Victim detection with Infrared Camera in a "Rescue Robot"

Saeed Moradi

Electrical Engineering Department
Sharif University of Technology
Azadi Ave., Tehran, Iran
E-mail: Moradi@ee.sharif.edu

Abstract

Nowadays the presence of robots is a must in many fields. One of these, is assisting and rescuing injured people in a disaster such as earthquake, flood, ... where the human assistance is difficult due to hard and dangerous conditions and long delays, in such cases robots can be used to detect victims and report their locations to rescuers.
One of the most useful tools is IR (InfraRed) Camera. The body of a human is usually warmer than the environment. Furthermore it has an IR radiation, which is different from other things. So a victim can be detected with an IR Camera. This paper proposes a method to detect victims with images taken by an IR camera in an intelligent way with neural networks.

Keywords: Robot, Victim detection, Infrared camera, Intelligent method, neural networks.

.

Introduction

Generally, detection of an object in an image is so complicated that an intelligent and optimal method must be used.
This paper first introduces an optimum method for initial processing of the images taken by an IR camera. Due to the difficulties, which will be introduced later, this paper shows that used method is optimum and practical results prove this.
It shows how to use neural network method for recognition of the body of a human in taken images. Real time processing and high efficiency are the most important features of this method, which make it different from similar works.
So far, many approaches have been proposed in machine vision which do several tasks with different complexity levels and some of them have special uses (e.g. edge detection.)
Below are some of the successful systems used for this purpose [2][3][4]:
1) EGI: Horn et al. used advanced gaussian image model for 3D objects.

2) 3D PO: Bells & Haraud proposed 3D PO system for object recognition.

3) Ikeuch: He proposed a method for object recognition

4) Fauger & Hebert: They designed a system to localize and recognize the 3D objects.

5) Oshima & Shirai: They proposed a model-based system for objects with flat surfaces.

6) ACRONYM: Brooks proposed a system in which advanced cylinders were used to describe the model and objects.

Introduction of proposed method

According to Figure 3, a system of image understanding includes six main parts [1]:
Image Acquisition, Initial Processing, Feature Extraction, Acquisition Storage, Knowledge Base and Recognition.
The first step is image acquisition. Image function means a 2D brightness function f(x,y) where x,y represent the position and f(x,y) is relevant to the brightness or gray level of the image in that position.
The next step is initial processing, which includes low level processing. The image is processed to improve its quality. First the image is convolved with a gaussian filter. This filtration removes unnecessary information in the image. The more standard deviation increases, the more the clarity of edges increases. Then the gradient vector is calculated all over the image. In the next step the Non-Maximal suppression method is used. This method keeps only the most important

edges of the image. Finally the hysteresis method is used to connect edges and form continuous objects. This method put threshold for edges in the image with hysteresis to eliminate spurious responses.

The next level is middle level processing, which tries to find the relation between the results achieved by low-level processing and extract significant information. One of the well-known middle level processing methods is feature extraction which maps a view vector to the feature space. The most important goal of feature extraction is reducing the dimension of information by measuring special features, which separates input models. The input view vector is obtained by sampling in order to reduce the dimension and keep most of the information. The view vector will be mapped to a feature domain and information will be ranked relevant to its significance. Features should be independent of rotation, position and size. The output from this level is a vector from feature domain used as input in the classification level. In this level, the classifier, on the basis of the complete set of vectors, determines which of them are recognizable features of each object's class. This task should be done in a way that new vectors are placed in the correct classes with minimum error.

In the method proposed in this paper, it is assumed that points of object have value “1” and the other points have value “0”. Consider a two-dimensional coordinate system and suppose that the interested model is a set of black points distributed with probability density function $\rho(x, y)$. Two-dimensional torque with “p+q” degree of function $\rho(x, y)$ is m_{pq} that

$$m_{pq} = \sum_{x} \sum_{y} x^{p} y^{q} \rho(x, y)$$

Low degree torques have a close relationship with the object model:

m_{00} introduces area, m_{10} and m_{01} introduce the gravity center of the object, m_{11}, m_{02}, m_{20} introduce the four cardinal points.

