the centroid tracking algorithm.

distances between each pair of existing object centroids and the input object centroid. The same ID of the old object will be assigned if the distance between the old object and the newly created object is significantly less. The primary idea behind this is that if the object moves a short distance from the previous centroid (here ‘object’ refers to a face), it does not suddenly appear at a different location. That is why it is given the same ID for which the Euclidean distance between the

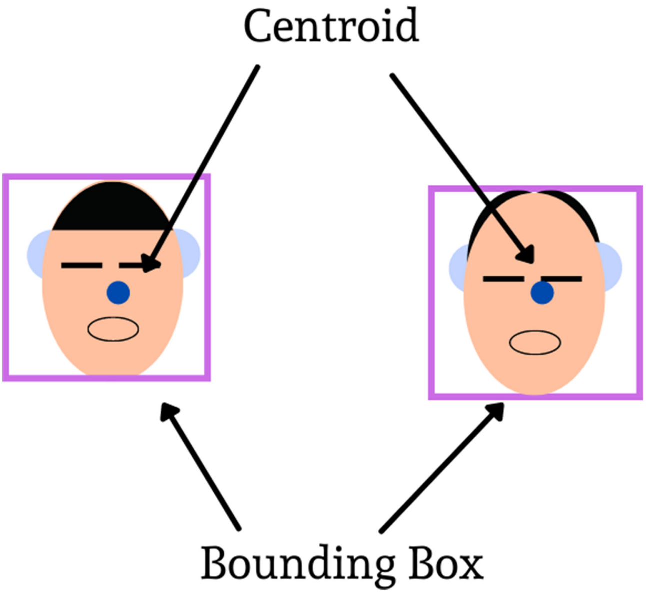


Figure 1.12. Centroid calculation using a bounding box.

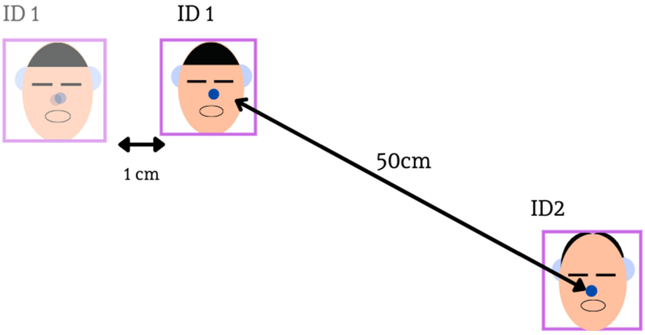


Figure 1.13. Euclidean distance calculation between the bounding boxes.

objects is very small, as shown in figure 1.13, where the face is shifted by 1 cm, but its ID is the same as the one which is closer to its new location.

Step 3. Registering of a new object.

Whenever a new face appears in the frame, to add that face to our tracking list we first assign a unique object ID, as shown in figure 1.14. We then calculate the centroid of the bounding box, store this information and repeat the process.

Step 4. Deregistering of an old object

The old object is deregistered if it cannot be matched with any existing face for *N* subsequent frames, as shown in figure 1.15. When a person leaves the frame, the algorithm keeps assigning the old IDs to the remaining people.

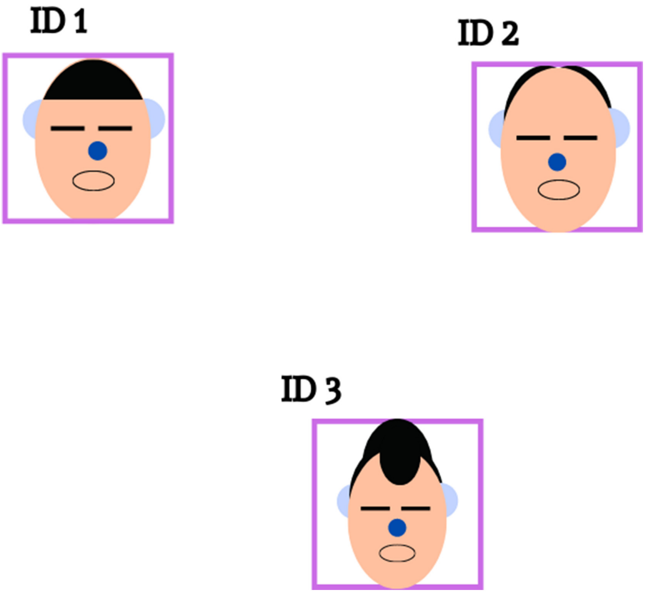
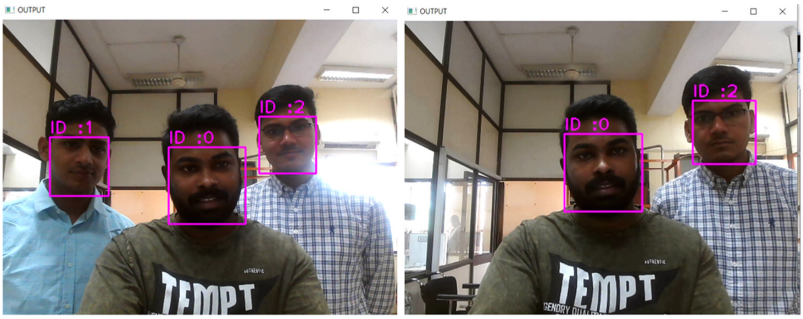


