Segmentation techniques for extracting humans from thermal images

J. S. Dickens
CSIR Centre for Mining Innovation
PO Box 91230
Auckland Park 2006
Johannesburg, South Africa
Email: jdickens@csir.co.za

J. J. Green
CSIR Centre for Mining Innovation
PO Box 91230
Auckland Park 2006
Johannesburg, South Africa
Email: jgreen@csir.co.za

***Abstract*—A pedestrian detection system for underground mine vehicles is being developed that requires the segmentation of people from thermal images in underground mine tunnels. A number of thresholding techniques are outlined and their performance on a number of thermal images is investigated. The thresholding techniques are evaluated on images in various ambient conditions and it is shown that a minimum error thresholding technique is the most effective.**

I. Introduction

A pedestrian detection system for underground mine vehicles is being developed to address the high number of fatalities caused by mine vehicles [1]. The system makes use of a combination of thermal and 3D imaging to identify and track people near the mine vehicle. The system will help improve drivers' awareness of people near their vehicles and also allow for the safe operation of autonomous mine vehicles in the presence of humans.

The system detects, classifies and tracks humans in the thermal images and then combines the thermal images with 3D images to provide actual position information. It will need to determine how far away from the vehicle the people are and track them to determine whether they are on a collision course with the vehicle. A commonly used paradigm for object detection and tracking in video is to first extract regions of interest and then classify or validate them [2–7], which is the methods that is being used for the pedestrian detection system, as shown in the system diagram, Fig. 1. This paper deals with the segmentation subsystem, the regions that have been segmented by this system will be further processed to remove small noise regions and then the remaining regions will be classified. Various methods for segmenting people from thermal images will be reviewed and compared.

Image thresholding takes in a multi-valued input image and outputs a binary image where one of the states represents foreground objects and the other represents the background. Image thresholding is used for a wide variety of applications from extracting printed characters for optical character recognition through identification of defects in automated inspection tasks, to segmenting computed tomography x-ray images.

At first glance segmenting humans from thermal images may seem trivial because we know that human core body temperature remains in a very narrow range. However human surface temperatures vary significantly depending on a number of factors such as the clothes worn and the naturally lower temperature of the extremities (arms and legs).

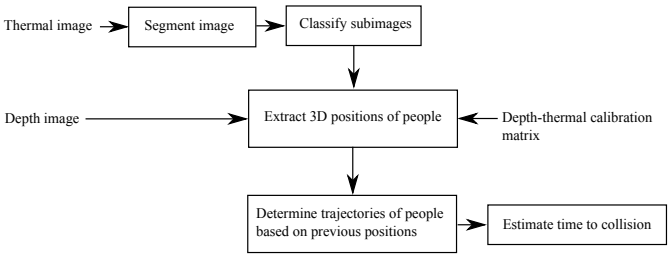


Fig. 1. A block diagram showing the subsystems making up the pedestrian detection system.

It is assumed that the temperature of people within a mine tunnel will always exceed the temperature of the tunnel itself. In deep South African gold mines the virgin rock temperatures can be as high as 60 $^{\circ}C$ however ventilation and other cooling brings the temperature within working areas (stopes) down to below 30 $^{\circ}C$ to allow work to be done [8]. Work conducted to model the heat flow from advancing stopes shows that the rock surface temperature can be assumed to be equal to the ventilation air wet-bulb temperature (T_{wb}) [9]. The wet-bulb temperature takes into account the relative humidity of the air and therefore the effects of evaporative cooling. Since the wet-bulb temperature takes into account the effect of evaporative cooling there will always be a positive temperature gradient between people and the environment to allow the dissipation of metabolic heat.

Since the people in the thermal images will always be warmer than the background the segmentation of the thermal images involves determining an optimal threshold to extract only the people as foreground objects. The camera used to capture the images used for evaluating the thresholding methods is a FLIR A300 providing a thermometric image.

II. Thresholding methods

There are a very large number of thresholding algorithms belonging to a number of categories, a good survey of a

Presented at the 4th Robotics and Mechatronics Conference of South Africa (ROBMECH 2011)
23-25 November 2011, CSIR Pretoria South Africa.

large number of them is provided by Sezgin and Sankur [10]. The methods evaluated here will be those identified as the best performing by Sezgin and Sankur as well as a number of other techniques chosen for certain characteristics. Thresholding methods falling into the following categories; clustering-based thresholding, entropy-based thresholding, locally adaptive thresholding and model-based thresholding will now be discussed.

