

Human Detection by Robotic Urban Search and Rescue Using Image Processing and Neural Networks

Fahed Awad¹, Rufaida Shamroukh²

¹Department of Network Engineering & Security, Jordan University of Science & Technology, Irbid, Jordan

²Department of Computer Engineering, Jordan University of Science & Technology, Irbid, Jordan

Email: fhawad@just.edu.jo, rufaida.b@gmail.com

Received 3 February 2014; revised 3 March 2014; accepted 15 March 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper proposes a new approach for detecting human survivors in destructed environments using an autonomous robot. The proposed system uses a passive infrared sensor to detect the existence of living humans and a low-cost camera to acquire snapshots of the scene. The images are fed into a feed-forward neural network, trained to detect the existence of a human body or part of it within an obstructed environment. This approach requires a relatively small number of images to be acquired and processed during the rescue operation, which considerably reduces the cost of image processing, data transmission, and power consumption. The results of the conducted experiments demonstrated that this system has the potential to achieve high performance in detecting living humans in obstructed environments relatively quickly and cost-effectively. The detection accuracy ranged between 79% and 91% depending on a number of factors such as the body position, the light intensity, and the relative color matching between the body and the surrounding environment.

Keywords

Urban Search and Rescue, Human-Robot Interaction, Autonomous Systems, Intelligent Interface, Neural Networks

1. Introduction

In light of recent catastrophes, which are either natural such as earthquake, hurricane, flood, and fire, or man-made such as wars, bombing, and terrorist attacks, the need for creating new ways to rescue the survivors who

get trapped under rubble, in the fastest way possible, is a must. According to the field of Urban Search and Rescue (USAR), the probability of saving a victim is high within the first 48 hours of the rescue operation, after that, the probability becomes nearly zero [1].

During the rescue operation, several human factors are deployed such as fire fighters, policemen, and medical assistants. All of them are exposed to very dangerous situations which are caused by the destructed environment they work in, such as collapsed buildings, landslides, and crater. Hence, the rescuer may become a victim who needs to be rescued. Therefore, the rescue operation imposes a substantial risk on the rescue personnel themselves. From this point of view, looking for alternatives to human rescuers has been of a great demand.

Trained dogs have also been used in this field because of their high sensitivity to any slight motion or human presence. However, in such situations, it is hard to depend totally on them because even if they can detect the presence of a living victim, they still cannot judge the situation or relay the information quickly and systematically. For that reason, dogs cannot work independently but rather side by side as assistants to human rescuer. Therefore, the need is for a totally or partially independent alternative to the human factor.

Meanwhile, robots were achieving relatively good progress in other fields such as education, industry, military, and medicine and have proven their robustness and efficiency in most of what they are designed for. Therefore, robots are expected to play an important role in replacing the human factor in many of these fields.

Even though the USAR field seems to be very challenging for robots, robots have already invaded this field like many other fields. The first real attempt to use USAR robots was during the rescue of the World Trade Center (WTC) disaster in New York in 2001. Since then, a worldwide interest in using robots in rescue operations has been rapidly increasing.

During the WTC rescue operation, six remotely-controlled robots were used to assist human rescuers in finding and locating trapped victims [1]. These robots were remotely-controlled by operators. Each robot was equipped with a camera and a wireless communications link and was supposed to transmit images of the scenes they traverse. However, the attempt faced several critical problems due to the remotely-controlled nature of the robots:

- 1) More than 25% of the communications between the wireless robot and the control unit was extremely noisy and therefore was useless. This eventually led to a loss of communication between the robots and the operators. As a result, the robots stopped working totally and got lost in the rubble because they do not have any autonomous intelligence to continue on their own.
- 2) A very high communication cost was incurred due to the large number of transmitted images to the operator, especially when the communication line is very error-prone.
- 3) A very high processing cost is also incurred by the capturing, storing, and transmitting a large number of images.
- 4) There was a continuous need for illumination due to the dark nature of these environments. Therefore, a continuous light source would be needed in order to be able to acquire and send images. This requires a relatively large power supply, which is not feasible in such situations.

As a result, the attempt of using robots at the WTC failed to accomplish even a small part of its mission [1].

This research is an attempt to tackle the problems faced the USAR robots during the WTC rescue operation via equipping the robot with some autonomous intelligence towards the ultimate goal of designing a fully autonomous robot.

In this research, a new approach for detecting surviving humans in destructed environments using an autonomous USAR robot is proposed. The proposed system uses a passive infrared sensor (PIR) to detect the existence of living humans and a low-cost camera to acquire snapshots of the scene as needed. Having detected a sign of a living human, the PIR sensor triggers a camera to acquire an image of the scene. The image is fed into a feed-forward neural network (NN), trained to detect the existence of a human body or part of it in the image for different positions within an obstructed environment. This approach requires a relatively small number of images to be acquired and processed during the rescue operation. In this way, the real-time cost of image processing and data transmission is considerably reduced. The transmitted information may include any relevant phenomena that the robot is prepared to sense such as the location, the temperature, the level of certain types of gases, etc. The information may also include the image itself in order for the rescue team to have a better evaluation of the situation.

The robot is assumed to have the capability to determine its current location in real-time, to wirelessly communicate with the rescue team, and to locally store the status and location information about the trapped victims

in case the wireless communications link is temporarily disconnected.

The environment is assumed to have obstacles around the surviving victims such as debris, dust, and collapsed walls such that the human body is partially or mostly covered.

In order to prove the concept and the functionality of the system, the autonomous robot was emulated by attaching a laptop PC (representing the brain of the robot) to a large toy car (representing the navigation mechanism of the robot). The car was fully controlled by the laptop PC via an external interface electronic circuit that was designed and built for such purpose. The laptop PC was equipped with a low-cost, medium-resolution web camera and a PIR sensor, and was running the MATLAB software.

