Human Victim Detection

Geert De Cubber, Gabor Marton
Royal Military Academy
Department of Mechanical Engineering (MSTA)
Av. de la Renaissance 30, 1000 Brussels
Geert.De.Cubber@rma.ac.be

Keywords

Person Detection, Victim Detection, Computer Vision, Recognition

Abstract

This paper presents an approach to achieve robust victim detection from color video images. The applied approach goes out from the Viola-Jones algorithm for Haar-features based template recognition. This algorithm was adapted to recognize persons lying on the ground in difficult outdoor illumination conditions.

1. Introduction

1.1 Problem Statement

In the event of a crisis situation, a primordial task for the fire and rescue services is to assess whether there are still human victims on the incident site, as the outcome of this assessment determines directly the risk the crisis management teams will be willing to take themselves. However, in many cases, one needs to enter the danger zones just to make this assessment. To this end, some tele-operated mobile robots have been put into service worldwide, allowing a human operator to look for victims from a distance. However, with the current advance in autonomous robotics, this task could be automated further, if there would be a reliable way to automatically detect human victims, even in difficult outdoor illumination conditions. This is one of the goals of the View-Finder project and the subject of this article.

The human victim detector presented here is part of a mobile robot control suite [4], and thus makes it possible to raise an alarm for a remote operator when a victim is detected. The remote operator can then determine whether the classification was correct and take the appropriate actions. As such, an incident site can be scanned for victims semi-autonomously by a mobile robot, relieving the already stressed human crisis managers from the tedious and dangerous task of searching for victims.

Keywords: Person Detection, Victim Detection, Computer Vision, Recognition
Funding: View-Finder FP6 IST 045541

1.2 Relation to previous work

Automated human victim detection is a very difficult task, especially in complex, unstructured environments. In order to detect a human victim, one or more physical parameters of the victim need to be perceived by a sensor. These physical parameters can be [2]: voice, temperature, scent, motion, skin color, face or body shape. In recent years, a number of research teams have developed human victim detection algorithms based on the detection of these physical parameters. In this section we'll give an overview of these activities and approaches used for human victim detection.

1.2.1 Voice

Kleiner et al. [6] perform audio-based victim detection by positioning two microphones with known distance. Given an audio source left, right or between both microphones, the time difference, i.e. phase shift, is measured between both signals. This is carried out by the Crosspower Spectrum Phase (CSP) approach, which allows to calculate the phase shift of both signals based on the Fourier transformation. Using this approach, the bearing of the sound source can be successfully determined, if there is not too much background noise. In the context of the View-Finder project, no a priori information about the sounds in the environment is given, so the voice-based approach was not followed.

1.2.2 Temperature (body heat)

Infrared cameras give a complete picture of the environment heat which is very useful in human detection. Although infrared cameras are very expensive, they seem the best solution to make the discrimination between human and non human presence and, as such, seem to be essential to a robust and efficient solution for human finding. Infrared cameras are used by Kleiner and Kummerle in [5], Nourbakhsh et al. in [7] and Birk et al. in [3]. A new development in this field is the hyperspectral IR imaging approach of Trierscheid et al. [9]. Hyperspectral IR images contain a contiguous spectrum in the bandwidth of IR in a spatial scanline in the scene and provide the technique to combine spectroscopy and imaging, which makes it very suited for human victim detection.

Pyroelectric sensors are another type of heat-based human detectors. These sensors are designed specifically for human detection. This sensor is made of a crystalline material that generates a surface electric charge when exposed to heat in the form of infrared radiation. It is calibrated to be sensitive to human heat wavelength (8 - 14 µm). These sensors are very sensitive, cheap and robust. They are composed of two infrared sensors, so they detect humans only if the human or the sensor is moving. Pyroelectric sensors have been used by Pissokas and Malcolm in [8] and Nourbakhsh et al. in [7]

In the context of the View-Finder project, the use of an infrared camera was initially foreseen, but this has been sadly cancelled during the course of the project, such that also the temperature-based human victim identification approach could not be followed.