These torques are being independent by geometric method with respect to rotation, size and position.

The final levels (Acquisition Storage, Knowledge base, Recognition) mean high level processing. The main features of Acquisition Storage is using different internal representation or understanding a complex representation from a small part of it.

Recognition level does the classification. In this level each object is related to a class by using the information provided.

Neural Network is a strong and robust tool for classification [1]. In a neural network classifier, descriptive methods are used. Descriptive methods are based on classification rules, which map the input feature vector to the output class. In this case classification rules have been stored in the "Knowledge Base". Each prior information about the image can be stored in the "Knowledge Base" too.

Learning in a neural network means finding a set of appropriate weights.

For finding the weights a set of appropriate vectors should be used as an input for finding the weights. These vectors are called “training vectors”.

For the method proposed in this paper, two training sets are used:

1. A set of the most probable two-dimensional images of the body of a human in several forms.

This set should be complete and includes several probable forms of body.

For example two 2D images of face from several points of view are the most important members of this set. Because the face isn't usually covered by clothes, so it is more detectable in the images taken by the IR camera.

In choosing the images (training vectors) noisy images should be taken into account (Figure 1).

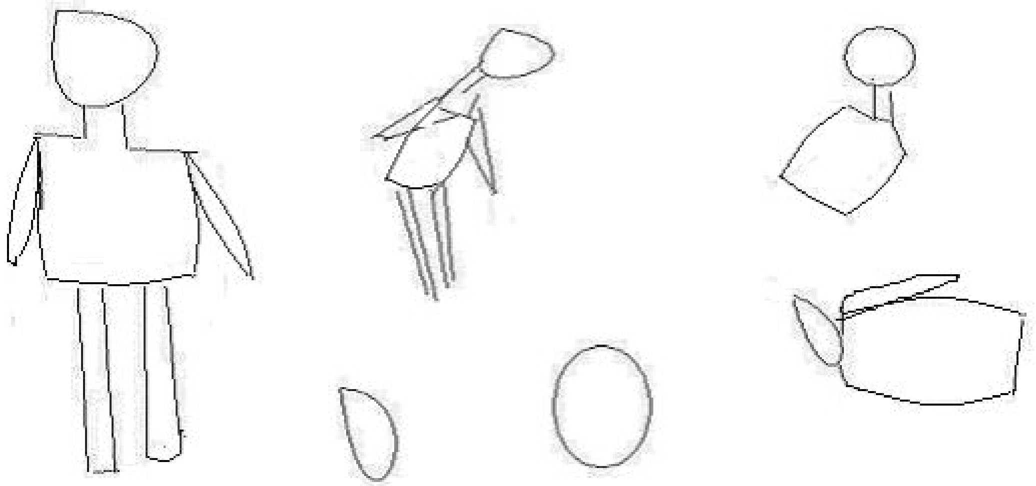

Figure 1

2. A set of 2D images that probably can't be an image of the human body.
For example, images which contain straight lines (Figure2).

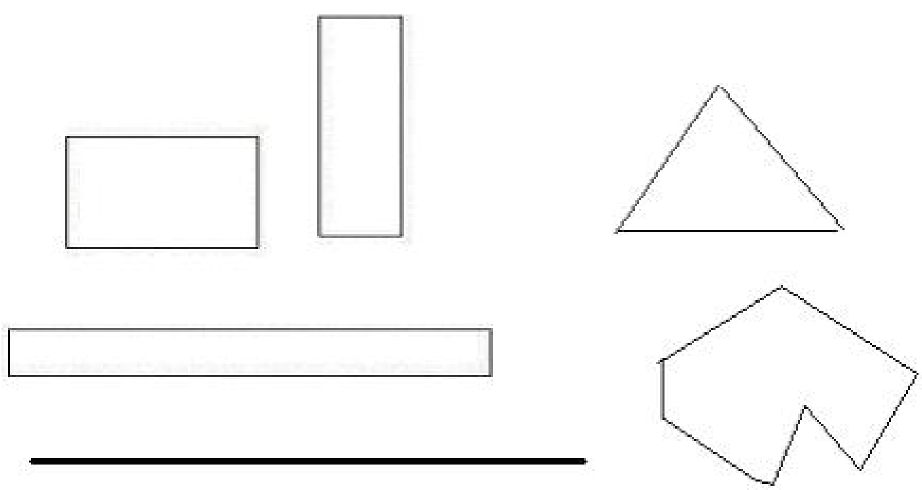

Figure 2

Now training vectors (in sets I, II) are used for training the neural networks.
The proposed neural network is Dystall neural network. Dystall network has these important properties:

