Machine Learning: Science and Technology

PAPER • OPEN ACCESS

Autonomous victim detection system based on deep learning and multispectral imagery

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
28 November 2022

REVISED
17 January 2023

ACCEPTED FOR PUBLICATION
27 January 2023

PUBLISHED
10 February 2023

PAPER

Autonomous victim detection system based on deep learning and multispectral imagery

Christyan Cruz Ulloa[*], **Luis Garrido, Jaime del Cerro and Antonio Barrientos**

Centro de Automática y Robótica, Universidad Politécnica de Madrid-Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain

[*] Author to whom any correspondence should be addressed.

E-mail: christyan.cruz.ulloa@upm.es

Keywords: artificial vision, multispectral image, convolutional neural networks, multispectral index, search and rescue robotics

Abstract

Post-disaster environments resulting from catastrophic events, leave sequels such as victims trapped in debris, which are difficult to detect by rescuers in a first inspection. Technological advances in electronics and perception have allowed the development of versatile and powerful optical sensors capable of capturing light in spectrums that humans cannot. new deep learning techniques, such as convolutional neural networks (CNNs), has allowed the generation of network models capable of autonomously detecting specific image patterns according to previous training. This work introduces an autonomous victim detection system to be deployed by using search and rescue robots. The proposed system defines new indexes based on combining the multispectral bands (Blue, Green, Red, Nir, Red Edge) to obtain new multispectral images where relevant characteristics of victims and the environment are highlighted. CNNs have been used as a second phase for automatically detecting victims in these new multispectral images. A qualitative and quantitative analysis of new indexes proposed by the authors has been carried out to evaluate their efficiency in contrast to the state-of-the-art ones. A data set has been generated to train different CNN models based on the best obtained index to analyze their effectiveness in detecting victims. The results show an efficiency of 92% in automatically detecting victims when applying the best multispectral index to new data. This method has also been contrasted with others based on thermal and RGB imagery to detect victims, where it has been proven that it generates better results in situations of outdoor environments and different weather conditions.

1. Introduction

Post-disaster scenarios resulting from natural disasters or attacks can involve thousands of victims trapped beneath the rubble due to the collapse of buildings or big infrastructures [1]. The search and rescue (SAR) robotics seeks to assist the first aid teams through the transmission of remote information, usually images (RGB and Thermal) and sound, to determine at an early stage of their actuation the existence of victims in high risk of difficult access areas.

This work has been developed as task of the TASAR project (Team of Advanced Search And Rescue Robots), which is focused on using ground SAR-Robots for Humanitarian Assistance and Disaster Relief [2].

Commonly, the state-of-the-art of multispectral imagery shows vegetative analysis applications focused on extracting indexes such as NDVI (as shown in some works developed in the state-of-the-art and also the authors) [3–5]. However, there are few works that use specific bands of multispectral images and neural networks to identify people [6–8]. however, it has not been studied at the level of post-disaster situations to detect victims, specially in case of body extremities occlusions (covered victims).

The works within the state-of-the-art use either Nir (near infrared from 750 to 2500 nm) or RedEdge (between the Visible Red and Near Infrared) bands cameras for people identification. However, no significant contribution has been found regarding the study or generation of new indexes from the combination of the bands captured by a multispectral camera (Blue, Green, Red, Nir, Red Edge) to detect limbs or human bodies so far.

The main contribution of this work focuses on proposing new indexes based on combining the multispectral bands captured by the Altum Camera. A qualitative and quantitative study (histogram analysis, intensity and pixel concentration) of the resulting multispectral images has been performed to define the best index for both indoors and outdoors.

Once the best index has been defined, a data-set of indoors and outdoors has been generated and manually labelled (head, hand, leg, torso, victim) from reconstructed post-disaster scenarios. From this data-set of 1454 images (with data augmentation), four models of Convolutional Neural Networks (CNN)(YOLOv5 x, l, m and s) have been trained to evaluate effectiveness.

The result has shown high efficiency in detecting victims both indoor and outdoor by using images generated with the proposed index and the mentioned CNNs.

The proposed method has been contrasted with methods based on Thermal and RGB images to detect victims. The main results show greater efficiency for scenarios and cases where the thermal or RGB image does not have a more significant effect, such as the presence of victims next to heat sources or clothing with colours similar to the environment.

This paper is structured as follows: section 2 shows the most relevant state-of-the-art works, section 3 describes the proposed methods. Proposed indexes and CNN are introduced in detail, followed by the Results and Discussion in section 4. To conclude, section 5 summarises the main findings.

2. Related works

Multispectral imagery is a relatively new field that suffered a boom last decade, especially for evaluating crop health in remote sensors applications through the calculation of vegetative indexes (such as NDVI, GNDVI) by combining the different bands (Blue, Green, Red, Nir, Red Edge) captured [9].

Nevertheless, multispectral imagery has been least applied to victim detection, which involves the study of new indexes that improve the extraction of characteristics by combining the different bands.

This section analyses the previous studies in the field of victim detection as well as those ones carried out by using multispectral information.

2.1. Multispectral images analysis for information extraction

Among the most relevant studies focused on extracting information and identifying objects through classical vision and multispectral images is the one carried out in [10], where the paradigm of the objects is explored, seeking to group pixels with a similar spectrum for later extraction of features.

There are studies focused on multispectral image processing to extract information from satellite images [11–15]. On the other hand, there are studies focused on predicting and preventing natural disasters by combining previous and updated information of specific areas provided by multispectral cameras on board the satellites [16].

Some works focus the detection on the physical conditions of subjects such as stress [17], Monitoring vulnerable people [18] or bio-metric analysis [19].

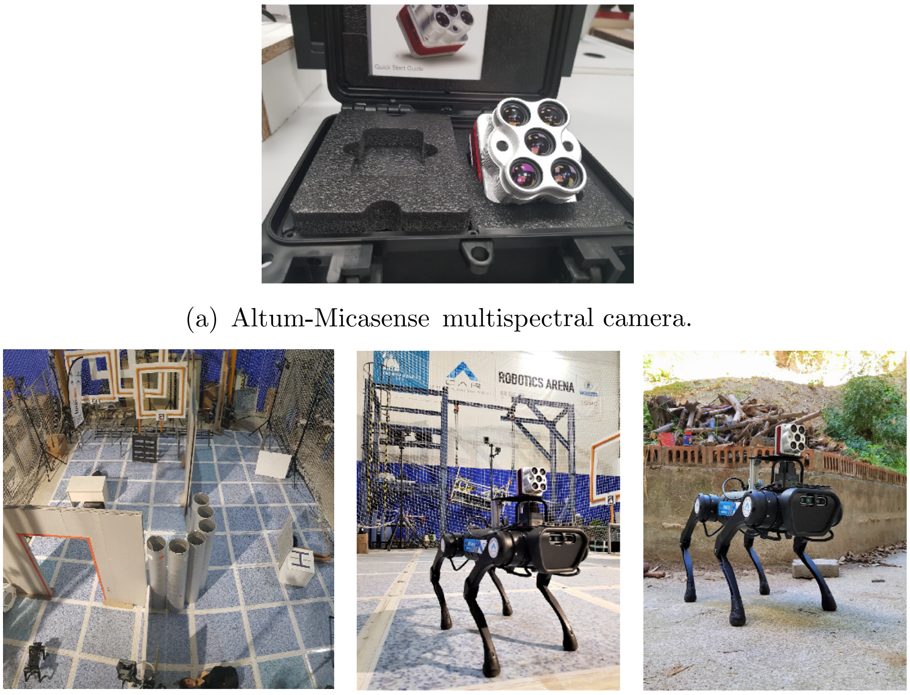
As previously, most studies with this type of images have been focused on agriculture, for the analysis of the vegetative state of crops [4, 5, 20–23].

