# Geolocation and Counting of People with Aerial Thermal Imaging for Rescue Purposes

Córdova C. Andrea[1,5], Jiménez Q. Byron[1], Pardo I. Jorge[2,4], Toalombo CH. Inti[1], and Wilbert G. Aguilar[3,6](✉)

[1] DEM, Universidad de las Fuerzas Armadas ESPE, Latacunga, Ecuador
[2] UGT, Universidad de las Fuerzas Armadas ESPE, Latacunga, Ecuador
japardo@espe.edu.ec
[3] CICTE Research Center, Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador
wgaguilar@espe.edu.ec
[4] Facultad de Ingeniería, Pontificia Universidad Católica del Ecuador, Quito, Ecuador
[5] Aerospace Engineering Department, The Pennsylvania State University, State College, PA, USA
[6] GREC Research Group, Universitat Politècnica de Catalunya, Barcelona, Spain

**Abstract.** Thermography has become more frequently used in rescue operations when used together with flight technologies such as unmanned aerial vehicles (UAVs). This is due to its non-invasive and powerful supervision characteristics in the spectrum range not perceivable for the human eye or for a standard camera. This paper presents a developed system based on the synergy of a UAV and a counting and geolocation algorithm that detects people with aerial shots in areas of difficult access. The system integrates a thermal camera to a UAV, thus being useful in different scenarios such as floods, fires or wooded areas. For this purpose, the UAV navigation paths are configured from an earth station using a telemetry-specific software. Thermal images will be recorded during the mission at a height determined by the operator, which will later be processed to filter, discriminate, count and geolocalize people at risk. The processing of the images is done by means of artificial vision tools combined with Artificial Neural Networks (ANN).

**Keywords:** Geolocation · ANN · Telemetry · Thermography UAVs

## 1 Introduction

The use of thermal imagery in military and industrial applications has awaken a growing interest in the research community, who have been experimenting with its uses in search and rescue applications. While image processing with artificial vision tools in the range of visible light has been studied for decades and recently combined with


L. T. De Paolis and P. Bourdot (Eds.): AVR 2018, LNCS 10850, pp. 171–182, 2018.
https://doi.org/10.1007/978-3-319-95270-3_12

machine learning techniques for the visual detection of objects with extremely fast image processing, research in the field of methodologies for developed applications in thermal imagery is becoming more popular [1, 6]. Likewise, the trend in recent years has been the development of flight technologies such as UAVs, and similarly to thermography, its initial applications were merely military. The benefits provided by aerial vehicles have enabled researchers to use these technological devices in various applications. Due to advances in the autonomous performance of the UAV, the intervention of an operator has been reduced and has resulted in these devices with some degree of intelligence being used in a wide range of assignments considered dangerous or difficult to perform by human beings [1]. This includes Search and Rescue (SAR) applications, which involve identifying and extracting information of people in emergencies caused by natural disasters.

The analysis of the thermal images is done on board the air vehicle, greatly reducing the need for a fast and stable communication with the earth station and avoiding delays in the decision-making process. The system is able to track several objects simultaneously, where the user has direct control over the types of objects that wants to follow [2]. The technique presented detects humans at a speed of 25 Hz, first analyzing the thermal image to find human temperature silhouettes and then using the regions corresponding to the color of the silhouette to classify human bodies. In addition, the system detects human positions, which are geolocalized to then build a map with points of interest [3].

SHERPA project has created collaborative robots to help in tasks of rescuing people in hostile environments or in emergencies in the Italian Alps. They work with small-scale airplanes, drones and autonomous small helicopters called "Hawks", they are used to transport technical materials like scanners and thermal cameras, but also food and water. The project has the advantage of not interrupting search activities during the night or when the visibility is limited. The operator establishes the mission, delimits the area of interest and with the push of a button gets the plans by scanning the area for postprocessing [4].

The Cotopaxi volcano is considered one of the most dangerous volcanoes in Ecuador. According to the Geophysical Institute (IG) of the National Polytechnic School, a research center in Ecuador for the diagnosis and monitoring of seismic and volcanic hazards, there have been five major eruptions since its activation in 1532 affecting more than 300,000 people [5]. The eventual reactivations of the Cotopaxi volcano have prompted efforts to develop a prototype of SAR; the first in Ecuador, of low cost that helps rescue agencies to optimize resources, in search missions and locating people in areas of difficult to access [5].

