

The detection and quantification of persons in cluttered beach scenes using neural network-based classification

Steve Green^{*}, Michael Blumenstein^{*}, Matthew Browne[‡] and Rodger Tomlinson[‡]

^{*}*School of Information and Communication Technology,*

[‡]*Griffith Centre for Coastal Management*

Griffith University, Gold Coast Campus, QLD 9726, Australia.

Email: { s.green, m.blumenstein, m.browne, r.tomlinson }@griffith.edu.au

Abstract

This paper presents an initial investigation into the detection and quantification of persons in real-world beach scenes for the automated monitoring of tourist sites. Aside from the obvious use of video and digital imagery for surveillance applications, this research focuses on the analysis of images for the purpose of predicting trends of tourist activities at beach sites in Australia. The proposed system uses image enhancement and segmentation techniques to detect objects in cluttered scenes. Following these steps, a newly proposed feature extraction technique is used to represent important information in the extracted objects for training of a neural network. The neural classifier is used to distinguish the extracted objects between "person" and "non-person" categories to assist in quantification. Encouraging results are presented for person classification on a database of real-world beach scenes.

1. Introduction

Quantifying the number of people on beaches can provide valuable information for local authorities to estimate the number of persons using a beach on a particular day. People use beaches for exercise, relaxation, and social activities. The study of local beach behavior can provide valuable information for calculating advertising requirements, and provide information on whether a beach environment is sufficient to meet the current human demand. At most major beaches around Australia, and the world, cameras have been installed to capture and provide local beach information through the Internet (these cameras are sometimes referred to as 'web cams'). Coastalwatch¹ Pty. Ltd provides one such service.

The web cams at Coastalwatch provide up to date information on surf and current weather conditions at many beach locations around Australia. The streamed images from Coastalwatch web cams are of a low resolution due to the amount of bandwidth required to stream the video. Therefore processing low quality images provides quite a challenge for any automated human quantification system. An important aspect of detecting people on beaches is that of safety. Beaches can provide a dangerous environment for people of all ages [1]. Currently lifeguards provide assistance for swimmers who encounter trouble between the flags. This requires human surveillance of the beach to notify a lifeguard when a swimmer is distressed or in trouble. If a system could be developed to monitor and detect erratic or uncharacteristic swimmer behaviour and notify the lifeguard, this would provide extra safety measures for swimmers.

¹ <http://www.coastalwatch.com>

2. Overview of existing techniques for person and object detection

Schofield *et al.* [2] proposed a system for analysing video imagery to count persons on the floors of buildings for improving the efficiency of elevator systems. They have used intelligent techniques for identifying the background of a scene [3]. The limitation of this system was that it was based indoors. However, a number of systems have been proposed for dealing with outdoor conditions, which have proved to be far more variable. Iketani *et al.* [4] propose a system for real-time detection of intruders (persons) in difficult, outdoor and cluttered scenes using information from video imagery over space and time. Sacchi *et al.* [5] present advanced image processing tools for remote monitoring of a tourist site involving the counting of persons in outdoor scenes. Bartolini *et al.* [6] propose a system for automatically counting persons getting in and off a bus using image sequence analysis for allocating appropriate resources on bus lines. Finally, Pai *et al.* [7] present a system for pedestrian tracking using vision-based techniques to prevent traffic accidents.

Aside from the quantification of persons, some studies have also been performed for the purpose of quantifying and tracking the behaviour of motor vehicles. Tai *et al.* [8] present an image tracking system, which locates motorcycles and other vehicles for traffic monitoring and accident detection at road intersections. Another system recently proposed by Ha *et al.* [9] has employed a neural network-based edge detector for vehicle detection and traffic parameter estimation (vehicle count, class and speed) in an image-based traffic monitoring system. Other systems such as that of Wohler and Anlauf [10], have been used to assist drivers in automobiles through the detection of overtaking vehicles. They employed an adaptable time-delay Artificial Neural Network (ANN) to analyse complete image sequences. Finally, some researchers have utilised additional indicatory information for the purpose of vehicle detection. For example, Altmann *et al.* [11] have developed a system to detect military vehicles using acoustic and seismic information for application in disarmament and peacekeeping.

3. System overview

This research describes a person quantification system model for detecting people on beaches. The proposed system uses a classification-based strategy, which searches grey-scale images of beach scenes for potential objects of interest. The sub-images extracted are then processed to determine whether a person object has been found. Each sub-image is processed by using an edge detector, feature extractors and finally a neural-based classifier. An overview of the entire system is presented in Figure 1.