2.2. Neural networks for people detection based on multispectral images

Several studies have been carried out to detect pedestrians in different light conditions through multispectral, RGB and thermal images, which have shown a first step in the effectiveness of people detection [24]. There are also studies focused on detecting people and objects from images captured in the NIR band [6, 7]. Moreover, some works combine multispectral information with thermal information for the detection of people [8].

Although there are works that make a first approach within the state-of-the-art, most directly use the Nir or RedEdge bands for processing and a neural network for person identification. However, no contribution has been made within the study or generation of new indexes from the different multispectral bands captured.

Although the studies carried out focus on the detection of people, the works within the state-of-the-art do not show cases or scenarios where the entire body of the person is not present, which is usually common in the case of victims covered by debris with limbs, torso, head covered.



(a) Altum-Micasense multispectral camera.

(b) Reconstructed post-disaster indoor scenario with moderate obstacles and victims.

(c) Quadruped robot equipped with the multispectral camera for indoors data acquisition

(d) Quadruped robot and the multispectral camera for outdoors data acquisition.

Figure 1. Materials used for the development of the work. Source: Authors.

3. Proposed methods

3.1. Materials

For this work, an Altum camera, developed by MicaSense, has been used. It is shown in figure 1(a); with dimensions of 0.82 × 0.67 × 0.67 m and a weight of 357 g. The camera relies on six lenses and provides with five simultaneous captures, images in different ranges of the visible spectrum and one in the thermal infrared.

The bands of the visible spectrum are Blue: (centre 475 nm, bandwidth 32 nm), Green (centre 560 nm, bandwidth 27 nm), Red (centre 668 nm, bandwidth 14 nm), Red edge (717 nm centre, 12 nm bandwidth), Near IR (842 nm centre, 57 nm bandwidth). The thermal band captures images in the 8000–14 000 nm wavelength range.

All these images have a single colour channel (grey scale), with values between 0 (black) and 255 (white) for each pixel according to the light captured in the corresponding wavelength by each image camera sensor.

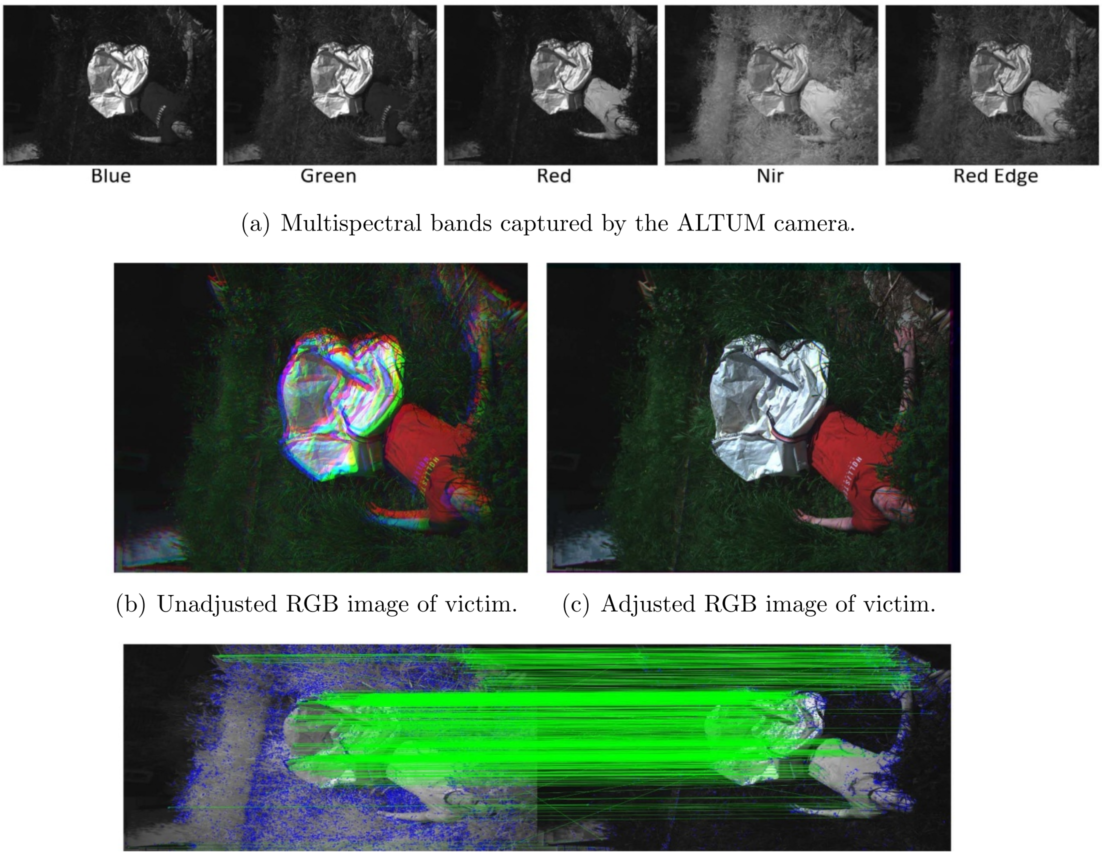
The camera's resolution is 2064 × 1544 pixels for the images taken by the five multispectral lenses, a field of view FOV of 50.2° × 38.4°.

Two post-disaster scenarios (indoor and outdoor) have been recreated for data acquisition purposes (figure 1(b)). The data acquisition has been carried out using a tele-operated quadruped robot with the Altum camera anchored (Outdoors figure 1(c)—Indoors figure 1(d)). The images are periodically stored in memory with a period of 1 s.

3.2. Adjustment of captured multispectral images

Python and OpenCV libraries have been used for image processing. Figure 2(a) shows an example of the multispectral bands captured by the Altum camera in grey scale, with pixel intensity in the range [0–255]. The images in figures 2(b) and (c) are the results of overlapping the camera's Red, Green and Blue bands. However, there is a noticeable difference in fit between the first and second.

The first case (figure 2(b)) corresponds to overlapping the RGB image without prior adjustment. This is due to the very arrangement of the camera lenses since they have a relative displacement. Second case (figure 2(c)) shows the adjustment avoiding distortion.



(a) Multispectral bands captured by the ALTUM camera.

(b) Unadjusted RGB image of victim.

(c) Adjusted RGB image of victim.

(d) Establishment of correspondences to relate the key-points in two images.

Figure 2. Example of data-set images with and without adjustment. Source: Authors.

Ignoring this effect and trying to combine the images without prior adjustment produces distortion in the final images. These effects can make it difficult to recognize victims. This factor has much influence in a neural network since the edges serve to identify patterns

A fixed image is considered as a base for the rest to perform the mentioned adjustment. Some key points o feature points are used to perform the correction parameters (figure 2(d)). This procedure is best described in work a previously presented by the authors [25].