The paper is organized as follows. Firstly, the selection of components such as the UAV and the Thermal Camera is done. Next, the development of the SAR system is described, which integrates the configuration of the autonomous trajectories of the UAV, the algorithm designed that allows to process thermal images, sampling for ANN training, counting and geolocation. Finally, the conclusion of the paper analyzes and evaluates the system performance, based on collected data from several flight tests.

# 2 Devices Selection

## 2.1 Thermal Camera

For our project it is necessary to acquire aerial thermal images, obtained with the help of a thermal camera without radiometric functions because in this particular case these functions are not relevant. The technical characteristics of three thermal cameras have been analyzed: the Zenmuse XT, Workswell Wiris, and Flir Vue. The three cameras share similar functionalities varying slightly in their characteristics. The main difference is that the first two cameras have radiometric functions allowing to make thermal point measurements of the image in the range of −40 °C to 1500 °C and the third does not have such functions. Therefore, the Flir Vue Pro thermal camera was selected, which has the necessary functionalities to be used in a SAR. Its technical characteristics are described in Table 1.

**Table 1.** Technical characteristics of thermal camera.

| Parameter | Value |
|---|---|
| Sensor | Uncooled VOx Microbolometer |
| Sensor Resolution | 640 × 512, 336 × 256 |
| Spectral Band | 7.5–13.5 um |
| Frame Rates | 7.5 Hz |
| Size | 63 mm × 44.4 mm × 44.4 mm |
| Weight | 92.1–113.4 g |
| Image Optimization for sUAS | Yes |
| Color Palettes | Yes-Adjustable Via PWM |
| Input Voltage | 4.8–6.0 VDC |
| Operating Temperature Range | −20 °C to +50 °C |
| Operative Altitude | +12000 m |
| Analog Video Output | Yes |

## 2.2 UAV

The most suitable device for the development of the project is the drone Phantom 3 Professional of the DJI brand, this is considered as the best option as it has optimal characteristics such as: Safety, Useful Load, Stability, Maximum Distance, Flight Modes.

One of the most important features that the Phantom 3 Pro has is flight safety, which proves a RETURN TO HOME option automatically, in case of an anomaly during the flight. It is also proven that the payload that can lift with great stability can be up to 2.5 lbs according to evidence gathered in the investigation. Through its internal sensors, its GPS system and positioning by vision allows a stability at each flight point that is located. Other important features are the maximum distance which has a range of up to 5 km, as well as a maximum speed of 16 m/s. In addition, it has several intelligent flight modes which are defined in its DJI GO application, which are: Follow Me, Course Lock, Waypoints, Home Lock, Point Of Interest.

## 3 System

The system follows the next sequence, it consists in identifying, filtering, counting and geolocating people by processing images obtained with a thermal camera mounted on a UAV as is described in the following image (Fig. 1).

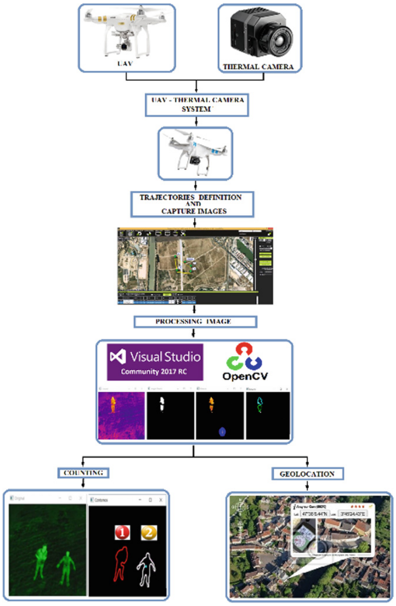


**Fig. 1.** SAR system diagram

### 3.1 Autonomous Trajectories

The DJI GS "Ground Station" Pro application is used for the definition of trajectories, which allows an autonomous flight of the UAV without the need of operating it by radio control, allowing us to load the Phantom 3 Pro trajectories defined by waypoints, delimiting in this way the area of search or interest. In addition, the application shows other flight characteristics such as takeoff parameters, actions to be taken at each point, speed, which according to tests performed for the acquisition of suitable images will be 2 m/s and finally return home or starting point.

The application allows to have programmed control of the UAV, it calculates and presents functions such as total distance traveled, flight time, and GPS coordinates during the mission. In addition, it divides the terrain with a sweep which allows to cover the entire search area, taking a parallel path or focus on a point of interest according to its configuration.

