5.2.5 Comparativa frente a imagenes RGB

The previously trained CNN Model 1 from Section 5.1 was employed to conduct this comparative analysis. RGB and thermal images were captured from the same perspective under varying lighting and environmental conditions to conduct this comparison. This approach ensured a consistent spatial viewpoint while introducing diversity in the illumination and environmental factors, enabling a robust evaluation of the model's performance across different scenarios.

In Figure 5.14-a-f, the assessment of victim detection in various scenarios is observed. Cases 1 and 2 depict situations with inadequate illumination, where victim detection fails for the RGB method, as it relies on adequate lighting conditions. Conversely, the thermal method remains unaffected and can identify victims without issues.

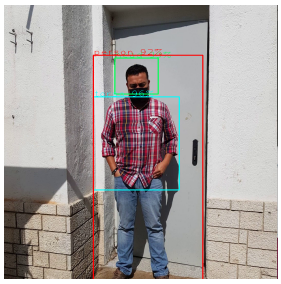
Case 3 presents a scenario where a first-responder is positioned in front of a material with high emissivity that has accumulated heat over a while, as defined in the materials analysis section of this chapter. This heat accumulation poses challenges for accurately measuring a person, rendering the thermal method unviable in this particular case. However, this differs from the RGB method, which successfully identifies the first-responder.

(a) Case 1: Bad detection of RGB method in absence of light

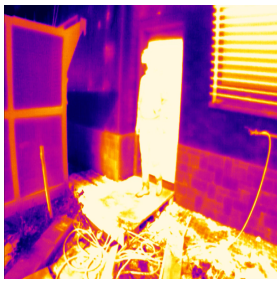
(b) Case 1: Good detection of Thermal method in absence of light

(c) Case 2: Bad detection of RGB method in low light

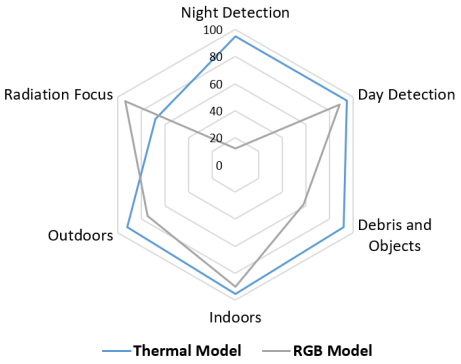
(d) Case 2: Good detection of Thermal method in low light



(e) Case 3: Good detection of people in front of heat sources for RGB method



(f) Case 3: Bad detection of people in front of heat sources for Thermal method



(g) Comparative radial graph of the RGB and thermal methods.

Figure 5.14: Assessment of the thermal and RGB methods in diverse scenarios.

In Figure 5.14-g, the radial graph presents a comprehensive visualization of the percentage parameters derived from practical data, serving as a comparative analysis between RGB and thermal methodologies for victim detection.

It can be inferred that, particularly in environments characterized by poor or absent illumination, the thermal method demonstrates superior efficiency compared to the RGB one. Nevertheless, a notable drawback of the proposed method becomes apparent when confronted with substantial heat sources. Consequently, a potential remedy would involve integrating both methods to establish a more resilient system capable of mitigating disruptions posed by such heat sources.

5.2.6 Conclusion

The proposed method has demonstrated robustness in victim detection based on thermal images and CNNs, considering a broad range of scenarios encompassing both indoor and outdoor environments and different environmental and lighting conditions. It achieved an average efficiency of 90% in detecting victims during the experimental phase.

The emissivity of materials plays a significant role in enhancing the detection of victims under conditions of zero visibility. This property enables specific materials, such as plastics or thin fabrics (materials with low emissivity), to allow the passage of radiation emitted by covered people, thereby facilitating their detection. This effect becomes particularly valuable in scenarios where traditional visual detection methods could be more effective due to a complete or severe lack of visibility.

Regarding the conditions under which victims are best detected, the CNN trained with the day dataset has exhibited better detection performance, yielding a 12% higher mean Average Precision (mAP) against the nighttime dataset and an 8% higher mAP compared to the combined dataset. Furthermore, it maintains nearly identical recall rates to the other datasets while achieving an 88% reduction in loss compared to the nighttime dataset and a 72% reduction compared to the combined dataset. Although CNN trained with the combined dataset shows lower values for these parameters, their increased robustness, allowing them to operate effectively in both nighttime and daytime environments, position them as viable and practical alternatives.

The classes 'victim,' 'head,' and 'leg' are the most effectively detected, achieving class precision levels of around 90%. Nevertheless, the 'arm' and 'torso' classes are correctly detected in most instances, with class precision levels of approximately 60%. The method that evaluates the length-to-width ratio within the 'victim' class to differentiate between first-responders and victims proves to be a viable approach, yielding satisfactory results.

5.3 Victims Detection in Multispectral Imagery

A third phase delved into the domain of victim detection embarks on a novel hypothesis: the feasibility of discerning victims from images acquired through a multispectral range. This distinctive approach not only encompasses a single band, as in alternative methodologies, but rather encompasses diverse bands. The principal bands enlisted encompass Green, Red, Red Edge, and Near-Infrared (NIR) wavelengths:

- Blue $[440nm - 510nm]$
- Green $[520nm - 590nm]$
- Red $[630nm - 685nm]$
- Red Edge $[690nm - 730nm]$
- NIR $[750nm - 2.5\mu m]$

Figure 5.15 visually represents the organized disposition of distinct spectral bands across the wavelength spectrum. Multispectral imagery has been predominantly channelled towards agricultural investigations, which serve as a powerful tool for assessing the vegetative condition of plants. The NDVI (Normalized Difference Vegetation Index) is an established metric that relies on the synergy between the Near-Infrared (NIR) and Red spectral bands defined by $(\frac{\rho NIR - \rho RED}{\rho NIR + \rho RED})$, offering insights into the health and vitality of vegetation.

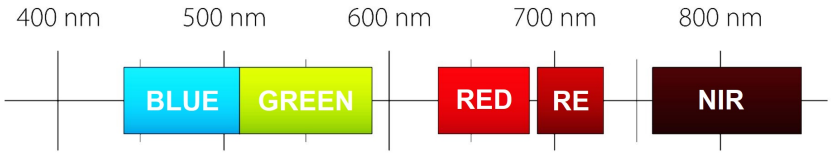


Figure 5.15: Arrangement of bands according to the spectral ranges.

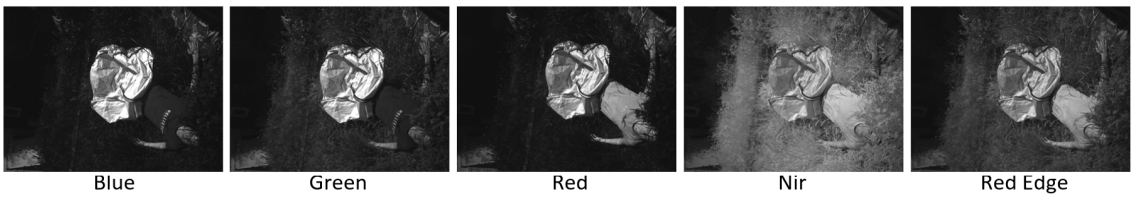
Nevertheless, the application of multispectral imaging has been expanded in recent years, with efforts extending towards exploring the Near-Infrared (NIR) range and the Red Edge band for victim detection purposes. While these initiatives have furnished partially conclusive outcomes, the impetus from these efforts has invigorated our pursuit to delve further. Hence, our research efforts aim to propose a pioneering index tailored to enhance victim detection's effectiveness within the multispectral imagery framework.

An experimental approach has been established, utilizing the MicaSense Altum camera to capture the predefined images across diverse spectral ranges. The intrinsic attributes of the resulting multispectral images have been examined through an investigation involving qualitative and quantitative dimensions. The qualitative dimension involves histogram analysis to decode pixel intensity distribution. The quantitative aspect quantifies intensity levels and pixel concentrations within each spectral band. The culmination of this integrated methodology facilitates the identification of an optimal index endowed with the capacity to identify victims efficiently across both indoor and outdoor scenarios.

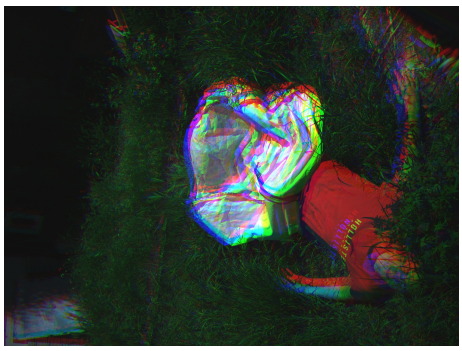
5.3.1 Images adjust and conbination

The initial phase preceding processing encompasses the critical task of image calibration. Unlike RGB or thermal cameras equipped with single lenses to capture specific spectral bands, multispectral cameras have more than one; the used camera has six lenses. This demands a pre-processing phase to align the images. Figure 5.16-a shows the images captured by the camera in its different bands. The comprehensive dataset of images compiled from indoor and outdoor environments is presented in the repository `https://drive.upm.es/s/xPKDp5Xyh1HTHWA`.

On the other hand, Figure 5.16-b is the result of the combination of R-G-B channels in the original capture state. As becomes apparent when looking at the edges that there is an irregularity, and the colour is not uniform because the RGB channels have yet to be adjusted (the image already adjusted in Figure 5.16-c is shown as a reference). Ignoring this effect and trying to combine the images without prior adjustment produces distortion in the final images. The edges of the objects are blurred, and their delimitation needs to be clearer. These effects can make it extremely difficult to recognize victims even if a human, not a neural network, is evaluating the photographs. For a convolutional neural network, distortion is even more detrimental since edge detection plays an important role, serving as a basis for learning more complex patterns and shapes.



(a) Multispectral bands captured by the ALTUM camera.



(b) Unadjusted RGB image of victim.

(c) Adjusted RGB image of victim.

Figure 5.16: Example of data-set images with and without adjustment.

Adjusting the images involves transforming the photographs captured by each lens to align with one chosen as a reference. Multispectral cameras are typically designed for drone mounting, capturing crop imagery from above at a consistent altitude. Adjustment becomes straightforward when all images are taken with a uniform camera-to-object distance. Selecting one lens as the reference, a trial-and-error approach is used to find pixel combinations aligning images from other lenses with this fixed reference. This yields coordinate values (x, y) representing displacement for each lens.

This procedure cannot be applied since the distance between the camera and the victims is not constant for the entire dataset. The photographs have been taken from different angles

and points of view, making it impossible to use the usual adjustment method since there is no fixed pixel value; it varies for each image. The solution lies in finding a way to calculate this number of pixels for each set of 6 images (one for each lens). For this, it has been necessary to create an algorithm that is capable of reading the images, understanding that the content of the image is the same in each photograph (it is displaced), finding points in common between them, and finally finding the values that make the images match.

The empirical algorithm SIFT (Scale-Invariant Feature Transform) belonging to the OpenCV library will be used to adjust the images. This algorithm tries to find pixels in the images representing elements invariant to image disturbances such as scale changes, distortions, rotations, etc. These elements will be called 'key points' [40]. The algorithm consists of applying a series of Gaussian filters to the images defined by equation 5.2:

$$G(x, y, k\sigma) = \frac{1}{2\pi(k\sigma)^2} e^{\frac{-(x^2+y^2)}{2k^2\sigma^2}} \tag{5.2}$$

Where (x, y) are the image pixels and $k\sigma$ is the blur value, the optimal values obtained experimentally being $\sigma = 0.6, k = \sqrt{2}$.

Once the key points and descriptors of the fixed and mobile image have been determined, an algorithm called 'matcher' is used, which, based on the descriptors, finds the key points of the two images that represent the same characteristic point to adjust an image to the other. The matcher "FLANN matcher" has been used, which goes through the key points looking for the minimum distance [287]; later, the best matches are selected, and the key points of the still and moving images are stored in a vector to calculate the Homography –a matrix of 3x3 transformation (translation, rotation)– from the mobile image to the fixed one. Figure 5.16-a shows the matches found by the algorithm for the RED and NIR bands of Figure 5.17; in green, the best key points from which the homography will be calculated are shown. While Figure 5.16-c results from the adjustment and overlapping of the R-G-B images.

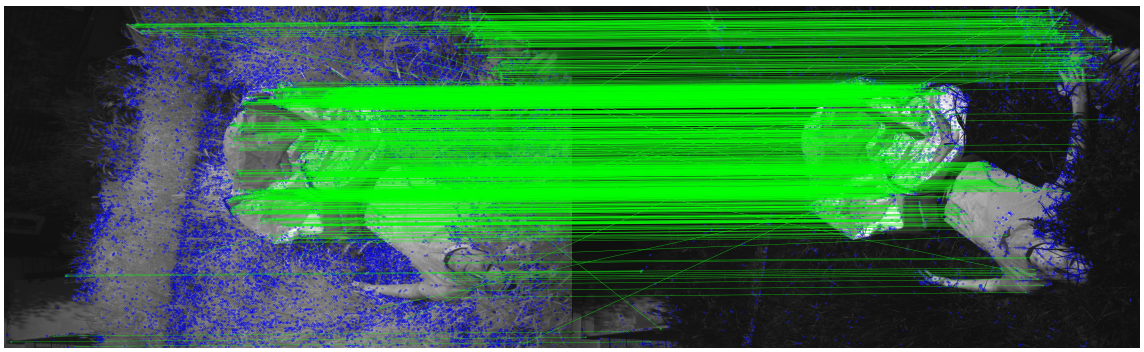
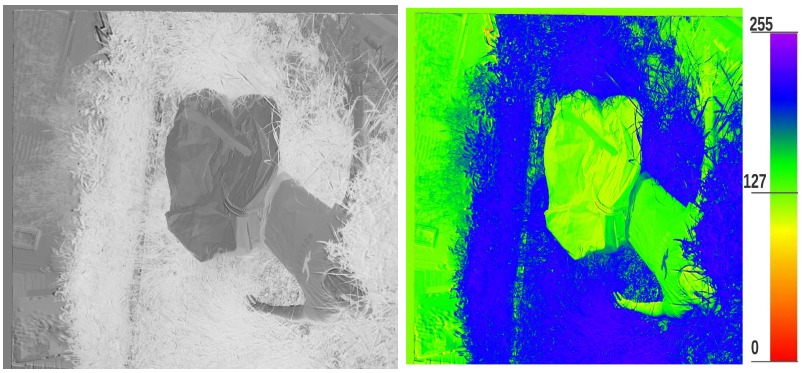


Figure 5.17: Flann matcher application to relate the key points of two images

The subsequent stage pertains to integrating the appropriately adjusted images into a composite image corresponding to the respective index (whether pre-existing like NDVI or novel). For this purpose, the images are treated as arrays of floating-point vectors, facilitating operations involving decimal values. As an illustrative instance, the combination will be applied to one of the established indices in the state of the art, the NDVI ($\frac{\rho NIR - \rho RED}{\rho NIR + \rho RED}$). The outcome of this operation must subsequently be scaled within the interval [0 - 255].

In this manner, all pixels are confined within the desired interval. The remaining step involves rounding the values to integers, considering that NDVI was originally a vector of floats, and the index computation operation typically entails inherent imprecision. A colour map is employed to enhance the visual representation of the index, which assigns a pseudocolour to each pixel intensity within the grayscale spectrum. This augmentation increases the image's depth, endowing it with three colour channels. The colour rendition imparted by this map concerning pixel intensities can be observed in Figure 5.18.



(a) $NDVI$ index applied to a victim scene of the Figure 5.16.

(b) $NDVI$ index and color map applied to a victim scene in Figure 5.16).

Figure 5.18: Band combination following the function described by the NDVI index.