1.2.3 Scent

CO_2 sensors allow to detect the carbon dioxide emission, and even the breathing cycle of a victim. It is thus possible to determine if he is still alive. These sensors have been used by a number of participants of the RoboCupRescue [13], but the disadvantage is that the response time of a CO_2 sensor is very slow and that the sensor has to be very close to the victim to have useful data because it is very directional

and depends much on the air conditions like humidity, temperature, wind, and dust. This makes it difficult to use it in a disaster area, reason why the scent-based approach was not pursued in this project.

1.2.4 Motion

Motion can be detected by a variety of sensors (sonar, laser, visual & IR camera, …) and can serve as an indication that somebody alive is present. However, motion analysis alone can never determine whether the cause of this motion field is a human being. Therefore, it is only used in combination with other characteristics[5][6] [7].

1.2.5 Skin Color

Skin color is a popular parameter in the computer vision community to detect humans. Visser et al. use skin color in [12] for human victim detection. They construct a 3D color histogram in which discrete probability distributions are learned. Given skin and non-skin histograms based on training sets, the probability that a given color value belongs to the skin and non-skin classes can then be learned. The problem with these approaches is twofold: 1) in unstructured outdoor environments, there is no a priori data on the colors present in the environment (which could lead to a large number of false positives), and 2) the field of view of typical outdoor robot cameras is quite large, which means that a person's face only consists of a limited number of pixels (which would reduce the detection rate).

1.2.6 Face & Body Shape Detection

Another popular approach in computer vision to detect persons is to perform face detection. Other detectors are specifically trained at detecting the upper body. Together, these detectors provide powerful cues for reasoning about a person's presence. The problem with these methods is that detecting victims lying on the ground using standard camera images is very different from standard person detection. These standard person detection algorithms, relying on face or upper body detection, assume that the person's face is clearly visible in the camera image and that the person is standing straight up, such that the upper body can be easily detected. Victims, however, do not tend to stand up and do not tend to look straight into the camera. Therefore, special techniques have to be applied, as proposed for example in [1]. The approach presented in this article aims to classify the body shape of lying human victims and thus falls into this category of victim detectors.

1.2.7 Combined approaches

With such a multitude of detection approaches, each having their advantages and disadvantages, it is evident that the integration of multiple cues can provide better results. Therefore, several teams have investigated hybrid approaches, mixing for example motion, sound and heat in [7] or motion sound and faces in [6]. Others [5] have focused specifically on determining the best way to integrate the information from all cues, leading to MRF-based approaches.

1.3 Proposed Approach

In this paper, we present an approach to achieve robust victim detection in these difficult outdoor conditions, by going out from the Viola-Jones algorithm for Haar-features based template recognition and adapting it to recognize victims. Victims are assumed to be human body shapes lying on the ground. The algorithm tries to classify visual camera input images into human body shapes and background items.

2. The Victim Detector

2.1 The Viola-Jones Detector

The basis for this work is a learning-based object detection method, proposed by Viola and Jones [11]. Viola and Jones originally applied this technique in the domain of face detection [10]. Their system yields face detection performance comparable to the best previous systems and is considered the fastest and most accurate pattern recognition method for faces in monocular grey-level images.

The method operates on so-called *integral images*: each image element contains the sum of all pixels values to its upper left allowing for constant-time summation of arbitrary rectangular areas.

During training, weak classifiers are selected with AdaBoost, each of them a pixel sum comparison between rectangular areas.

The object detection system classifies the images using simple features. The features are able to encode ad-hoc domain knowledge and the features-based system operates much faster than a pixel-based one. The Haar-wavelets are single-wavelength square waves, which are composed by adjacent rectangles. The algorithm does not use true Haar-wavelets, but better suited rectangle combinations. This is why the called Haar-features instead of Haar-wavelets. The Haar-features detection procedure works by subtracting the average dark-region pixel value from the average light-region pixel value. If the difference is above a threshold, which is set during the training, the feature is present. Figure 1 shows some example for these rectangle combinations.