3.3. Proposal of new multispectral indexes

3.3.1. Initial testing with state-of-the-art indexes

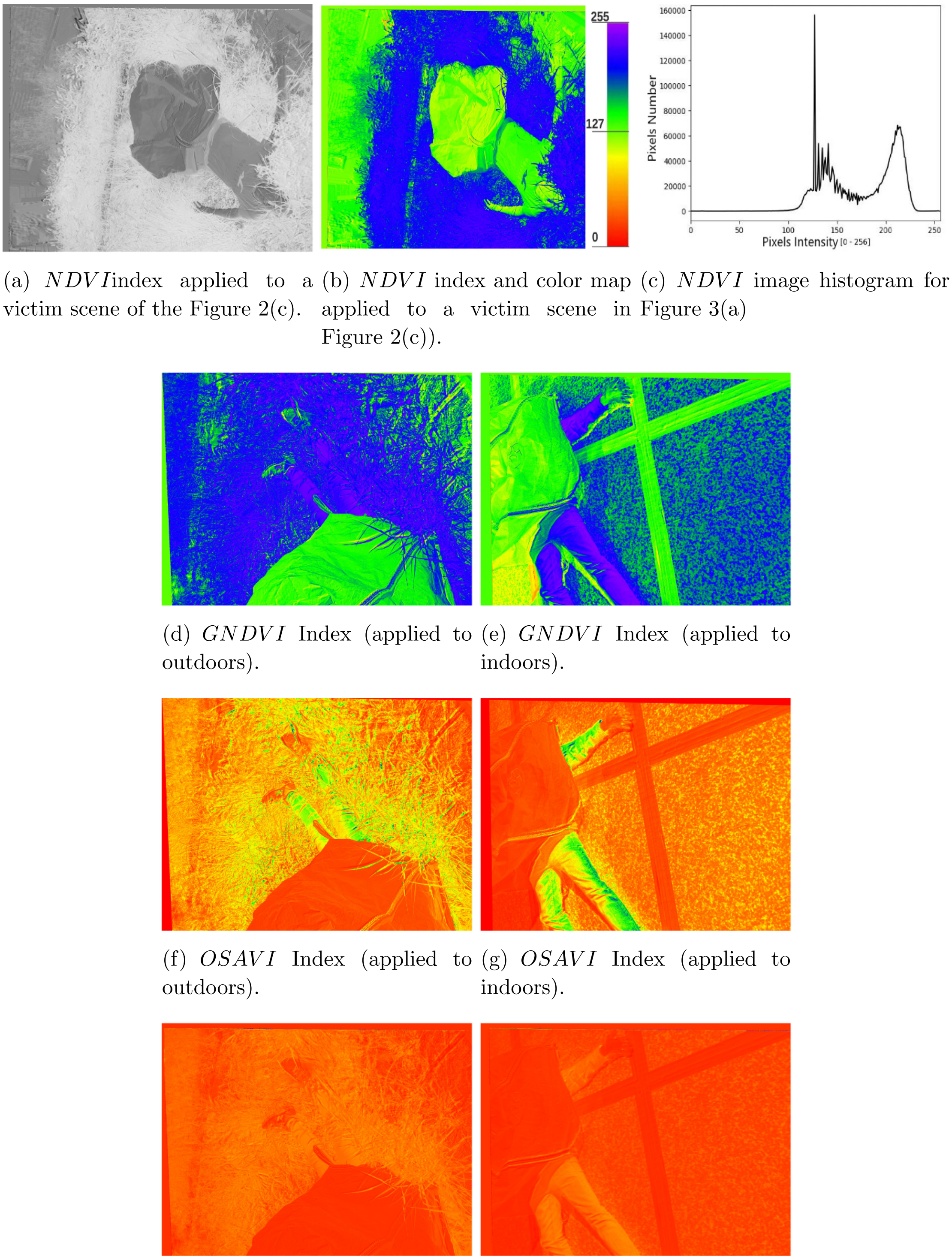
Once the methodology for adjusting the images has been determined, the existing indexes (applied to agriculture) are preliminary evaluated to determine if any of them can be used as a starting point for detecting victims. Table 1 describes the indexes used for this purpose. Works to detect people in the state-of-the-art directly use one of the Nir or RedEdge bands, but analysis of indoor-outdoor scenarios is not contemplated, as well as an adequate environment differentiation in the obtained image.

Figure 3 shows the corresponding application of the indexes presented in table 1 to indoor and outdoor scenarios.

The resulting image for the first case (NDVI) is show in figure 3(a) by using gray scale. For visualization purposes, a colour map has been applied, which assigns a false colour to each pixel intensity in the gray scale, as shown in figure 3(b). This has been applied to the rest of the images obtained from the indexes throughout the entire article.

A histogram is used to visualize the distribution to visualize the distribution of pixels throughout the image (figure 3(c)), which shows the number of pixels present in the image for each intensity value [0, 255] and allows to evaluate in a first inspection the concentration of pixels corresponding to the environment and the victim.

For instance, the grass present in the figure 3(b) obtains very high NDVI values, as expected.



(a) $NDVI$index applied to a victim scene of the Figure 2(c).

(b) $NDVI$ index and color map applied to a victim scene in Figure 2(c)).

(c) $NDVI$ image histogram for Figure 3(a)

(d) $GNDVI$ Index (applied to outdoors).

(e) $GNDVI$ Index (applied to indoors).

(f) $OSAVI$ Index (applied to outdoors).

(g) $OSAVI$ Index (applied to indoors).

(h) $SIPI$ Index (applied to outdoors).

(i) $SIPI$ Index (applied to indoors).

Figure 3. Common used vegetative indexes applied as a starting point in the identification of victims. Source: Authors.

The victim stands out against background, and her outline can be distinguished against the grass. However, it is not distinguishable from the bag used as rubble or the brick background. All of these elements have the same pixel intensity as the victim. For these reasons, NVDI cannot be considered as an appropriate index for detecting the victim in this set of images.

Something similar is shown for the figures 3(d)–(f), (i) and (h), where it is challenging to differentiate the environment of the victim. An improvement can be observed in figure 3(g), but the information corresponding to hands and the head has been lost.

Table 1. Common used vegetative indexes. Source: [4].

Index	Expression	Description
NDVI	$\frac{\rho NIR-\rho RED}{\rho NIR+\rho RED}$	Measures health, density and crop development
GNDVI	$\frac{\rho NIR-\rho GREEN}{\rho NIR+\rho GREEN}$	Sensitive to chlorophyll levels
OSAVI	$\frac{\rho NIR-\rho RED}{\rho NIR+\rho RED+0.16}$	Difference terrain from vegetation
SIPI	$\frac{\rho NIR-\rho RED}{\rho NIR+\rho RED+0.16}$	Estimate the relationship between carotenoids and chlorophyll

Table 2. Tested Indexes for victim identification. Source: Authors.

Index	Expression
Index 1	$\frac{\rho GRE-\rho RED+0.25*\rho NIR}{\rho GRE+\rho RED}$
Index 2	$\frac{\rho GRE}{0.5*\rho GRE+\rho RED+0.5*\rho NIR}$
Index 3	$\frac{\rho GRE+\rho RED-\rho NIR-\rho REG}{\rho GRE+\rho RED+\rho NIR+\rho REG}$
Index 4	$\frac{-\rho GRE-\rho RED+\rho NIR+\rho REG}{\rho GRE+\rho RED+\rho NIR+\rho REG}$
Index 5	$\frac{\sqrt{\rho NIR}}{\rho GRE+0.5*\rho REG}$
Index 6	$\frac{\rho GRE-1.5*\rho RED+\sqrt{\rho NIR}}{\rho GRE+\rho RED}$
Index 7	$\frac{\rho GRE-\sqrt{\rho RED+\rho NIR}}{\rho GRE+\rho RED}$
Index 8 (InVD)	$\frac{\rho GRE-\rho RED-\sqrt{\rho NIR}}{\rho GRE+\rho RED}$
Index 9	$\frac{\rho RED}{0.5*\rho GRE+0.5*\rho NIR+\rho RED}$
Index 10	$\frac{\rho GRE+\rho RED-\rho NIR-\rho REG}{\rho GRE+\rho RED+\rho NIR+\rho REG}$
Index 11	$\frac{-\rho GRE-\rho RED+\rho NIR+\rho REG}{\rho GRE+\rho RED+\rho NIR+\rho REG}$
Index 12	$\frac{\sqrt{\rho NIR}}{\rho GRE+0.5*\rho REG}$
Index 13	$\frac{\frac{\rho RED}{0.1*\rho GRE+\rho RED}}{\sqrt{0.1*\rho GRE+\rho RED}}$
Index 14	$\frac{\rho RED}{\rho GRE+\rho RED}$
Index 15	$\frac{\rho GRE-\rho RED+0.75*\rho REG}{\rho GRE+\rho RED}$

3.3.2. Proposed indexes

Based on the inefficiency obtained after applying conventional indexes to establish the first phase of separation of the victim and the environment, this section proposes a series of indexes that will be evaluated in the results chapter to determine their effectiveness.