1) Local learning.
2) Calculation complexity in algorithm is from order O(N). It means that learning time in a parallel machine doesn't depend on network's size. So, the number of necessary models for network learning doesn't increase incrementally with the image's complexity.
3) Monotonic convergence: The network can reach its stable values with out fluctuation.
4) It has the ability to learn, store and remember of noisy models.
5) Convenient implementation for learning in a VLSI chip.

The number of nodes and layers depends on the desired accuracy. If the network converges, we can test the network with fewer layers. Now we compare the difference of outputs of neural networks I and II with a threshold. If the difference is more than the threshold, the detected object is a human. So, we can implement this method with fuzzy logic.

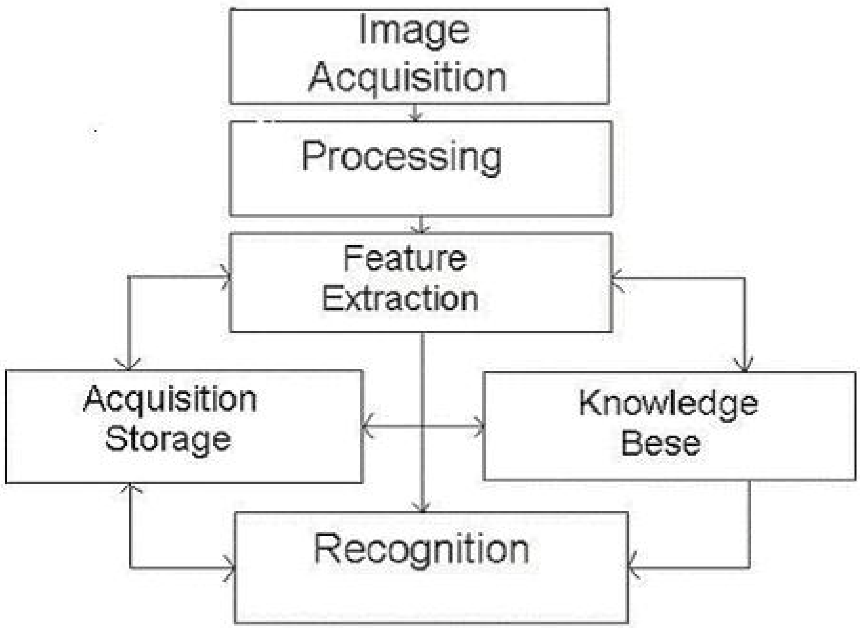

Figure 3

Difficulties of Implementation

There are many factors which complicate the detection and the recognition process:
Absorbency of thermal energy by the environment and victim's clothes, existence of identical thermal sources such as fire, another live being (e.g. an animal.) Furthermore, the images taken by the IR Camera usually have low resolution. All of the above factors make it difficult to analyze of taken images. (Figure 4).