For all of the following discussions the following notation will be used. Each picture has a total of N pixels that fall into L grey-levels. The number of pixels that fall into each grey-level (i) of the image histogram is denoted by n_i. The normalised grey-scale histogram can be considered an estimate of the probability distribution of pixel intensities i.e.

$$p_i = n_i/N \tag{1}$$

Where:
p_i is the probability that a pixel belongs to the i^{th} grey level
N is the total number of pixels

The cumulative probability function for the k^{th} grey-level is defined as

$$P(k) = \sum_{i=1}^{k} p_i \tag{2}$$

A. Clustering-Based Thresholding

1) Otsu's Method: The first thresholding method that is evaluated is Otsu's threshold selection method [11]. Otsu's method is evaluated due to its popularity as a thresholding method, being one of the most cited thresholding methods [10]. Otsu's method finds a threshold that minimises the within-class variances of the foreground and background classes. Minimising the within class variance is equivalent to maximising the between class variance.

The zeroth- and first-order cumulative moments of the image histogram up to the k^{th} grey-level are:

$$\omega(k) = \sum_{i=1}^{k} p_i \tag{3}$$

and

$$\mu(k) = \sum_{i=1}^{k} ip_i \tag{4}$$

The total mean level of the original picture is:

$$\mu_T = \mu(L) = \sum_{i=1}^{L} ip_i \tag{5}$$

It can be shown that the between class variance, σ_b^2, is:

$$\sigma_b^2 = \frac{\left(\mu_T\omega(k) - \mu(k)\right)^2}{\omega(k)\left(1-\omega(k)\right)} \tag{6}$$

Otsu's method selects the optimal threshold T_{opt} in order to maximise the between class variance. The optimal threshold is the value of k that maximises Equation 6, ie.

$$T_{opt} = \underset{k}{argmax}\ \sigma_b^2(k) \tag{7}$$

2) Iterative Clustering: Iterative clustering assumes that the intensity histogram has two peaks, one for the foreground objects and another for the background objects. The algorithm starts with the threshold set to the centre intensity level, the peak of the histogram on either side of the threshold is then determined. The threshold value is moved to the midpoint of the two peaks and the peaks are found again. The process is repeated until the change in the threshold is sufficiently small.

3) Minimum Error Thresholding: Minimum error thresholding assumes that the image is made up of foreground and background objects with normally distributed intensities. The method of minimum error thresholding is that of Kittler and Illingworth [12], their method minimises a criterion function which gives the approximate minimum error threshold. The criterion function derived by Kittler is

$$\begin{aligned} J(k) \quad = \quad & 1 + 2\left[P(k)ln\left(\sigma_1(k)\right) + \left(1-P(k)\right)ln\left(\sigma_2(k)\right)\right] \\ & -2\left[P(k)ln\left(P(k)\right) + \left(1-P(k)\right)ln\left(1-P(k)\right)\right] \end{aligned} \tag{8}$$

Where:
$\sigma_1(k)$ is the standard deviation of the background up to grey level k
$\sigma_2(k)$ is the standard deviation of the foreground, from k to L

The criterion function shown in Equation 8 gives a measure of the overlap of the two distributions, so the method estimates the parameters of the two normal distributions on either side of the threshold and then calculates the overlap of the two estimated distributions. Using Equation 8 the optimal threshold is easily determined.

$$T_{opt} = \underset{k}{argmin}\ J(k) \tag{9}$$

Since the distributions overlap, the estimation of the parameters will contain a bias, however this is assumed to be small. The bias does indeed appear to have little effect of the result. Another advantage of the minimum error thresholding technique is that the criterion function will not have a minimum for a unimodal distribution, so an image that does not contain any people can be detected and not segmented.

B. Entropy-Based Thresholding

Entropic thresholding methods exploit the entropy distribution of the grey-levels in the scene. Maximising the entropy of the thresholded image maximises the information between the foreground and background distributions in the image [10, 13]. For a threshold at grey-level k the entropy of the background up to grey-level k is

$$H_b = -\sum_{i=1}^{k} \frac{p_i}{P(k)} ln \frac{p_i}{P(k)} \tag{10}$$

and the entropy of the foreground is

$$H_f = -\sum_{i=k+1}^{L} \frac{p_i}{(1-P(k))} ln \frac{p_i}{(1-P(k))} \tag{11}$$

Defining the sum of the two entropies as $\phi(k)$ we get

$$\phi(k) = -\sum_{i=1}^{k} \frac{p_i}{P(k)} ln \frac{p_i}{P(k)} - \sum_{i=k+1}^{L} \frac{p_i}{(1-P(k))} ln \frac{p_i}{(1-P(k))} \tag{12}$$

Maximising $\phi(k)$ gives the maximum information between the two distributions. So the optimal threshold is

$$T_{opt} = \underset{k}{argmax}\ \phi(k) \tag{13}$$

C. Locally Adaptive Thresholding

Locally adaptive thresholding adapts the threshold for each pixel in the image, instead of having one threshold (T) the threshold is an matrix the same size as the image ($T(x, y)$). The adaptive thresholding method is that of Sauvola and Pietikäinen [14] which is adapted based on the mean and standard deviation of the pixels in a window around each pixel. The threshold is calculated according to the formula

$$T(x, y) = m(x, y) \cdot \left[1 + k\left(\frac{s(x, y)}{R} - 1\right)\right] \tag{14}$$

