Title : Real Time Human Detection In Post of Conflict Area (UNIFIL) using UAV Platform

Group 3: Qiao Ren, Jiahe Wang, Intan Antasari, Sheeba Lawrence

Introduction

The United Nations Interim Force in Lebanon or UNIFIL is a demilitarized zone created by the UN. Following the conflict between Lebanon and Israel, the UNIFIL was created in 1978. After the 2006 Lebanon war, the UN enhanced their mandates. One of their main mandates is to provide humanitarian aid in the conflict area. Lebanon is still one of the top ten conflict regions for the year 2018.

Wars have many side effects on the environment and most importantly on human beings. Rescuing the injured humans post-conflict has been one of the toughest jobs due to many reasons. Recent developments in the fields of Unmanned Aerial Vehicles have paved the way for human search and rescue during post-disaster and conflicts. Human detection is crucial for the rescue operation. With the help of optical cameras and thermal cameras, the precise location of the humans can be detected in real time. When the precise locations of the humans are known, troops can be deployed to carefully relocate the humans to safety in an effective and efficient way.

Objective

To develop the algorithm which is used to identify the human body in the post-conflict zone using the UAV platform by acquiring visual and thermal imageries.

Role

We are a Research Institute who has a fleet of UAVs specifically designed for search and rescue operations. Our primary goal is to create a near perfect CNN algorithm for detecting humans both at daytime and night time since the rescue operation will be carried out 24/7. We also aim at providing a short course on how to use the drones for the rescue mission and how to interpret the dot map (which will be our final deliverable) to the troops. Our work plan can be seen in (dia)

Once the algorithm is created it will be tested and validated. The algorithm will be specifically made keeping in mind the local conditions of Lebanon. This algorithm will be handed out to the UN. A training program of one month will also be arranged to teach the volunteers how to operate the UAVs. We will also be actively helping the UNIFIL to build up the entire infrastructure for this mission.

Requirement

From the objective, our project needs to acquire real-time images which can be used to detect human. UAV platform can acquire images in a fast way. An unmanned aerial vehicle (UAV), commonly known as a drone, is an aircraft without a human pilot aboard. UAVs are a component of an unmanned aircraft system (UAS); which include a UAV, a ground-based controller, and a system of communications between the two. The flight of UAVs may operate with various degrees of autonomy: either under remote control by a human operator or autonomously by onboard computers. UAS are nowadays becoming more and more popular in several applications. Better batteries, miniaturized payload and, in general, affordable prices, has made available small UAVs with very good performances. These technologies can bring benefits also to the search and rescue field.

Not only visual images from UAV, but thermal images have to be acquired. Thermal images is helpful to detecting human. Thermal imaging can be seen as a method of improving visibility of objects by detecting the objects' infrared radiation and creating an image based on that information. In nature, all objects radiate infrared light, so using the detector to measure the infrared difference between the target itself and the background, you can get different infrared images, called thermal images. The thermal image of the same target is different from the visible light image. It is not a visible light image that can be seen by the human eye, but an image of the target surface temperature distribution. Or it can be said that it is that the human eye cannot directly see the surface temperature distribution of the target, but becomes a thermal image representing the target surface temperature distribution that can be seen by the human eye. Using this method, it is possible to image and measure the long-distance thermal state of the target, and to perform intelligent analysis and judgment.

Infrared thermal imaging technology is a passive infrared night vision technology. Its principle is based on the fact that all objects in the natural world above the absolute zero (-273 °C) radiate infrared rays at all times, and the infrared radiation is carried at the same time. The characteristic information of the object provides an objective basis for using infrared technology to discriminate the temperature level and heat distribution field of various measured objects. By using this characteristic, after the photoelectric infrared detector detects the power signal radiated from the heating part of the object into an electrical signal, the imaging device can simulate the spatial distribution of the surface temperature of the object one by one, and finally process the system to form a thermal image video. The signal is transmitted to the display screen, and a thermal image corresponding to the heat distribution on the surface of the object, that is, an infrared thermal image is obtained.

In our project, we are going to use Dji Phantom 4 (Figure A) as the UAV platform and DJI GS Pro (Figure C) as the corresponding communication system between the UAV and controller. The parameter of Dji Phantom 4 shows below (Figure B).