Table 2 shows a list of the most outstanding indexes proposed empirically by applying different operations between the spectral bands: GRE, RED, NIR and REG. The BLUE band has been ruled out since the elements that make up the environment (floor, grass, bricks, walls) have a very low incidence of this band.

The best-determined index whose functionality turned out to be suitable both indoors and outdoors is $Index8$ or $InVD$ (Index for People Detection), which is determined based on the difference between the average values, the concentration of pixels for the victim, and its correct differentiation with the environment. The better indexes obtained after the evaluation are detailed in the section 4.1 of results.

3.4. Convolutional neural network training

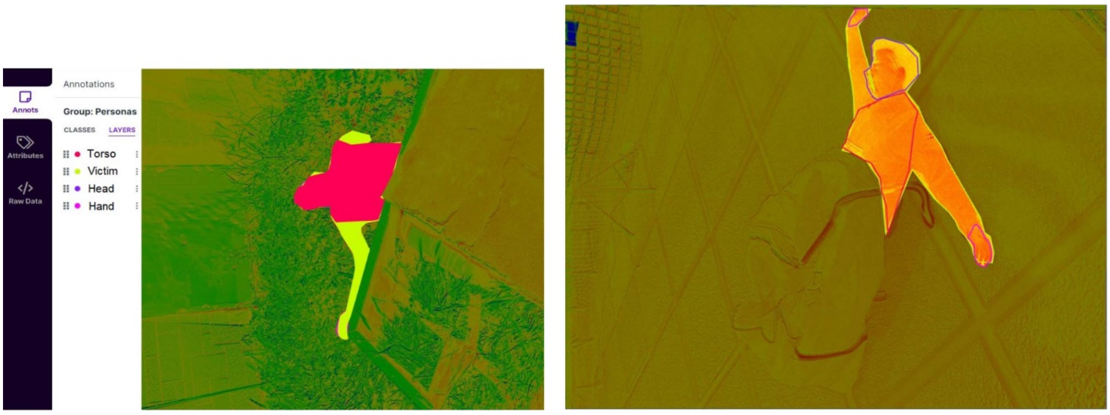
3.4.1. Data-set generation

After determining the combination of bands ($Index8$) that will be used to train the model,the index was calculated for all the images in the data-set. The total number of photographs is around 4200 between the five lenses, leaving us with 840 different scenarios.

The complete data-set used for this development is available in the authors' repository at https://drive.upm.es/s/xPKDp5Xyh1HTHWA.

A labelling phase has been carried out with the following tags types: Victim, Head, Torso, Leg, and Hand, for the training of the CNN. Figure 4 shows the tagging process for two images containing victims with exposed torsos.

Since this number of images could be considered short for training a CNN, a technique called 'Data Augmentation' was used to expand the data-set.



(a) Labeling of torso and head of a victim

(b) Labeling of a victim.

Figure 4. Victim tagging using Roboflow. Source: Authors.

This process involves taking already labelled images and applying them to disturbances such as rotations, blur effects, noise introduction, and cropping. The effects introduced include noise between 0–5% and clipping between 0–20% of the pixels. In this case, it is unfeasible to modify the colours since it would modify the previously extracted indexes. Versions of the original 840 images have been used to generate a data-set of 1454 images in this manner.

The data-set has been split up into three groups for training, validating and testing, according to the following percentages: Training 82%, Validation 12%, and Test 6%.

3.4.2. Convolutional neural network selection

The YOLO v5 Convolutional Neural Network is used for the automatic detection of victims, which was previously tested against other relevant models, such as Faster R-CNN or SSD; showing high performance. The performance of this comparison considering the older version YOLO v3 was presented in the previous work developed by the authors [26].

YoloV5 has four main models that have been trained with the COCO data-set [27], and they differ in the number of layers of the architecture and the depth of the dense layers. the four models of YOLOv5 are referred as ‘s’, ‘m’, ‘l’ and ‘xl’ according to the size (small, medium, large, xl).

Each of these models was trained to check which one performed better in terms of precision (in absolute value) and the precision-time relationship.

It is necessary to define specific parameters related to neural network learning to evaluate the mode’s effectiveness: True positive (TP): Accuracy in predicting the detection of the object. False positive (FP): Error in predicting an object that does not exist. True negative (TN): Success in predicting the non-detection of an object that does not exist. False negative (FN): Error in the prediction by not finding an object.

Once these metrics are known, efficiency meters are defined, such as the confusion matrix, in addition to the precision, recall and accuracy of the network described according to the equation (1), following [28].

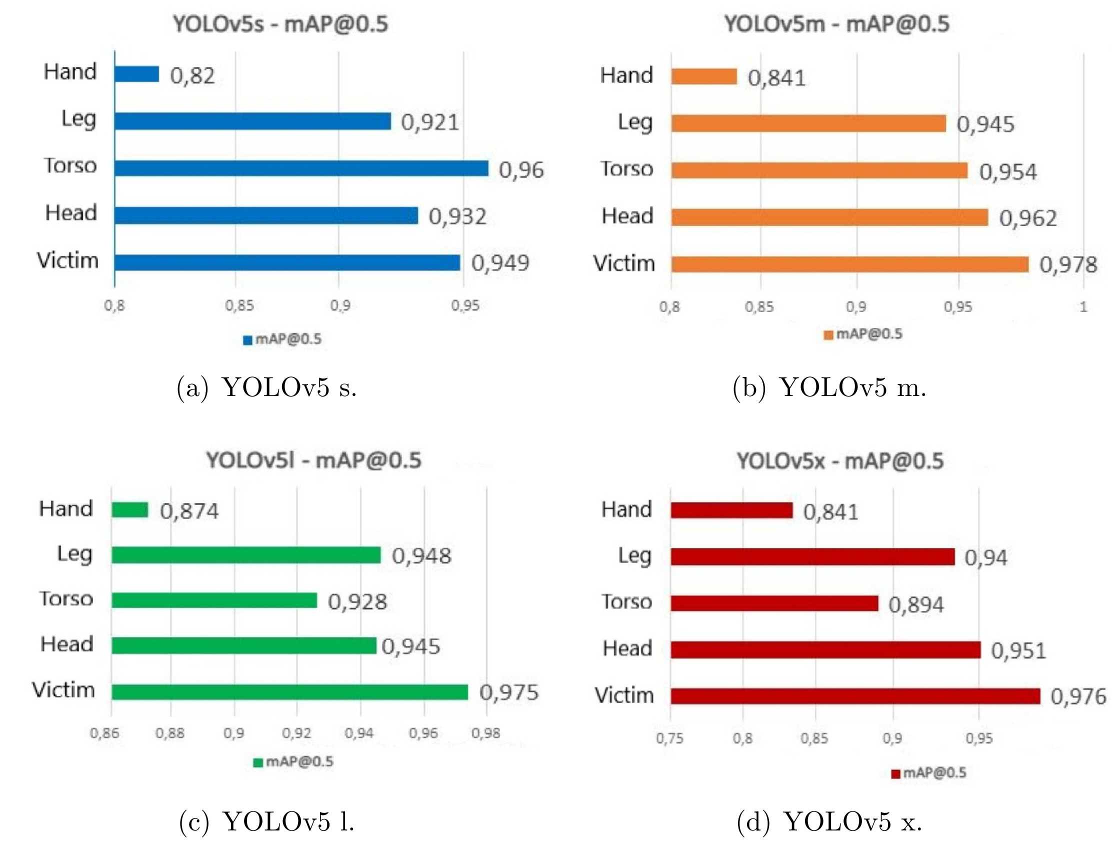
$$\text{Precision} = \frac{TP}{TP + FP} \tag{1a}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{1b}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1c}$$

The parameters for training the networks were: image input (resize to 928x928 pixels), batch 8, the epochs set were 300, and learning rate 0.001. With these values, the average times and epochs reached for the training of the different network models were: ‘s’ (2.5 h–200 epochs), ‘m’ (4.1 h–150 epochs), ‘l’ (5.2 h–125 epochs), ‘x’ (13 h–88 epochs). The mean average precision (mAP) values obtained during the evaluation are shown in figure 5.