5.3.2 Analysis for victims detection using traditional indexes

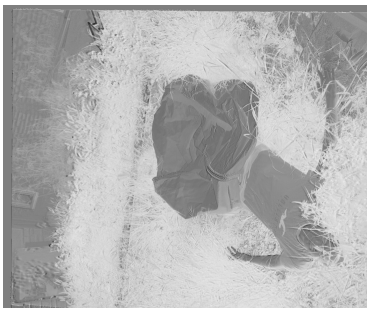
A preliminary evaluation of established agricultural indexes is conducted to ascertain their suitability as approaches for victim detection. Table 5.3 expounds the indexes pertinent to this investigation. Notably, works in the state-of-the-art for victim detection often employ individual Nir or RedEdge bands; however, these works do not account for indoor-outdoor distinctions nor facilitate adequate environmental delineation within the resultant images.

The aim is to determine whether the pixel concentrations associated with a victim (a partially covered or uncovered person lying down) are adequately distinguishable from the environment context, necessitating a phase of human judgment to validate this discernment. Histograms have been employed to visualize the pixel distribution across the image, illustrating the pixel quantity for each intensity value [0, 255] present within the image.

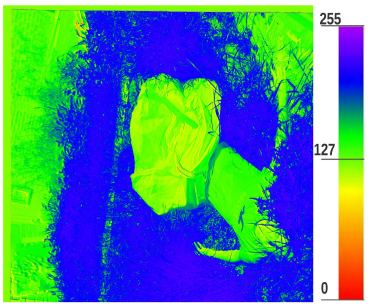
Table 5.3: Common used vegetative indexes. Source: [43].

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

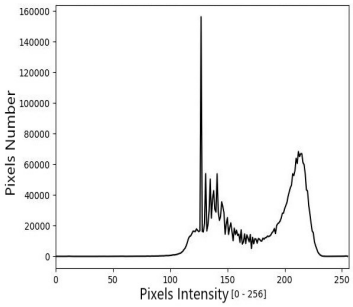
Figure 5.19 illustrates the application of the indices from Table 5.3 to a series of sample images captured in both indoor and outdoor scenarios. As depicted in Figure 5.19-a, the high pixel intensity in the grass-covered region results in significantly elevated NDVI



(a) $NDVI$ index applied to a victim scene of the Figure 5.16.



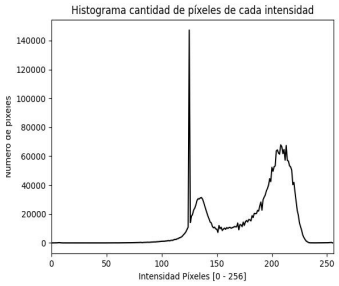
(b) $NDVI$ index and color map applied to a victim scene in Figure 5.16).



(c) $NDVI$ image histogram

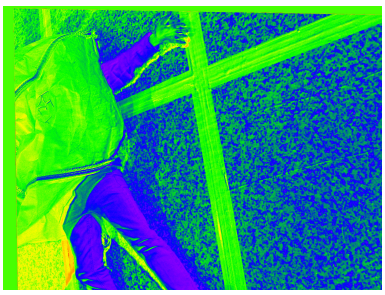
(d) $NDVI$ index applied to a victim outdoors scene.

(e) $NDVI$ index and color map applied to a victim outdoors scene.

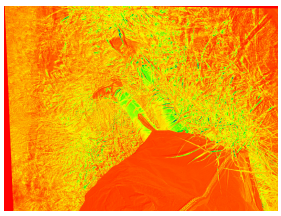


(f) $NDVI$ image histogram.

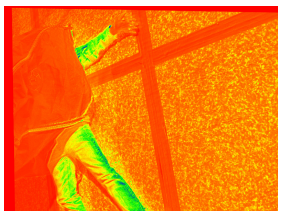
(g) $GNDVI$ Index (applied to outdoors scene).



(h) $GNDVI$ Index (applied to indoors scene).



(i) $OSAVI$ Index (applied to outdoors scene.).



(j) $OSAVI$ Index (applied to indoors scene.).

(k) $SIPI$ Index (applied to outdoors scene.).

(l) $SIPI$ Index (applied to indoors scene.).

Figure 5.19: Vegetative indices commonly utilized as primary reference markers in victim identification.

values, as anticipated. It should be noted that high pixel intensity values in grayscale translate to white colour (255), while low values tend towards black (0).

The victim stands out against this background, and its outline can be discerned concerning the grass. However, it is noticeable that the victim cannot be distinguished from either

the bag used as debris or the brick background. All these elements share the same pixel intensity as the victim. Additionally, the green colour introduced after colouring is the predominant hue in Figure 5.19-b, as evidenced by the histogram (Figure 5.19-c), with a presence in nearly 160000 pixels.

Due to these reasons, NDVI cannot be considered a suitable index for victim detection in this image set. A similar situation arises in Figure 5.19-d-e-f, where the outcome is even less favourable as the legs are scarcely distinguishable from the grass, displaying markedly elevated intensities. Hence, it is confirmed that NDVI is inadequate, prompting the search for another index that distinctly accentuates the victim against both the grass and other elements within the image.

Subsequently, the following indices are evaluated: GNDVI, OSAVI, and SIPI, with results shown in Figures 5.19-g-l. There is no clear demarcation of the victim from the surroundings, except in the case of the OSAVI index. However, this distinctiveness is limited to indoor environments; even so, the victim's hand remains inconspicuous. Consequently, these indices are not suitable for the stated objective.

5.3.3 New Indexes Proposed

The objective is to amalgamate the adapted images to accentuate the victim against the backdrop, facilitating enhanced detection and recognition for the convolutional neural network. To achieve this, the index $GRVI = \frac{\rho NIR}{\rho GRE}$ is employed as the starting point, with the inclusion of the RED band, thus defining equation 5.3:

$$Index1 = \frac{\rho NIR}{\rho GRE + \rho RED} \tag{5.3}$$

Figure 5.20-a-b displays the outcome, wherein adding the red band to the numerator does not significantly alter the GRVI index. Elements in the image featuring reddish hues, such as the victim's clothing, show diminished distinction from the surroundings. This distinction arises due to their substantial reflection of red light, resulting in elevated intensity values within the denominator and, consequently, lower values within the overall fraction. The differentiation between the victim's skin and the background remains discernible within the grassy setting but not indoors.

By integrating the red band into the denominator, by equation 5.4, the result is depicted in Figure 5.20-c-d. In this scenario, the reddish elements attain heightened intensity. However, the victim's outline combines with the grass in the initial image. As for the images portraying the legs and the indoor environment, improved outcomes are observed.

$$Index2 = \frac{\rho NIR + \rho RED}{\rho GRE} \tag{5.4}$$

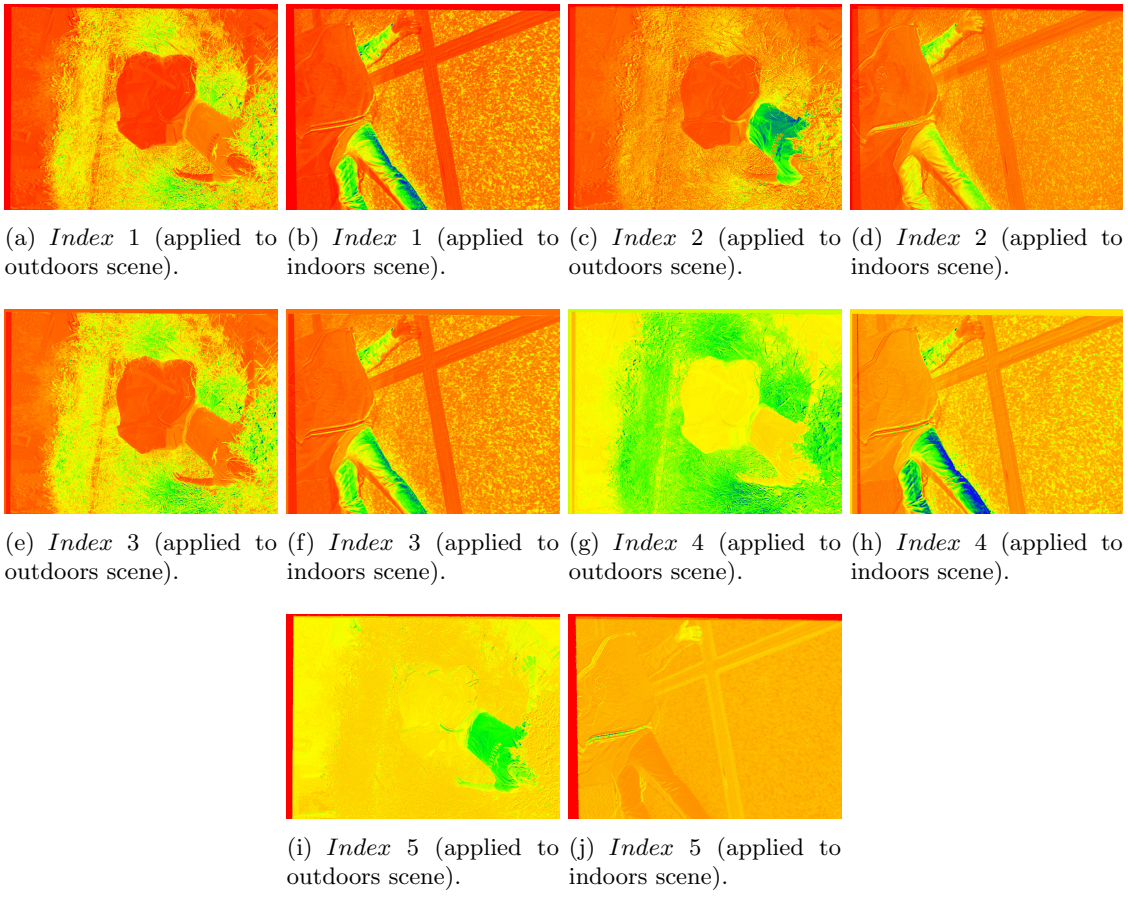
To compensate for this mixing effect with the grass, the red bandpass to subtract in the numerator (equation 5.5) in the same way as it does in the NDVI:

$$Index3 = \frac{\rho NIR - \rho RED}{\rho GRE + \rho RED} \tag{5.5}$$

In Figure 5.20-e, the result of Index 3, heightened contrast is observed between the grass and the victim. Nonetheless, challenges persist in effectively distinguishing the skin from debris or the ground indoors (Figure 5.20-f). Consequently, an approach to address these challenges involves the incorporation of the Red Edge (REG) band within the preceding expressions of the equation, rendered in different ways, equations 5.6-5.7:

$$Index4 = \frac{\rho NIR - \rho RED - \rho REG}{\rho GRE + \rho RED} \tag{5.6}$$

$$Index5 = \frac{\rho RED + \rho NIR + \rho REG}{\rho GRE + \rho NIR} \tag{5.7}$$



(a) *Index* 1 (applied to outdoors scene). (b) *Index* 1 (applied to indoors scene). (c) *Index* 2 (applied to outdoors scene). (d) *Index* 2 (applied to indoors scene).

(e) *Index* 3 (applied to outdoors scene). (f) *Index* 3 (applied to indoors scene). (g) *Index* 4 (applied to outdoors scene). (h) *Index* 4 (applied to indoors scene).

(i) *Index* 5 (applied to outdoors scene). (j) *Index* 5 (applied to indoors scene).

Figure 5.20: Vegetative indexes commonly employed as initial reference points in identifying victims.

With this latest index, the most favourable results to date are achieved within outdoor settings, although this is not the case indoors. An observable issue pertains to the framing surrounding the resulting images upon applying these two indices. This effect arises from the superposition of four distinct bands (GRE, RED, NIR, REG), of which three have undergone a displacement, resulting in pixel absence (black pixels, intensity 0) along the

edges. Employing four or even all five bands yields broader instances of this framing artefact. Consequently, it is prudent to identify an index employing the fewest bands.

Until now, all proposed indices stem from straightforward operations involving the bands. An additional consideration involves the introduction of coefficients that amplify or diminish the effect of certain bands. Non-linear elements have also been introduced in the operations, and an extensive array of variations has been tested, selecting those that, at first glance, exhibit the most favourable outcomes in terms of victim-environment differentiation, as elucidated in Table 5.4:

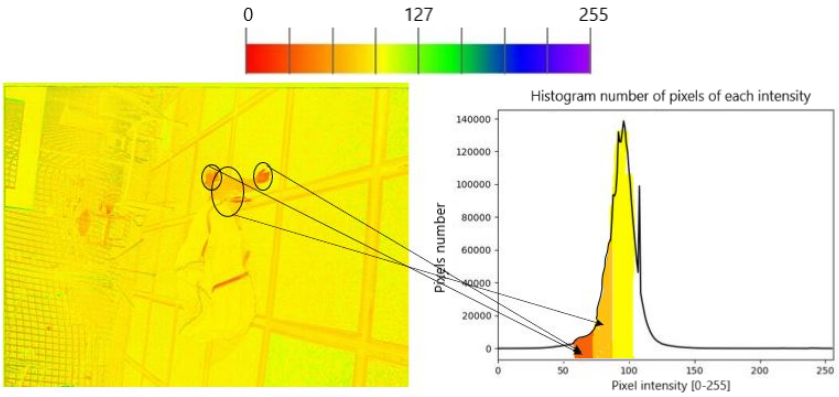
Table 5.4: Tested Indexes from multispectral bands for victim identification.

Index	Expression
Index 1	$\frac{\rho NIR}{\rho GRE+\rho RED}$
Index 2	$\frac{\rho NIR+\rho RED}{\rho GRE}$
Index 3	$\frac{\rho NIR-\rho RED}{\rho GRE+\rho RED}$
Index 4	$\frac{\rho NIR-\rho RED-\rho REG}{\rho GRE+\rho RED}$
Index 5	$\frac{\rho RED+\rho NIR+\rho REG}{\rho GRE+\rho NIR}$
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	$\frac{\rho GRE-\rho RED+0.75*\rho REG}{\rho GRE+\rho RED}$

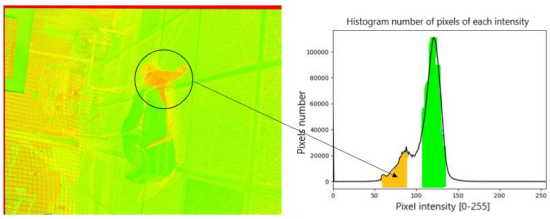
Once a preliminary selection of indices that best emphasize the perceptible properties has been conducted (indices 6 to 15), a shift from qualitative to quantitative foundation is pursued. It is considered that an index's efficacy is directly proportional to the extent to which it accentuates the victim from the surrounding environment. This quantification manifests as a higher discrepancy in pixel intensities between the victim's location and the neighbouring environment, leading to an enhanced index performance. Consequently, an analysis of these intensities becomes imperative. Histograms corresponding to the colour map of the new images have been employed to facilitate this analysis, delineated across ten intervals within the 0-255 scale.

Figure 5.21 illustrates the application of indices 6 to 10 to an image representative of the process. In sequence, the colours associated with the victim are correlated with the histogram's region range alongside those corresponding to the environment. Subsequently, the relative percentage difference is computed by employing the mean values obtained for both the victim and the surroundings, as per Equation 5.8. The quantified results of this procedure are succinctly summarized in Table 5.5.

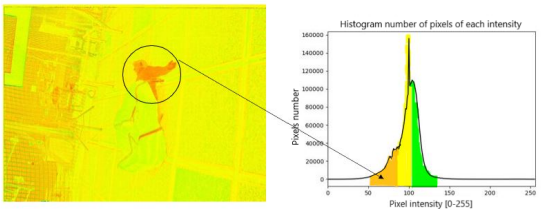
$$RelativeDiff = \frac{\mid mean\_vict\_pixel - mean\_env\_pixel \mid}{255} * 100 \tag{5.8}$$



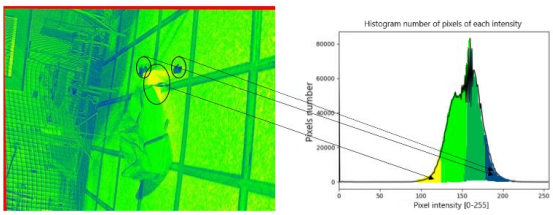
(a) Index 6 and Histogram.