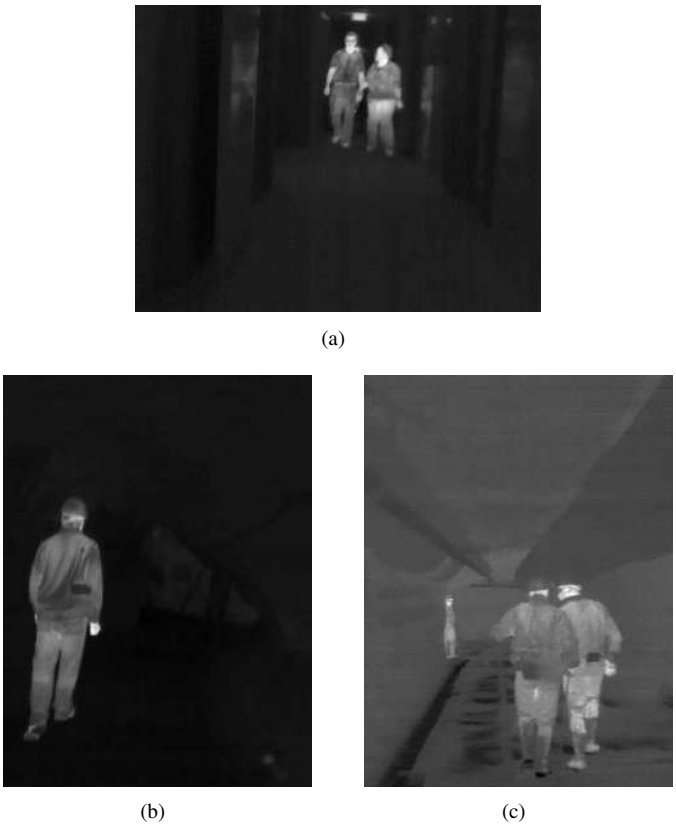
Where:
$m(x, y)$ is the mean of the window centred on pixel xy
$s(x, y)$ is the standard deviation of the window centred on pixel xy
R is the range of the standard deviation
k is a user defined constant

In our experiments the value of k was chosen to be $k = -0.02$ and the window for calculating the mean and standard deviation is 15×15 pixels. The value of k is negative because we are attempting to extract higher intensity (warmer) objects from a darker background while Sauvola was attempting to extract dark text from a light background.

III. Thresholding results

The methods were evaluated on thermal images containing people in a variety of conditions. The background temperature of the images varies from about 11 $^{\circ}C$ to 25 $^{\circ}C$. Due to the difficulty in establishing ground truth for testing the thresholding, the methods are tested qualitatively. Qualitative testing is sufficient due to the fact that the results are very sensitive to the threshold chosen so mostly the results are binary, the method provides an acceptable threshold or not. The test images used for the testing of the thresholding methods are shown in Fig. 2 below.

The images in Fig. 2 represent typical images from three datasets. The corridor provided a good dataset to test the classification algorithm because of the presence of warm objects that were not people (the lights and reflections off doors). The mine in b provides one end of the spectrum, it is a shallow mine with a cold air temperature. The tunnel in c shows an example of a problem case, the air temperature was fairly high but there was a very high ventilation air velocity, this high velocity air reduces the temperature difference between the people and surroundings. The area in image c was part of the training area of the mine and does not represent the typical conditions that would be present in the mine.



(a)

(b) (c)

Fig. 2. Test images for the thresholding algorithms: (a) a corridor at 25 $^{\circ}C$; (b) a mine tunnel at 11 $^{\circ}C$ and (c) a tunnel at 21 $^{\circ}C$.

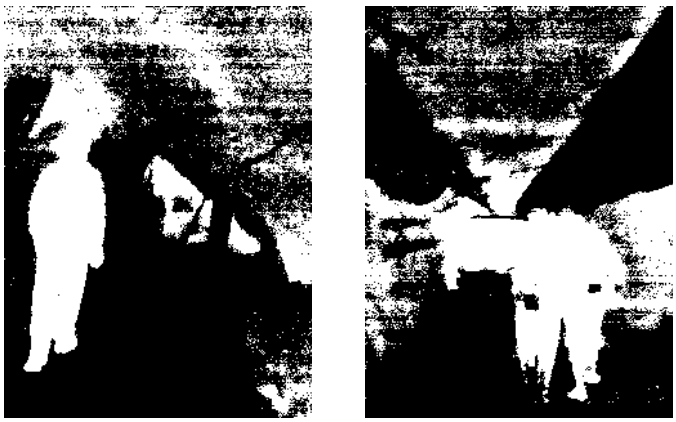
A. Clustering-Based Thresholding

All the clustering based thresholding methods suffer from a similar problem, they assume that the foreground and background objects have intensity distributions that are well separated which is not the case in the thermal images in this work.