Depending on the training, it can be established that for the different values of mAP, the lowest results were obtained in model ‘s’ (figure 5(a)). Although the detection results are comparable in the other three cases (figures 5(b)–(d)), training time has been prioritized to select the network, opting for the model ‘m’.



(a) YOLOv5 s.

(b) YOLOv5 m.

(c) YOLOv5 l.

(d) YOLOv5 x.

Figure 5. Evaluation of trained YOLOv5 models. Source: Authors.

If we analyze the structure of the network, first of all, it should be noted that the difference between the s, m, l, and x models lies in the number of layers and parameters used by the network. Thus the 's' and 'x' models have, respectively, 7.2 and 86.7 Million parameters (which correspond to the sum of elements in each layer of the model). For its part, the model m has 21.2 Million parameters, a quarter of the model 'x' and three times more than the 's' model. On the other hand, the Depth and Width (D, W) parameters of the network are directly influenced by the scale of the number of filters and layers of the network. In this sense, the model 's' has (0.33, 0.5), 'm' (0.67, 0.75), 'x' (1.33 , 1.25). In this context, the model 'm' shows to have a number of parameters and layers greater than 's' and less than 'l' and 'x', which allows optimization of inference time and optimization of computational resources.

4. Results and discussion

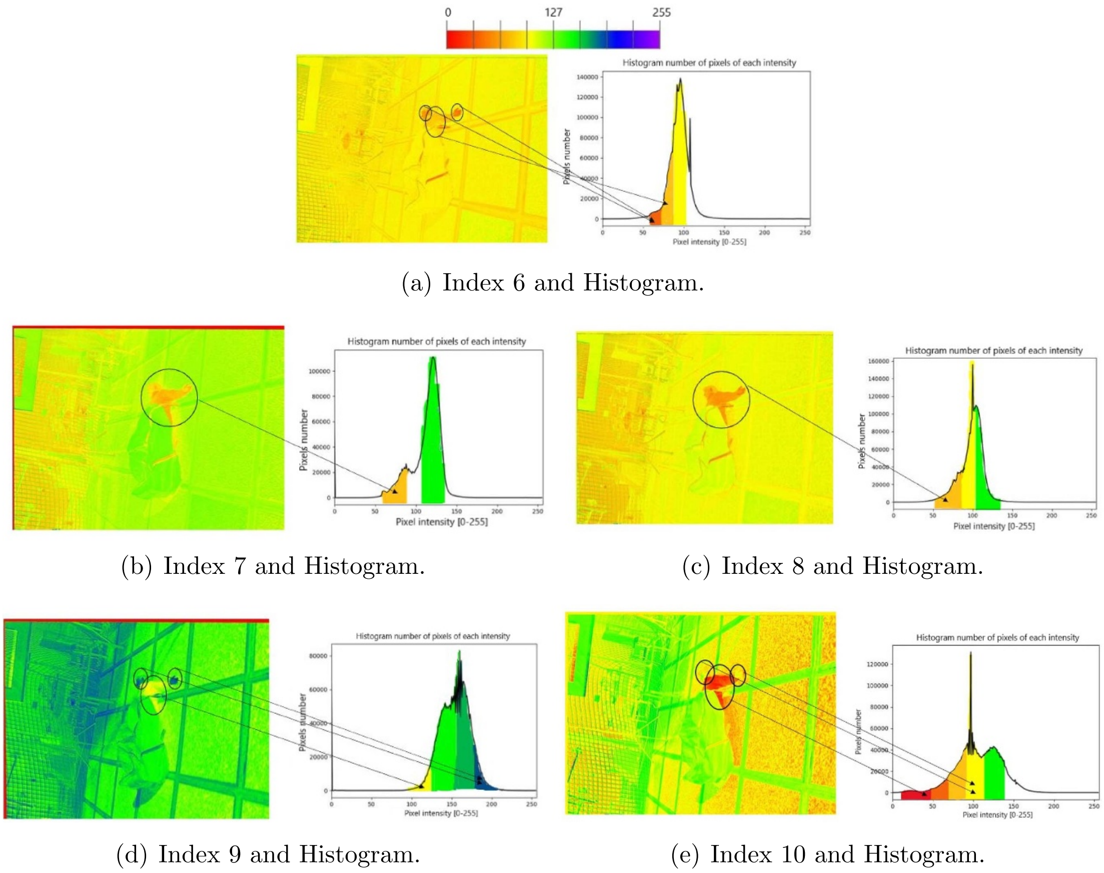
4.1. Analysis for the best proposed index

For this phase, 20 representative images of the data-set have been defined (indoors, outdoors, with and without occlusions). The indexes of table 2 have been evaluated by using these images and discarding by visual inspection those indexes that did not show effective results to highlight representative characteristics of the victim concerning the environment (Indexes 1, 2, 3, 4, 5).

Histograms have been extracted from the remaining ten indexes as an auxiliary tool to quantify the difference in pixel intensity between the victim and the environment. Figure 6 shows a sample of the indexes 6, 7, 8, 9, 10 applied to indoors scene (taken in figure 1(b)). From these data, table 3 has been generated, which is representative of the process carried out. That table shows the intensity of the mean colour of the victim and the environment, as well as their difference divided by 255 (maximum possible victim-environment difference).

The complete results (images and histograms) of this process obtained from the indexes [6–15] are available in the authors' GitHub repository https://github.com/ChristyanCruz11/Multispectral-Images.git.

After applying this procedure to the 20 sample images for ten indexes, 200 results were obtained, which can be observed in table 4. The analysis of the results highlights that *Index*8 is the one that presents the best results in terms of consistency, despite not being the best either in the open or in the closed scenarios separately. Figure 7 shows the quantification of this table's total.



(a) Index 6 and Histogram.

(b) Index 7 and Histogram.

(c) Index 8 and Histogram.

(d) Index 9 and Histogram.

(e) Index 10 and Histogram.

Figure 6. Evaluation of the relative difference between victim and environment for indexes 6–10. Source: Authors.

Table 3. Evaluation of the relative difference between victim and environment for indexes 6–10 from figure 6. The mean value of the pixels colour have been obtained from the histograms in figure 6.

Ind	Victim colour mean pixels value	Environment colour mean pixels value	Difference Env-Vict	Relative diff / 255 %
6	64	96	32	12,54
7	64	127	63	24,70
8	64	96	32	12,54
9	96	127	31	12,1568
10	96	96	0	0

Compared to the rest of indexes, this index (*InV D*) allows distinguish the victim from the environment in all the foreseen scenarios and presents the best total value of relative difference. It should be noted that although the value of 18.6% of Relative Difference obtained for the best index may seem low, it must be taken into account that it is more interesting to find which index has the highest value rather than the value itself. Thus, if a full scale (100%) were used, a 255 difference in pixel intensity would be obtained. This is equivalent to having the victim represented in pure white in gray scale and the one around her in black, or vice versa, values unattainable in practice. It is worth noting that, the most restrictive value has been used to calculate this difference in cases where the victim has different pixel intensities.