- 1) By adding such capabilities to the robot, the problems that faced the USAR robots during the WTC rescue operation are tackled as follows:
- 2) Even if the robot loses communication with the operator temporarily, it still can continue with its mission of looking for survivors until the communication is resumed. The critical collected information is saved in an internal memory and is transmitted as soon as the communication is resumed. The robot may also have the capability to maintain and dynamically update internal coverage map of the wireless communications quality.
- 3) The communication cost is significantly reduced by mostly sending small-size encoded information to the rescue team about the location, and perhaps the status, of the victim rather than transmitting heavy streams of images. In the worst case, a single image of the scene is transmitted for better evaluation of the situation.
- 4) The image processing cost is reduced by acquiring only a single image of a scene as needed rather than a stream of images.
- 5) As a result, the power consumption is considerably reduced easing the requirements on the power source specifications. In the worst case, the light source (e.g. a flash light) is used only for taking the snapshots when the illumination at the scene is insufficient.

A large number of experiments were conducted in order to evaluate the performance of the system. The results of the experiments demonstrated that the system has the potential to achieve high performance in detecting living humans in obstructed environments relatively quickly and cost-effectively. The detection accuracy of the system ranged between 79% and 91% depending on a number of factors such as the body position, the light intensity of the scene, and the relative color matching between the body and the surrounding environment.

The rest of the paper is organized as follows: Section 2 discusses the related work; Section 3 presents the implementation of the proposed system; Section 4 explains the image processing and neural network strategies used; Section 5 presents the experimentation methodology and the test results; and Section 6 provides the conclusion.

2. Related Work

Image processing, neural networks (NN), and passive infrared sensing were used in detecting the presence or shape of the human body for several applications such as surveillance and emergency rescue services. However, to the best of our knowledge, a combined use of all of them in the way proposed here has not been cited.

In this section, a brief discussion of some of the previously proposed techniques in the literature is presented.

Moradi presented a method that uses the infrared radiation emitted by the human body using an infrared camera that acquires thermal images of the surrounding environment [2]. This system includes six main parts: Image Acquisition, Initial Processing, Feature Extraction, Acquisition Storage, Knowledge Base, and Recognition. The features within the images are then classified using the Dystall NN. The number of nodes and layers in this system depends on the desired accuracy. The difference between the training stage and the testing stage of the NN is compared with a pre-defined threshold. If the difference is more than that threshold, then a human is detected. This method has a number of advantages that distinguish it from other similar methods such as: real-time processing speed, high efficiency, and fast learning. However, there are several factors that complicate the process of the detection and recognition such as: the absorpency of thermal energy by the environment and the victim's clothes, the existence of identical thermal sources such as fire or another living creature (e.g. an animal), and that the images taken by the IR camera usually have relatively low resolution. In addition to that, the cost of the thermal camera is relatively very high.

Burion presented a project that aimed to provide a sensor suite for human detection for the USAR robots [3]. This study evaluated several types of sensors for detecting humans such as pyroelectric sensor, USB camera,

microphone, and IR camera. The pyroelectric sensor was used to detect the human body radiation, but its limitation was its binary output. The USB camera was used for motion detection, but its limitation was its sensitivity to changes in light intensity. The microphone was used for long duration and high amplitude sound detection, but it was severely affected by noise. Lastly, the IR camera was used to detect humans by their heat image, but it was affected by other nearby hot objects. The algorithm was based on collecting data from all these sensors as an attempt to improve the robustness of the final results. The main idea was to detect a change in the image scene by checking the values of the pixels. Several images for the scene were acquired and subtracted from each other to discover if a motion has occurred. The used technique was fairly efficient in detecting the victims. However, the robot was not fully autonomous and was dependent on the operator. No neural network was used in the control program, and no detection of the human shape was used.

Nakajimaa *et al.* proposed a system that learns from examples in order to recognize the person's body in the images taken indoors [4]. The images were represented by color-based and shape-based features. The recognition process is carried out by using the linear Support Vector Machine (SVM) classifiers. This system works in real-time and it can achieve high recognition rate on normalized color histograms of people's clothes. However, the main limitation of this system is that it demonstrated high performance rates only when both the training and test images were recorded during the same day. When the test set contained images of a day that is not represented in the training set, the performance of the system drops down to about 53% due to the change of clothing the person might have every day.

Mohan *et al.* presented a hierarchical technique for developing a system that locates people in images [5]. In this Adaptive Combination of Classifiers (ACC) technique, the learning occurs in multiple stages. The system is first trained to find the four components of the human body: the head, legs, left arm, and right arm separately from each other. After ensuring that these parts are present in a proper geometric configuration, the system combines the results in order to classify a pattern as either a "person" or a "non-person". The obtained results demonstrated that this system performed significantly better than a similar full-body person detector [4]. Moreover, the system handles the variations in lighting and noise in an image better than a full-body detector.

Engle and Whalen presented a simulation study for a multiple robots foraging in a simulated environment [6]. The robots were supposed to gather two different sets of resources from a walled environment containing obstacles. Each set of resources must be returned to a specific goal. The robots were equipped with laser range finder for long range obstacle detection, a thermal sensor which detects heat around resources, and a color sensor to discriminate between different types of resources. The robots were supposed to search for humans by their body heat, and distinguish humans from non-human objects by using the color sensor. The final simulation was done using just one robot due to time constraints and implementation complexity and the used navigation algorithm could not cover the whole environment. In addition, the financial cost of the robot was relatively high. Cavalcanti and Gomes presented an approach for determining the presence of people in still images [7]. It starts with the segmentation of the skin areas in the image using YCbCr skin filter using a color space with a set of experimental thresholds. Then, the detected skin areas are grouped together to form body regions. This technique was been applied on a database of still images and achieved 77.2% correct classification. The advantage of this approach is its simplicity and relative low processing cost. However, higher-level analysis would be needed in order to achieve higher detection performance.