Figure 1.14. Assigning new IDs to new faces in the frame.



Location: Power & Control Lab (PDPM IIIT Jabalpur)

Figure 1.15. Deregistering of faces from the frame.

1.3.3 Locating and capturing human faces

Thus far the detection of faces and tracking them in the frame have been discussed. However, this section will demonstrate how to capture a face in the image and locate where it was taken.

A global positioning system (GPS) can be utilized to determine the drone's coordinates. However, for this work DJI's Tello is employed, which lacks GPS functionality, and if the operations are performed indoors GPS will not function. Thus, for estimating the drone's position, velocity data are used, and its angle is determined by the angular speed data; this procedure is called odometry. Using this information, the position of the drone can be calculated relative to its take-off position. There have been other advanced methods for estimating position [21–24], but in this case, odometry is employed.

For example, if the drone is at its origin, then given a velocity of 20 cm s^{-1} and a time of 1 s, after which a measurement of distance is taken, the distance is 20 cm, and a grid is created, as shown in figure 1.16.

Now if the velocity is fixed and by knowing how many seconds the forward button is pressed one can easily estimate the position of the drone relative to the take-off position.

When the drone is provided with only an angular velocity of 45° s^{-1}, it will head 45° away from the original position. If both a fixed forward velocity and a fixed angular velocity are provided, for example 20 cm s^{-1} and 45° s^{-1} for 2 s, as shown in figure 1.17 after 1 s the drone is at (14.1, 14.1) relative to its original position with a heading angle of 45°. After another second it is at (28, 28.2) relative to its original position with a heading angle of 90°.

In this way the drone position is estimated and this information is used in naming the image. Thus, one can see where a person is in relation to the take-off position. Result of which can be seen from figure 1.22.

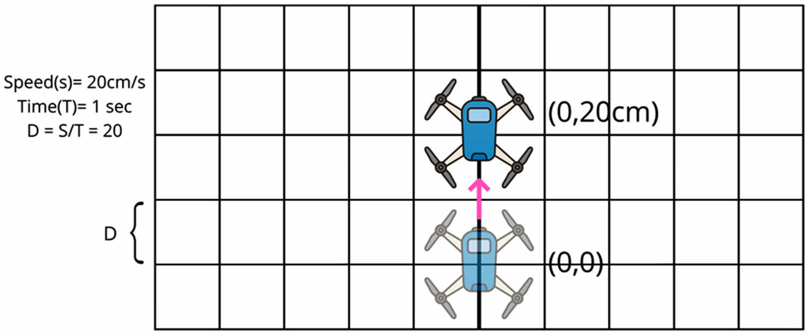


Figure 1.16. Grid cell localization.

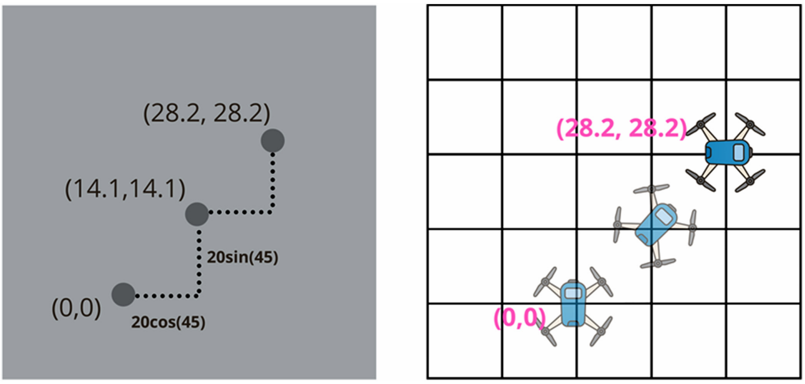
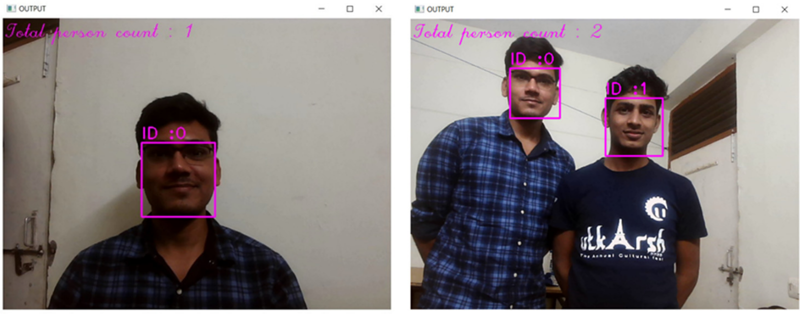