### 3.2 Image Acquisition

With the thermal camera mounted on the Phantom3 and with defined trajectories, several flyovers were carried out over areas of interest at a height of 20 m above the ground. The camera is set to recording mode and one of the nine palettes available in the camera is selected, for our project the GreenHot and IronBow palettes are used. Additionally the frames are configured to 10 fps and the scene where it is going to work in this particular case is outdoors.

During the flight the camera is able to capture images of the possible victims which are of our interest, but also captures any type of body that radiates heat, such as animals, automobiles, fire, etc. Once the UAV finishes its route returns to the ground station where information is extracted from the memory card to a computer.

### 3.3 Segmentation

Segmentation consists in distinguishing and separating each of the objects present in the scene and filtering only the objects of interest. In the particular case of this project, it is necessary to separate thermal silhouettes of people from the background of the image. Since an image sequence is to be processed, that is to say a video, the histogram will be different in each frame because the UAV records different scenes during its course. In Fig. 2(a) and (b) thermal captures of a person with their respective histogram are observed using the GreenHot palette where the variation of intensities of the pixels can be seen. It should be noted that the same thing happens when using the IronBow palette.

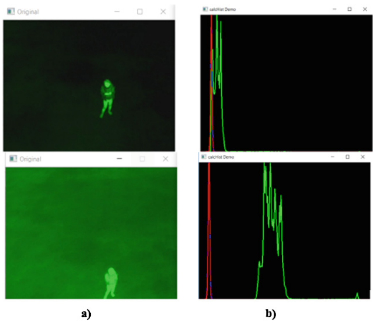


**Fig. 2.** (a) Thermal captures, (b) Histograms (Color figure online)

To facilitate segmentation, the image with color is converted to gray levels. This technique consists in calculating the average of each intensity for the matrices of the colors Red (R), Green (G) and Blue (B). The equation that rules this transformation is as follows.

$$I = Round\left\{\frac{1}{3}(R + G + B)\right\} \quad (1)$$

Later binarization is used as segmentation technique specifically the Otsu method, which is based on the variation of intensity between the pixels of the object of interest and the pixels of the background, assuming that the image contains two classes of pixels. The Otsu method calculates the optimal threshold that minimizes the intra-class variance and maximizes the inter-class variance automatically, so it does not need supervision. According to research by J. Kittler and J. Illingworth the ruling equations of the Otsu method are described below [6].

The process starts from a grayscale image with N pixels and L possible different levels.

$$p_i = f_i/N \tag{2}$$

Where $f_i$ is the repetition frequency of the ith gray level with $i = 1, 2, \ldots, L$.

In the case of thresholding in two levels, or also called binarization, the pixels are divided into two classes, $c_1\,[1, 2, \ldots, t]$ and $c_2\,[t+1, \ldots, L]$. Where the probability distribution of gray levels for the two classes is as follows.

$$c_1 = \frac{p_1}{\omega_1(t)}, \ldots, \frac{p_t}{\omega_1(t)} \tag{3}$$

$$c_2 = \frac{p_{t+1}}{\omega_1(t)}, \frac{p_{t+2}}{\omega_2(t)}, \ldots, \frac{p_L}{\omega_2(t)} \tag{4}$$

Where

$$\omega_1(t) = \sum\nolimits_{i=1}^{t} p_i \quad \& \quad \omega_2(t) = \sum\nolimits_{i=t+1}^{L} p_i \tag{5}$$

The average for class $c_1$ and class $c_2$ is

$$u_1 = \sum\nolimits_{i=1}^{t} \frac{iP_i}{\omega_1(t)} \quad \& \quad u_2 = \sum\nolimits_{i=t+1}^{L} \frac{iP_i}{\omega_2(t)} \tag{6}$$

If is $u_T$ the average intensity of the whole image, it is shown that

$$\omega_1.u_1 + \omega_2.u_2 = u_T \quad \omega_1 + \omega_2 = 1 \tag{7}$$

Using discriminant analysis Otsu defines the variance between two classes of a threshold image as follows.

$$\sigma_B^2 = \omega_1(u_1 - u_T)^2 + \omega_2(u_2 - u_T)^2 \tag{8}$$

For two level thresholding, the optimal threshold t* is chosen so that $\sigma_B^2$ will be maximum, this is

$$t^* = Max_t\left\{\sigma_B^2(t)\right\} \quad 1 \le t \le L \tag{9}$$

Considering that thermal images have a clear difference between the objects to be extracted or filtered from the background of the scene, the Otsu method described mathematically is very useful for our purpose because it has a good response in situations of the real world. After applying the segmentation to the infrared image, the following results are obtained, with the IronBow palettes (Fig. 2(a) and (b)) and GreenHot (Fig. 3(a) and (b)).