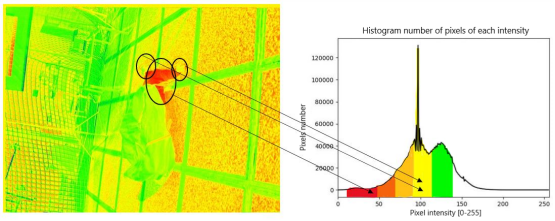
(b) Index 7 and Histogram.



(c) Index 8 and Histogram.



(d) Index 9 and Histogram.



(e) Index 10 and Histogram.

Figure 5.21: Relative difference evaluation about victim and environment for indexes 6 - 10.

Table 5.5: Assessment of the Relative Discrepancy Between Victim and Environment: Examining Indexes 6 - 10 from Figure 5.21. The average pixel color values were extracted from the histograms showcased in Figure 5.21.

Ind	Victim Color Mean Pixels Value	Environment Color 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

5.3.4 Selection of the best index

To select the best index from those proposed in the previous section, 20 samples of representative images of the dataset have been taken both indoors and outdoors. This dataset is available in the repository `https://drive.upm.es/s/xPKDp5Xyh1HTHWA`. The procedure of relative difference computation was then executed for the images and indices ranging from 6 to 15, culminating in the derivation of 200 resultant values. These outcomes are presented in Table 5.6 [277].

Table 5.6: Table of relative differences between victim and environment divided by 255, images of indoors (grey) and outdoors (blue) have been applied new indexes, complete results are available in `https://github.com/ChristyanCruz11/Multispectral-Images.git`.

	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	16,15	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
Mean	15,50	16,72	18,60	14,27	9,33	8,76	10,58	16,11	15,64	11,25

Subsequently, the absolute difference between mean values obtained for indoor and outdoor environments was determined for each index, and the quantified outcome is depicted in Figure 5.22. Upon scrutiny of the results, it becomes evident that Index 8 exhibits superior consistency despite not ranking as the foremost performer in isolated outdoor or indoor settings.

This index, subsequently referred to as InVD (Index for Victims Detection), in contrast to its counterparts, showcases a distinctive ability to discern the victim from the surroundings across all envisaged scenarios. Furthermore, it attains the highest overall mean value. Consequently, the forthcoming dataset of images earmarked for neural network training

will be generated based on this index.

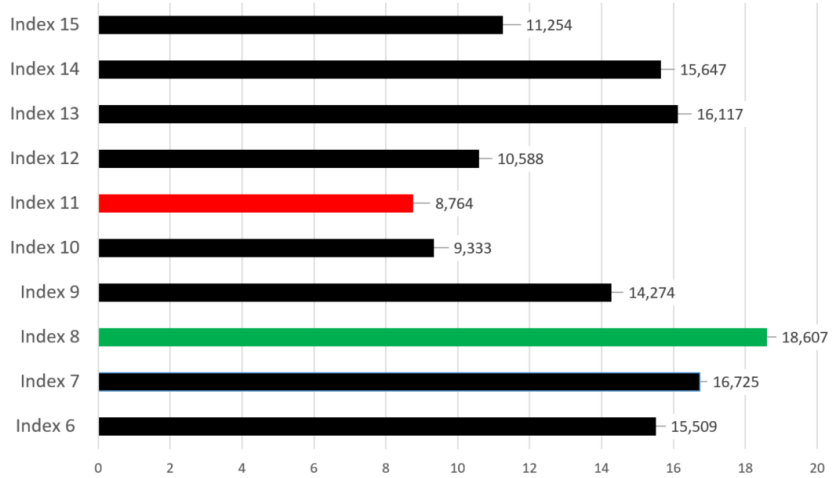


Figure 5.22: Relative mean value between environment and victim in evaluated indoors and outdoors.

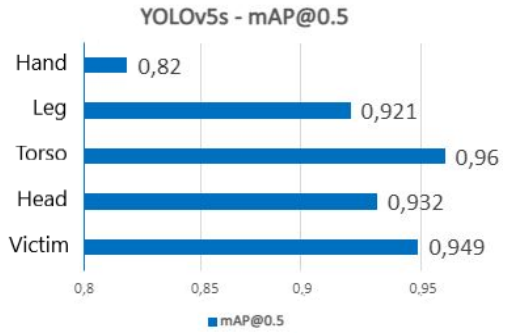
5.3.5 Automatic Victims Detection using CNN

In this phase, analogous to the section outlined in 5.1, the YOLO model has been employed to detect victims. The training parameters of the network encompassed an image input resized to dimensions of 928x928 pixels, a batch size of 8, and a learning rate set at 0.001. Using these settings, four versions of YOLOv5 were trained to evaluate the performance; below is the number of epochs and time required for training: model “s” (200 epochs - 2.5 hours), model “m” (150 epochs - 4.1 hours), model “l” (125 epochs - 5.2 hours), and model “x” (88 epochs - 13 hours).

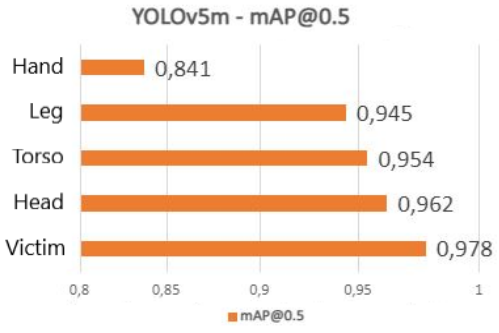
The dataset derived from the InVD index is presented within the GitHub repository, comprising 840 images. The employed labels encompass Victim, Leg, Hand, Head, and Torso. In the same way, data augmentation was applied, a total of 1454 images was obtained, subsequently allocated into the Training (82%), Validation (12%), and Test (6%) subsets.

The Mean Average Precision (mAP) values obtained during the evaluation are presented in 5.23. Based on the training outcomes, it becomes evident that varying the mean average precision values yielded the least favourable outcomes in the “s” model (Figure 5.23-a). While the detection outcomes demonstrated comparability among the remaining three instances (Figures 5.23-b-c-d), the choice of the network was determined by prioritizing training time, ultimately leading to the selection of the “m” model.

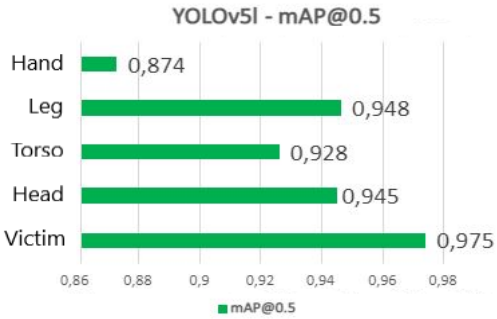
Within the confusion matrix (Figure 5.24-a), conspicuous high-confidence values along the main diagonal, corresponding to True Positives, signify the network’s adeptness in classifying objects into their respective categories. Conversely, the Precision-Recall curve (Figure 5.24-b) encompasses a substantial portion of the domain beneath its curve, approximating the unit rectangle for all classes except ’Hand’. Surpassing the 0.9 threshold of mAP@0.5 signifies a strong relationship between precision and recall, indicating that nearly every detected positive instance within an image is a true positive (exhibiting sound precision), and conversely, nearly all actual positives present within images are deemed positive (demonstrating robust recall).



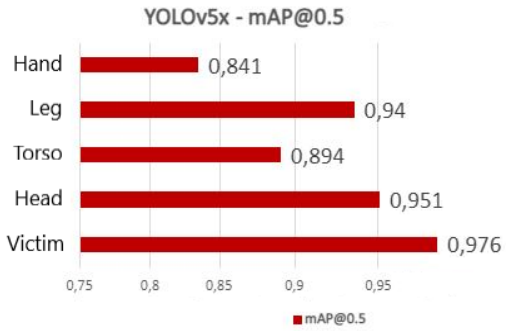
(a) YOLOv5 s.



(b) YOLOv5 m.

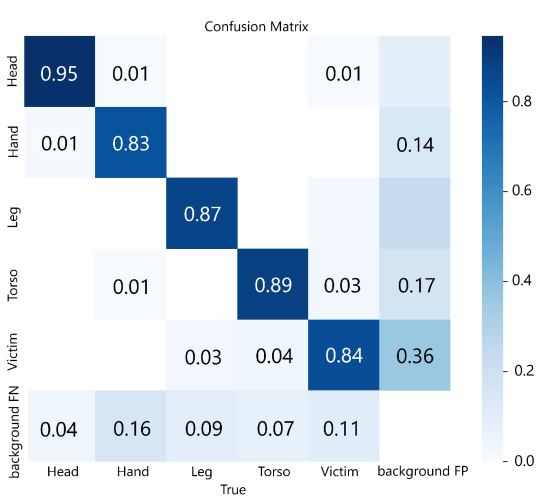


(c) YOLOv5 l.

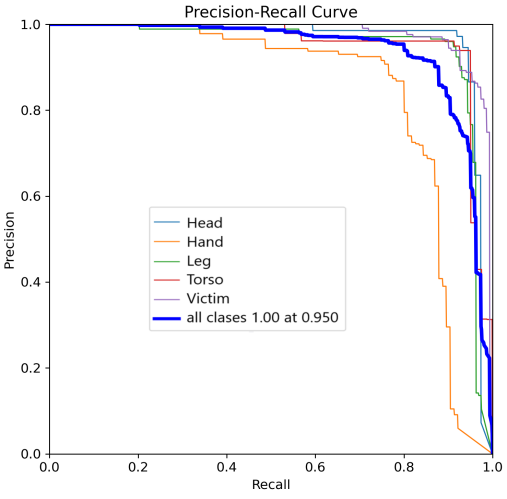


(d) YOLOv5 x.

Figure 5.23: Evaluation of trained YOLOv5 models.



(a) Confusion matrix for the trained YOLOv5m model.



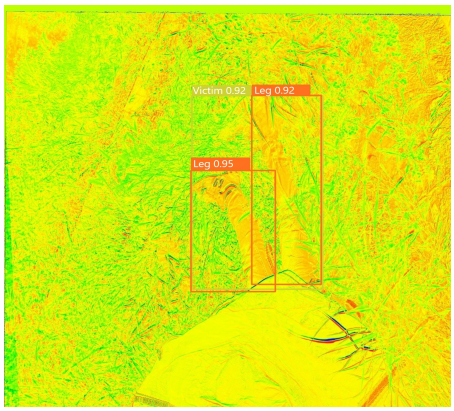
(b) Precision-Recall curve for the trained YOLOv5m model.

Figure 5.24: Evaluation of the YOLOv5m.

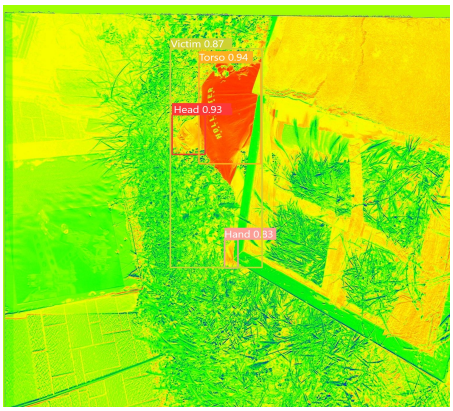
In Figure 5.25, the outcomes of the detection process on new images are presented. Distinctly coloured bounding boxes have been superimposed on the regions where victims were identified. These bounding boxes are accompanied by corresponding labels and detection percentages within the range of [0-1].

Figures 5.25-a to 5.25-b showcase scenarios situated in outdoor environments. In the initial instance, a victim with a covered torso is depicted, yet her lower limbs are identified with accuracy, boasting an efficiency surpassing 92%. The subsequent depiction features a victim whose legs are concealed under debris. Accurate classification is achieved for classes encompassing "head," "torso," "victim," and "hand." Particularly noteworthy is the performance in the "head" and "torso" categories.

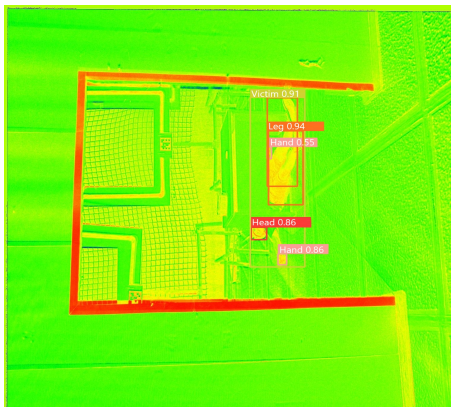
Progressing to Figures 5.25-c to 5.25-d, these images illustrate indoor settings. The initial indoor scenario presents a victim with the torso hidden. Notably, all labels are successfully discerned, maintaining an average detection efficiency of 89%. In the subsequent indoor scenario, a classification of classes is realized, with an average efficiency exceeding 91%.



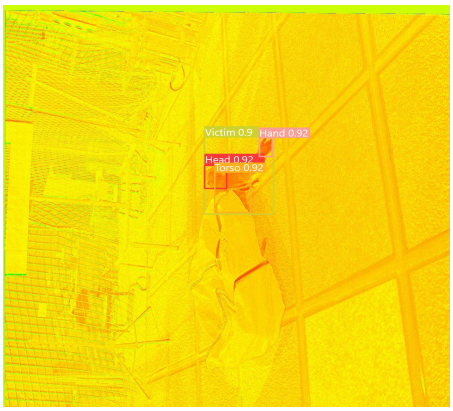
(a) Detection of victim covered by bag and legs in outdoor environment.



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



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



(d) Indoors partially covered victim detection.

Figure 5.25: Victim Detection across Varied Environments in Post-Processed $Index8$ Images.

5.3.6 Conclusion

This study substantiates the efficacy of utilizing multispectral images in victim detection by amalgamating diverse spectral bands obtained via a multispectral camera. This is achieved by implementing novel indices introduced in this PhD thesis in conjunction with convolutional neural networks.

After exploring various permutations of multispectral bands (indexes) and subsequent qualitative and quantitative assessments, the index that most effectively discriminates a victim from the surrounding milieu (encompassing indoor and outdoor scenarios) is identified as InVD. This index is predicated upon the bands Green (GRE), Red (RED), and Near-Infrared (NIR), which constitute the dataset employed for training the neural network.

$$InVD = \frac{\rho GRE - \rho RED - \sqrt{\rho NIR}}{\rho GRE + \rho RED} \tag{5.9}$$

Multispectral cameras are positioned as a viable substitute for thermal and RGB technologies in victim detection. This position is underpinned by their capacity to engender an array of iterations for each primary image by effecting spectral band amalgamation. This adaptability confers multifarious benefits for victim detection, particularly in outdoor contexts, thereby contrasting them favorably against conventional RGB and thermal camera modalities.

The automatic recognition of victims using the inVD index-derived image dataset has demonstrated remarkable efficacy. Notably, classes such as 'victim' and its corresponding extremities consistently achieved a mean precision of 85%, while the 'head' category displayed exceptional precision at 95%. These outcomes underscore the index's and CNN's ability to discern victims and their vital anatomical regions accurately.

5.4 Implemented Methods Discussion

This section introduces a qualitative and quantitative comparative analysis based on relevant criteria established throughout this research to evaluate the advantages and disadvantages of different victim identification methods from RGB, thermal, and multispectral image data.

This section introduces a comprehensive set of metrics, delineated in Table 5.7, designed to evaluate the efficacy of individual vision systems in the context of victim detection. Aligned with contemporary advancements in the field and guided by the author's established framework, these metrics encapsulate the fundamental evaluation criteria within the domain of search and rescue. They systematically examine each system's performance under distinct functional conditions while concurrently appraising the overall effectiveness of the proposed detection system

Table 5.7: Quantitative metrics proposed for the evaluation of vision systems.