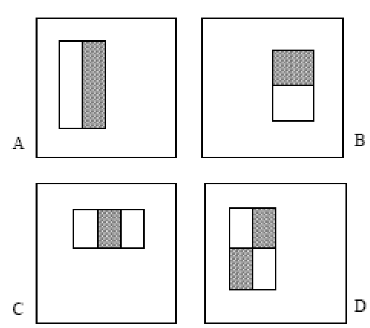

Figure 1 – Example Haar-features in the detection window

Hundreds of these classifiers are then arranged in a multi-stage cascade. Lazy successive cascade evaluation and the constant-time property allow the detector to run fast enough to achieve an overall low latency.

2.2 Face & Upper Body Detection

In a first attempt at victim detection, we used the standard Viola-Jones detector for face and upper body detection. The first tests were executed on indoor and good quality images. These tests were very successful, 90% of the faces and 80% of the upper bodies were detected. All together the hit rate reached the 95% while the false alarm rate stayed under 25%. However, the target hardware, the RobuDem, is going to operate in outdoor environment where the background is various and the illumination is unpredictable. So, outdoor experiments were strongly suggested. Although, the results were better than expected, the false alarm was increased dramatically while the hit rate was decreased to 70% for the upper body and to 30% for the face detection.

The conclusion from these tests is that in outdoor environment the face detection based person detection is not viable. Usually it only consumes the computation time without giving any results or any correct results. If the detection was more detailed, the system became too slow with minor success. If the detection was tuned to be faster, the hit rate decreased under 10%.

The upper body detection is more robust, it adopts itself to different illuminations much better. However, it gives much more false alarms.

Figure 2 – Upper body and face detection in an outdoor image

Our first idea was to fuse the face and upper body detector for having a more robust system. Unfortunately, the face detector does not really improve the performance of the upper body detector. Also, these detectors work only for standing, sitting or crouching person detection. In the case of the person lying on the ground, all of the detectors based on the existing Haar-cascade classifiers fail. At this point, we decided to fuse the upper body detector with a new detector which has to be trained for victims.

2.3 Extension to Victim Detection

For the victim-detection application, we adapted the Viola-Jones technique, by training the algorithm with bodies, lying on the ground. To deal with the huge number of degrees of freedom of the human body and the camera viewpoint, the configuration space for human victims was reduced to victims lying face down and more or less horizontally in front of the camera. This case has been chosen because in real disasters this pose has the highest probability. The people try to protect their head and their ventral body parts which are the most vulnerable. Another reason is that in this position, the possible positions of the limbs form a relatively small pool comparing to the other cases. Also the orientation of the body must be considered because the legs have a different shape than the upper body and the head. To handle this, the sample images were taken with the both body orientations (left-to-right and right-toleft). To enlarge the data-set, the images were then later flipped horizontally and re-used during the Haar-training.

2.4 Database training

Figure 3 shows some example of the sample images. They were taken in an outdoor environment, featuring human victims in several orientations and under varying illumination. These images were taken with an on-board stereo camera system. In total, 800 positive scenes were recorded, in each case the color rectified image of the left and right camera were recorded.

Figure 3 - Example training images with lying bodies (positive samples)

Furthermore 500 pairs of negative images were recorded outside and 100 pairs inside. These images contain no humans but the variety of the background is high in order to make the AdaBoost learning method set up good thresholds. In Figure 4, there are three examples of the outdoor negative images.

Figure 4 - Examples for negative training images

3. Results

Theoretically, the cascaded Haar-classifier for victim detection has a 100% detection rate and less than 10^{-6} % false alarm rate. Of course, this is only true in the case of the positive and negative sample images. With new test images – which were taken in similar illumination conditions as the sample images but mainly in different positions –, the correct detection rate was approximately 65%.