Several research papers [8]-[12] were published regarding the performance of the National Institute of Standard and Technology (NIST) Standards. The papers list the criteria for designing standard arenas and victim models and their characteristics. The arenas are supposed to simulate the destruction caused by the real catastrophes. In addition, they are supposed to provide tangible, realistic, and challenging environments for mobile robot researchers interested in urban search and rescue applications. The arenas were modeled for buildings in various stages of collapse, where the robots were allowed to traverse and were tested for obstacle avoidance and victim detection by repeatedly testing their sensory, navigation, mapping, and planning capabilities. These arenas differ in their level of complexity. They are divided into three levels, the Yellow arena, the Orange arena, and Red arena. The Yellow arena, which is the simplest one, simulates a home after an earthquake or a similar disaster. The floor of this arena is easy to navigate so that most robots can handle it. There is no loose paper, no ramps, etc. The Orange arena is harder than the Yellow arena. The major distinguishing feature of the Orange arena is a ramp with a chicken-wire. In addition, the floor is covered with paper, which is much harder to traverse by small wheeled robots. The Red arena is much harder than the other two. There are rubble and obstacles all over the place. There is also enough metal to cause serious problems to wireless communication. Moreover,

simulated victims with various signs of life can be found throughout the three arenas. Some of these signs are human form, motion, audio, thermal, and chemical signatures that represent victims of various states. The robots are awarded points for accurate mapping and victim detection according to the predefined performance metrics.

Trierscheid *et al.* proposed a new technique for detecting presence of a victim [13]. The technique depends on the spectral imaging that analyzes an invisible light radiation of the human body, which is a near infrared (NIR) light that has shorter wavelengths. This light has been used too to solve the problem of dust coverings in rescue environments. The advantage of this technique is that it does not depend on temperature, unlike thermal imaging, such that cold bodies (*i.e.* dead persons) can also be detected. The algorithm was responsible for locating victims and marking them in a given image.

Zhao *et al.* presented a technique for detecting the breathing and heartbeat signals of a living human [14]. This system can detect passive victims who are either completely trapped or too weak to respond to the existing traditional detection systems. The system depends on sending a low-intensity microwave beam to penetrate through the rubble. If a living human exists, his/her small amplitude body-vibrations due to the breathing and heartbeat will modulate the back scattered microwave signal. After that, the reflected wave from the body is demodulated and the breathing and heartbeat signals can be extracted in-order to locate the buried human. This system was able to detect buried humans under rubbles of nearly 90cm of thickness.

3. System Implementation

In order to emulate the proposed search and rescue robotic system and evaluate its performance, a laptop PC, which represents the main processing and control unit (MPCU) of the system, was attached to a navigation mechanism (e.g. a large toy car) via a navigation control unit (NCU).

The NCU consists of electronic circuitry that interfaces the sensors (as system inputs) and the navigation mechanism (as a system output) to the main processing and control unit. The system was equipped with three types of sensors:

- 1) A PIR sensor to detect the presence of a living human based on body radiation.
- 2) An Infra-Red (IR) range sensor to detect the obstacles in the way of the robot.
- 3) An image sensor, which is a traditional web camera to acquire still images according to orders from the control program when a need arises.

The MPCU consists of the following components (see **Figure 1**):

- 1) The navigation algorithm, which is responsible for steering the robot throughout the search area and around obstacles, based on the readings of the PIR and the IR sensors while searching for survivors.
- 2) The location tracking algorithm, which is responsible for calculating the real-time coordinates of the robot in order to report the locations of the survivors to the rescue personnel. The real implementation of a search and rescue robot is assumed to be equipped with a Global Positioning System (GPS) or a GPS-like system for location identification.

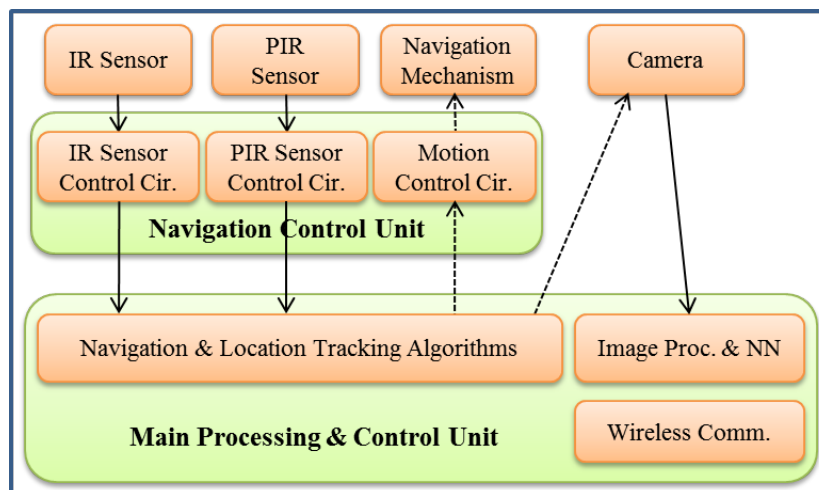


Figure 1. Block diagram of the system.

- 3) The image processing and neural networks unit (IPNNU), which is responsible for detecting a human being in a picture received from the camera. This unit is the core of this research and will be discussed in details in the following sections.
- 4) The wireless communication unit is responsible for providing reliable communication link with the rescue personnel in order to provide the locations of the survivors in addition to any other information the system may be designed to provide such as a snapshot of the scene so that the rescue personnel may be able to assess the situation of the survivor and prioritize their rescue tasks. In our system, the laptop PC was equipped with a wireless local area networking adapter, which was used for communications.