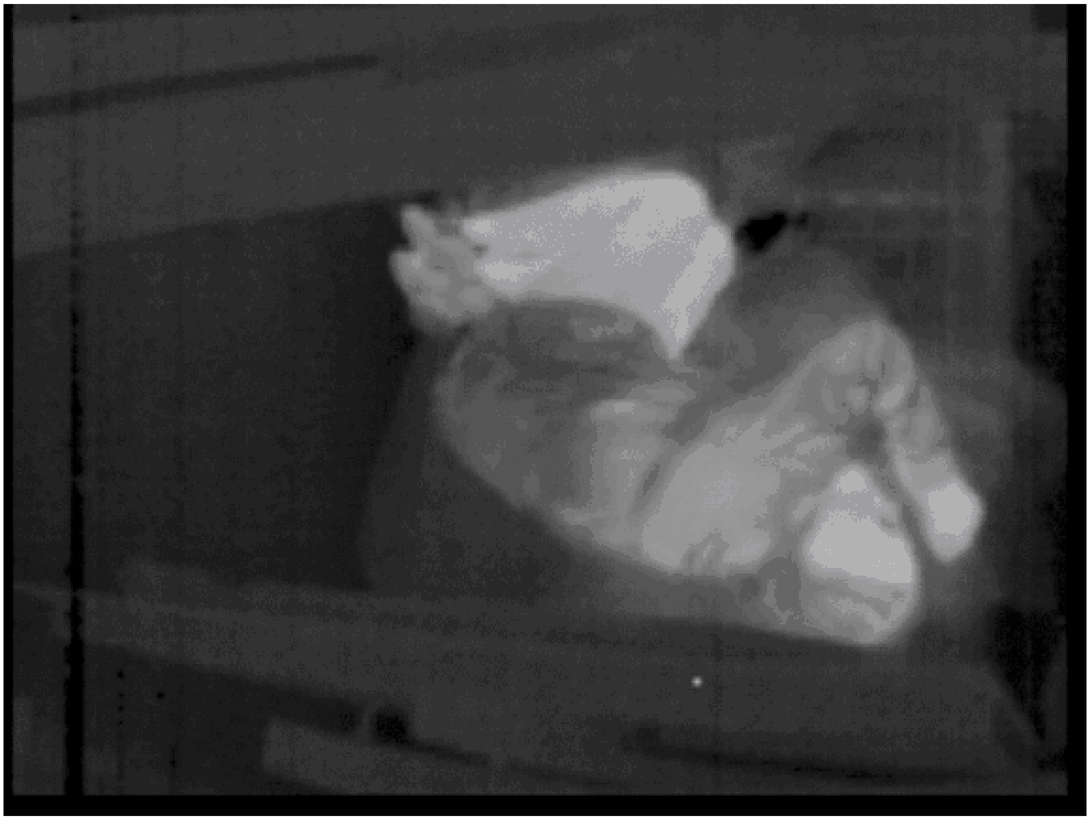

Figure 4

Advantages of proposed method

Many similar methods present the image as a structure in a linear vector domain and need a strong method for finding similarities between images. Vector domain methods need to solve the similarity finding problem: finding the points with similar features in two or more available images. This needs high computation. Furthermore, available methods for finding similar points don't have enough accuracy or involve complicate computation. This problem causes difficulties in achieving real-time system.

Because of the mentioned problems, we propose illustrated method for finding image features and using neural network for classification.

This system has many features that distinguish it from previous methods. E.g. high speed of the system, simulation in a parallel system, high efficiency and high-speed learning.

Experimental Results

For testing the proposed algorithms, this method was used in several experiments:

1) Change in scale: Images were scaled three times and their feature vectors were calculated. Then the average error was calculated for each image. The error percentage was less than %9.
2) Rotation: Each image was rotated 30° each time. Then their diversion from the feature vectors in the original images was calculated. The error percentage was less than % 10 .
3) Noise Response: In this test, for each image, 4 noisy images were created with 10db, 50db and 80db signal to noise ratio with random selection of some points in image and changing its value from 0 to 1 or from 1 to 0. Signal to noise ratio calculated from equation 20 log((L– H)/H), that L is the number of the original image points and H is the number of different points in the original image and the noisy image. For testing the noise response, the

feature vector of each noisy image was calculated and compared with feature vectors of the original image. The error percentage was less than similar methods described previously.

References:

[1].A. D. Kkulkarni, “Artificial neural networks for image understanding’” Van Nostrand Reinhold, NY, 1994

[2].T. J. Fan, G. Medioni, & R Nevatia, “Recognizing 3D objects using surface descriptors,” IEEE Trans. On PAMI, Vol. 41, No. 11, 1989.L.

[3].R. M. Haralic, & L. G. Shapiro, “Computer and robot vision,” Addison-Wesley, Reading, MA, 1993

[4].M. Snoka, V. Hlavac, & R. Boyle, “Image processing, analysis and machine vision,” Chapman & Hale Computing Great Britain, 1993