Proposed Individual Evaluation Metrics	
α	Outdoors
β	Poor Light Conditions
γ	Indoors
δ	Processing Time
ϵ	Heat Sources Presence
θ	Totally covered victim
ϕ	Partially covered victim
ω	Clothes colour
λ	Summer/Fire conditions
τ	Changing Light Conditions

In a broad context, this study proposes a direct evaluation of three sensor types under generic conditions across diverse environments. This evaluation incorporates normalized coefficients [0-100] based on functional efficacy, as defined by Equation 5.10 (Thermal-RGB-Multi). The temporal parameter is omitted in this context, and coefficients related to victims are addressed generically as objects.

$$Score_{General}(Thermal - RGB - Multi) = \alpha + \beta + \gamma + +\epsilon + \theta + \phi + \omega + \lambda + \tau \quad (5.10)$$

Conversely, in a Search and Rescue (SAR) approach, systems undergo evaluation based on their functionality in specific scenarios. Parameters such as processing time receive negative penalties, reflecting the critical role of time in exploration. Conversely, factors like robustness in changing light conditions contribute more substantially to the overall score. Notably, the study highlights the importance of identifying concealed victims and detecting individuals in low-light conditions—frequent occurrences in post-disaster environments.

Equation 5.11 succinctly summarizes the weighted relationships for coefficients in each type of image detection. These modified coefficients, derived from conducted experiments, assign greater significance to factors considered more pertinent in Search and Rescue operations.

$$Score_{SAR}(Thermal - RGB - Mult) = \alpha + 1.75 * \beta + \gamma + (100 - \delta) + 0.8 * \epsilon + 2 * \theta + \phi + \omega + \lambda + 1.5 * \tau \tag{5.11}$$

5.4.1 Systems Individual

The metrics presented in Table 5.7 have been assessed using victim detection quality data acquired from conducted missions. The evaluation involves mean precision values (mAP) across ten specified conditions, considering repetitions of missions conducted indoors and subsequent evaluations outdoors.

Figure 5.26 consolidates normalized percentage values (ranging from 0 to 100) for each of the three image types across the ten established indices in this study. The figure offers an intuitive overview of the strengths exhibited by each image type relative to the others. Noteworthy differentials emerge, particularly in scenarios with fully covered victims and poor lighting conditions, where thermal cameras demonstrate a significant advantage. Conversely, in aspects such as heat source detection or summer weather conditions, RGB and multispectral images emerge as preferable alternatives.

While mean effectiveness values of the three cameras remain similar for indoor, outdoor, and changing light conditions, moderate differences arise in scenarios involving partially covered victims or individuals wearing "camouflage" clothing that blends with the surroundings, with discrepancies of up to 20 percent.

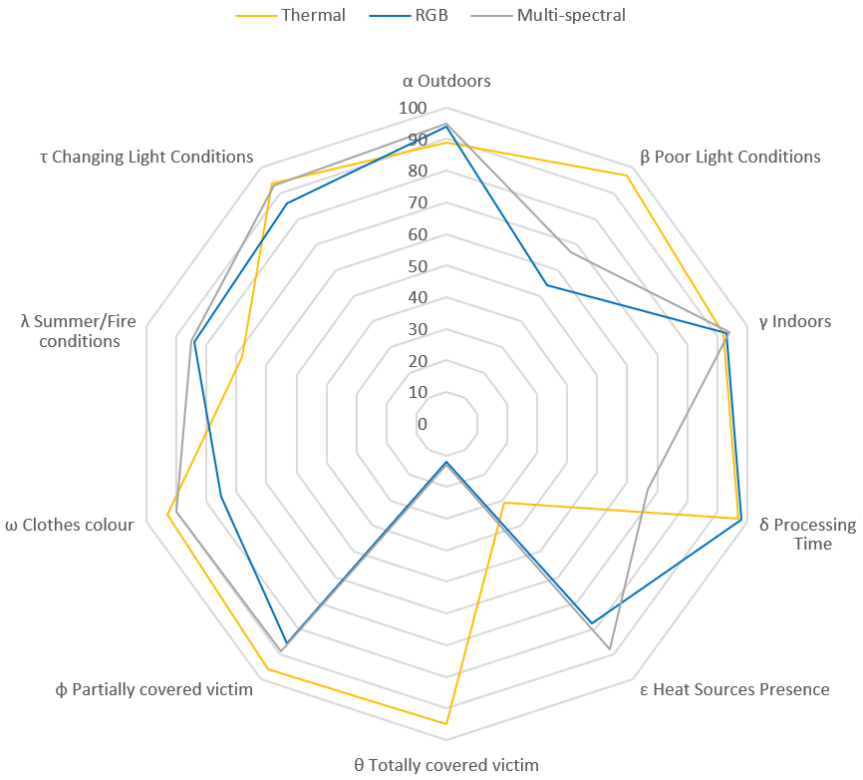


Figure 5.26: Radial graph employed for comparative analysis, illustrating distinct coefficients that assess the indication of mission success across various light spectrum ranges.

The rounded mean values, derived from the radial Figure 5.26 and linked to the coefficients of situational analysis for environmental conditions, are consolidated in Table 5.8 within the Metrics Analysis section. Furthermore, the outcomes of Equation 5.10 are detailed, illustrating the values for each image type in a generic context and specifically in accordance with Equation 5.11 for the Search and Rescue case.

Table 5.8: Metrics derived from the coefficients proposed for each image type within the three ranges of the light spectrum.

	Parameter	Thermal range	RGB range	Multi-spectral range
Metrics Analysis	α	89.3	97.1	95.4
	β	97.2	54.6	67.4
	γ	92.1	93.4	94.8
	δ	97.1	98.2	67.5
	ϵ	31.7	78.1	88.4
	θ	95.1	12.7	13.8
	ϕ	96.7	89.2	89.4
	ω	93.1	75.5	90.7
	λ	68.1	84.7	85.2
	τ	94.2	86.5	93.4
	General Score	755	662	714
	SAR Score	966	744	839
Indoors Experiments	Victims detection success rate %	92.4	84.7	86.5
Outdoors Experiments	Victims detection success rate %	91.1	79.4	81.2
Time Evaluation	Inference time f.p.s	26	28	8
Individual area covered %		85.2	76.0	78.1
Total Covered Area of Analysis %		93.5		

For generic detection scenarios during exploration missions, the calculated scores rank highest for Thermal (755), followed by Multispectral (714) and RGB (662), with a variation between extremes of 93 points. A similar pattern is evident in the SAR Score, maintaining the same order. However, the specific incidence difference for **Search and Rescue** is notably more pronounced, with a substantial gap of up to 222 points, considering the weighting factors assigned to the proposed coefficients.

Consistently across all three cases and for experiments conducted both indoors and outdoors, the thermal range emerges as the most effective for victim detection, followed by multispectral and RGB.

In Figure 5.27, a boxplot diagram is presented, illustrating the percentage values of victim detection across different configurations of coefficients $(\alpha, \ldots, \tau)$. Notable differences are prominently observed within the coefficients associated with ϵ and θ, signifying their substantial impact on victim identification. Conversely, τ and γ demonstrate minimal influence from their respective parameters on the detection of victims using either type of camera.

These observations highlight the sensitivity of detection outcomes to variations in specific coefficients, particularly those linked to environmental and contextual considerations. Notably, coefficients β, ϵ, δ, and θ emerge as pivotal factors influencing the performance of victim detection algorithms. This underscores the necessity for meticulous consideration and optimization of these parameters within the designed system.

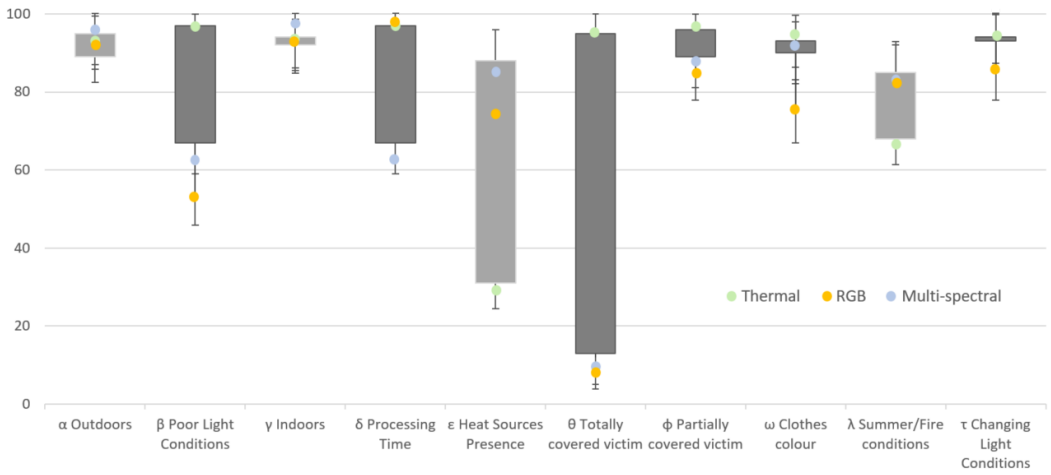
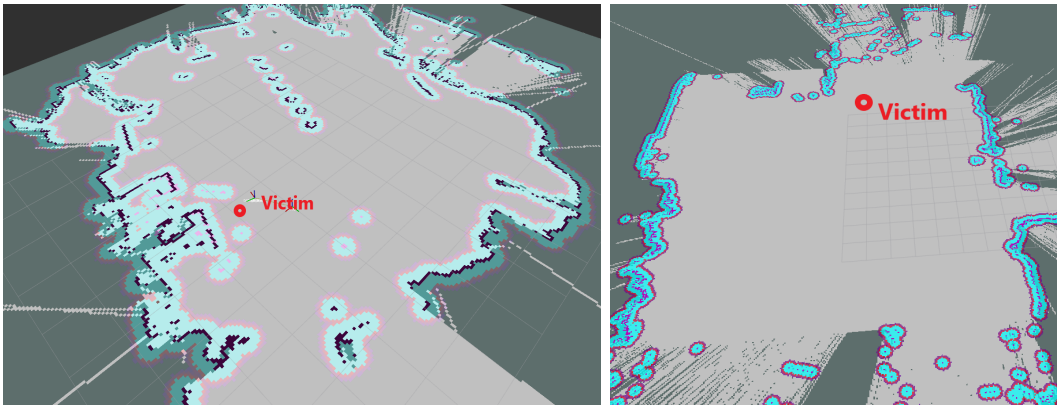


Figure 5.27: Assessment of errors in the mean values of the proposed coefficients depicted in the Figure 5.26.

As previously stated in the preamble of this section, all these victim detection methods could be applied to real-time systems, and the information regarding the detected ones can be incorporated into environment maps to enhance their content. This information proves particularly valuable to first-responders during their subsequent entry into disaster-stricken areas following the robot's exploration. Search and direct assistance times to potential victims identified can be optimized by doing so. Figure 5.28 provides an illustrative example of a generated map with markers indicating the locations of potential victims.



(a) Indoors Map generated with a victim detected.

(b) Outdoors Map generated with a victim detected.

Figure 5.28: 2D environment maps generated with potential victims marks placed.

5.4.2 Conclusions

- Under low-light conditions, Thermal Imaging exhibits superior performance in victim detection due to its reliance on thermal footprint, compared to methods such as RGB or multispectral.
- In terms of obstacle penetration, Thermal Imaging proves to be particularly better in detecting victims hidden by slender obstructions. Conversely, RGB-based techniques may confront challenges when endeavouring to identify individuals hidden

behind barriers, as they primarily rely on visual information and may struggle to differentiate victims in such scenarios.

- The RGB method's performance is influenced by clothing and the surrounding environment's colouration. Therefore, when the person wears less distinctive clothing, accurate identification becomes problematic using this method. In contrast, the Multispectral method demonstrates robust detection under such challenging conditions.
- Both the Multispectral method and the Thermal one stand out in this comparative analysis. However, it is crucial to note that the Thermal Imaging method relies on a single measurement spectrum. In contrast, the Multispectral method offers a broader spectrum of combinations, enhancing its adaptability and robustness in diverse scenarios.
- The RGB method, in conjunction with Multispectral, exhibits a substantial advantage in scenarios containing heat sources or specific cases like fires, surpassing the thermal method.

Chapter 6

Active Search in Unknown Environments

"If I don't know something, I will investigate it."

- Louis Pasteur

Exploring unknown environments is crucial for maximizing the probability of locating victims. This chapter introduces a concept of informed, active search, emulating, with a robot and AI algorithms, the decision-making criteria of a "first-responder" under these circumstances during exploration.

Scientific publication minor revision pending related: **Active Robotic Search for Victims using Ensemble Deep Learning Techniques.**

Preamble

This Chapter aims to emulate human behaviour, precisely that of a first-responder, during the exploration phase of an unknown environment to search for victims. ARTU-R robot and its sensory equipment, including the LIDAR system and RGB-D camera, were employed to achieve this development. This setup allows for acquiring spatial mapping data with high reliability in a 2D environment and high-resolution images for real-time object identification.

The contribution of this chapter proposes the execution of a non-semantic search in an unknown environment. In contrast to traditional environmental exploration methods that employ predefined strategies such as zigzag or radial exploration to cover the majority of terrain, the implemented method aims to conduct an efficient exploration of the environment to maximise the probability of locating a person. To achieve this purpose, a method based on an active search algorithm has been implemented that combines state-of-the-art techniques in indirect search, Next Best View (NBV) selection, and random forest (Ensemble Deep Learning Techniques).

The decision-making process controlling the advancement of the robot in an unknown environment relies on analyzing the robot's immediate surroundings (local environment). In this context, it is inherent to identify candidate points on the 2D map generated from the LIDAR system using exploration techniques like Next Best View (NBV). In this case, three points of interest are defined, corresponding to potential areas for advancement, such as corners or entrances to new instances. The robot orients itself toward each defined point to capture visual information and, utilizing pre-trained Convolutional Neural Networks (CNNs), extracts environmental features such as the presence of doors, instance type, and objects in the area. These three environmental characteristics are input for a previously trained decision tree (random forest), resulting in a value representing the highest likelihood of human presence.

A configurable weight function has been implemented to determine the best point for advancement. It considers the random forest's output and the information about the initially defined candidate points (distance to them and unexplored area around them). Upon reaching the new point (and during the advance), victim presence is evaluated –thematic addressed in the previous chapter and applicable within this new one–.

The system is structured as a master-slave architecture, with the algorithm running on the main computer and the robot handling mapping and movement options. Communication between them is facilitated via velocity directives. This controller was combined with the internal planner of the Unitree A1 robot to execute long trajectories efficiently. Extensive testing, both at the individual module level and within a complex, unstructured environment, demonstrated the system's success.

In this chapter of the PhD thesis, the implemented intelligent search-exploration method will be described in detail, its different components: detection modules, indirect search system, evaluation of the environment, and finally, the control and planning of the robot. In addition, the proposed method was contrasted with human criteria to validate the results obtained using the robot and AI algorithms. The individual testing phase evaluates both the subsystems separately and together in different controlled environments.

6.1 System Architecture

The central autonomous exploration system proposed in this section exhibits some similarity (in the block structure and information flow) to the one described in the preamble of the previous Chapter 5, as depicted earlier in Figure 5.2. However, it can be interpreted as a modification in the teleoperation layer. The robot still receives velocity commands for its movement in the environment. However, instead of originating from a joystick or teleoperation element operated by a human, these commands now come directly from the system that coordinates the autonomous exploration zones, the collision-free path planner, and the controller. This implementation in Figure 6.1 involves a master-slave system while continuing the decentralized processing architecture in a command centre.

This adjustment represents an evolution in the control structure, where the robot's movements are autonomously guided based on exploration objectives, path planning, and collision avoidance, streamlining the teleoperation aspect and enhancing the system's overall autonomy. This change in the control scheme allows for more efficient and adaptive exploration in various environments, further improving the system's capability in autonomous exploration missions.

The implemented master-slave architecture is designed to allocate computational processes efficiently, emphasizing relocating tasks demanding substantial computational resources to the command centre situated within a secure zone. Simultaneously, the robot executes activities such as mapping, acquisition of sensor data, and transmission to the command centre. Additionally, it receives movement commands for navigation within the environment.