Accordingly, we conclude after performing the analysis that *Index*8 or *InV D* is the most appropriate for CNN training.

4.2. Analysis of the neural network effectiveness

Figure 8(a) shows the confusion matrix for the selected CNN model and images generated with *InV D*. Where the main diagonal values are high, reflecting a high confidence level for the detections, in the True Positive. This graph allows us to establish that the network can effectively classify the objects in the corresponding classes. Among the values eternal to the diagonal, we observed a greater error in the detection and differentiation of objects concerning the landscape than in their classification. The most difficult class to detect is *Hand*, probably due to its reduced size and morphology.

Table 4. Table of relative differences divided by 255, between victim and environment, 20 images of indoors (grey) and outdoors (blue) have been selected for the application of the indexes, the resulting images and histograms can be found in https://github.com/ChristyanCruz11/Multispectral-Images.git. Source: Authors.

	Relative difference / 255 (%), between victim and environment									
	In6	In7	In8	In9	In10	In11	In12	In13	In14	In15
Img 0	12,54	24,70	12,54	12,15	0	0	0	12,53	25,09	12,55
Img 1	12,15	24,71	12,15	0	12,16	12,55	0	0	12,54	12,54
Img 2	37,25	24,7	37,25	24,70	37,25	25,09	12,54	24,70	24,71	25,09
Img 3	0	12,54	12,54	0	12,15	0	12,55	12,15	12,55	12,56
Img 4	24,70	0	12,55	24,70	25,09	12,55	0	24,70	12,54	24,71
Img 5	12,54	12,55	12,54	0	0	0	0	12,55	12,54	12,54
Img 6	12,54	24,71	24,70	0	0	12,55	12,54	12,54	12,54	12,54
Img 7	12,54	0	12,54	12,55	0	0	12,55	12,54	25,09	0
Img 8	24,70	12,54	24,70	24,70	12,54	25,09	24,70	24,71	25,09	12,54
Img 9	37,25	25,09	37,26	49,80	12,54	12,55	12,54	24,70	25,09	12,54
Img 10	0	24,70	12,54	24,70	25,09	25,09	37,25	24,70	0	0
Img 11	12,54	24,70	24,70	0	0	0	12,54	12,54	12,55	12,55
Img 12	12,15	24,70	24,70	0	0	0	24,70	0	12,54	0
Img 13	12,54	12,55	12,54	12,55	12,15	0	0	12,54	0	12,54
Img 14	12,54	0	12,55	12,54	0	0	12,55	24,70	25,09	0
Img 15	24,70	0	24,71	37,25	25,09	12,54	12,55	24,70	25,09	12,54
Img 16	0	24,70	12,54	0	0	0	0	12,54	12,54	0
Img 17	24,70	12,15	12,15	49,80	12,54	12,54	0	24,70	24,70	24,70
Img 18	0	24,7	12,15	0	0	0	24,70	12,15	0	12,54
Img 19	24,70	24,71	24,70	0	0	24,70	0	12,54	12,55	12,54
Mean Value Indoor	9,96	21,05	12,15	4,94	6,15	7,49	7,49	12,47	11,29	10,03
Mean Value Outdoor	21,05	12,39	21,05	23,60	12,50	10,03	13,68	19,76	20	12,47
Difference	15,50	16,72	18,60	14,27	9,33	8,76	10,58	16,11	15,64	11,25

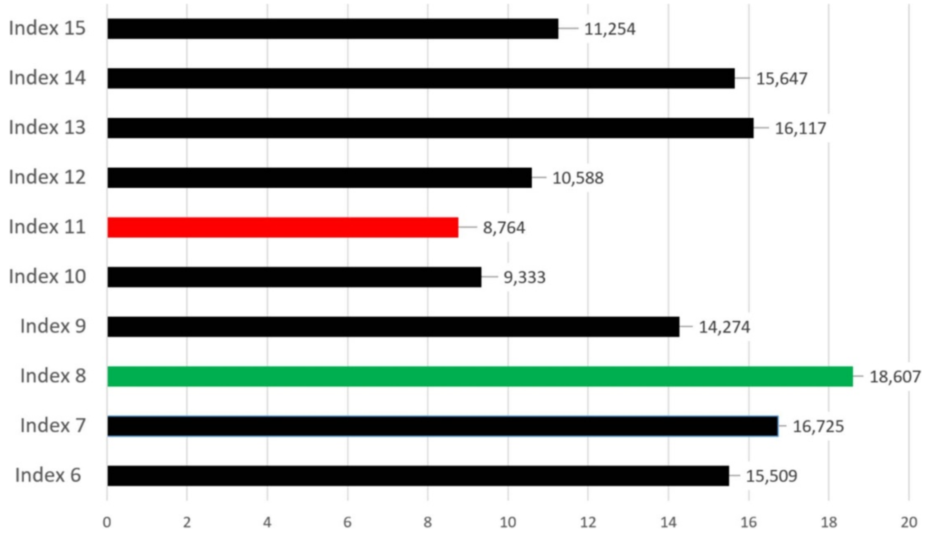
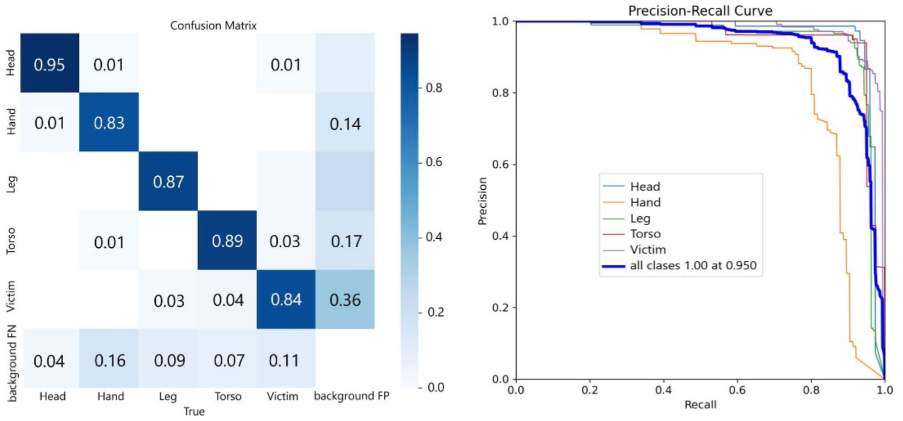


Figure 7. Graph of relative differences between victim and environment in the set (evaluated indoors and outdoors). Source: Authors.

Figure 8(b) shows the Precision-Recall curve, where precision degrades for mean values of 0.81. This curve encloses a large amount of the domain under its area, approaching the rectangle for all classes except 'Hand'.