Figure A Dji Phantom 4

Dji Phantom 4 achieves 28 minutes flight time, which is a good choice for our project. Besides, we used FLIR Vue ™ Pro 336 thermal camera (Figure D) to acquire the thermal images. It is preconfigured to generate an optimal thermal image for drone applications. This thermal camera is light enough and will not affect the drone's center of gravity and will have little to no impact on flight time. With simple power-in and video-out connections, the thermal camera is easy to integrate on almost any platform. The parameter shows below (Figure E).

Figure D FLIR Vue ™ Pro 336 thermal camera

Approach and Conceptual Design

1. Why we choose CNN?

Convolutional Neuron Network (CNN) is a highly efficient method, compared with other traditional methods like HOG (histogram of gradient). Because traditional methods use a dictionary to store all the human in the training data. It has two drawbacks. Firstly, its speed is slow, because it compares with all the objects in the dictionary with the unknown object in the image. Secondly, if the algorithm sees a human which does not exist in the dictionary, the it is not able to detect it. In contrast, CNN uses feature extraction to capture the core of human shape.

2. What is CNN?

CNN in object detection consists of two main processes: feature extraction and classification. Feature extraction aims at extracting the distinctive features from images, such as a line, a curve, a tail and a nose. Classification means to compute the degree of certainty of whether an unknown object belongs to a certain category. If the degree of certainty is high, then the outcome will be it belongs to a class. An example of the CNN structure can be seen in the example of figure 1.

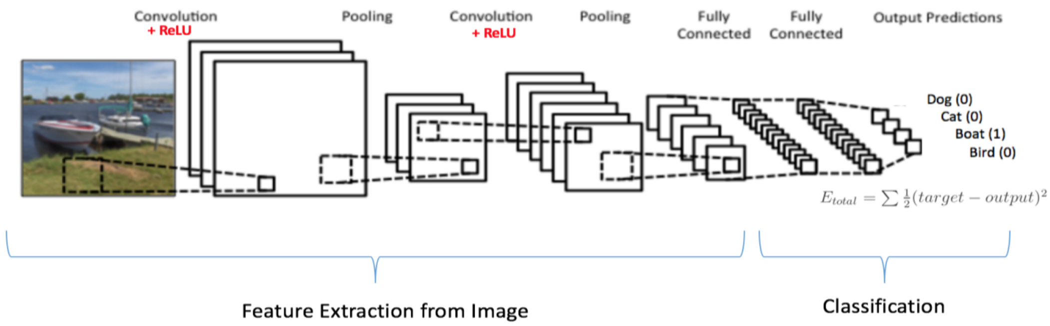

Figure 1 CNN architecture of detecting dog, cat, boat and bird from a visual image. CNN architecture contains feature extraction and classification.

2.1 feature extraction

The basic principle of feature extraction is to detect features from simple feature to complex feature. Simple feature contains edges and colours. Complex feature means a pattern that is assembled by multiple simple features. For example, in face recognition (figure 2), the first layer of CNN detects simple features, such as a vertical line, a horizontal line and an oblique line. The second layer detects middle level features, such as a nose, an eye or an ear. The final layer detects the most complicated feature which is the whole face in this case.

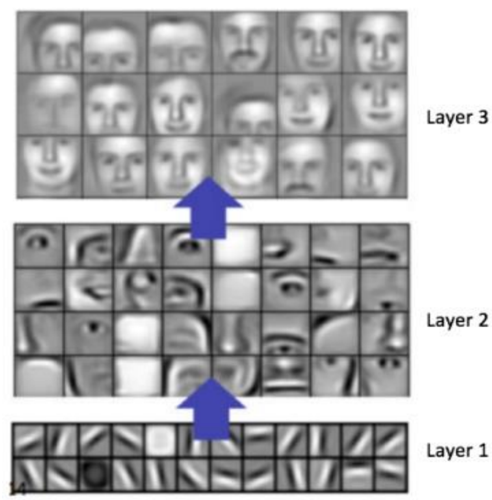

Figure 2 three layers of filters that are used to detect face features. Layer 1 aims at detecting simplest features. layer 2 aims at detecting the most complicated features

Feature extraction consists of four steps:

Step 1. **Convolutional Layer** (CONV): Convolutional layer applies filters on the input images. Filters are also called as kernels. Each kernel slides through the whole input image. A dot product of a kernel and a subset of the input image generates a pixel value in the output image(figure3). Different layer uses different kernels. In the layer of simple feature extraction, kernels are designed to detect lines and edges. In the layer of complex feature extraction, kernels are designed to detect complex patterns. The extraction of simple features contributes to the extraction of complex features. This can be seen in the hierarchy (figure 4).