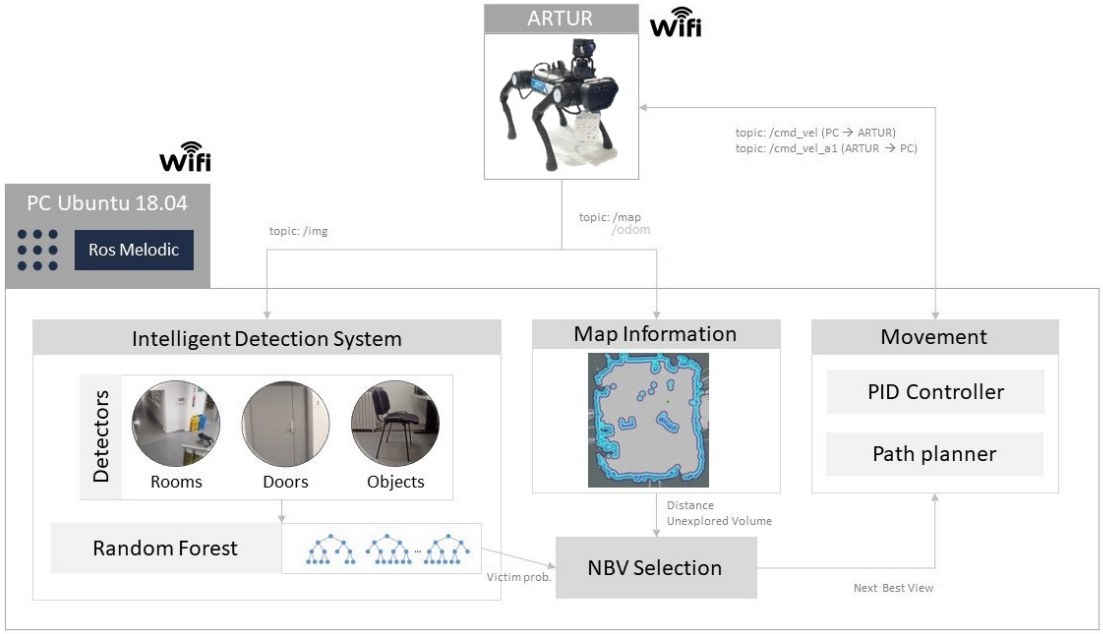


Figure 6.1: Subsystems implemented architecture.

This architectural framework presents a notable advantage in terms of computational resource management and task distribution. The robot's onboard resources are predominantly allocated to real-time navigation, sensor data acquisition, and local decision-making tasks by centralizing computationally intensive operations at the command centre in a safe environment.

The master comprises a series of subsystems structured as presented in Figure 6.1. Three different detectors have been set up for the Intelligent Detection System: a first one to determine the category of the explored room, a second one to detect objects and a third one in charge of detecting doors. These entities share a standard structure based on CNNs, with the only variation being the weight parameters associated with the network, which differ among the individual detectors. The data originating from the first two detectors is used to feed into a trained random forest model, which computes the probability of the presence of a victim within the spatial area captured by the image. Meanwhile, the information from the door detector is integrated with the map generated by the robot. This integration facilitates seamless navigation from one instance to the next once the exploration of a given instance is completed.

Upon entering a post-disaster environment, ARTU-R initiates door detection, moves towards the nearest door, and subsequently explores the instance beyond the door. In the context of each instance (bedroom, living room) exploration, determining each exploration point following a Next Best View (NBV) strategy involves a combination of the probability of victim presence and the Map Information gathered from the ARTU-R mapping system. A fitness function assesses various point candidates, considering factors such as the available free space around them, the distance required to reach them (derived from the map), and the probability of encountering a victim.

Upon selecting the next exploration point, the robot initiates movement toward it. Velocity commands are computed by the master using a Proportional-Integral-Derivative (PID) controller based on the new path calculated considering the robot position, goal destination and current map. The Planner continuously monitors the robot's real-time speed and replanning when it detects deceleration, indicating proximity to an unforeseen obstacle. This systematic approach ensures efficient and adaptable navigation for autonomous exploration missions.

In the subsequent sections, the system's components will be addressed as the integration and coordination among these components.

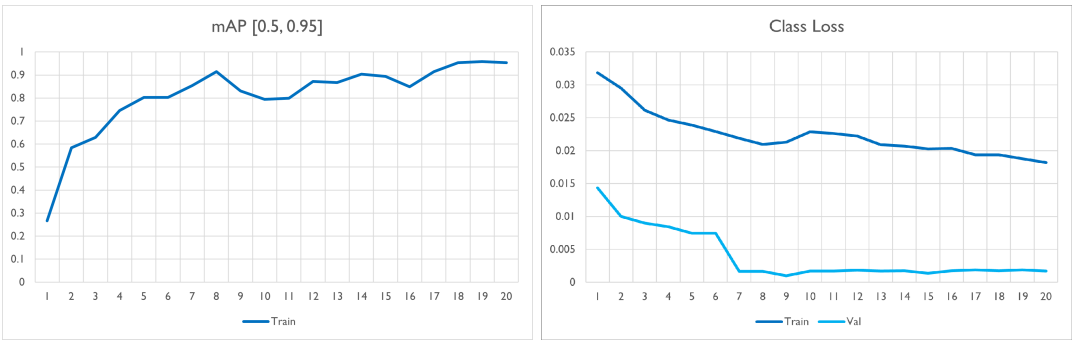
6.2 Intelligent detection system

6.2.1 Instance Detection.

The initial phase of system implementation involved training a detector to discern the specific type of room in which humans were being searched, a critical factor in enhancing success rates during indirect search operations, as discussed in Pronobis et al. [240]. To accomplish this, a dataset containing diverse room images was required for exploration tasks, and the Reni et al. dataset [249], comprising 1158 house room images, met this criterion. Interestingly, no data augmentation was deemed necessary in this context, but data labelling was essential to facilitate subsequent training procedures.

Initially, the system was designed to recognize five distinct rooms: bedrooms, bathrooms, dining rooms, living rooms and kitchens. Including room types more closely associated with professional environments, such as offices, computer rooms, workshops, or meeting rooms, was initially absent. Although there was a contemplation of merging the dataset with other containing elements of these professional settings, as will be elaborated upon subsequently, it was observed that the system's performance remained commendable with the limited set of room categories, rendering additional categories unnecessary.

The room detection process utilized the YOLO architecture, with 80% training and 20% testing data split over 20 epochs. The training comprised three stages: initial approximation and two refinement phases, achieving a 96.99% mean Average Precision (mAP), a precision of 92.66%, and a recall of 93.2%. Although overfitting concerns arose, subsequent sections demonstrate the network's adaptability to images from various contexts. Figure 6.2 shows the training curves.

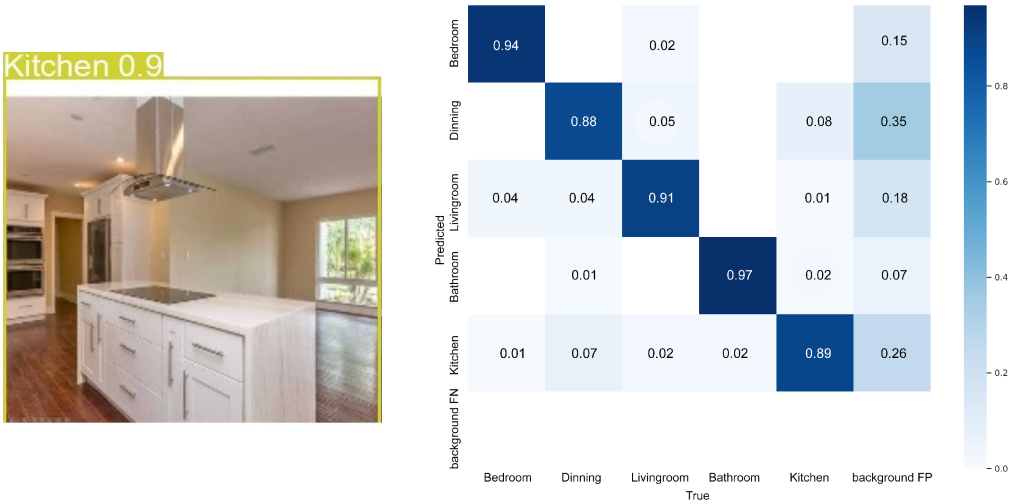


(a) The mAP evolution, calculated across IoU thresholds ranging from 0.5 to 0.95.

(b) The evolution of class loss observed in the training and validation sets.

Figure 6.2: The progression of mean Average Precision and class loss during the CNN training.

Figure 6.3-a provides an illustrative excerpt from the training data. The associated confusion matrix, presented in Figure 6.3-b, showcases the system's accuracy in accurately detecting room types, achieving a minimum accuracy rate of 88%. Notably, the most prevalent sources of confusion appear in distinguishing between kitchens and dining rooms, accounting for 7% and 8% of misclassifications, respectively.



(a) Validation results of kitchen detection.

(b) The confusion matrix for the instances detection system.

Figure 6.3: Instances CNN evaluation.

The model, upon utilization, yields a two-dimensional array characterized by a number of rows equivalent to the detected objects. Each row comprises bounding box coordinates and the associated detection probability for the detected object. As delineated in Algorithm 4, the pertinent information is condensed into a one-dimensional array denoted as *detRooms*.

Within this array, element i signifies the probability that the room corresponds to type i. Table 6.1 provides a mapping between vector indices and their corresponding room types.

Table 6.1: The mapping of vector indices within the *detRooms* array to their respective room types.

Id.	Instance
0	Bathroom
1	Bedroom
2	Dinning Room
3	Kitchen
4	Living Room

Algorithm 4: Algorithm for Instances Detection.

Function *DetectRoom(image)* **is**
  boundBoxes ← [ ]
  RoomsNet ← CNN trained model
  detRooms ← [ ]
  boundBoxes ← Evaluate *image* using *RoomsNet*
  foreach *i* of *boundBoxes* **do**
    t ← type(*i*)
    detRooms[*t*] ← Detection probability related to *i*
  end
  return *detRooms*
end

6.2.2 Victims Detection based on Random Forest and Indirect Search

The subsequent configuration focused on setting up the object detector subsystem, which required no additional training. The YOLO network was employed with pre-trained weight files encompassing detection capabilities for various entities and everyday objects from downloading. Similar to the approach applied for the Room Detector, the output from the CNN was processed and incorporated into an analogous vector structure, as illustrated in Algorithm 5.

Algorithm 5: Algorithm for Object Detection.

Function *DetectObjects(image)* **is**
  boundingBoxes ← [ ]
  ObjectsNet ← Trained CNN model
  detObjs ← [ ]
  boundingBoxes ← Evaluate *image* using *ObjectsNet*
  foreach *i* of *boundingBoxes* **do**
    t ← type(*i*)
    detObjs[*t*] ← Detection probability associated to *i*
  end
  return *detObjs*
end

Considering that the system's primary aim revolves around indirectly detecting individuals, the central challenge entailed identifying which everyday objects and items could indicate a person's presence within a post-disaster environment. In certain scenarios, what might appear as an intuitive correlation—such as expecting a remote control to be near a television—proved less specific, as individuals engage in diverse activities at various times throughout the day.

Subsequently, a foundational premise was established, positing that following a disaster, such as an earthquake, individuals tend to remain in positions that closely mirror their pre-disaster locations. Consequently, developing intelligent systems necessitates training only on everyday images rather than post-disaster data. While this assumption may initially appear impractical, research studies like [141, 184] assert that sensible responses often need to be addressed amidst the commotion caused by natural disasters.

The selected dataset for this task was the COCO (Common Objects in Context) dataset [178], developed by Microsoft, comprising a substantial collection of 330,000 images depicting various facets of daily life. From this dataset, a database was constructed to correlate the presence of humans with objects. Both of the previously two described detectors were applied to each image in COCO, and the associated probabilities for objects and rooms were stored within this database. A corresponding tag was appended to the respective database entry whenever a human presence was detected within an image. In the context of Algorithms 4 and 5, the inputs to the random forest model are represented by the vectors $detRooms$ and $detObjs$, while the output pertains to the probability of a human being present in that particular region.

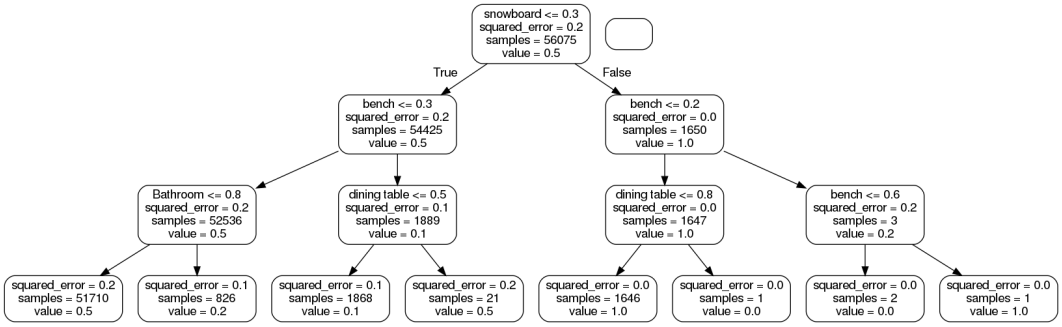
While the potential use of neural networks was contemplated, the decision was made to employ decision trees, primarily due to the system's collaborative nature with humans. This choice offers the advantage of aligning more closely with human decision-making processes. The probability of a victim's presence in a particular region is determined by iteratively inquiring whether a series of objects, prioritized by importance, are located within that region. Although random forests may appear less intuitive than simpler decision trees, they can perform exceptionally well with extensive datasets, exhibiting reduced susceptibility to overfitting. In this study, random forests achieved outcomes akin to those attainable with a CNN.

Table 6.2: Hyperparameters of the Random Forest.

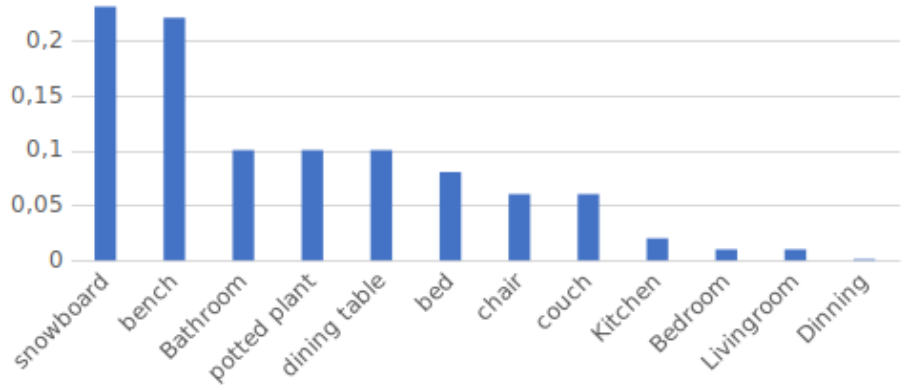
Maximum depth	15
Number of estimators	10
Maximum number of features	All

Model hyperparameters were optimized through cross-validation (Table 6.2). To improve generalization, variables with significance rates under 1% were removed, reducing them to 13. The model achieved 70% mAP precision and 40% MSE loss, although these values may seem low compared to typical Machine Learning accuracy rates.

Figure 6.4-a displays a subset of the decision trees, consisting of 13. As previously discussed, it is feasible to quantify the significance of individual variables. These significance values elucidate the impact on the loss function if a specific explanatory variable were to be omitted from the model. Figure 6.4-b presents the significance values of the final trained model.



(a) The initial three tiers of one of the trees within the random forest.



(b) The most influential variables within the random forest model. The assignment of a value of 0.1 to the "dining table" feature signifies that the inclusion of this variable in the model results in a 10% reduction in model error.