4.3. Automatic victims detection

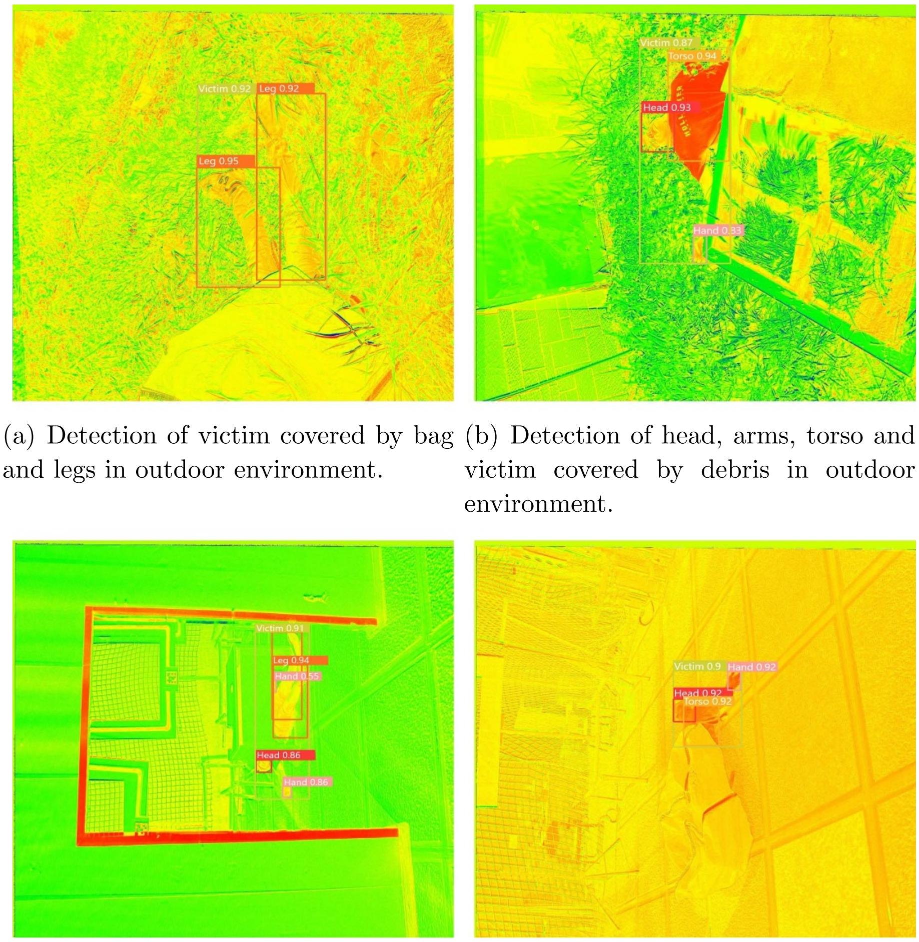
Once the trained neural network model has been obtained, the inference has been made on new images processed with the *Index*8.



(a) Confusion matrix for the trained YOLOv5m model- Source: Authors.

(b) Precision-Recall curve for the trained YOLOv5m model - Source: Authors.

Figure 8. Evaluation of the best model (YOLOv5m) obtained and used for inference. Source: Authors.



(a) Detection of victim covered by bag and legs in outdoor environment.

(b) Detection of head, arms, torso and victim covered by debris in outdoor environment.

(c) Victim detection (all tags) partially covered indoors.

(d) Indoors partially covered victim detection.

Figure 9. Detection of victims in different environments in post-processed images with *Index*8. Source: Authors.

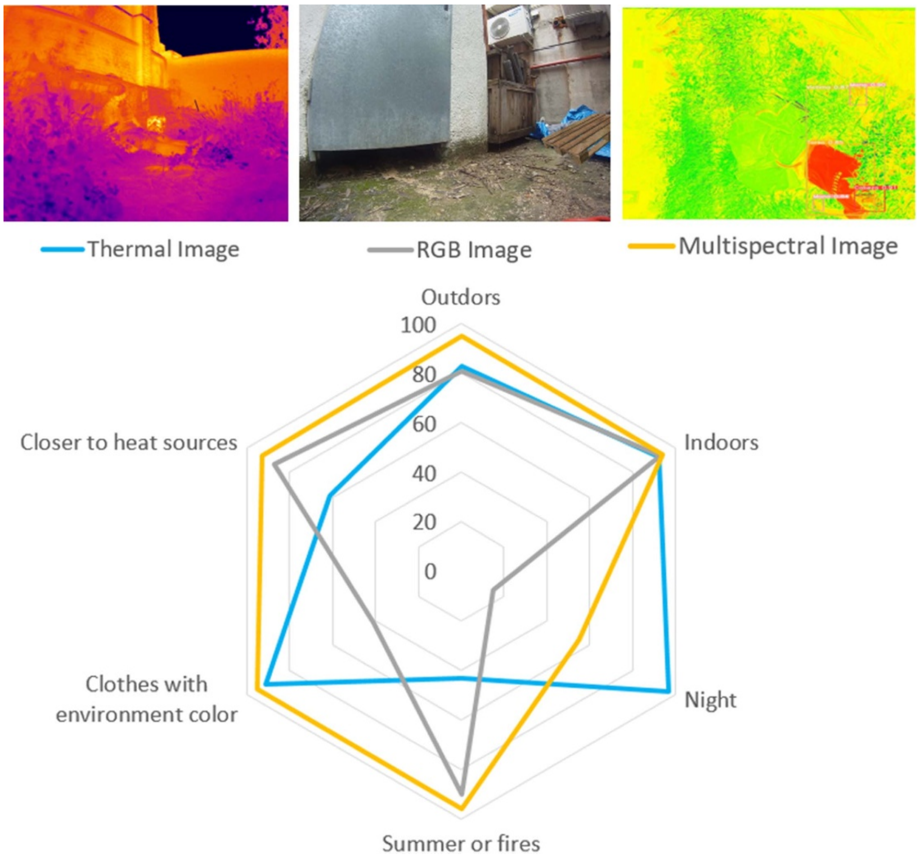


Figure 10. Comparison between the proposed method (Multispectral) versus others developed in the state of the art (Thermal and RGB) for victims detection. Source: Authors.

Figure 9 shows the detection result on the new images. Bounding boxes with different colours have been placed on the parts of the identified victims. The labels and the detection percentage [0-1] have been placed along with the bounding boxes.

Figures 9(a) and (b) show outdoor scenarios. The first case corresponds to a victim whose torso has been covered, and her legs have been correctly identified with an efficiency greater than 92%. In the second case, a victim is shown where her legs have been covered by rubble. The classes head, torso, victim and hand are correctly identified. Of these classes, the best identified are head and torso.

Figures 9(c) and (d) correspond to indoors scenarios. In the first case, a covered victim is shown in the torso area. In this case, all the labels with an average detection efficiency of 89% and a false correspondence of 'hand' have been identified. The second case shows a correct identification of the classes with an average efficiency greater than 91%.

Due this camera model (and with some others that we have worked with the Parrot Sequoia) does not allow real-time image transmission, in the case of the MicaSense, although it connects to its address web, the protocol is closed, and a minimum of 30 fps is not achieved for real-time processing.

4.4. Comparison with previous methods for victims detection

In this section, a comparison will be made between conventional methods for victim detection (Thermal Imaging and RGB) and the proposed method (Multispectral). In previous work, the authors established a first comparison between the method based on the thermal image and RGB, where they established as first conclusions that the thermal image is better under low light conditions, as well as to identify victims behind thin objects. However, it had disadvantages when there were heat sources next to the person [26].

In this next phase, the multispectral image is introduced into the comparison. Figure 10 shows a radial comparison of potential conditions that may occur and may represent a challenge when detecting a victim, which has been evaluated in percentage terms.

One of the remarkable points of the proposed method over the others is the versatility for detecting victims indoors and outdoors, thanks to the exhaustively selected index. Another advantage to highlight is the identification of victims near heat sources, for which or in specific seasonal conditions (fire). Figure 10 shows a thermal, RGB and multispectral image of a person simulating being a victim outdoors in its upper part. In the first case, detection is difficult since the person's heat signature is confused with the environment.

The RGB Method is directly influenced by the clothing and colour of the environment, so if the person had little distinctive clothing, it could not be correctly identified with this method. However, the multispectral method respond adequately in detection under these conditions.

The proposed hyperspectral method and the method based on thermal imaging are the most notable in this comparison. However, the latter has a single measurement spectrum, while the multispectral has a wide range of combinations.