It should be noted that the design of the navigation and location tracking algorithms is beyond the scope of this research. Therefore, simple obstacle avoidance navigation algorithm and simple step counting location identification were used in the system.

The work flow of the system is as follows:

- 1) The MPCU continuously reads the IR range sensor via the NCU. When the IR sensor is triggered (*i.e.*; an obstacle is detected), the MPCU signals the NCU to change the direction of the movement based on the navigation algorithm and keeps track of the current location and the status of the communication link.
- 2) The MPCU periodically reads the PIR sensor. If the PIR sensor is triggered, the MPCU signals the NCU to stop the robot and the camera to acquire an image.
- 3) The image (*i.e.*; a still frame) is passed to the IPNNU, where it is preprocessed by the image processing algorithm (*e.g.*; converting to gray scale, resizing, and canceling some background effects) and is then fed to the pre-trained NN.
- 4) The NN determines whether a human shape or a body part exists in the image and returns the result back to the MPCU by categorizing the image based on the training sets used.
- 5) If a human is detected, the MPCU reports the location to the rescue center. If the communication link is not available, the MPCU may decide to buffer the information and continue its normal operation until the communication is resumed. It may choose to move back to a nearby location where the communication link is known to be usable before it continues. The MPCU might also decide to send the image as well if a need arises. However, the cost of transmitting an image is relatively high in terms of delay, processing, and communication. Therefore, sending an image must be well justified. Otherwise, the battery lifetime of the robot will be reduced significantly. The MPCU may also have the capability to compress the image before sending it, which may reduce the cost significantly, depending on the resources available at the MPCU.

Figure 2 shows the simplified flowchart of the system operation.

Before the experiments were conducted, the detection ranges and coverage angles of the PIR sensor and the camera had to be well-tested and unified in order to make sure that the camera is capable of “seeing” the survivor that has triggered the PIR sensor with enough quality for processing and detecting. The test results were as follows:

- The maximum detection range of the PIR sensor was found to be 5 m at 180°.
- The maximum distance between the camera and the object, D_{\max} , of which the image is to be acquired with an acceptable image quality, was found to be 2 m.
- The radius, R , of the image taken at 2 m by the camera was found to be about 1.15 m (see **Figure 3(a)**).
- Based on the detection range measurements of the camera, the coverage angle, θ , was calculated as follows:

$$\theta = 2 \times \tan^{-1} \left(\frac{R}{D_{\max}} \right) = 60^\circ \quad (1)$$

Due to the incompatibility in the detection range and the coverage angle between the PIR sensor and the camera, the following procedure was performed in order to obtain a unified coverage of both:

- 1) In order to reduce the coverage angle of the PIR sensor, an aluminum cap was used. The length of the cap, L , was determined as related to the diameter of the PIR sensor lens, d , which was 9.5 mm, as follows (see **Figure 3(b)**):

$$\tan \left(\frac{\theta}{2} \right) = \frac{d}{L} \quad (2)$$

$$L = \frac{9.5 \text{ mm}}{\tan(30^\circ)} = 16.5 \text{ mm} \quad (3)$$

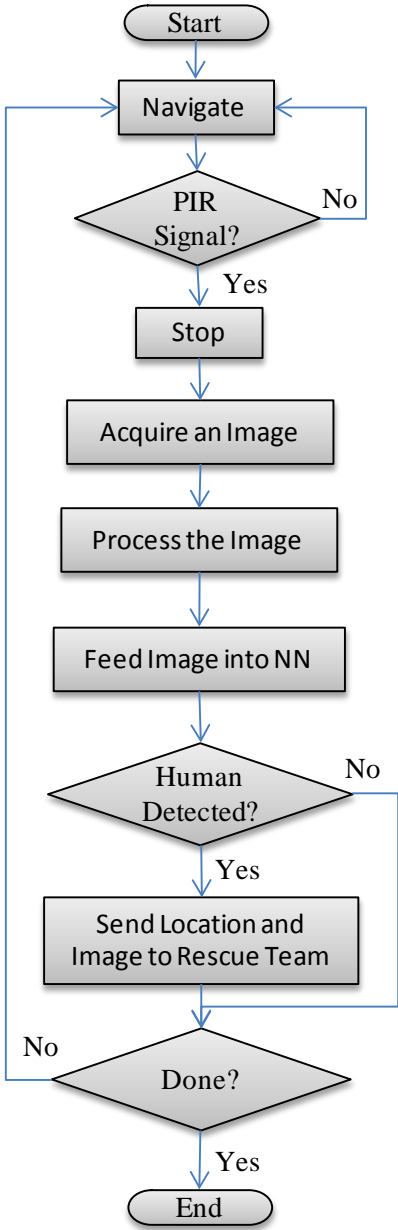


Figure 2. The simplified system operation flowchart.

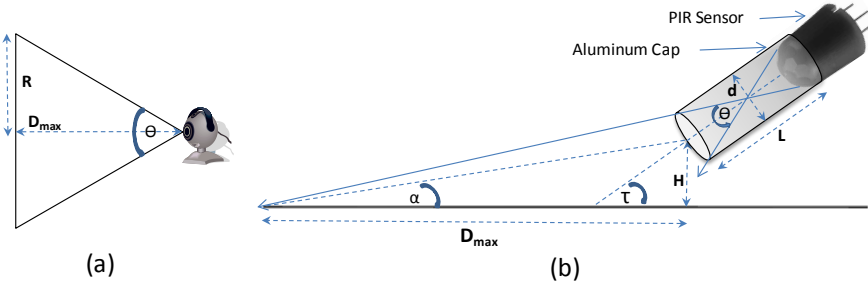


Figure 3. The cover angle of the camera.