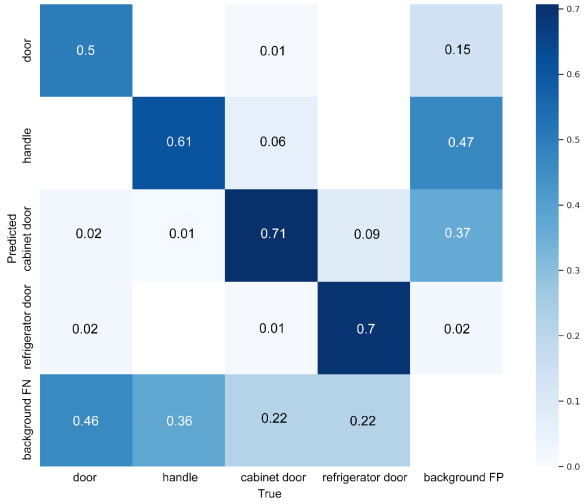
Figure 6.4: Graphs derived from random forest.

6.2.3 Doors detection system

The third necessary component facilitating initial exploration involves the door detector. Upon arrival in a post-disaster environment, ARTU-R starts its operational sequence by questing for doors, subsequently advancing towards the nearest door. Upon completing exploration within a room, the robot proceeds to seek out new doors within the same room. In cases where no doors are detected, ARTU-R exits the room and searches for other doors identified during the initial inquiry.

Door detection is accomplished by employing a CNN trained on the dataset referenced as [6]. This dataset encompasses 1213 annotated images featuring various types of doors, including room doors, shelf doors, Etc. Figure 6.5-a presents the confusion matrix, revealing that doors are accurately identified in 50% of the cases. In the remaining instances, they are categorized as background, signifying an absence of a specific classification. Notably, this performance aligns with the reported results from the dataset's creators [6]. Furthermore, it is necessary to emphasize the low likelihood of misclassifying any environmental element as a door, as evidenced by the minimal event of false positives, which stands at a mere 15%. Illustrations showcasing the accurate detection of a partially occluded door are presented in Figure 6.5-b.

After door identification within the image, the system determines their respective locations in the 2D-generated map. This localization process is achieved by estimating the door's relative angular position concerning the robot and comparing this estimate with the obstacles discerned on the map using lidar data. Despite the CNN's door detection rate



(a) The confusion matrix for the door detection system.

(b) An instance of a successful door identification accomplished by the door detector.

Figure 6.5: Doors CNN evaluation.

being limited to 50%, it is important to note that each door is typically captured in the image twice, significantly enhancing the probability of detection. Consequently, ARTU-R accurately stores the positions of all doors within the environment, achieving a level of precision on the order of centimetres. A comprehensive pseudocode outlining the entire door detection process is provided in Algorithm 6.

6.2.4 Map Information

Environmental data acquisition relies on ARTU-R's integrated lidar sensor. Within ARTU-R's sensor processing framework, a Simultaneous Localization and Mapping module is seamlessly integrated, generating a cell-based map as its output. This map, denoted by the `/map` ROS topic, assumes the form of a two-dimensional array, with each constituent element signifying the status of a corresponding cell. To accommodate ARTU-R's specifications, a resolution of 0.1 m has been meticulously chosen. Cell states are quantified in terms of occupancy probability, ranging from 0 (indicative of unoccupied space) to 1 (representative of occupied space).

In the NBV selection context, cells registering values below 0.5 are classified as unoccupied, while unexplored regions are assigned a value of -1. During exploration, the array dimensions expand dynamically to accommodate newly identified regions. To provide operators with valuable insights, information about obstacles is also visually represented in the RViz interface, a feature designed for enhanced operator situational awareness, as exemplified in Figure 6.10.

6.2.5 Control and Planning

Control of ARTU-R involves the transmission of linear and angular velocity commands, simplifying its holonomic capabilities to focus on these velocity components. The control process comprises three stages: when rotational adjustments are needed, angular velocity

Algorithm 6: Algorithm for Detecting Doors.

```
Function SearchForDoors() is
    doorTol ← 60°
    nPossDoors ← 15
    pos ← robot position
    doors ← [ ]
    nAttempts ← 0
    do
        nAttempts ← nAttempts + 1
        Rotate doorTol
        Capture an image and initiate a door search.
        if door detected then
            foreach door do
                Determine the door's position on the map.
                door.dist ← ‖pos − (door.x, door.y)‖
                doors.push(door.x, door.y, door.dist)
            end
        end
    while isEmpty(doors) and nAttempts < nPossDoors
    if isNotEmpty(doors) then
        Navigate to the nearest door
    else
        Single-Room Environment.
    end
end
```

commands are sent from the master controller, fine-tuned through a PID control with saturation limits (parameters in Table 6.3), primarily used for tasks like capturing images to assess Next Best View (NBV) candidates.

Table 6.3: Controller and Limiter Parameters.

Parameter	ω controller	v controller
K_p	0.2	
K_i	0.2	
K_d	0	
τ	0.1	
Max. error	0.1 rad	0.3 m
Max. speed	0.6 rad/s	0.3 m/s

In cases where a translational movement is necessitated, but the distance to be traversed is less than 1.5 m, a comparable PID controller is employed, with its associated parameters detailed in Table 6.3. In this scenario, the system initiates by orienting ARTU-R towards the designated target point and subsequently issues commands for linear motion until the specified location is reached.

The angular and linear velocity PID controllers are constrained to align with the robot's physical attributes. As a safety precaution, velocities exceeding 0.3 m/s and 0.6 rad/s are prohibited. Likewise, a point or orientation is deemed reached when measurement errors fall below 0.3 m and 0.1 rad. The controller's block diagram is illustrated in Figure 6.6.

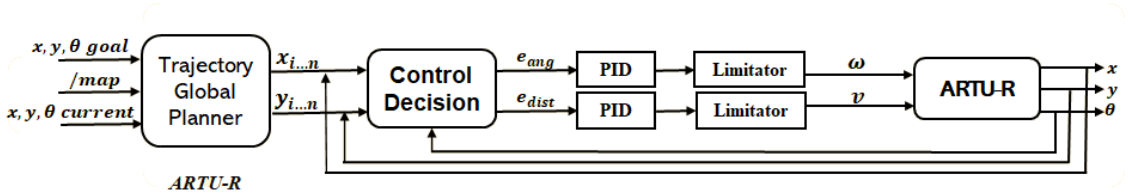


Figure 6.6: Block diagram illustrating the controller for robot movement.

In subsequent stages, it becomes imperative to employ planning strategies, albeit at the cost of increased computational time, owing to the inherent limitation of the controller in ensuring obstacle avoidance. The planning system is orchestrated through the master controller, which meticulously calculates a navigational path from ARTU-R's current position to the specified target destination, exclusively considering map cells designated as unobstructed. Upon path determination, velocity commands are transmitted to the robot. In the event of obstacle detection, the costmap undergoes real-time updates, prompting local path recalibration to circumvent encountered obstacles.

6.3 NBV Selection based on Image-Laser Fusion Sensor

The three subsystems were integrated to form the complete system, whose operational sequence is delineated in Figure 6.7 and Algorithm 7. Key configuration parameters for the comprehensive algorithm are summarized in Table 6.4.

Algorithm 7: Complete algorithm.

```
SearchForDoors()
foreach room do
    image ← Take an image
    detRoom ← DetectRoom(image)
    while room is not explored do
        NBV ← SelectNBV(detRoom)
        Move to NBV
    end
    SearchForDoors()
end
```

Following the process in Figure 6.7, once arriving at the designated starting point, whether it's an entrance in a house or a foyer in an office building, the system activates its door search protocol. This involves capturing an image and applying the relevant detector to it. If a door is detected, ARTU-R immediately moves towards the nearest one and explores the room beyond. In cases where no door is detected, the robot performs rotational manoeuvres as defined by the `doorTol` parameter (as specified in Table 6.4), all while continuously repeating the search process. If a complete revolution happens without door detection, ARTU-R relocates to a random point and resumes its quest for new doors. After a certain number of repetitions, determined by the `nPossDoors` parameter, the environment is categorized as a single room, prompting exploration to commence.

Behind room exploration, ARTU-R initiates a similar process to identify doors within the room. If additional doors are found, the robot explores them; otherwise, it exits the room and moves on to the second door identified in the previous phase. The room exploration

Table 6.4: Algorithm's Main Parameters.

Parameter	Description	Value
`nPoints`	New Best Views (NBV) Candidates Generated per Iteration.	3
`dist`	Maximum Permissible Distance Between Candidates and ARTU-R's Current Position.	5 m
`candRadius`	Radius for Evaluating Cell States around an NBV Candidate.	2 m
`doorTol`	Rotation Angle When No Door is Detected.	60°
`nPossDoors`	Threshold for Deeming the Environment as Single-Room After Multiple Attempts.	15

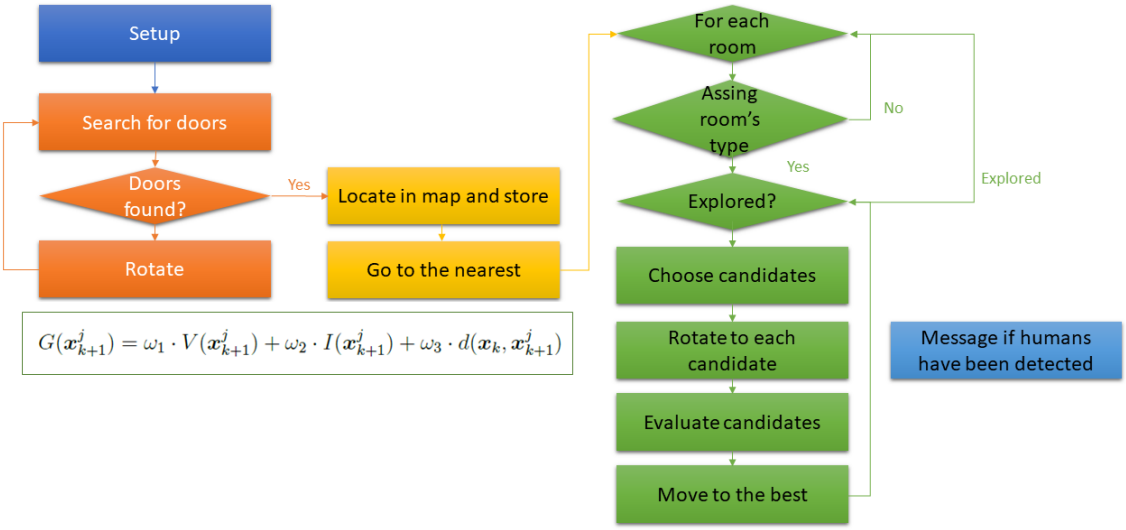


Figure 6.7: Descriptive synthesis of the complete algorithm implemented.

procedure begins with the classification of room type, a task executed only once until a new room necessitates exploration. Following this, the system navigates the space, progressing step by step.

At step k, when the robot is positioned at $\boldsymbol{x}k$, it proceeds to randomly generate a set of candidate positions $\boldsymbol{x}k+1^j$ (where $j = 1, \ldots, nPoints$). These candidates are situated within a distance no greater than `dist` from $\boldsymbol{x}_k$. The SLAM tool must categorize the corresponding cells of these candidates as "free" based on information acquired from the laser sensor. Subsequently, each candidate position is assigned a fitness value, as determined by Equation 6.1.

$$G(\boldsymbol{x}^j_{k+1}) = \omega_1 \cdot V(\boldsymbol{x}^j_{k+1}) + \omega_2 \cdot I(\boldsymbol{x}^j_{k+1}) + \omega_3 \cdot d(\boldsymbol{x}_k, \boldsymbol{x}^j_{k+1}) \tag{6.1}$$

Equation 6.1 defines the upcoming advancement point for exploration. It evaluates the three candidate points defined by the NBV, and the one with the highest weighting is considered for advancement. This function comprises three terms, each composed of a

constant w and variable parameters described as follows:

- Candidate's Surrounding Explorable Volume ($V(\boldsymbol{x}_{k+1}^{j})$): This metric quantifies the proportion of unoccupied cells within a square region of dimensions ($2\cdot$`candRadius`)$\times$ ($2 \cdot$ `candRadius`) centred on $\boldsymbol{x}_{k+1}^{j}$. Maximizing this value is essential since a clear environment enhances the robot's chances of object detection via its camera. The measurement is expressed as a percentage rather than an absolute count, considering that the number of free cells may naturally decrease towards the map edges.
- Information Gain ($I(\boldsymbol{x}_{k+1}^{j})$): To quantify information gain, the robot orients itself towards $\boldsymbol{x}_{k+1}^{j}$ and captures an image. Subsequently, the random forest model is employed to incorporate the probabilities associated with object detection and the current room. The value of $I(\boldsymbol{x}_{k+1}^{j})$ is a direct outcome of this intelligent process.
- Distance to Target Point ($d(\boldsymbol{x}_k, \boldsymbol{x}_{k+1}^{j})$): This measure represents the Euclidean distance, computed using the norm-2, between $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k+1}^{j}$. The associated coefficient ω_3 is negative, emphasizing exploring nearby points as a priority.

The values of $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ have been established through empirical experimentation. In this specific case, the values of $\boldsymbol{\omega}$ are set as $\boldsymbol{\omega} = (0.1, 2, -0.7)$.

Equation (6.1) shares a structural resemblance with the equation introduced in [213]. Nevertheless, this model diverges in its considerations, incorporating aspects such as distance to the candidate position, aligning with references like [31] and [20]. An implementation of this function is provided in the pseudocode detailed in Algorithm 8.

The victim identification procedure is conducted concurrently with object identification to compute information gain. This approach is feasible as the object detector can also detect humans. When injured individuals are detected, the system marks their location on the interface map with an arrow indicator.

Algorithm 8: Algorithm for Selecting New Best Views (NBV).

```
Function SelectNBV(detRooms) is
    dist ← 5[m]
    nPoints ← 3
    ω ← [0.1, 2, −0.7]
    candRadius ← 3[m]
    m ← 0
    pos ← robot position
    while m < nPoints do
        do
            cand.x ← random point in interval [pos.x − dist, pos.x + dist]
            cand.y ← random point in interval [pos.y − dist, pos.y + dist]
        while isNotFree(cand)
        m ← m + 1
    end
    foreach cand do
        V ← % of unexplored cells in a square of side candRadius cells around
         cand
        Rotate ARTU-R towards cand
        image ← Take an image
        detObjs ← DetectObjects(image)
        I ← Random forest evaluation of detRooms and detObjs
        D ← ‖cand − pos‖
        if human detected in the image then
            Publish an arrow
        end
        cand.G ← ω_1 · V + ω_2 · I + ω_3 · D
    end
    NBV ← cand with higher G
    return NBV
end
```

6.4 Individual systems validation

6.4.1 Instances Classification

As noted earlier, during its training phase, the room detection system achieved a ninety-seven mean average precision. Experiments were conducted using real images from various training scenarios to confirm that this high accuracy was more than just a result of overfitting. The qualitative test results, including detection examples presented in Figure 6.8, showcased similar levels of accuracy for each room class compared to the training phase. An empirical value probability threshold of 30% was established to disregard detections with lower probabilities, effectively mitigating false positives. Confusion between room classes remained minimal, manifesting in fewer than 10% of cases.

(a) Living room identification. (b) Bedroom identification.

Figure 6.8: Different room detections with their associated probabilities.

Identification and placement of victims on the map.

The previously described RGB method from Subsection 5.1 of the chapter has been employed in this section. Consequently, the following stage aims to provide visual assistance to the operator by placing an arrow pointing toward the victim's location from the robot's position and orientation.

Once victims are identified, the system places an arrow on the map, indicating their location. This visual representation is readily visible to system operators, as demonstrated in Figure 6.9. The arrow's position and orientation correspond to ARTU-R's precise position when the image was captured. It is important to note that while the victims themselves are not explicitly mapped onto the system—no semantic SLAM or similar techniques have been employed—the consistently high human recognition rates enhance the likelihood of detecting the same victim from various perspectives, resulting in multiple arrows directing attention to their respective positions.

