A Scrubbing Technique for the Automatic Detection of Victims in Urban Search and Rescue Video

Martin Gerdzhev
Ryerson University
350 Victoria St
Toronto, Ontario, Canada
+1-416-979-5000 ext. 2758
mgerdzhe@ryerson.ca

Jimmy Tran
Ryerson University
350 Victoria St
Toronto, Ontario, Canada
+1-416-979-5000 ext. 2758
q2tran@ryerson.ca

Alexander Ferworn
Ryerson University
350 Victoria St
Toronto, Ontario, Canada
+1-416-979-5000 ext. 6968
aferworn@scs.ryerson.ca

Kevin Barnum
Provincial Emergency Response Team
P.O Box 728,
Bolton, Ontario, Canada,
+1-905-575-582
kevin.barnum@jus.gov.on.ca

Mike Dolderman
Provincial Emergency Response Team
P.O Box 728,
Bolton, Ontario, Canada,
+1-905-575-582
Mike.dolderman@jus.gov.on.ca

ABSTRACT

In the discipline of Urban Search and Rescue (US&R), the faster a live human can found the more likely their rescue will be successful with success being measured in lives saved. We have been working to augment trained US&R dogs with technology to help first responders in the US&R effort and give them a better understanding of the condition of the disaster area being searched and the trapped people who are found. As one can imagine, the video feed from a dog can be quite jittery. We have been exploring ways to speed the process of video "scrubbing" by automatically discarding segments of video which show nothing interesting and concentrating on segments that are critical. This paper discusses one of these techniques.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition**]: Applications; I.4.9 [**Image Processing and Computer Vision**]: Applications; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding -- Intensity, color, photometry, and thresholding, Video analysis

General Terms

Algorithms, Experimentation, Human Factors, Theory

Keywords

US&R; neural network; victim detection; CAT; CRDS; image recognition; scrubbing

1. INTRODUCTION

When a building collapses because of any unforeseen incident, people may become trapped in voids formed in the resulting rubble [6, 7]. The first major objective for Urban Search and Rescue (US&R efforts is to find the trapped people who are alive and the second objective is to rescue them. In urban disasters, the rubble field is often too dangerous for human access after the collapse and may remain so until stabilized. Unfortunately, inspection and stabilization takes time and reduces the probability that people will be found alive [1]. This problem can be addressed by using special dogs or robots [8, 9, 10, 11, 12, 13, 14] for the search process. However we should consider that utilizing a robot in this situation is highly problematic, since the environment is often well beyond the mobility capabilities of a ground robot to traverse. As a result, the robot most likely will have mobility problems which lead to slowing the search process. Alternatively, US&R dogs are capable of finding trapped people in rubble environments because of their strong sense of smell, tolerance to darkness, accuracy, determination, and agility. US&R dogs can often traverse rubble very quickly and indicate where they have detected live humans through sustained barking. Until recently, the barking indicated location but little else.

Our research involves using US&R dogs augmented with additional sensing equipment on specially constructed harnesses. The Canine Augmentation Technology (CAT) [3, 4, 5] system includes video, audio, and a variety of other sensors that communicate wirelessly with "off-dog" receivers. We have developed a protocol involving three steps [2] when a handler cannot accompany the dog in its search:

- A dog is sent without any equipment to search and find a human in the regular way.
- If the first dog gets a "hit", a second dog is sent with a Canine Remote Deployment System (CRDS) [2] that is capable of automatically dropping a marker when the dog barks. The bag is orange (containing supplies,

radio, etc.) and is usually 2 or 3 feet from the scent plume the dog has detected.

- Finally, a third dog is with CAT to collect video of the area and possibly of the trapped person and their surroundings and condition.

The dropped bag is the key to video processing as the bag is a colour which is different from the surrounding debris. Thus a technique used by a human scrubbing the video is to look for the orange bag to appear - the video depicting the victim is likely to be near this location in the video stream.

The scrubbed images can then be analyzed by structural engineers and rescuers to plan how to best save the person. We should take into account that going through the video and analyzing the video frames is a very time-consuming process which adds to the delay before a rescue can be attempted.