Figure 1.17. Estimating the position of drone using odometry.



Location : Power & Control Lab (PDPM IIIT Jabalpur)

Figure 1.18. Counting the people in the frame.

1.3.4 Counting the number of people

The information about how many unique IDs is generated is used in this work to count the number of people. As depicted in figure 1.18, in the left image, when there is one person in a frame ID 1 is assigned and the total person count is 1. When another person comes in ID 2 is assigned and the total person count increases to 2. So just by counting the number of unique IDs one can estimate the number of people.

1.3.5 Drone deployment and testing

Throughout this work a DJI Tello drone has been used, and for the deployment of all the methods discussed above the DJI Tello drone is programmed using the DJI Tello Python SDK. Figure 1.19 shows the flow chart of the system.

Figure 1.20(a) shows how the DJI Tello drone is locating the number of people in the frame and also assigning a unique ID to the people in the frame and, by counting the number of unique IDs assigned, it displays the number of people counted in the entire mission. It also displays the path and location where it is in reference to the take-off point, as shown in figure 1.20(b) so that rescuers know where the people are and can plan their rescue mission accordingly. It also saves the facial image of the tracked people, as shown in figure 1.20(c), so that the rescue team can make sure that all the people have been saved.

1.4 Conclusion

In this chapter an efficient surveillance system using a drone bases system is proposed and implemented. During disaster situations, the system will efficiently search areas that are beyond the reach of humans, as well as calculate the number of people trapped in these areas and their locations, so that the rescuers can perform their operations efficiently and provide first aid to these people. In addition to our system's performance during natural disasters, it can also be utilized for military surveillance to track a particular area and send useful intelligence to the military.

The system is deployed on a drone and the communication with the drone from the ground station is performed using Wi-Fi. As the distance between the drone and