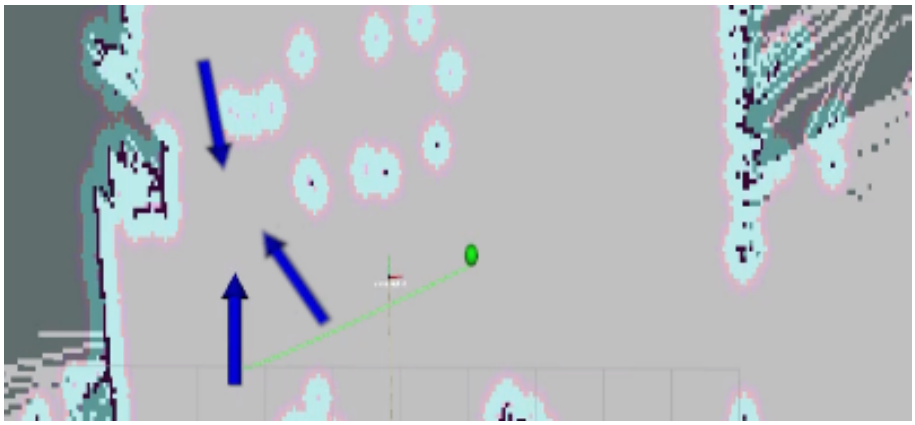


Figure 6.9: The interface visually represents victim detection at three stages: blue arrows pinpoint the human's presence based on robot-captured images aligned with the robot's current orientation. Additionally, green dots indicate the subsequent point of interest for ARTU-R's exploration.

6.4.2 Door detection and map matching

The assessment of the door detection took place within an authentic setting, specifically the building employed for the validation procedure. Different tests were conducted wherein the robot was tasked with autonomously seeking and mapping doors within this environment. Subsequently, the system's door placement accuracy was verified by cross-referencing it with the actual door locations on the scaled plan of the site in the facilities of CAR-UPM. To facilitate a comprehensive evaluation, diverse initial points were designated, enabling the assessment of ARTU-R's capability to identify doors in varying scenarios, including extended distances, image corners, and blind spots.

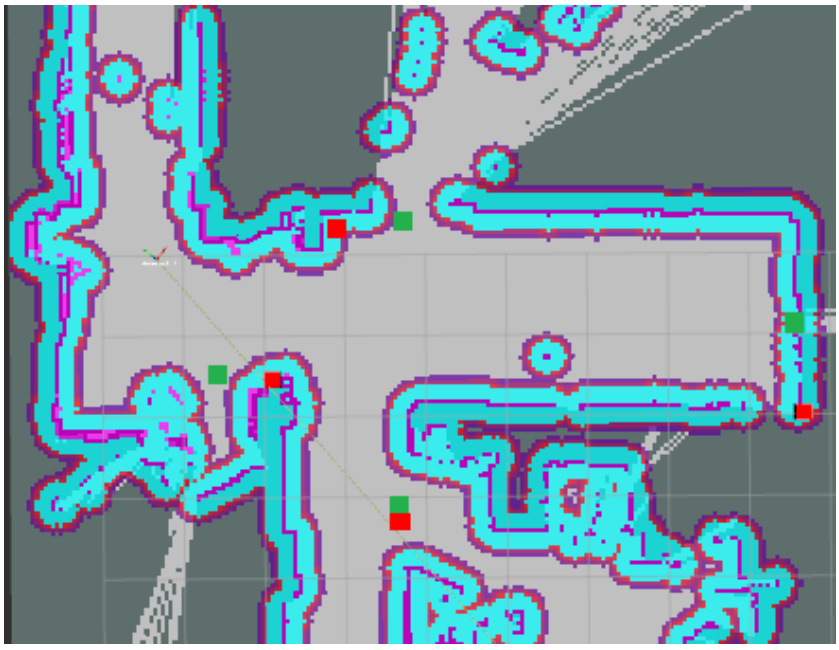


Figure 6.10: Detection and localization of doors within actual building structures are indicated on the map by green dots. Points recognized as doors by ARTUR are highlighted with red dots. All four doors in that region have been accurately identified.

Figure 6.10 illustrates the performance of the subsystem responsible for door detection,

showing an average door detection rate of 72%. The system exhibits a relatively low false positive rate, with only 6% of objects mistakenly identified as doors. Additionally, detected doors are placed with a spatial error of less than 0.92 meters.

This spatial error represents a significant source of uncertainty within the presented work and can be attributed to two primary factors. Firstly, it relates to the precision of the Door Detector, as previously discussed. Secondly, it stems from the methodology employed for gate localization on the map. This methodology estimates the gate's angle relative to the robot's position through a two-step process. Initially, this angle is approximated from a planar photograph, followed by alignment with an obstacle within the lidar-generated map. A potential solution to mitigate the second source of error involves improving angle estimations by accounting for image deformation during projection.

6.5 Results: General experiments and system validation

6.5.1 NIST Orange zone

Following the NIST standards, a test environment has been reconstructed for this experimental phase, precisely conforming to the Orange Zone specifications. This reconstruction encompasses the placement of mannequins and simulated victims within the environment. For each case, empirical data has been collected to facilitate a comparative assessment of the algorithm presented against existing methodologies documented in the state-of-the-art. Precisely, the dimensions of unexplored and explored areas throughout the mission were continuously monitored. Additionally, within the explored regions, the extent of space classified as "free" or "occupied" throughout the testing phase and victim detection time was noted.

Figure 6.11 depicts the trajectory of ARTU-R during the mission. Initially, ARTU-R searched for doors, discovering only one during the initial series of camera snapshots. Consequently, it proceeded directly towards that door (P1). Upon entering the room (P2), it navigated to the rear behind a desk (P3). This decision was driven by identifying a substantial unexplored area and recognising a table, which the random forest model had associated with a high likelihood of human presence.

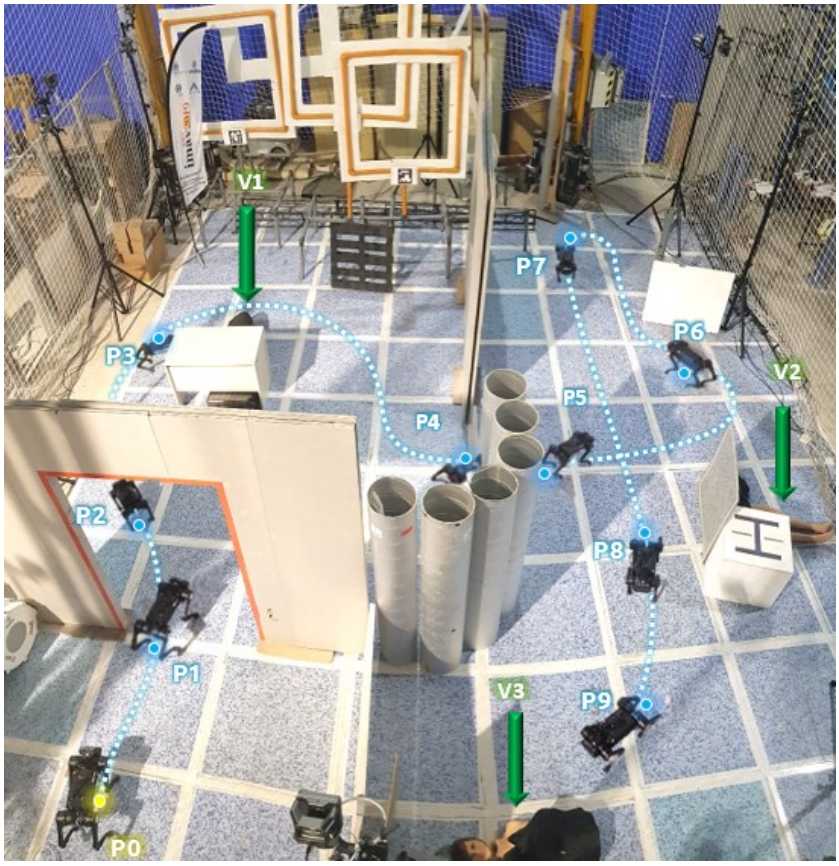


Figure 6.11: The trajectory followed by the robot throughout the rescue mission test.

ARTU-R initially identified and correctly detected the victim positioned at V1. Subsequently, it moved to the adjacent room (P4), bypassing exploration around a pallet since its see-through nature allowed for a direct inspection of potential victims hidden behind it.

The exploration of the second room started immediately upon passing through the door (P5). The first chosen Next-Best-View (NBV) was positioned near a table and a grey panel (P6), where another victim (V2) was successfully located and detected by ARTU-R. Following this, the robot transitioned to the background area (P7).

Even though no victims were found in the background area, it was a logical choice given its proximity to the previous position and limited visibility. Finally, the last explored point was the bottom-left corner of the room (P8), where the last remaining human (V3) was discovered.

The entire test, covering a 49 m^2 area (6.5 x 7.6 m), took 4 minutes and 15 seconds, resulting in an exploration speed of 0.192 m^2/s. Figure 6.12 shows a graph of time evolution and events during the exploration. When ARTU-R gets in a new room, new zones are mapped, achieving speeds exceeding 0.5 m^2/s. Then, it focuses on searching for victims, which were found early in both rooms, demonstrating the importance of camera information and the intelligence system.

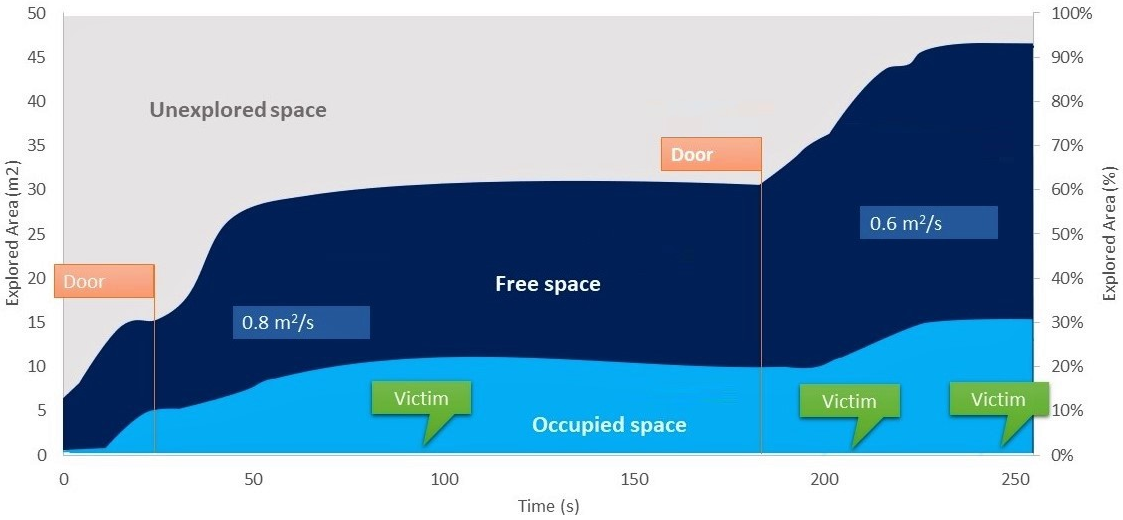
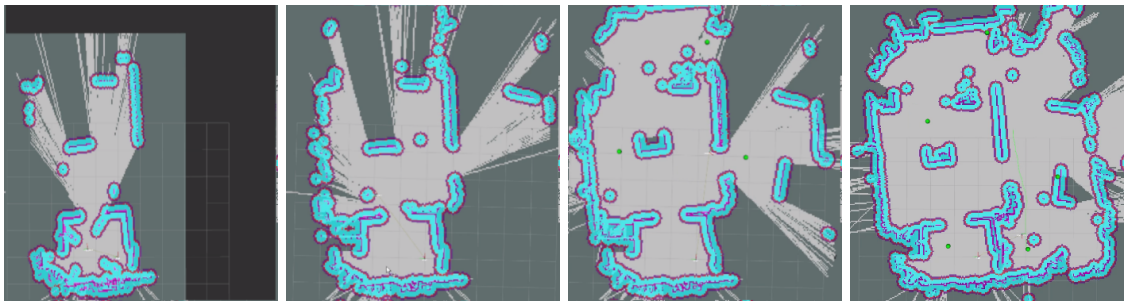


Figure 6.12: The advancement of the exploration procedure within the orange environment throughout the mission.

Figure 6.13 shows the environment 2D map reconstructed by ARTU-R throughout the exploration. All hidden victims were successfully located during the test.



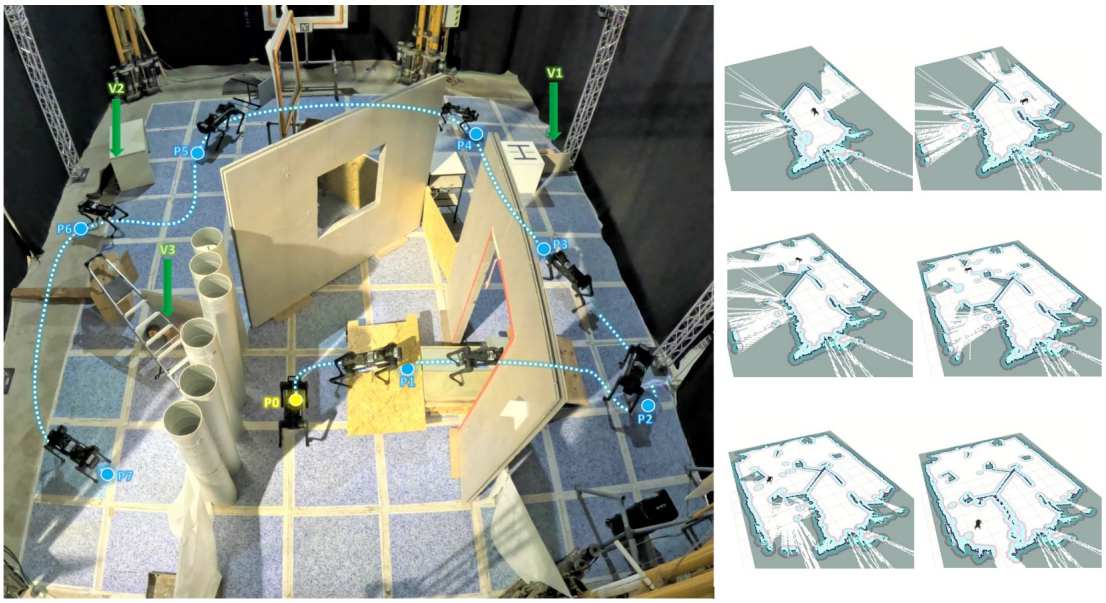
(a) t = 0 s (b) t = 188 s (c) t = 234 s (d) t = 255 s

Figure 6.13: A series of 2D maps illustrating the robot's advancement during exploration.

The test underwent ten repetitions, with mission durations ranging from 212 to 308 seconds. While the robot's route was not precisely identical in each repetition, it closely

resembled the path depicted in Figure 6.11, consistently exploring from nearby starting points. Notably, in 3 out of the ten repetitions, the exploration sequence in the second room followed the pattern P8-P9-P6-P7. This can be considered reasonable, as upon entering the room, there was a significant expanse of uncharted territory surrounding point P8, even though no objects indicative of human presence were identified.

In Figure 6.14, the layout of the test environment for the second scenario is depicted. It features two entrance ramps, one ascending with initial step dimensions of 14 centimetres in height, 87.5 centimetres in length, and an inclination of 11.8, and the other descending with a slope of 12.4.



(a) Path followed by the robot.

(b) Map generated.

Figure 6.14: During the first test rescue mission, the robot traversed a path within a scene that featured several areas with structurally compromised floors.