5. Conclusions

This work demonstrates the effectiveness of multispectral images for detecting victims based on the combination of their different spectral bands captured by a multispectral camera—the application of new indexes proposed by the authors and the convolutional neural networks.

After trying different combinations of the multispectral bands (indexes) and carrying out a qualitative and quantitative analysis, the proposed index best differentiates a victim from the environment (indoors and outdoors) was *InV D*. Which involves the bands (GRE, RED, NIR) from which the data-set images were generated for the neural network training.

$$InV\,D = \frac{\rho GRE - \rho RED - \sqrt{\rho NIR}}{\rho GRE + \rho RED}. \tag{2}$$

Multispectral cameras are presented as an alternative to thermal and RGB technologies for victim detection, thanks to the versatility offered by generating an indefinite number of versions of each original image by combining their spectral bands, presenting advantages for victims detection especially outdoors compared to RGB and thermal cameras.

The Convolutional Neural Network model selected was YOLOv5m since it provides with the best precision values for the data-set used in the analysis and the selected parameters. The YOLOv5l and YOLOv5x models obtain worse metrics and training times. Despite their higher complexity, the 'Victim' class obtains the best results in all the trained detection models with values around 0.92 from mAP@0.5.

As future works, it is proposed to execute the process in real-time (since due to the size of the images this is still complex), optimizing the capture and image size, executing the processing on board the explorer SAR robot and issuing early warning alarms when victims are detected.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

Special thanks to the LPF-Tagralia UPM Research Group for providing us with the Mica Sense Multispectral camera, without this camera, this study would not have been possible.

Funding

This work has received funding from the RoboCity2030-DIH-CM Madrid Robotics Digital Innovation Hub 'Robótica aplicada a la mejora de la calidad de vida de los ciudadanos, fase IV'; S2018/NMT-4331), funded by 'Programas de Actividades I+D en la Comunidad de Madrid' and cofunded by Structural Funds of the EU. and TASAR (Team of Advanced Search And Rescue Robots) PID2019-10 5808RB-I00, funded by "Proyectos de I+D+i del Ministerio de Ciencia, Innovacion y Universidades.

ORCID iD

Christyan Cruz Ulloa https://orcid.org/0000-0003-2824-6611

References

[1] United Nations Office for Disaster Risk Reduction 2019 UNDRR Home (available at: www.undrr.org/) (Accessed 6 October 2022)
[2] Barrientos A 2022 TASAR - Team of advanced search and rescue robots (available at: www.car.upm-csic.es/?portfolio=tasar)
[3] Noguera M *et al* 2021 Nutritional status assessment of olive crops by means of the analysis and modelling of multispectral images taken with UAVs *Biosyst. Eng.* **211** 1–18

[4] Cardim Ferreira Lima M, Krus A, Valero C, Barrientos A, del Cerro J and Roldán-Gómez J J 2020 Monitoring plant status and fertilization strategy through multispectral images *Sensors* **20** 435

[5] Cruz Ulloa C, Krus A, Barrientos A, del Cerro J and Valero C 2022 Robotic fertilization in strip cropping using a CNN vegetables detection-characterization method *Comput. Electron. Agric.* **193** 106684

[6] Konig D, Adam M, Jarvers C, Layher G, Neumann H and Teutsch M 2017 Fully convolutional region proposal networks for multispectral person detection *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops* pp 116–28

[7] Takumi K, Watanabe K, Ha Q, Tejero-De-Pablos A, Ushiku Y and Harada T 2017 Multispectral object detection for autonomous vehicles *Proc. of the on Thematic Workshops of ACM Multimedia 2017 (Thematic Workshops '17)* (New York: Association for Computing Machinery) pp 35–43

[8] Chen Y and Shin H 2020 Multispectral image fusion based pedestrian detection using a multilayer fused deconvolutional single-shot detector *J. Opt. Soc. Am.* A **37** 768–79

[9] Fawakherji M, Potena C, Pretto A, Bloisi D D and Nardi D 2021 Multi-spectral image synthesis for crop/weed segmentation in precision farming *Robot. Auton. Syst.* **146** 103861

[10] Navulur K 2006 *Multispectral Image Analysis Using the Object-Oriented Paradigm* (Boca Raton, FL: CRC press)

[11] Kumar M and Singh R K 2013 Digital image processing of remotely sensed satellite images for information extraction *Proc. Conf. on Advances in Communication and Control Systems (CAC2S 2013)* (Atlantis Press) pp 406–10

[12] Huang X, Wen D, Xie J and Zhang L 2014 Quality assessment of panchromatic and multispectral image fusion for the ZY-3 satellite: from an information extraction perspective *IEEE Geosci. Remote Sens. Lett.* **11** 753–7

[13] Townshend J and Justice C 1981 Information extraction from remotely sensed data *Int. J. Remote Sens.* **2** 313–29

[14] Paoletti M E, Haut J M, Plaza J, Plaza A Comparative A 2019 Study of techniques for hyperspectral image classification *Revista Iberoamericana de Automática e Informática industrial* **16** 129–37

[15] Landgrebe D 1998 Information extraction principles and methods for multispectral and hyperspectral image data *Information Processing for Remote Sensing* (Dartmouth: World Scientific)

[16] Serpico S B, Dellepiane S, Boni G, Moser G, Angiati E and Rudari R 2012 Information extraction from remote sensing images for flood monitoring and damage evaluation *Proc. IEEE* **100** 2946–70

[17] Hong K, Liu X, Liu G and Chen W 2019 Detection of physical stress using multispectral imaging *Neurocomputing* **329** 116–28

[18] Al-Temeemy A A 2019 Multispectral imaging: monitoring vulnerable people *Optik* **180** 469–83

[19] Rowe R K, Corcoran S P, Nixon K A and Ostrom R E 2005 Multispectral imaging for biometrics *Proc. SPIE* **5694** 90–99

[20] Baluja J, Diago M P, Balda P, Zorer R, Meggio F, Morales F and Tardaguila J 2012 Assessment of vineyard water status variability by thermal and multispectral imagery using an unmanned aerial vehicle (UAV) *Irrig. Sci.* **30** 511–22

[21] Diago M P, Tardaguila J, Barrio I and Fernández-Novales J 2022 Combination of multispectral imagery, environmental data and thermography for on-the-go monitoring of the grapevine water status in commercial vineyards *Eur. J. Agron.* **140** 126586

[22] Lu N *et al* 2019 Estimation of nitrogen nutrition status in winter wheat from unmanned aerial vehicle based multi-angular multispectral imagery *Front. Plant Sci.* **10** 1601

[23] Romero M, Luo Y, Su B and Fuentes S 2018 Vineyard water status estimation using multispectral imagery from an UAV platform and machine learning algorithms for irrigation scheduling management *Comput. Electron. Agric.* **147** 109–17

[24] Nataprawira J, Gu Y, Goncharenko I and Kamijo S 2021 Pedestrian detection using multispectral images and a deep neural network *Sensors* **21** 7

[25] Krus A, Valero C, Ramirez J, Cruz C, Barrientos A and del Cerro J 2021 Distortion and mosaicking of close-up multi-spectral images *Precision Agriculture'21* (Netherlands: Wageningen Academic Publishers) pp 33–46

[26] Cruz Ulloa C, Prieto Sánchez G, Barrientos A and Del Cerro J 2021 Autonomous thermal vision robotic system for victims recognition in search and rescue missions *Sensors* **21** 7346

[27] Lin T *et al* 2014 Microsoft COCO: common objects in context *CoRR* (arXiv:1405.0312)

[28] Goutte C and Gaussier E 2005 A probabilistic interpretation of precision, recall and F-score, with implication for evaluation *Advances in Information Retrieval D E* Losada and J M Fernández-Luna (Berlin: Springer) pp 345–59