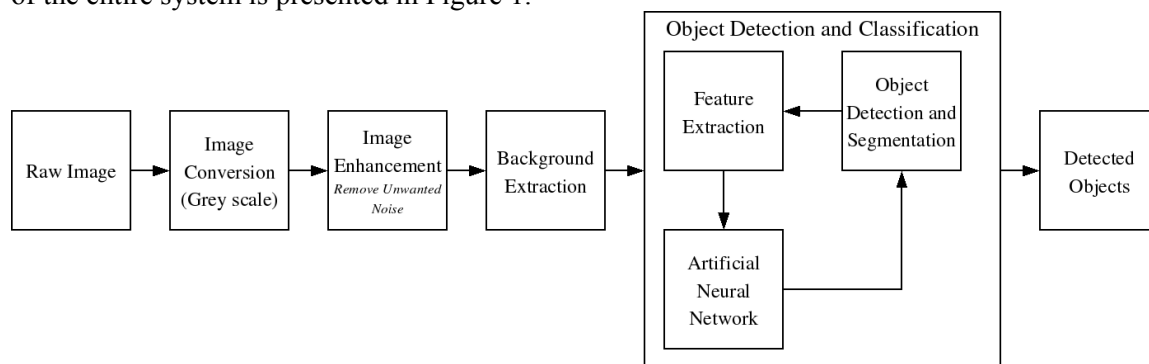


Figure 1: System Overview – Image Processing, Feature Extraction and Object Detection

3.1 Object detection and segmentation

The first important stage for segmenting objects out of a complex scene is to detect the object boundaries in the image. Boundaries encompass an object and are regions where there is a high change in luminance values. Beach scenes mainly contain large areas of sand and sea. This is of great benefit when detecting people, as people in the beach scenes contrast quite significantly and therefore are easier to detect, than say, a forested scene.

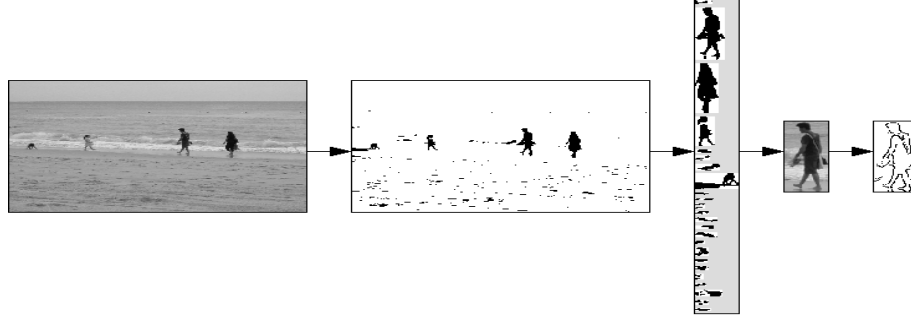


Figure 2: Object detection and segmentation process shown from left to right. The original image, the image after applying the moving average algorithm, segmented images, image segment containing the grey-scale image, and the edge detected image (not to scale).

In the initial step for detecting people in beach scenes, a moving average algorithm was used on the beach images to extract background information (see Figure 2). The moving average algorithm was originally designed for highlighting text in images [12] but can also be applied to beach image analysis. The moving average algorithm detects the mean grey level of the last n pixels (see Equation 1).

$$M_{i+1} = M_i - \frac{M_i}{n} + g_i + 1 \quad (\text{Eq. 1})$$

Where M_{i+1} is the estimated moving average for pixel $i + 1$ having a grey-scale value of $g_i + 1$ and M_i is the previous moving average value. If a pixel is less than the moving average, it is set to black, otherwise it is set to white. As may be seen in Figure 2, the black and white image obtained from the moving average algorithm provides an encouraging result. The next stage is to segment the objects out of the now binary scene.

Objects were segmented out of the binary image using a breadth first search, starting at the top left of the image. Pixels were grouped to form objects based on their radius r from other pixels. In Figure 3, the image contains 7 distinct groups of pixels, with some groups having only one pixel. If the above process is applied with a radius $r = 2$ and a minimum number of pixels $n = 4$, then only the three objects displayed on the right in Figure 3 will be returned.

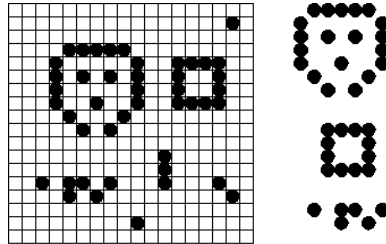


Figure 3: Object segmentation. The original image is shown on the left, and the segmented objects on the right. The segmentation settings were $r = 2$ and $n = 4$.