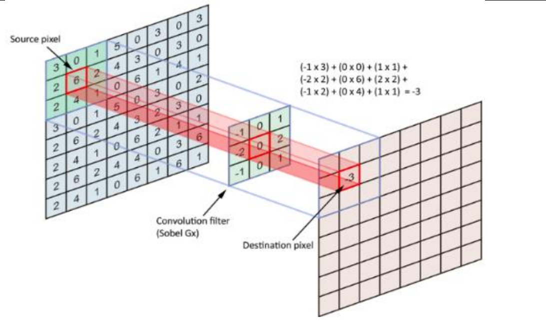

Figure 3 dot product of a kernel and a subset of image. The kernel slides through the whole image

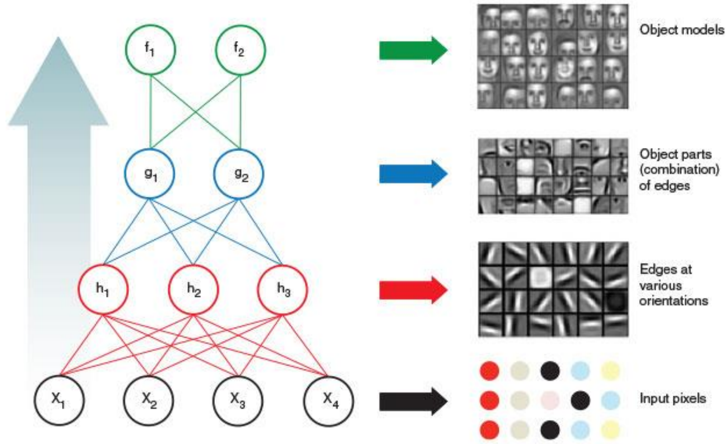

Figure 4 a hierarchy that shows the extraction of simple features (e.g.red neurons) contributes to the extraction of complex features (e.g. blue neurons)

Step 2. **Rectifying Linear Units Layer** (RELU): In order to remove negative values in the output images of a convolutional layer, a threshold is applied.

Step 3. **Maximum Pooling Layer** (POOL)**:** The goal of maximum pooling layer is to decrease the size of the output images of each convolutional layer and do not distort them. Therefore, the speed of further computation will not slow down.

Step 4. **Iterate the above 3 steps.** The more amount of iterations there are, the finer the features will be extracted.

2.2 Classification

The classification process calculates the degree of certainty that an object belongs to a class. The classification is based on the output of feature extraction. For instance, boat body is one of the features in the output of feature extraction. If the presence of boat body is clearly extracted, then it is unlikely that this is a dog or a cat or a bird, because they do not have a boat body. In the example of figure 6, the probability that this object is a boat is 0.94 (figure 5). Therefore, the output is that this is a boat.

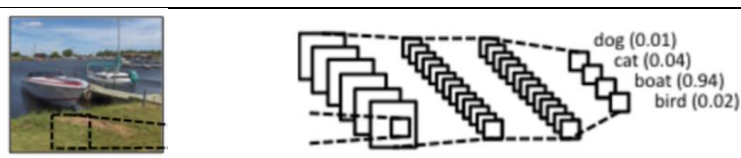

Figure 5 The left image is an input image of CNN. The right image shows that, after feature extraction, the classification process gives the output: The probability that an object in this image is a dog, a cat, a boat and a bird is 0.01, 0.04, 0.94 and 0.02 respectively.

3. Workflow Of This Project

Thermal images and visual images will be integrated in CNN. Because these two types of images provide complementary detection decisions. When illumination is sufficient, such as in a daytime, it is easy to distinguish objects by visual images. When illumination is insufficient, such as during the night or in a bad weather, thermal images have a better performance. So the integration of visual and thermal images will provide the most accurate detection, compared with using one of them. The workflow of this project is the following.