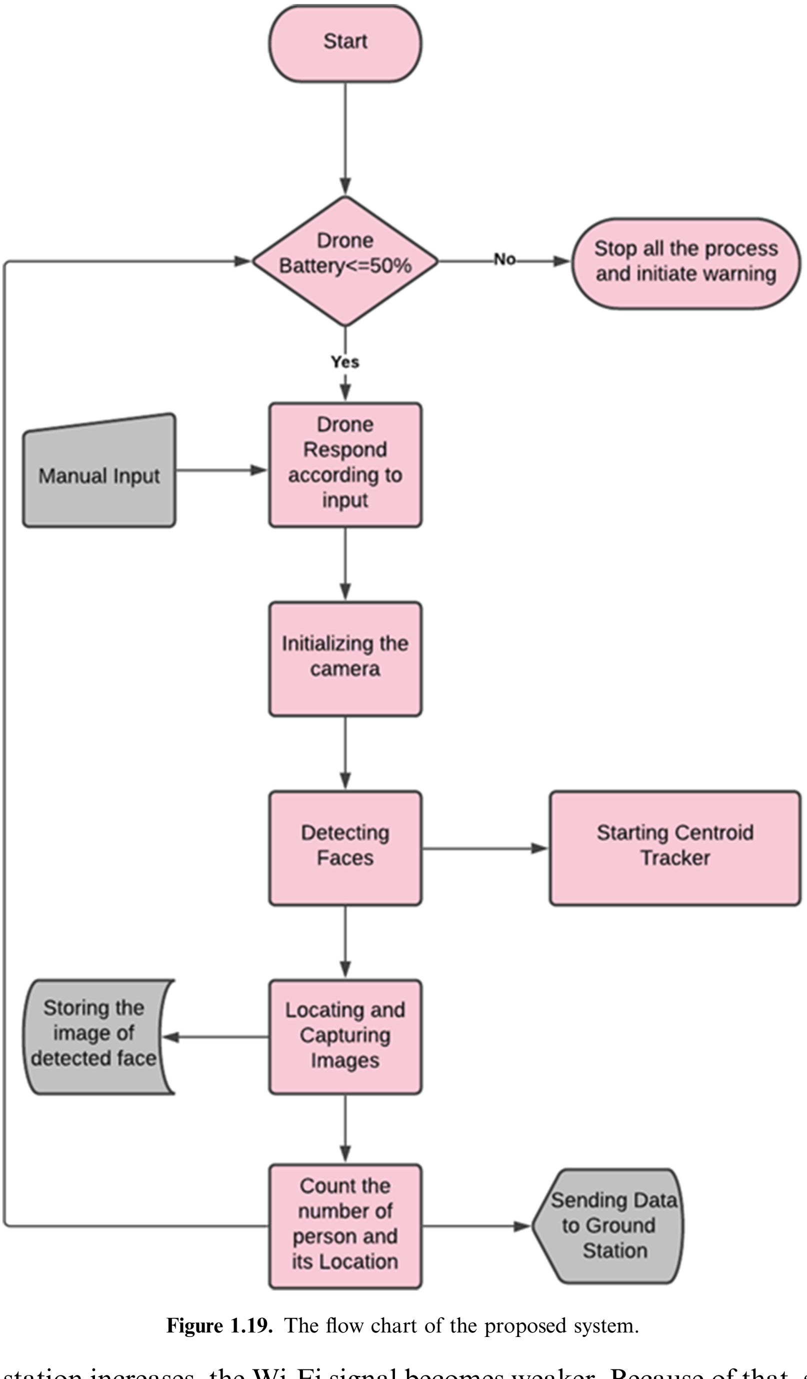
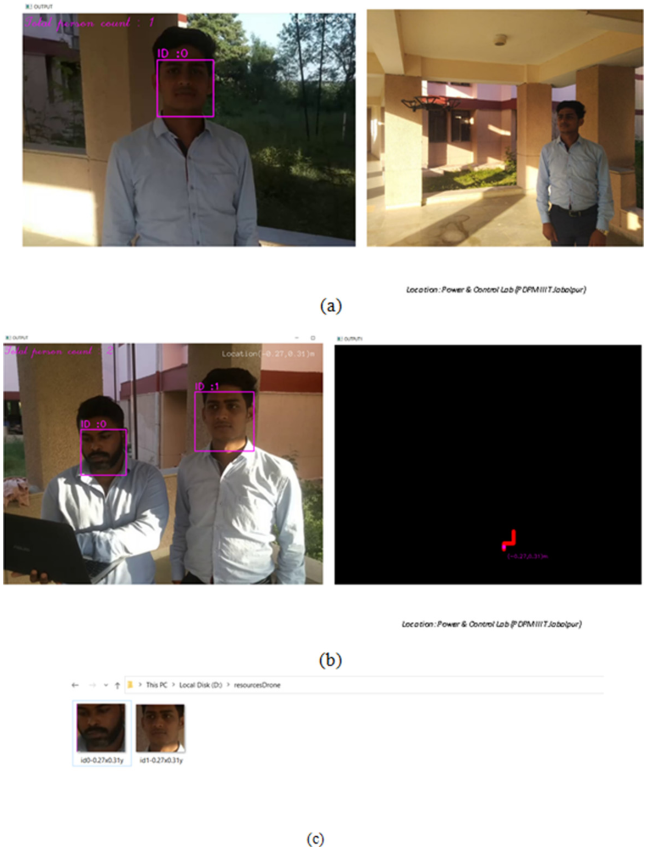

Start
Drone Battery<=50%
No
Stop all the process and initiate warning
Yes
Manual Input
Drone Respond according to input
Initializing the camera
Detecting Faces
Starting Centroid Tracker
Storing the image of detected face
Locating and Capturing Images
Count the number of person and its Location
Sending Data to Ground Station

Figure 1.19. The flow chart of the proposed system.

ground station increases, the Wi-Fi signal becomes weaker. Because of that, some of the frames can be missed, which will cause an error in counting the number of people. This can be rectified by using Wi-Fi boosters or by using the LoRaWAN technology for communication between drones and ground stations.