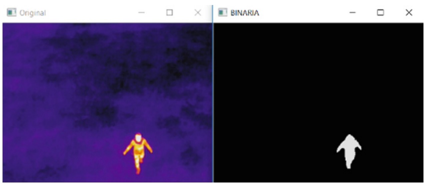


**Fig. 3.** (a) Thermal image, (b) Segmented image

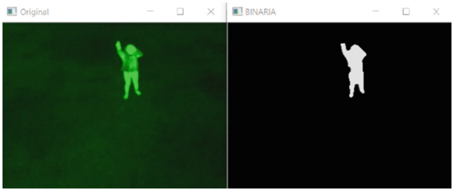


**Fig. 4.** (a) Thermal image, (b) Segmented image

### 3.4 Sampling

After segmentation of the objects of interest, it is required to obtain binary samples of all types of bodies such as: human, animal, automotive, or any other object that radiates heat. Samples of people are stored as positive samples and anybody other than a person is saved as a negative sample. Negative and positive samples are required because ANN needs to learn which silhouette corresponds to a person and which does not (Fig. 4).

To obtain the samples an algorithm was developed that consists of detecting contours of the silhouettes of the bodies. From the silhouettes found by the program, areas of all contours are calculated and the largest area is selected. Lastly, a Region of Interest (ROI) is created from the biggest area and only that region is clipped to save the sample to the computer's hard disk.

### 3.5 ANN Structure

The ANN select for this project is a multilayer perceptron whit retro propagation because this network is ideal for real world situations and the recognition is faster. Thus, the model was conformed by an input layer where a binary image of $32 \times 32$ pixels is received, next the net has eight hidden layers where the input information is processed and finally has an output layer which will give the final result of whether or not it is a person. The activation function of the perceptron is the standard sigmoid symmetric function and the activation values range from $-1$ to 1. Based on experimental evidence in this particular case the probability of prediction returned by network has been set at 0.5, if the values exceed this value the silhouette is identified as person otherwise not. The resulting model of the ANN is showing in Fig. 5.

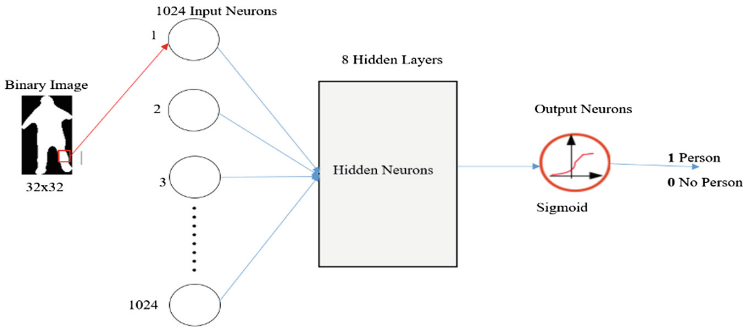


**Fig. 5.** Artificial Neuronal Network structure

### 3.6 ANN Learning or Training

The process of training or learning is completely analogous to teaching something to a child the algorithm has to be taught what to learn and what not, therefore there has to be a set of training pairs, enter them into the network and wait for the answer, if the answer is not correct you must perform the training again and so on until you achieve the desired result, the process of performing the training several times is called the period [1].

Previous to ANN training, samples must be prepared for which two algorithms are used. The first algorithm allows to give the mirror effect to all the images to obtain a total of 3288 positive samples from 1664 original samples, the same procedure is done with the negative samples, obtaining a total of 300.

The other algorithm allows to resize the sample images of any size to a size of $32 \times 32$ pixels, resulting in images with a total of 1024 pixels, this is done in order to have a fixed number of pixels in all images. The same algorithm allows to generate and save in the computer a file called "training.ocv" which is a file format distributed in rows and columns that stores the training data of both the positive samples as well as the negative samples in the form of text flat.

For training and recognize person silhouette it was necessary to use the multilayer perceptron using backpropagation algorithm, that model was created with the library "ml" of OpenCV that allows to develop common models of machine learning. Back-propagation algorithm works by determining the loss or error at the output and the propagating it back into the network. To minimize the error the weights are updated in each iteration until an acceptable error is obtained [11].