1) Data collection: Thermal images and visual images will be collected from several conflict areas which has a similar type of conflict as Lebanon, a similar terrain as Lebanon and a similar urban infrastructure as Lebanon.
2) Data annotation: Human in both of these two stacks of images will be labeled (also called annotated) manually. A label is a bounding box of a human. These labels are seen as ground truth.
3) Data splitting: The labeled thermal images and visual images will be splitted into two groups. One group is the training data. The other group is the test data.
4) Training of CNN:

The training process contain 5 steps.

① Input of CNN: Thermal images and visual images will be the input layer.

② Feature Extraction: Feature will be extracted from simple ones (edges and lines), to complicated ones (an arm, a leg, a foot, a head), to more complicated ones (upper body and lower body) and finally to the whole human body. The process is expected to be very similar as figure 14.

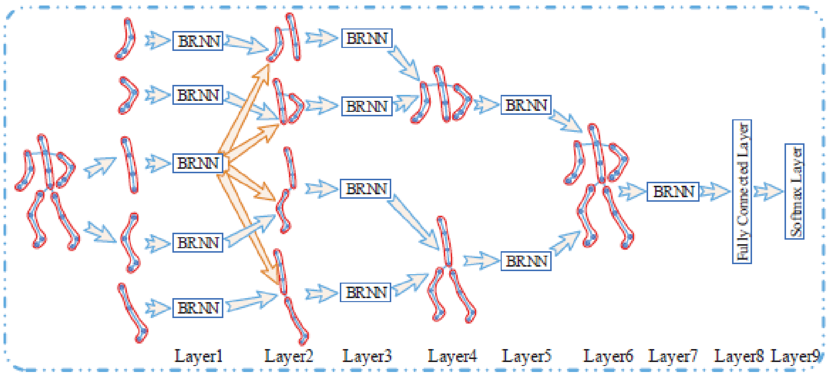

Figure 14 CNN architecture of human detection. Features visualized with red contours are the features that each layer aims at extracting.

③ Classification: The probability of there is a human or not on a certain location will be calculated.

④ Output: Make a conclusion whether there is human or not on each location.

⑤ Draw a bounding box on each detected person in visual images and thermal images (figure 15 and 16).

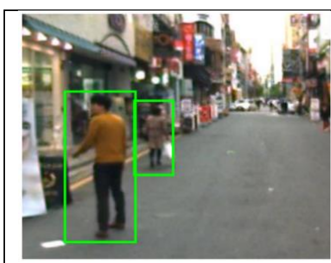

Figure 15

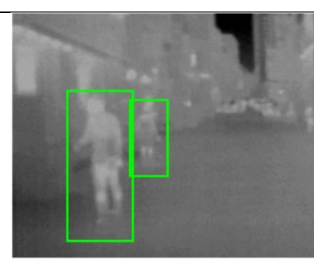

Figure 16

5) Testing of CNN: After the CNN has been trained, it will be tested. The bounding box drawn by CNN will be compared with the bounding box drawn in ground truth. The requirement of the accuracy will be made based on the agreement of us and our user.
6) Applying CNN to Lebanon conflict area: If the accuracy satisfies the requirement, the trained CNN will be applied to conflict area.

Deliverables

The human detection procedure is successfully implemented especially in open area. This method use fusion between visual images and thermal images for human detection. The thermal images help to determine the human temperature. In (Figure 1), the body can be recognized in thermal imagery.

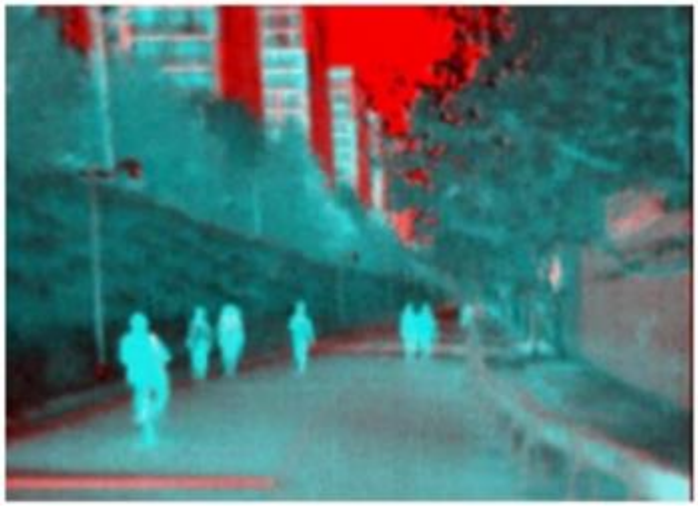

Figure 1 Thermal image

This research focuses on human detection aiming to help UNIFIL search and rescue the human body in the post conflict area. However, the result of this project will focus on extraction of the bounding box and the location of the human body, see in (Figure 2).