2. Problem Statement and Proposed Solution

Currently, the fastest and most accurate technique for finding live people trapped in rubble is through the use of US&R dogs. When the rubble makes access impossible for a human handler to follow it is often possible for the dog to search. The US&R dog finds the path to the person and barks. However, the handler does not know where the people are trapped and cannot determine the location of the victims from barking. Also the dog cannot verify the condition of the person or the environment after it locates the trapped victim.

The goal of this research is to provide video information to find casualties and to gather information about the surroundings which will help with the rescue process. As was discussed in the previous section, analyzing this video manually is time-consuming and requires a high degree of accuracy—two commodities in short supply in a disaster search. One must consider that spotting the orange bag by a person in the video might be very hard when the bag appears in one or two video frames for a fraction of a second. We therefore propose a software solution that scans and analyzes the received video frames automatically to spot the orange bag and thus allowing key frames to be analyzed faster.

The proposed solution is based on a pattern recognition technique in which the software uses specific recognition features to spot the orange bag in the video stream using a neural network approach.

Artificial Neural Networks (ANNs) play a major role in the development of different applications because of their ability to learn. An ANN is configured for a specific application such as data classification or pattern recognition, through a learning process. An application of an ANN in image pattern recognition is presented in [16]. In this project an ANN has been used to filter the result of the software to get more accurate end results. A different approach to a similar problem is attempted in [15].

3. Approach

We have implemented software that is able to analyze the video stream, either in real time or after the video has been extracted, and to tag the video where the orange bag appears. To accomplish this, each frame is checked for the presence of the orange bag. Thus the problem becomes finding the bag in an image, as a video is a sequence of images. Because the bag is not a rigid object (shape recognition cannot be applied) and can appear in the image in different orientations and distances from the camera, and in different light conditions, thus the problem becomes quite difficult. (See Figures 1 and 2)

Figure 1. Bag outside in good light conditions

Figure 2. Bag in an enclosed space in poor light conditions

In solving this problem several approaches were tried. One such approach was the use of the SIFT method [18] to find similar features between the database of trained images and the test images. The problem with this was that many images of rubble, especially in dark places look the same, and the key points selected normally would not contain the orange bag at all. This led to incorrect classification every time and the method was abandoned.

Another approach that proved to be effective is described below.

The problem was split into three parts: preprocessing, isolating regions of the image that contain an orange coloured blob (based on a threshold), and checking if that region contains the orange bag (using an Artificial Neural Network). The Software was implemented in the C language. The libraries used were the OpenCV (Open Computer Vision) library for image processing and FANN (Fast Artificial Neural Network) library for the ANN part.

3.1 Preprocessing

The preprocessing consisted of converting from the RGB colour space to the HSV colour space. This was done because the HSV colour space describes perceptual colour relationships more accurately than RGB, and allows us to more easily find colours that are very similar [17].

3.2 Isolating regions

For finding the regions with the orange blob, thresholds were selected for the Hue, Saturation and Value that best capture the colour that was sought in different light conditions. The values were found to be the following:

Table 1. HSV minimums and maximums

Hue		Saturation		Value	
Min	*Max*	*Min*	*Max*	*Min*	*Max*
0°	24°	0.44	0.75	0.35	1.00

A second black and white image is created so that it can be shown which parts of the image are detected in the colour that is sought. We then go through each pixel in the image and if it has values within that threshold, we change the corresponding pixel in the black/white image white, and black otherwise. We then perform an "erosion" on the black and white image to remove any stray pixels that might be detected by accident and then several dilations, to get a slightly bigger region that we can use for the ANN step.

What we are left with at the end of this step is one or more regions that were determined as probably showing the orange bag. If no regions were detected, then the orange bag was not found, we discard the image, and we check the next image. If one or more regions are found, they are each scaled to 30x30 pixels and the HSV values for the 900 pixels are given as an input to the Neural Network in the next step.

3.3 Checking isolated regions

A standard back propagation Artificial Neural Network was trained on 200 30x30 images to recognize the orange bag. It consists of 2700 inputs (900 pixels with HSV values for each) and 1 output – whether it is the bag or not. Best results were achieved with 2 hidden layers of 150 neurons each. The Sigmoid Symmetric activation function and a learning rate of 0.7 were used for all the neurons. When a 30x30 region is given to the ANN, it provides a floating point result between -1 and 1 depending on how sure the ANN is of the match. If the result is greater than 0 we assume that it is a positive match and when it is less than or equal to 0 it is a negative match. Based on these results, we can filter each image and remove some or all of the regions that were given as part of the previous step and therefore produce a more accurate result concerning the presence or absence of the orange bag in the image.