After performing the training it is necessary to save the characteristics of the ANN, for this purpose, a file called "parametros.xml" is generated. With this file format it is possible to arbitrarily store complex OpenCV data structures, as well as data types such as integers and floating-point numbers, as well as chain-text chains that are part of the designed ANN characteristics.

The results returned by the network are presented in Fig. 6 and after the training an accuracy of 95.4% was obtained in the recognition of the silhouettes of people.

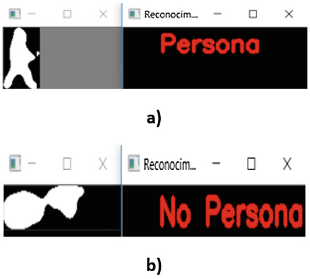


**Fig. 6.** (a) Person detection, (b) Other object detection

### 3.7 Counting and Geolocating

The detection of people through an ANN was used to perform the counting of human bodies. When a person was detected a counter set at zero at the beginning of the route was increased, so that by the end of the route it will have the total number of people detected.

Thanks to the GPS positioning system of the Phantom 3 Pro, precise parameters of latitude, longitude and height can be determined by means of coordinates that are shown on the screen in real time, thus indicating the exact position of the device in every second of Flight, as well as a histogram with all the information of the mission which after an analysis will facilitate the geolocation of each person.

For the tests performed it is possible to observe in Fig. 6 thanks to the DJI GS application, the defined trajectory that the drone must follow. Within the tour 10 people were positioned who are animated in the image as 8 green circles (people who were later identified as such) and 2 red circles (unidentified people), since the application image available is an offline map. Also, in the lower part of the image you can observe the people with their respective label (P) and geolocation (LAT and LON), as well as those that were not counted by the ANN that are in red color which are P5 and P8 (Fig. 7).

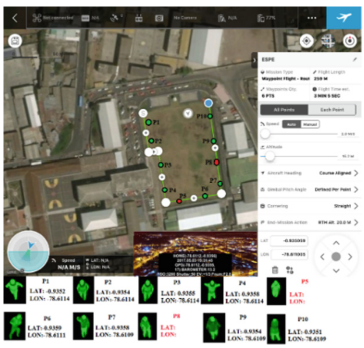

**Fig. 7.** Map of trajectories of geolocation (Color figure online)

Finally, the geolocation of each person or a group of people is delimited by the area of the vision range or focal length of the thermal camera [9]. The camera is positioned at an angle α = 45° with respect to the drone height. The triangulation of two images taken from different positions in the air, but focused towards the same point, allows us to find the distance between these two vectors for the geolocation of the point of interest [10]. Based on these investigations and considering that the height of the drone is $h = 15$ m, a right triangle is formed between the surface of the ground, the UAV and the identified person, as it is shown in the Fig. 8 below.

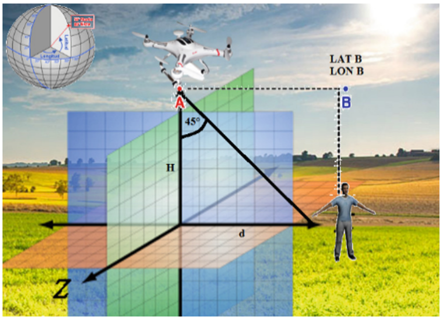


**Fig. 8.** Calculation of coordinates

This involves finding a distance $d$, which will serve as a reference to locate the real position of the person at the point (B − h) for which the following formula was applied.

$$d = tg\,\alpha * h \tag{10}$$

Once $d = 15$ m is obtained, an estimation of how long it will take the drone to travel this distance and to overfly the point B was done. The same point gives the coordinates in latitude and longitude (LAT B and LON B) knowing that the device travels At a speed $v = 3$ m/s.

$$t = \frac{d}{v} \tag{11}$$

After having completed a t = 5 s trip, the GPS coordinates of the person or the people identified are accurately known. Moreover, as shown in the lower part of Fig. 6, the variation in latitude and longitude is minimal and does not change its value significantly according to where the body is. This will also depend on the magnitude of the area to be studied.