Figure 2 Bounding box in the human detection

In order to help and carry the human body out from the post conflict area, the location of the human needs to be found by determining the position of the body. The location was defined by GPS that mounted in UAV platform. This position of the body will be represented as dot map.

Limitation

The major limitation of our study would be the algorithm has been created only for open space keeping in mind the limitation of the drones and the camera's capability. Further research can be done to improve this.

Appendix

PHANTOM 4

SPECS VIDEOS DOWNLOADS FAQ GO TO TOP

Max Speed	S-mode: 20 m/s
Max Tilt Angle	S-mode: 42° A-mode: 35° P-mode: 15°
Max Angular Speed	S-mode: 200°/s A-mode: 150°/s
Max Service Ceiling Above Sea Level	19685 feet (6000 m)
Max Wind Speed Resistance	10 m/s
Max Flight Time	Approx. 28 minutes
Operating Temperature Range	32° to 104°F (0° to 40°C)
Satellite Positioning Systems	GPS/GLONASS
Hover Accuracy Range	Vertical: ±0.1 m (with Vision Positioning) ±0.5 m (with GPS Positioning) Horizontal: ±0.3 m (with Vision Positioning) ±1.5 m (with GPS Positioning)

VISION SYSTEM

Vision System	Forward Vision System Downward Vision System
Velocity Range	≤10 m/s (2 m above ground)
Altitude Range	0 - 33 feet (0 - 10 m)
Operating Range	0 - 33 feet (0 - 10 m)
Obstacle Sensory Range	2 - 49 feet (0.7 - 15 m)
FOV	Forward: 60°(Horizontal), ±27°(Vertical) Downward: 70°(Front and Rear), 50°(Left and Right)
Measuring Frequency	Forward: 10 Hz Downward: 20 Hz
Operating Environment	Surface with clear pattern and adequate lighting (lux>15)

CAMERA

Sensor	1/2.3" CMOS Effective pixels:12.4 M
Lens	FOV 94° 20 mm (35 mm format equivalent) f/2.8 focus at ∞
ISO Range	• 100-3200 (video) • 100-1600 (photo)
Electronic Shutter Speed	8 - 1/8000 s
Image Size	4000×3000
Still Photography Modes	Single shot Burst shooting: 3/5/7 frames Auto Exposure Bracketing (AEB): 3/5 bracketed frames at 0.7 EV Bias Timelapse HDR
Video Recording Modes	UHD: 4096×2160 (4K) 24 / 25p 3840×2160 (4K) 24 / 25 / 30p 2704×1520 (2.7K) 24 / 25 / 30p FHD: 1920×1080 24 / 25 / 30 / 48 / 50 / 60 / 120p HD: 1280×720 24 / 25 / 30 / 48 / 50 / 60p
Max Video Bitrate	60 Mbps
Supported File Systems	FAT32 (≤32 GB); exFAT (>32 GB)
Photo	JPEG, DNG (RAW)
Video	MP4, MOV (MPEG-4 AVC/H.264)
Supported SD Cards	Micro SD Max capacity: 64 GB Class 10 or UHS-1 rating required
Operating Temperature Range	32° to 104°F (0° to 40°C)

Figure B parameters of Dji Phantom 4

Article number:	19-00625
Resolution:	336 x 256 pixels
Spectral range:	7.5 - 13.5 µm
Framerate:	30 Hz (NTSC); 25 Hz (PAL) US only, not for Export
Export Framerate:	7.5 Hz (NTSC); 8.3 Hz (PAL)
Size:	2.26 "x 1.75" (57.4 x 44.4 mm) (including lens)
Weight:	3.25-4 oz (92.1 - 113.4 g) (depending on configuration)
Confirmation:	Two M2x0.4 on each side & one 1 / 4-20 screw point on the top
Voltage:	4.0 - 6.0 VDC
Measuring range:	-20 ° C to + 50 ° C

Figure E the parameter of FLIR Vue ™ Pro 336 thermal camera

Figure C DJI GS PRO