- 2) In order to decrease the effective detection range of the PIR sensor from 5 m to 2 m, the PIR sensor was tilted down towards the ground by a calculated degree. Assuming that the PIR sensor is positioned on the robot such that the far edge of the cap is at a certain height, H , from the ground, the tilting angle, τ , of the sensor is calculated as follows assuming that $H = 0.5$ m:

$$\alpha = \tan^{-1} \left(\frac{H}{D_{\max}} \right) \quad (4)$$

$$\tau \approx \frac{\theta}{2} + \alpha \quad (\text{since } d \ll D_{\max}) \quad (5)$$

$$\alpha = \tan^{-1} \left(\frac{0.5 \text{ m}}{2 \text{ m}} \right) = 14^\circ \quad (6)$$

$$\tau = \frac{60^\circ}{2} + 14^\circ = 44^\circ \quad (7)$$

4. The Image Processing and Neural Networks Unit

As stated previously, when the PIR sensor is triggered, the MPCU signals the camera to acquire an image of the scene, which is passed to the IPNNU for processing and detection. In this section, the IPNNU is discussed.

4.1. The Main Features

The type of NN used for the IPNNU is the feed-forward with a supervised learning method [15] [16].

The process started by training the neural network on a large number of images for human hands, feet, and bodies of different sizes and positions as follows:

- 1) The set of training images, Tr , contained 500 RGB images of 172×140 pixels each.
- 2) At each training cycle (or epoch), six images were fed as input to the training stage, of which two are for hands, two for feet, and two for bodies.
- 3) Before feeding an image into the NN, some simple enhancements were first applied in order to reduce the NN processing cost as much as possible [17]. These enhancements were performed as follows:
 - 1) The image was converted into gray-scale.
 - 2) The resolution was adjusted by down sampling and deleting some unwanted background effects.
 - 3) The brightness was adjusted by increasing the contrast.
 - 4) The image was converted from a 172×140 matrix into a $24,080 \times 1$ vector in order to feed it as an input to the NN.

The NN used had three layers (see **Figure 4**):

- 1) The input layer consists of 256 nodes with each representing one of the 8-bit gray scales. The pixels of each image were fed to this layer represented as a vector.
- 2) The hidden layer consists of 256 nodes with each also representing a gray degree.
- 3) The output layer consists of three nodes that represent the desired values. Each node fires only if its desired input is given (*i.e.*; foot, hand, or body) as input to the first stage.

4.2. Neural Network Training and Testing

Before being used in the system, the NN had to go through an iterative training and testing phase. The training set, T_r , was divided into 10 groups of 50 images each. The testing set, T_e , consists of 50 images that the NN was never been trained on. The training and testing procedure was as follows:

- 1) Train the NN on a new set of 50 images from Tr .
- 2) Test the NN using T_e by feeding the images of T_e as inputs to the NN and recording the result for each.
- 3) Calculate the detection accuracy of the NN, which represents the percentage of successful detections.

Repeat the process until all training sets are used.

Figure 5 depicts the detection accuracy, $A(n)$, obtained as a function of the number of training images, n . It is noticed that the detection accuracy monotonically increases with the n . However, it slows down and starts to

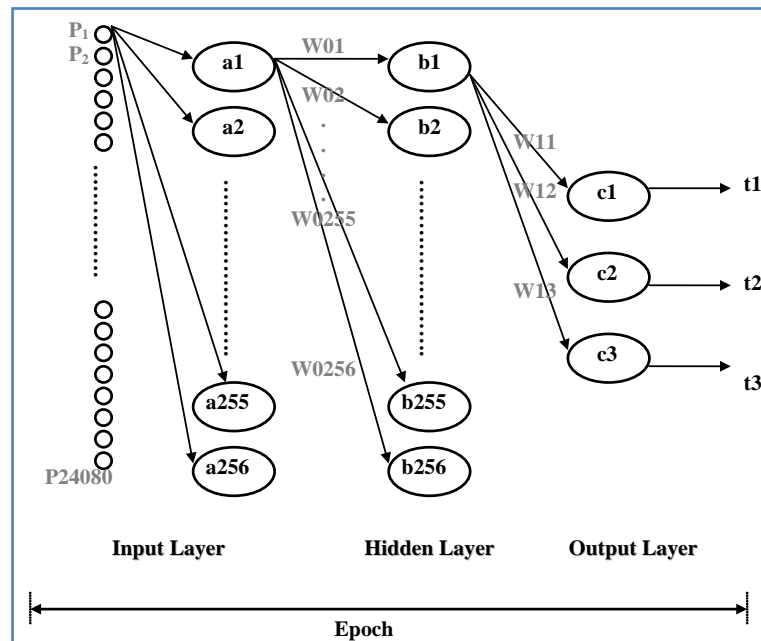


Figure 4. The architecture of the Neural Network used.