Location : Power & Control Lab (PDPM IIIT Jabalpur)

(a)

Location : Power & Control Lab (PDPM IIIT Jabalpur)

(b)

(c)

Figure 1.20. Results of the proposed system. (a) Localization and counting of people performed by the proposed system. (b) Tracing the path of the drone. (c) Saving images of people to the database.

Acknowledgements

This research work was supported by Jagadish Chandra Bose Research Organisation (JCBRO).

References

[1] Yang J and Li J 2017 Application of deep convolution neural network *14th Int. Computer Conf. on Wavelet Active Media Technology and Information Proc. (Chengdu)* pp 229–32

[2] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)

[3] Forsyth D and Ponce J 2011 *Computer Vision: A Modern Approach* 2nd edn (Englewood Cliffs, NJ: Prentice Hall) p 792

[4] Browning W M, Olson D S and Keenan D E 1999 High-altitude balloon experiment *Proc. SPIE* **3706** 187–95

[5] Wilkinson J 2013 Animalizing the apparatus: pigeons, drones and the aerial view *Grad. J. Vis. Mater. Cult.* **6** 1–21

[6] Mishra B, Garg D, Narang P and Mishra V 2020 Drone-surveillance for search and rescue in natural disaster *Comput. Commun.* **156** 1–10

[7] Jain A, Basantwani S, Kazi O and Bang Y 2017 Smart surveillance monitoring system *Int. Conf. on Data Management, Analytics and Innovation (ICDMAI) (Pune)* pp 269–73

[8] Menezes V, Patchava V and Gupta M S D 2016 Surveillance and monitoring system using Raspberry Pi and Simple CV *Proc. 2015 Int. Conf. Green Comput. Internet Things* pp 1276–8

[9] Alajrami E, Tabash H, Singer Y and Astal M T E 2019 On using AI-based human identification in improving surveillance system efficiency *Proc.—2019 Int. Conf. Promis. Electron. Technol.* pp 91–5

[10] Rawat M S and Dobhal R 2021 Study of flash flood in the Rishiganga and Dhauliganga Catchment in Chamoli District of Uttarakhand, India *Int. J. Georesources Environ.* **6** 84

[11] Ummer O, Scott K, Mohan D, Chakraborty A and Lefevre A E 2021 Connecting the dots: Kerala's use of digital technology during the COVID-19 response *BMJ Glob. Heal.* **6** e005355

[12] Norris C and Armstrong G 2020 *The Maximum Surveillance Society: the Rise of CCTV* (London: Taylor and Francis)

[13] Kroener I 2016 *CCTV A Technology Under the Radar?* (London: Routledge)

[14] Custers B (ed) 2016 *The Future of Drone UseInformation Technology and Law Series* vol 27 (The Hague: Asser)

[15] Kumar A, Kaur A and Kumar M 2019 Face detection techniques: a review *Artif. Intell. Rev.* **52** 927–48

[16] Bazarevsky V, Kartynnik Y, Vakunov A, Raveendran K and Grundmann M 2019 BlazeFace: sub-millisecond neural face detection on mobile GPUs ArXiv:1907.05047

[17] Liu W *et al* 2016 SSD: single shot multibox detector *European Conference on Computer VisionLecture Notes in Computer Science* vol 9905 (Berlin: Springer) pp 21–37

[18] Chollet F 2017 Xception: deep learning with depthwise separable convolutions *Proc.—30th IEEE Conf. Comput. Vis. Pattern Recognition* vol 2017 pp 1800–7

[19] Traore B B, Kamsu-Foguem B and Tangara F 2018 Deep convolution neural network for image recognition *Ecol. Inform.* **48** 257–68

[20] Yan W, Liu T, Liu S, Geng Y and Sun Z 2020 A lightweight face recognition method based on depthwise separable convolution and triplet loss *39th Chinese Control Conf. (CCC) 2020* pp 7570–5

[21] Suleiman A, Zhang Z, Carlone L, Karaman S and Sze V 2018 Navion: a fully integrated energy-efficient visual-inertial odometry accelerator for autonomous navigation of nano drone *IEEE Symp. VLSI Circuits (Honolulu, HI)* pp 133–4

[22] Nist D and Bergen J 2004 Visual odometry *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (Washington, DC) vol 1 p I-1

[23] Kozák V and Pivo T 2021 Robust visual teach and repeat navigation for unmanned aerial vehicles *European Conf. on Mobile Robots (Bonn)* pp 1–7

[24] Kamsani M N L and Mohd M N 2021 Implementation of deep learning and motion control using drone *J. Electr. Electron. Eng.* **2** 57–68