1) Otsu's Method: Otsu's method produces acceptable results for images where the number of foreground and background pixels are approximately equal [10]. This is not the case in the thermal images investigated where the number of background pixels is significantly larger than the number of foreground pixels. When there are a significantly larger number of pixels in one class than the other, then Otsu's method tends to split the larger mode in half [12], which is exactly what is seen in Fig. 3, the background has been split by a threshold dividing the background mode of the histogram.

2) Iterative Clustering: The results of the iterative clustering method test are shown in Fig. 4, the results for image a are acceptable and the results on image b are good but the result on image c is unacceptable. The reasons for the difference in the performance between the different images can be seen by looking at the image histograms shown in Fig. 5 and Fig. 6.

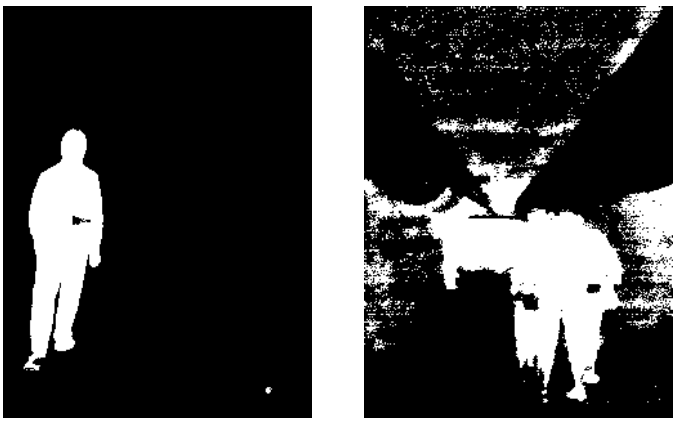
(a)



(b) (c)

Fig. 3. Thresholding results using Otsu's method

(a)



(b) (c)

Fig. 4. Thresholding results using iterative clustering method

It is evident in Fig. 5 that the histogram consists of two distributions that are fairly well separated, while in Fig. 6 the distribution appears to simply taper off to the right of the main peak. Without a well separated second peak, this method will obviously not work.

3) Minimum Error Thresholding: The results of the minimum error thresholding algorithm, shown in Fig. 7, indicate that the minimum error thresholding technique performs well on all of the input images. The result on image *c* shows incomplete segmentation of the two people close to the camera. While unfortunate it is not possible for a single threshold method to perform better since parts of the people (their hard-hats, gum-boots and cap-lamp batteries) are at the same temperature as the background.

B. Entropy-Based Thresholding

The entropy based threshold performs well on all of the images with only a small amount of noise, see Fig. 8. This makes sense since the entropy-based method is segmenting the images without making any assumptions about the underlying distributions of the foreground and background objects.

C. Locally Adaptive Thresholding

The locally adaptive thresholding method produces some interesting results. The method extracts part of the people and a fair amount of noise from the background. The reason for the poor performance of the adaptive thresholding method is that

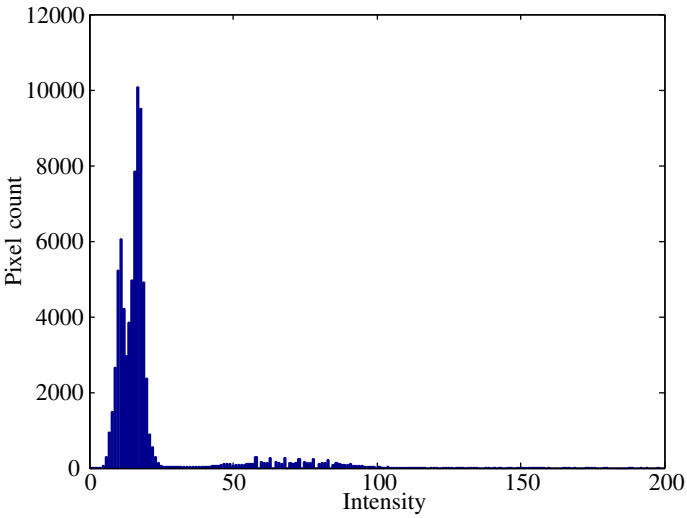


Fig. 5. Histogram for image in the cold mine tunnel (image *b*)

unlike text which the algorithm was originally intended for, the foreground objects in the thermal images are large in extent. In Sauvola's work each character being thresholded is smaller than the window used to calculate the mean. In the images used for these experiments the people (foreground objects) are larger than the window so the mean value is increased near the center of the object where the window encloses the whole object. The increasing mean towards the center