Figure 5 shows the result of the victim detection algorithm. The red rectangles are the hits of the detector for the victims whose head is at the left, the yellow ones for those whose head is at the right. In the first image of Figure 5, the victim was correctly found besides of a lot of false positives. These false alarms are eliminated by merging the adjacent rectangles of correct posture.

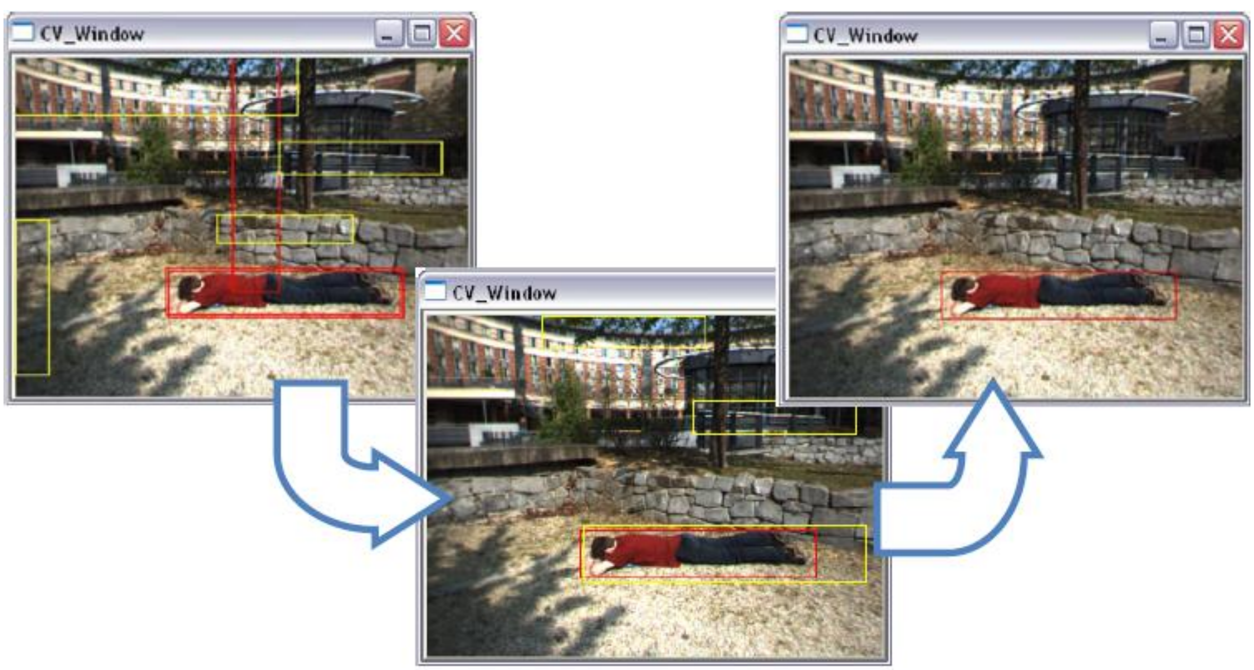

Figure 5 - An example test image for Victim Detection

In the case of Figure 6, it is more difficult to decide whether a correct classification is performed. In the first picture a smaller and bigger rectangle cover the victim. The smaller rectangle is a true positive, butthe bigger rectangle is a false alarm which may have some typical features of a victim. As it is showed in the second picture, these rectangles are considered neighbors and they were merged together. The merging is done by computing an average rectangle; this is why the marked area is bigger than the actual victim.

Figure 6 - Before and after merging the neighbor detection areas

The processing time for running the victim detector is between 60 and 80 milliseconds, which means 13-16 frames per second. This is a very good result, as it allows near real-time reactions in the robot control scheme and it also allows integrating the results of multiple detection runs over time by means of a tracking scheme, to enhance the detection rate and reduce the false positive rate.