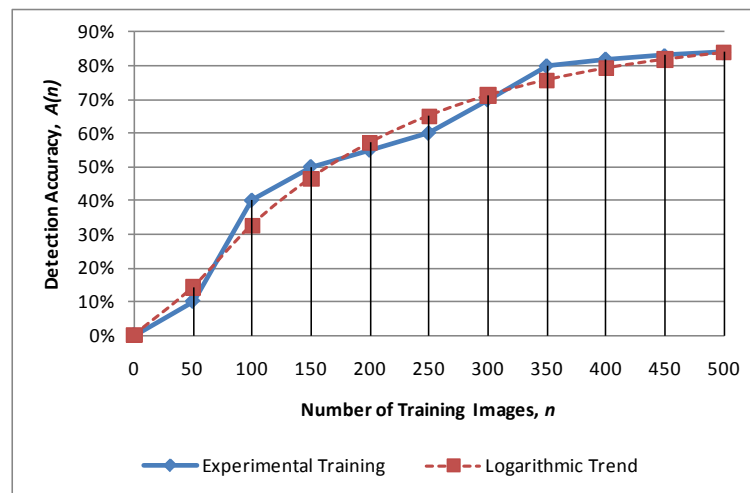


Figure 5. Neural Network detection performance curve.

saturate for $n > 350$ to about 84% accuracy at 500 images, which indicates a logarithmic behavior. Therefore, experimental regression analysis was performed on the training results and the trend of $A(n)$ was estimated as:

$$A(n) = 1 - e^{-\left(\frac{n}{180}\right)} \quad (8)$$

5. Experiments and Test Results

In order to evaluate the performance of the system, the rescue scenarios were performed by a number of designed experiments. Each experiment consisted of a combination of the following characteristics:

- 1) The human body position such as: the body or part of it is laid down in the middle of the picture or to the side of it, facing forward, facing backward, etc. The different body positions used are listed in **Table 1** and **Figure 6** shows a sample of the body positions used during the experiments. It should be noted that for Cases 5 and 6, the PIR sensor is triggered, but the IPNNU is not expected to detect a human in the picture because



Figure 6. Sample body positions used for experiments.

it either almost totally covered (only allowing the body radiation to be sensed) or is in a position, where none of the human body features is clearly visible. Therefore, the detection of a human in these cases is considered as a false positive type of decision.

- 2) The environmental characteristics around the body such as the lighting level and whether there are features or objects around it. The features and objects were emulated by pieces of furniture for indoor experiments and rocks and trees for outdoor experiments. The different types of environments used and their characteristics are listed in **Table 2**. This combination of features was used in order to test the robustness of the system

Table 1. The cases of body position.

Scenario	Position Description
Case 1	A full body was laid down facing the camera with no hidden parts
Case 2	A full body was laid on the side of the detection area with no hidden parts
Case 3	Only the hand of a human body is visible
Case 4	Only the foot of a human body is visible
Case 5	A human body was totally covered so that the shape is invisible but the body radiation is allowed to be sensed
Case 6	Human head was placed against the camera such that most other parts are invisible

Table 2. The environments and their characteristics.

Env. Name	Environment Description
ENV1	Very dark environment without any features or objects
ENV2	Very bright environment without any features or objects
ENV3	Fairly lit environment without any features or objects
ENV4	Fairly lit environment with features and objects

in detecting the human body when it is either surrounded or partially covered with debris or rocks or when it is laid in a dark area, etc.

- 3) The color levels of clothes dressed by the human being rescued; ranging from very dark to very bright. The reason this factor was added is to test the robustness of the system in recognizing the features of the human body under different combinations of color contrast levels between the body itself and the surrounding environment. The color contrast levels used are listed in [Table 3](#).

5.1. Analysis of the Results

All combinations of body positions, environments, and color contrast levels were tested with 5 trials per experiment for a total of 120 experiments and 600 trials.

The detection accuracy test results of all experiments are listed in [Table 4](#). For each experiment, the IPNNU takes a decision as either T, when a human is detected, or F, when no human is detected. For each Case/Environment combination, the total number of T and F decisions is listed.

- 1) The summary of the results for each Case/Environment combination is listed in [Table 5](#) and depicted in [Figure 7](#). The results of this part provide the following observations:
- 2) For Cases 5 and 6, where the system is expected to make only F decisions, as discussed above, the system made a relatively small number of false-positive decisions with an error rate of 7% to 8%.
- 3) Cases 1 and 2 produced a relatively high level of detection accuracy with 11% to 13% error rate since most of the body features were clearly visible.
- 4) For Cases 3 and 4, even though only the hand or the foot of the body was visible, the system performed relatively well in detecting them with 21% to 25% error rate.
- 5) Environments 3 and 4 provided very good conditions for accurate detection with about 9% error rate, while environments 1 and 2 provided moderate conditions for accurate detection with 17% to 21% error rate. The reason is that in Environments 3 and 4, the light intensity was moderate providing enough light to grasp the body features, while in environments 1 and 2, the lack of enough light and the excess of it seem to have blurred the body features and made them less visible.
- 6) It is worth noting that the impact of the features and objects surrounding the body was fairly minimal as there is a negligible difference in performance between Environments 3 and 4.

The summary of the results for each Case/Clothes Color combination is listed in [Table 6](#) and depicted in [Figure 8](#). The results indicate that the very bright and very dark clothes color contrast seem to provide the worst

Table 3. The contrast levels of clothes colors.

Symbol	Color Contrast Level
VD	Very dark
D	Dark
M	Moderate
B	Bright
VB	Very Bright