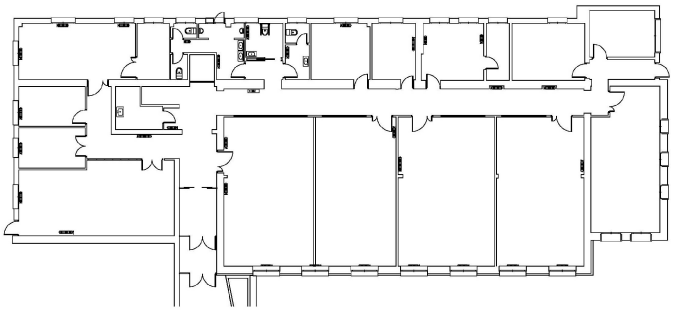
Moreover, an insurmountable obstacle in the form of a step for medium-sized wheeled robots, referred to as a beam, has been positioned between points P6 and P7. The passages between P3 and P4 and P4 and P5 are narrow and necessitate a robot with relatively modest width and agile manoeuvring capabilities.

ARTU-R's path sequentially explores the scene. After overcoming the ramp (P2), it seeks objects of interest (P3) and discovers a table. The area behind the table is explored (P4), leading to the discovery of a victim (V1). The robot continues its journey to the following table (P5-P6), where another victim (V2) is found. Ultimately, it traverses the beam (P7) and locates the final victim (V3). Access to the last region is crucial for discovering this last victim.

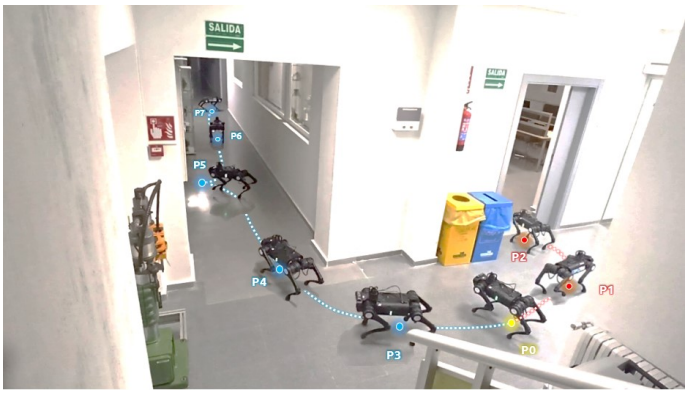
6.5.2 Open Controlled Environment

Additional tests were conducted in an open Controlled environment at CAR facilities (The 2D plane is shown in Figure 6.15-a).

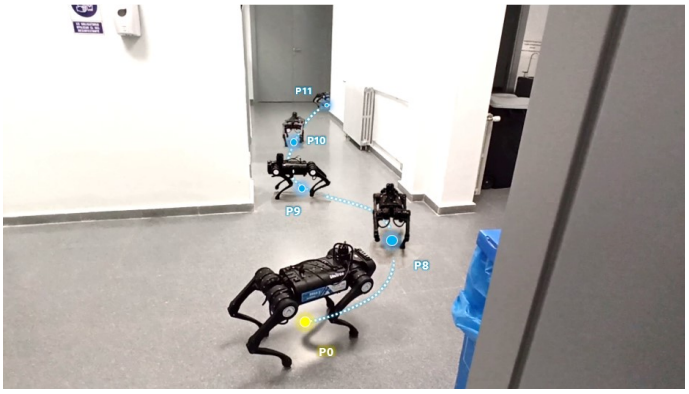
The rescue mission began at point P0 in Figures 6.15-b and 6.15-c, signifying the building's entrance. The door detection process identified three doors: the open door in Figure 6.15-b leading to a computer room, the corridor in the same scene, and the door in Figure 6.15-c granting access to a small, obstacle-filled laboratory.



(a) CAR's primary test floor in controlled open environment.



(b) The initial segment of the trajectory.



(c) The second segment of the trajectory.

Figure 6.15: The trajectory traversed by ARTU-R during the controlled test mission in an open environment.

As expected, the system chose to explore the area behind the nearest door, corresponding to the entrance of the computer room. ARTU-R's path to reach this door is depicted in red (P1-P3). Upon entering, ARTU-R efficiently explored the room, requiring only a few

steps to cover the entire area and detect the human presence.

Following this, a corridor exploration, indicated in blue in Figure 6.15-b (P3-P7), was conducted. ARTU-R's behaviour during this phase mirrored that of a human: it advanced down the corridor, refrained from entering open rooms due to adequate visibility at the entrances, and retraced its steps upon reaching the end.

The third phase of the mission is depicted in Figure 6.15-c. The system searched for victims within the laboratory beyond the door (P9). Upon completing the room exploration and detecting no additional doors, the mission was deemed finished (P10-P11).

6.6 Analysis of Efficiency for the Proposed Method

A comparative analysis against diverse methodologies from the literature has been performed. These methods vary in environmental settings and robot characteristics and serve broader exploration purposes beyond rescue missions. Consequently, the results offer a qualitative rather than definitive quantitative assessment. The primary objective is to qualitatively compare the performance of the presented methodology to that of similar existing methods, acknowledging their shared characteristics.

The selection process for comparative works in this study was guided by two primary criteria. Firstly, the chosen works were required to apply to ground robots of various types, including both quadruped and wheeled robots. Secondly, the selected scenarios had to closely resemble the orange NIST environment used in the study, with a preference for real-world scenarios that shared a similar spatial footprint (about 50 square meters) and included objects distributed throughout the area, allowing for robot navigation between them. It's worth noting that not all test scenarios met this last requirement, posing a challenge in the selection process.

This study analysed exploration speed, quantified in m^2/s, for both Reinforcement Learning and Next-Best-View techniques. This analysis involved the conversion and synthesis of data derived from various sources, leveraging the information available within the respective literature. It is important to note that some of these sources do not directly pertain to Search and Rescue (SAR) tasks; thus, statistical information about victim detection is not explicitly provided in this analysis. The comprehensive results of this exploration speed analysis are presented in Table 6.5.

The methodology outlined in this study, which calculates the Next-Best-View (NBV) utilizing the fitness function as defined in Equation 6.1, demonstrates a noteworthy exploration speed. Specifically, it achieves an average speed across the two evaluated scenarios of 0.198 m^2/s.

Table 6.5: Exploration Speed Comparison Among Various Methods in the Literature

Reference	Method	Expl. Area (m^2/s)
[247] (Sim. #2)	Classical RL	0.34
[295] (Real-world #1)	NBV (Boundary)	0.2
[213]	NBV (Fitness Function)	0.108
Present work	**NBV (Fitness Function)**	**0.198**

In Rasouli et al.'s work [247], a classical Reinforcement Learning approach was employed to evaluate a set of algorithms across diverse scenarios. Their most efficient solution located

all human entities within a 20x20 m^2 post-disaster environment in an average of 1194.4 s, resulting in an exploration speed of 0.34 m^2/s. This outperformed the results in the current study, albeit with a 99% confidence level. Two notable distinctions between their work and ours include their use of a simulated environment, lacking real-world challenges, and the employment of a pioneer wheeled robot, which may have higher traversal speeds but limited capability for unstructured terrains.

In Wang et al.'s study [295], a generic wheeled robot explored both simulated and real-world environments. Like Rasouli et al., their exploration speeds in simulated environments exceeded real-world counterparts. We compared our work to their first real-world scenario due to its size similarity to the orange NIST scenario, albeit with fewer obstacles within the rooms. In this work, speeds of 0.2 m are reached, but the algorithm does not consider using images (which increases processing time).

The work by Naazare et al. in 2022 [213] utilizes a mobile robot equipped with a manipulator arm to conduct inspections in various outdoor environments. Notably, the research focuses exclusively on outdoor scenarios, all exhibiting a comparable object density to the one addressed in this study. In this investigation, a fitness function plays a pivotal role in determining the Next Best View (NBV) at each step of the inspection process. The remarkable outcome is that it takes approximately 7200 seconds to map 80% of a 20 x 20 m^2 area, yielding an average mapping speed of 0.108 square meters per second. Notably, the speed of arm manipulation is estimated to be slightly slower than the movement of ARTU-R, although the difference is not substantial.

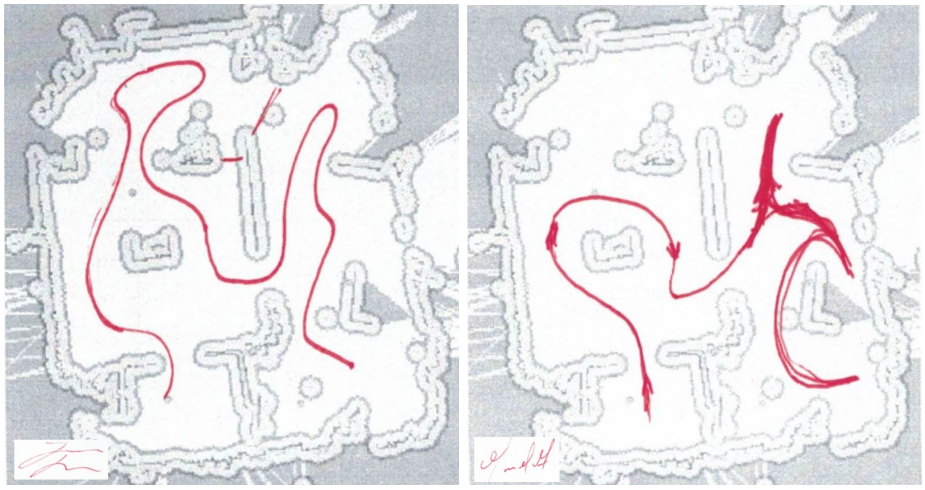
Moreover, it's essential to acknowledge that Naazare et al. (2022) have explicitly demonstrated the superior performance of their methodology compared to the AEP algorithm, as originally proposed by Selin (2019) [264] for UAV exploration. While direct time-based comparisons between terrestrial and aerial robots pose inherent challenges, the observed outperformance of this study over Naazare et al. [213] and, consequently, the AEP algorithm strongly suggests its potential superiority over similar or less effective techniques, such as those outlined in [31] and [20].

6.7 Experimental evaluation against human criteria

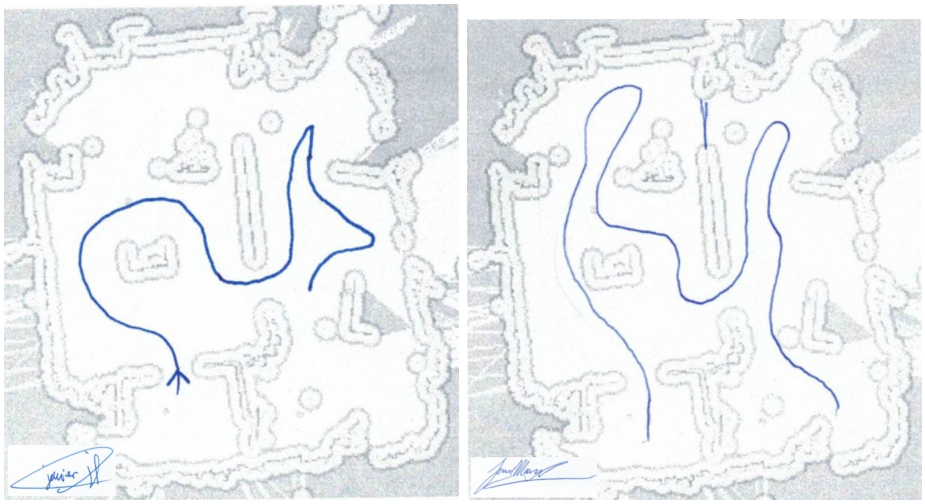
This section applied a test to twenty individuals to analyze the route and priority zones they would propose (according to their criteria) for exploring the environment depicted in Figure 6.11. The primary objective of this test is to determine the degree to which both the path followed by the robot and the key points explored align with human criteria. This process involves a scientific assessment of the congruence between the robot's actions and the preferences and decisions that individuals would make in a similar situation.

Figure 6.16 shows the exploration routes defined by four out of the twenty participants in the evaluation. All the tests corresponding to the participants are available in the following repository `https://blogs.upm.es/tasar/surveys/`. Participants were tasked with defining an exploration route based on their criteria as "rescue personnel." To achieve this, they were shown the initial scenario and the map generated by the robot of the environment upon which they were asked to chart the exploration route. This test concerned the starting point from which the robot, denoted as P0 (as referenced in Figure 6.11), initiated its journey.

The analysis will be conducted by delineating it into two distinct components, each serving



(a) Exploration path defined by subject 4. (b) Exploration path defined by subject 5.



(c) Exploration path defined by subject 6. (d) Exploration path defined by subject 18.

Figure 6.16: Routes defined by random four participants for exploring the area within the scenario depicted in Figure 6.11.

a specific purpose in the overall evaluation process. By combining these two components, a comprehensive assessment can be made regarding both the navigation strategies employed and the points of significance within the environment, contributing to a deeper understanding of the exploration process and its alignment with the objectives set for the evaluation.

The first component, which can be described as comparative and descriptive, entails a detailed examination of the sequences of environmental exploration. This entails examining how each participant or the robot, in the case of autonomous exploration, navigated and interacted with the environment. This comparative-descriptive phase is essential for understanding the diverse approaches taken by individuals or robots in their search to explore the environment efficiently and effectively.

The second component is related to the specific areas of interest explored during the evaluation. It delves into an in-depth analysis of the zones or points within the environment that participants or the robot identified as critical during their exploration. This phase

seeks to uncover patterns in the areas prioritized for exploration, evaluating the frequency of visitation, the rationale behind these choices, and any commonalities or disparities among the participants.

Starting from point "P0," the most evident scenario involves moving through the door located at "P1," a situation contemplated in all evaluation cases. In the subsequent stage, a table is centrally positioned, which could potentially hide a victim on its rear side. Ninety-five percent of the test participants explored this table (circumventing it) before proceeding to the next stage. Eighty-five percent of these participants followed a counterclockwise route, similar to the path taken by ARTU-R. Towards the rear of this area, translucent objects such as a pallet offer a direct line of sight, facilitating the continuation of exploration.

Upon crossing through point "P4" to the next stage, the environment expands, and a decision must be made to explore either to the right or left. Eighty percent of the participants charted an exploration route from left to right, similar to ARTU-R's approach. In 95% of the cases, the points of interest "P7-P8" are considered, as they may conceal hidden victims. In 50% of the cases, the exploration concludes in proximity to point "P9," near the last victim, similar to ARTU-R's behaviour.

In all cases, the described routes have effectively facilitated a comprehensive exploration of the entire environment, mainly encompassing critical zones that could potentially contain hidden victims. In most instances, these routes closely correspond to the path followed by ARTU-R (as in order sequence and explored points) in Figure 6.11, thereby validating the initial premise of this subsection of the doctoral thesis, which aims to emulate human search behaviour through a robot and artificial intelligence algorithms.

6.7.1 Conclusion

In this chapter, it has been demonstrated how combined sensory systems along with intelligence algorithms enable the emulation of human behaviour during an exploration phase in an unknown environment.

This work has integrated multiple subsystems into a unified exploration algorithm. This algorithm utilizes advanced strategies in indirect search, mainly due to the victims often being out of direct view, and Next Best View selection to optimize exploration. The integration process followed a cascade approach, ensuring each subsystem operates when necessary. This approach not only enhances operator comprehension of the process but also streamlines the development process. The modular design based on subsystems allows for future incremental enhancements without altering the overall scheme.

Upon completion of the training process, the accuracy of all neural networks exceeds 90%, except for the door detection network, where the results align closely with state-of-the-art performance. These results, while not a significant impediment, effectively enable the accurate mapping of door locations.

The random forest achieves a 70% accuracy in human presence prediction, enabling a 40% improvement in exploring regions with a higher probability of victim presence compared to blind searches. Notably, the system's decision-making process is not a black box; key variables from the random forest are extracted, and operators can access information about selected Next Best View (NBV) candidates. ARTU-R's movements are guided by its objective function's three components: unexplored volume around candidates, potential information gain, and distance to candidates, making them understandable from a human perspective.

The integrated system performs well during rescue missions, successfully passing tests in a simulated NIST environment and a real workplace (CAR-UPM facilities). In both scenarios, it effectively identified all human victims and prioritized their path, similar to human decision-making. Despite not relying on image processing delays, the system achieved exploration speeds comparable to state-of-the-art works. Its efficiency in victim rescue is notably enhanced thanks to its improved information utilization.