## 4 Test and Results

System tests were performed at different heights and a UAV displacement speed of 2 m per second, with 10 people located at different points of the defined trajectory. The error $e$ in detection of the people is calculated as follows, where $N_t$ is the theoretical number of people that the system must detect and $N_r$ is the real number of people that the system detects.

$$e = \left| \frac{N_t - N_r}{N_t} \right| * 100\% \tag{12}$$

**Table 2.** People detection

| Height | Detected people | Geolocation | % error |
|---|---|---|---|
| 15 m | 10 | Yes | 0 |
| 17 m | 10 | Yes | 0 |
| 19 m | 12 | Yes | 20 |
| 21 m | 9 | Yes | 10 |

Therefore, the results obtained for the counting of people are shown in Table 2. The tests performed in terms of geolocation were verified by means of the GPS that the Phantom 3 has. Once the mission is finished, the drone is placed in the location of each person to take the data of Latitude and Longitude, obtaining an error less than 1% as is shown in the Table 3.

**Table 3.** Latitude and longitude test and error

| Person | Calculate value | Real value | % error |
|---|---|---|---|
| P1 | LAT: −0.9352 | LAT: −0.9350 | 0.02 |
| | LON: −78.6114 | LON: −78.6115 | 0 |
| P2 | LAT: −0.9352 | LAT: −0.9352 | 0 |
| | LON: −78.6114 | LON: −78.6115 | 0 |
| P3 | LAT: −0.9355 | LAT: −0.9354 | 0.01 |
| | LON: −78.6114 | LON: −78.6114 | 0 |
| P4 | LAT: −0.9358 | LAT: −0.9358 | 0 |
| | LON: −78.6114 | LON: −78.6114 | 0 |
| P6 | LAT: −0.9359 | LAT: −0.9360 | 0.01 |
| | LON: −78.6111 | LON: −78.6111 | 0 |

## 5 Conclusions

The Search and Rescue System (SAR) was able to identify and geolocate people from a UAV, for which a total of 3588 samples were used, being one of the investigations with the largest number of images used in the training of an Artificial Neural Network for identification of people for rescue purposes. Furthermore, the algorithm designed segmented thermal images, independently of the scenario in which people were. The Otsu method applied in the segmentation of thermal images allowed to obtain suitable samples to train the ANN in order to identify and classify people correctly. Based on several experimental flight tests, it was determined that the range of heights at which the lowest error rate in the identification of persons is obtained is from 15 m to 17 m and the appropriate speed of the UAV is 2 m/s, using a 9 mm lens for the thermal camera. Considering the curvature of the earth and knowing that the system was tested near the equatorial line, a large displacement is required in order to have major changes in length and latitude parameters. The results obtained by the tests were compared with coordinates of Google Maps and drone GPS, which show a minimal variation in the order of the fourth decimal. This precision allows us to ensure that the method proposed can be applied efficiently in the geolocation of people.

## References

1. Portman, J., Lynen, S., Chli, M., Siegwart, R.: People detection and tracking from aerial thermal views. In: IEEE International Conference on Robotics & Automation (ICRA), Hong Kong, China (2014)
2. Leira, F.S., Johansen, T.A., Fossen, T.I.: Automatic detection, classification and tracking of objects in the ocean surface from UAVS using a thermal camera. In: 2015 IEEE Aerospace Conference, Big Sky, MT, USA (2015)
3. Rudol, P., Doherty, P.: Human body detection and geolocalization from UAV search and rescue missions using color and thermal imagery. In: 2008 IEEE Aerospace Conference, Big Sky, MT, USA (2008)
4. Sherpa: sherpa-project.eu, 23 March 2017. http://www.sherpa-project.eu/sherpa/workshop-SR-2017. Último acceso: 20 Feb 2017
5. Instituto Geofísico: igepn.edu.ec (2016). http://www.igepn.edu.ec/cotopaxi. Último acceso: 16 Mar 2017
6. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, Hi, USA (2003)
7. Kittler, J., Illingworth, J.: On threshold selection using clustering criteria. IEEE Trans. Syst. Man Cybern. **15**(5), 652–655 (1985)
8. Izurieta, F., Saavedra, C.: Redes Neuronales Artificiales. U. d. C. Departamento de Física, Ed., Concepción (2006)
9. Gibbins, D., Roberts, P., Swierkowski, L.: A video geo-location and image enhancement tool for small unmanned air vehicles (UAVs). In: Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, Melbourne, Vic., Australia (2005)
10. Okello, N., Musicki, D.: Emitter geolocation with two UAVs. In: 2007 Information, Decision and Control, IDC 2007, Adelaide, Qld., Australia (2007)
11. Vladimir, V., Rauf, I.: Knowledge transfer in SVM and neural networks. Ann. Math. Artif. Intell. **81**(1–2), 3–19 (2017)