4. Conclusions

In this paper, we have first overviewed the different existing approaches towards human victim detection. In the scope of this research project, the detection of human body shapes from visual input data was chosen as a recognition method, a decision mainly due to the available sensor equipment on board of the victim detection robot. We have presented an approach, based upon the Viola-Jones (face) detector, which was adapted, such that human victims lying on the ground can be detected. The first results of this approach are encouraging, but future research is required to increase the detection rate and reduce the number of false positives. This will be done by integrating the human victim detector in a tracking scheme. A mobile robot equipped with such a human victim sensing system can be a valuable aid for human crisis managers, as it could scan a designated area for human survivors semi-automatically.

References

[1] S. Bahadori, L. Iocchi. *Human Body Detection in the RoboCup Rescue Scenario*. First International Workshop on Synthetic Simulation and Robotics to Mitigate Earthquake Disaster, July 5th, 2003

[2] S. Burion, *Human Detection for Robotic Urban Search and Rescue*, Diploma Work, EPFL, February 2004

[3] A. Birk, S. Markov, I. Delchev and K. Pathak. *Autonomous Rescue Operations on the IUB Rugbot*, International Workshop on Safety, Security, and Rescue Robotics (SSRR), IEEE Press, 2006

[4] D. Doroftei, G. De Cubber, E. Colon and Y. Baudoin. *Behavior Based Control For An Outdoor Crisis Management Robot.* Third International Workshop on Robotics for risky interventions and Environmental Surveillance-Maintenance, RISE'2009- Brussels, Belgium. January 2009

[5] A. Kleiner and R. Kuemmerle. *Genetic MRF model optimization for real-time victim detection in Search and Rescue*. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2007). San Diego, California 2007.

[6] A. Kleiner, B. Steder, C. Dornhege, D. Höfer, D. Meyer-Delius, J. Prediger, J. Stückler, K. Glogowski, M. Thurner, M. Luber, M. Schnell, R. Kümmerle, T. Burk, T. Bräuer, and B. Nebel. *RoboCupRescue - Robot League Team RescueRobots Freiburg (Germany)*. In *RoboCup 2005 Team Description Paper, Rescue Robot League*, (Osaka, Japan), 2005.

[7] I.R. Nourbakhsh, K. Sycara, M. Koes, M. Yong, M. Lewis, and S. Burion. *Human-Robot Teaming for Search and Rescue*. IEEE Pervasive Computing 4, 1 (Jan. 2005), 72-78.

[8] J. Pissokas and C. Malcolm, *Experiments with Sensors for Urban Search and Rescue Robots*, International Symposium on Robotics and Automation, 2002

[9] M. Trierscheid, J. Pellenz, D. Paulus, D. Balthasar: *Hyperspectral Imaging for Victim Detection with Rescue Robots*. 2008. Proceedings of the 2008 IEEE Int. Workshop on Safety, Security and Rescue Robotics.

[10] P. Viola and M. Jones. *Robust Real-Time Face Detection*. International Journal of Computer Vision, Vol. 57, No. 2. (May 2004), pp. 137-154.

[11] P. Viola and M. Jones. Robust *Real-time Object Detection*. In Intl. Workshop on Statistical and Computational Theories of Vision, July 2001.

[12] A. Visser, B. Slamet, T. Schmits, L.A. González Jaime and A. Ethembabaoglu. *Design decisions of the UvA Rescue 2007 Team on the Challenges of the Virtual Robot competition*, Proc. of the Fourth International Workshop on Synthetic Simulation and Robotics to Mitigate Earthquake Disaster, July 2007, Atlanta, GA (pp. 20-26).

[13] Intelligent Systems Division National Institute of Standards and Technology, *Urban Search and Rescue Robot Competitions*, 2005, http://www.isd.mel.nist.gov/projects/USAR/competitions.htm