4. Results

A series of tests were conducted on the pattern recognition software for the thresholding part, the ANN part and the combination of both. The thresholding part and the combination of thresholding and ANN were tested on a series of 20 images, while the ANN part was trained on 200 images and tested on 135 images. The error for the ANN converged quickly after only 41 epochs and produced minimal errors on the testing data as well. The results are presented in Table 2 and Table 3.

Table 2. Approaches and corresponding error

	Thresholding	Thresholding + ANN
Error	15%	15%

Table 3. Neural Network's convergence and errors

Epochs	Training Error	Testing Error
41	0.001813	0.067022

While the thresholding approach yielded a 15% error, it actually recognized all the images that had the orange bag in them, but also picked up some false positives. The two approaches combined still exhibited 15% errors which means that 15% of the time it produced false indications of an orange bag that was not present at a particular location in the image. Figures 3-11 provide examples of correct classification, false negative, and false positive.

Figure 3. Correct classification - Original image

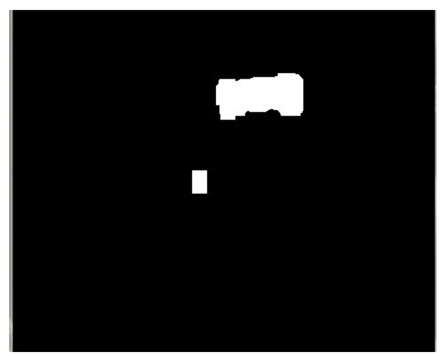

Figure 4. Image after thresholding

Figure 5. Image after the ANN filtering

In Figures 3-5 the thresholding function detected 2 regions, one of which was a false positive. After filtering the 2 regions through the ANN only the correct region remains.

Figure 6. False negative - Original Image

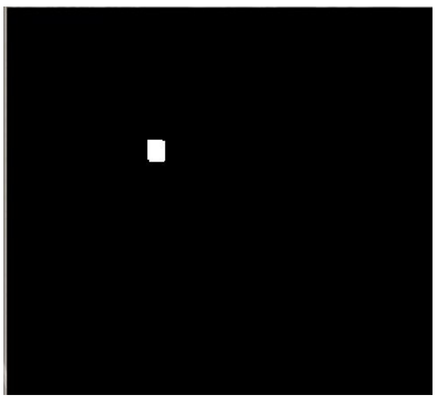

Figure 7. Image after thresholding

Figure 8. Image after the ANN filtering

In Figures 6-8 the thresholding function detects 1 region which shows the bag. After the ANN filtering that region is ignored and gives a false negative.

Figure 9. False positive - Original image

Figure 10. Image after thresholding

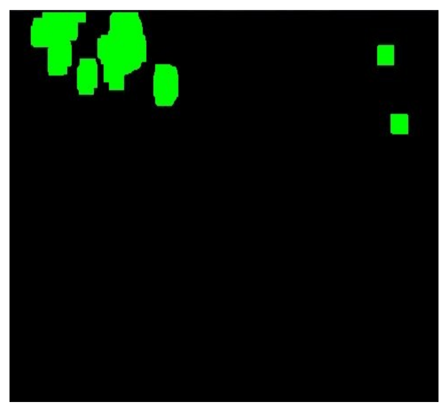

Figure 11. Image after the ANN filtering

In Figures 9-11 the thresholding function detects 8 regions, all of which are false positive. These are due to orange colour of hanging insulation in a partially collapsed parking garage. After the ANN filtering, 2 of these regions are ignored but still provide a false positive. This image is particularly difficult to process in order to find the orange bag, because most of the image has orange shapes most of which look like the bag. It is almost impossible even for a human looking at the image to figure out whether the bag is present in it.

5. Conclusions

From the results it can be concluded that with the approach that was taken, the recognition system successfully recognized the bag 85% of the time while in some difficult cases it failed to give a correct indication. The thresholding function alone gives a very good recognition rate and doesn't produce any false negatives, but generates a substantial number of false positive images. Thus combining it with an Artificial Neural Network we can successfully eliminate a large portion of the false negatives, but in doing so sometimes eliminate the correct region as well.