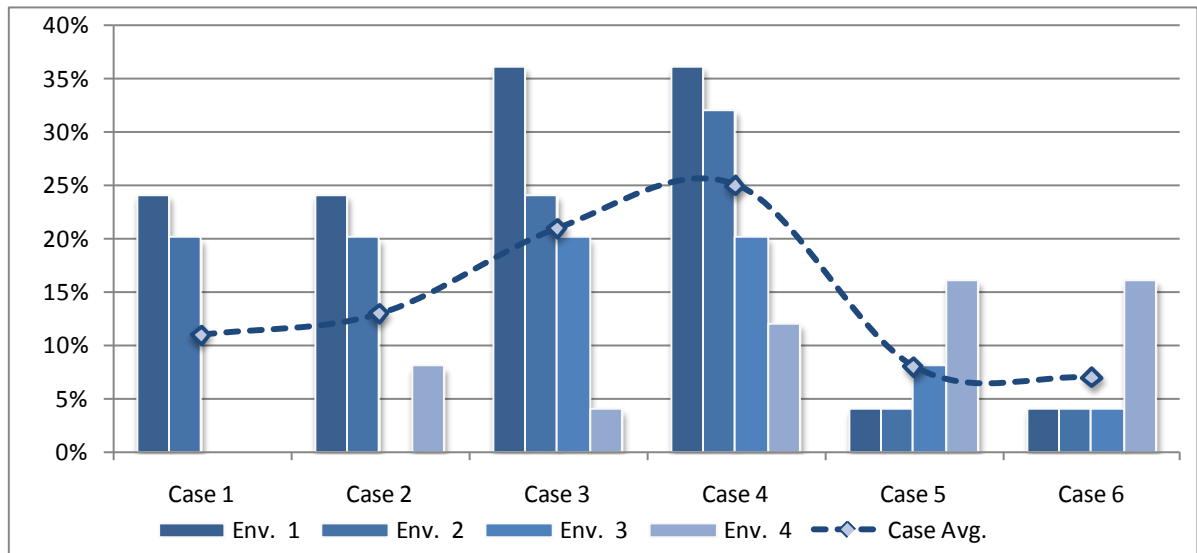


Figure 7. Overall detection error rate statistics per environment type.

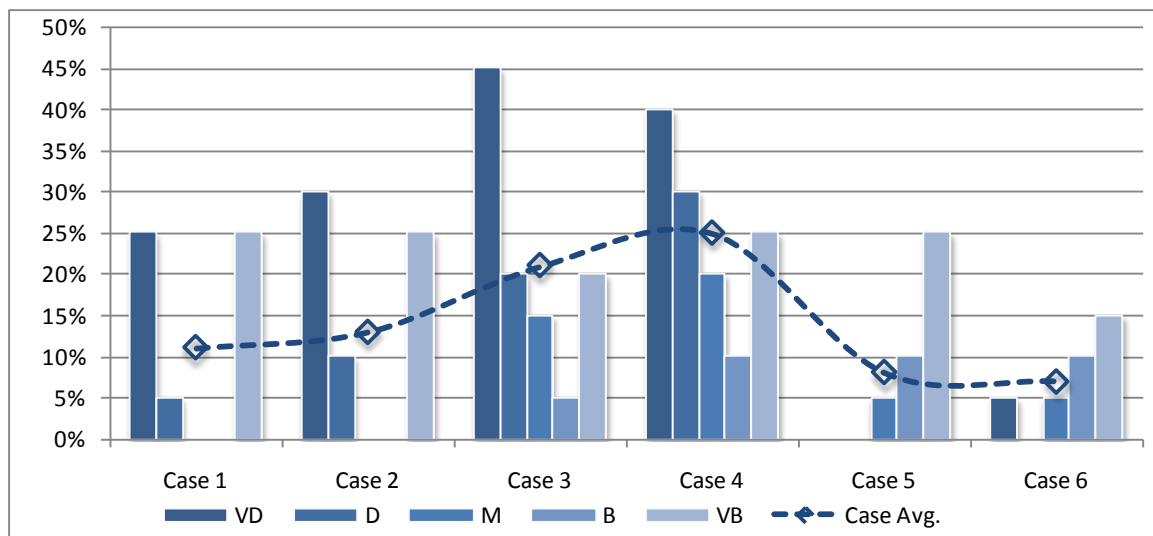


Figure 8. Overall detection error rate statistics per clothes color level.

conditions for detection accuracy with 22.5% to 24% error rate, while the medium to dark clothes color contrast provide the best conditions for detection accuracy is all cases and environment conditions with only 6% to 11% error rate.

Table 4. Experimental detection accuracy results.

Env.		ENV1		ENV2		ENV3		ENV4	
Decision		T	F	T	F	T	F	T	F
Case 1									
Clothes Color	VD	0	5	5	0	5	0	5	0
	D	4	1	5	0	5	0	5	0
	M	5	0	5	0	5	0	5	0
	B	5	0	5	0	5	0	5	0
	VB	5	0	0	5	5	0	5	0
	SUM	19	6	20	5	25	0	25	0
Case 2									
Clothes Color	VD	0	5	5	0	5	0	4	1
	D	4	1	5	0	5	0	4	1
	M	5	0	5	0	5	0	5	0
	B	5	0	5	0	5	0	5	0
	VB	5	0	0	5	5	0	5	0
	SUM	19	6	20	5	25	0	23	2
Case 3									
Clothes Color	VD	0	5	5	0	2	3	4	1
	D	2	3	5	0	4	1	5	0
	M	4	1	4	1	4	1	5	0
	B	5	0	4	1	5	0	5	0
	VB	5	0	1	4	5	0	5	0
	SUM	16	9	19	6	20	5	24	1
Case 4									
Clothes Color	VD	0	5	5	0	3	2	4	1
	D	2	3	4	1	4	1	4	1
	M	4	1	4	1	4	1	4	1
	B	5	0	4	1	4	1	5	0
	VB	5	0	0	5	5	0	5	0
	SUM	16	9	17	8	20	5	22	3
Case 5									
Clothes Color	VD	0	5	0	5	0	5	0	5
	D	0	5	0	5	0	5	0	5
	M	0	5	0	5	0	5	1	4
	B	0	5	0	5	1	4	1	4
	VB	1	4	1	4	1	4	2	3
	SUM	1	24	1	24	2	23	4	21
Case 6									
Clothes Color	VD	0	5	1	4	0	5	0	5
	D	0	5	0	5	0	5	0	5
	M	0	5	0	5	0	5	1	4
	B	0	5	0	5	1	4	1	4
	VB	1	4	0	5	0	5	2	3
	SUM	1	24	1	24	1	24	4	21

Table 5. Summary of the experimental detection error rates per environment type.