In most cases the orange bag would be present in several consecutive frames which would improve the chances of successfully recognizing it in at least 1 or 2 frames. To improve the success rate further, as future work, we can incorporate other sensory data such as audio for bark detection since we know that the dog would bark when it finds a person's scent and the bag from the second dog will be close by.

Taken as a whole, the software can substantially decrease the time taken to scrub the video and thus decrease the time needed for making an informed decision about how to rescue the person. Since time is critical in US&R situations, there is no doubt that using this software would increase the chances of survival of the person to be saved.

6. REFERENCES

[1] McGuigan K. and Friedman J., Canine-firefighter teams search for survivors, Fire Engineering 159 (2006), no. 5, pp. 197-198.

[2] Ferworn A., Ostrom D., Barnum K., Dallaire M., Harkness D., and Dolderman M., Canine Remote Deployment System for Urban Search and Rescue, JHSEM: Vol. 5 [2008], No. 1, Article 9

[3] Ferworn A.., Sadeghian A., Rahnama H., Pham H., Erikson C., Barnum K., Ostrom D. and Dell'Agnese L., Urban search and rescue with canine augmentation technology, 2006 IEEE International Conference of Systems of Systems (SoSE'06), 2006

[4] Ferworn A., Ostrom D., Sadeghian A., Rahnama H., Barnum K. and Woungang I., Rubble search with canine augmentation technology, System of Systems Engineering, 2007. SoSE '07. IEEE International Conference on, 2007, 1-6

[5] Tran J., Ferworn A., Ribeiro C., Denko M., Enhancing Canine Search, IEEE Systems of Systems Engineering (SoSE'08), June 2-5 2008, Monterey CA USA

[6] Hnatko B., Building breakdown: Key concepts & tactics in responding to a structural collapse, Fire Rescue Magazine 25 (2007), no. 1, pp. 66-68.

[7] Collins L., Assessing structural collapse from acts of terror, Fire Engineering 159 (2006), no. 11, pp. 55-66.

[8] Murphy R. R., "Trial by fire (rescue robots)," IEEE Robotics & Automation Magazine, vol. 11, 2004, pp. 50-61.

[9] Casper J. and Murphy R. R., Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center, IEEE Transactions on Systems, Man and Cybernetics, Part B 33 (2003), no. 3, 367-385.

[10] Murphy R. R., Marsupial and shape-shifting robots for urban search and rescue, IEEE Intelligent Systems and Their Applications 15 (2000), no. 2, 14-19.

[11] Burke J. L. and Murphy R. R., Human-robot interaction in usar technical search: Two heads are better than one, IEEE International Workshop on Robot and Human Interactive Communication, 2004, p.^pp. 307-312.

[12] Davids A., Urban search and rescue robots: From tragedy to technology, IEEE Intelligent Systems and Their Applications 17 (2002), no. 2, 81-83.

[13] Ferworn A., Hough G., Manca R., Expedients for marsupial operations of usar robots, IEEE International Workshop on Safety, Security and Rescue Robotics (SSRR06), 2006, p.^pp.

[14] Matsuno F., Tadokoro S., Rescue Robots and Systems in Japan, Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics August 22 - 26, 2004, Shenyang, China

[15] Sumimoto T., Kuramoto K., Okada S., Miyauchi H., Imade M., Yamamoto H. and Arvelyna Y., Detection of a Particular Object from Motion Images under Bad Condition, IEEE Instrumentation and Measurement Technology Conference Budapest, Hungary, May 21-23,2001.

[16] Zhang H., Guan J., and Sun G. C., 1992. Artificial neural network-based image pattern recognition. In Proceedings of the 30th Annual Southeast Regional Conference (Raleigh, North Carolina, April 08 - 10, 1992). ACM-SE 30. ACM, New York, NY, 437-441. DOI= http://doi.acm.org/10.1145/503720.503776

[17] Schwarz M. W., Cowan W. B., and Beatty J. C.,1987. An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models. ACM Trans. Graph. 6, 2 (Apr. 1987), 123-158. DOI= http://doi.acm.org/10.1145/31336.31338

[18] Mokhtarian F., Suomela R. and Chan K.C., "Image Point Feature Detection through Curvature Scale Space", IEEE Press, Dec. 1998 p22-23