	C1	C 2	C 3	C 4	C 5	C 6	Avg.
E 1	24%	24%	36%	36%	4%	4%	21%
E 2	20%	20%	24%	32%	4%	4%	17%
E 3	0%	0%	20%	20%	8%	4%	9%
E 4	0%	8%	4%	12%	16%	16%	9%
Avg.	11%	13%	21%	25%	8%	7%	14%

Table 6. Summary of the experimental detection error rates per clothes color level.

Color	C 1	C 2	C 3	C 4	C 5	C 6	Avg.
VD	25%	30%	45%	40%	0%	5%	24%
D	5%	10%	20%	30%	0%	0%	11%
M	0%	0%	15%	20%	5%	5%	8%
B	0%	0%	5%	10%	10%	10%	6%
VB	25%	25%	20%	25%	25%	15%	23%
Avg.	11%	13%	21%	25%	8%	7%	14%

6. Conclusions

In this paper, a new system and methodology for detecting surviving humans in destructed environments using an emulated autonomous robot is proposed. The proposed system applies existing image processing and neural networks techniques for detecting human body or body parts within the camera's range of view. The performance of the system was validated and evaluated via experimentation.

A number of experiments were designed and conducted and a large number of scenarios were tested. For each scenario, the tests were conducted within a number of designed environments that differ in their illumination degree and existing surrounding features. In each environment, different positions of the human body were tested.

The experiments demonstrated that the proposed system can achieve a relatively high detection accuracy of up to 91% in certain conditions and an overall average detection accuracy of 86% for all experiments and environments tested.

This research demonstrates that using relatively simple image processing and neural networks techniques in critical applications such as the urban search and rescue is conceptually efficient and has the potential for playing a significant role in more sophisticated urban search and rescue systems.

References

- [1] Shah, B. and Choset, H. (2004) Survey on Urban Search and Rescue Robotics. *Journal of the Robotics Society of Japan*, **22**, 582-586. <http://dx.doi.org/10.7210/jrsj.22.582>
- [2] Moradi, S. (2002) Victim Detection with Infrared Camera in a Rescue Robot. *IEEE International Conference on Artificial Intelligence Systems*. http://ce.sharif.ac.ir/~rescuerobot/downloads/victim_detection.pdf
- [3] Burion, S. (2004) Human Detection for Robotic Urban Search and Rescue. Infoscience Database of the Publications and Research Reports, Technical Report.
- [4] Nakajimaa, C., Pontilb, M., Heiselec, B. and Poggioc, T. (2003) Full-Body Person Recognition System. *The Journal of the Pattern Recognition Society*, **36**, 1997-2006. [http://dx.doi.org/10.1016/s0031-3203\(03\)00061-x](http://dx.doi.org/10.1016/s0031-3203(03)00061-x)
- [5] Mohan, A., Papageorgiou, C. and Poggio, T. (2001) Example-Based Object Detection in Images by Components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**, 349-361. <http://dx.doi.org/10.1109/34.917571>
- [6] Engle, S. and Whalen, S. (2003) Autonomous Multi-Robot Foraging in a Simulated Environment. UC Davis Report.
- [7] Cavalcanti, C. and Gomes, H. (2005) People Detection in Still Images Based on a Skin Filter and Body Part Evidence.

The Brazilian Symposium on Computer Graphics and Image Processing,

- [8] National Institute of Standards and Technology (NIST) (2011) Performance Metrics and Test Arenas for Autonomous Mobile Robots. <http://www.isd.mel.nist.gov/projects/USAR/>
- [9] Jacoff, A., Messina, E., Weiss, B., Tadokoro, S. and Nakagawa, Y. (2003) Test Arenas and Performance Metrics for Urban Search and Rescue Robots. Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, 27-31 October 2003, 3396-3403. <http://dx.doi.org/10.1109/IROS.2003.1249681>
- [10] Jacoff, A., Messina, E., and Evans, J. (2002) Performance Evaluation of Autonomous Mobile Robots. *Industrial Robot—An International Journal*, **29**, 259-267.
- [11] Jacoff, A., Messina, E. and Evans, J. (2001) Experiences in Deploying Test Arenas for Autonomous Mobile Robots. *Proceedings of the 2001 Performance Metrics for Intelligent Systems*, Mexico, 4 September 2001, 1-8.
- [12] Jacoff, A., Weiss, B. and Messina, E. (2003) Evolution of a Performance Metric for Urban Search and Rescue Robots. *Proceedings of the 2003 Performance Metrics for Intelligent Systems*, Gaithersburg, 16-18 August 2003, 1-11.
- [13] Trierscheid, M., Pellenz, J., Paulus, D. and Balthasar, D. (2008) Hyperspectral Imaging for Victim Detection with Rescue Robots. *Proceedings of the 2008 IEEE International Workshop on Safety, Security and Rescue Robotics*, Sendai, 21-24 October 2008, 7-12. <http://dx.doi.org/10.1109/SSRR.2008.4745869>
- [14] Zhao, K., Wang, X., Li, Y. and Yu, X. (2006) A Life-Detection System for Special Rescuing Robots. The 9th International Conference on Control, Automation, Robotics and Vision, Singapore, 5-8 December 2006, 1-5.
- [15] Stergiou, C. and Siganos, D. (2006) Neural Networks. Surveys and Presentations in Information Systems Engineering. SURPRISE 96 Journal.
- [16] Russell, S. and Norving, P. (2003) Artificial Intelligence—A Modern Approach. 2nd Edition, Prentice Hall, Upper Saddle River.
- [17] Gonzalez, R. and Woods, R. (2002) Digital image processing. 2nd Edition, Prentice Hall, Upper Saddle River.