The segmented components are now used to locate the objects in the original (grey-scale) image scene. These objects are subsequently processed by the Canny edge detection algorithm [13] and a feature vector for each object (created using the Modified Direction Feature - MDF - extractor) is sent to a Multi-layer Perceptron (MLP) for classification.

3.2 Modified direction feature (MDF)

MDF has been detailed elsewhere [14] and will only be briefly described here. MDF feature vector creation is based on the location of transitions from background to foreground pixels in the vertical and horizontal directions of an image represented by edges. When a transition is located, two values are stored: the Location of the Transition (LT) and the Direction Transition (DT) (see Figure 4). An LT is calculated by taking the ratio between the position where a transition occurs and the distance across the entire image in a particular direction. The DT value at a particular location is also stored.

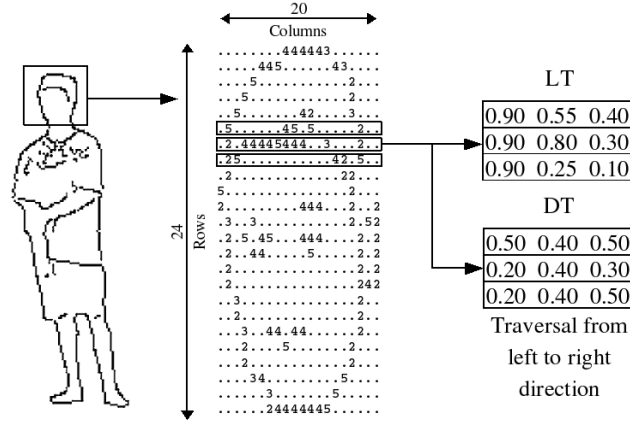


Figure 4: In the above image the MDF extraction technique looks at the person from left-to-right direction. The head section shown only part of the process to find the LT and DT values.

The DT is calculated by examining the stroke direction of an object's boundary at the position where a transition occurs (as defined in Blumenstein *et al.* [14]). Finally, a vector comprising the [LT, DT] values in each of the four possible traversal directions is created

3.3 Neural-based classification

In this research, an interconnected, feed-forward Multi-layer Perceptron was adopted for the process of classifying objects into person and non-person categories. The MLP was trained using the backpropagation (BP) algorithm. In the experiments conducted, the BP-MLP was trained with images processed by the MDF extraction technique.

4. Results

4.1 Image database

To test the person detection system on beach scenes using the MDF feature extraction technique, a small database was created including people and non-people objects. The image database consisted of 580 grey-scale images of varying size. These images were extracted from grey-scale beach images of 648×480 pixels. The image database was classified by a human operator into two categories; "person" and "non-person". The image database was then broken into two sets. Set 1, the training set, consisted of 200 people and 200 of non-person images. Set 2, the test set, consisted of 90 people and 90 non-person images (see Figure 5).



Figure 5: Person/Non-Person Image Database

4.2 Classification results

In this section, results for the classification of person objects are presented in tabular form. Table 1 presents the results obtained using MDF for processing the test image set. The MLP parameters for training were modified to optimise the training result. The test set classification rate was calculated based upon the number of successfully recognised person objects (true positives). Also listed below is the number of "non-person" items incorrectly labeled as person objects (false positives). The table shows that overall the system gave a fairly good result considering the high complexity of the beach scenes.

Table 1: BP-MLP classification rates using the MDF extraction technique

# Hidden Units	Person object classification rate [%]	
	True positive	False positive
6	88	14
8	88	13
10	86	16
12	84	12
16	87	17
20	87	11
24	90	12
30	91	13
34	83	18

4.3 Discussion

As may be seen from the table above, the best result was a classification rate of 91% for people with a false positive rate of 13% using the MDF technique and an MLP with 30 hidden units. Figure 6 shows people correctly identified, with a white rectangle highlighting the actual segmented image. As may be seen, some of the people not identified were very small and therefore it was probably unrealistic for the system to detect them. Figure 7 shows images that the system did not recognise as people. This seems mainly due to poorly segmented images, other connected objects and background interference.

5. Conclusions and future research

In this paper, a person detection system for quantifying people in beach scenes has been described. The system segmented the beach scenes into person and non-person images using a moving average algorithm, which then incorporated a neural-based classification system for the detection process. At this stage, the features extracted are not invariant to object rotation and future research will be directed to address this point. Early results are encouraging for classifying people in beach scenes, but future work will examine different object segmentation techniques, and determine ways to further examine images that are not correctly classified. Future work will also address methods to ignore or reject objects below a certain pixel range.



Figure 6: Person detection results

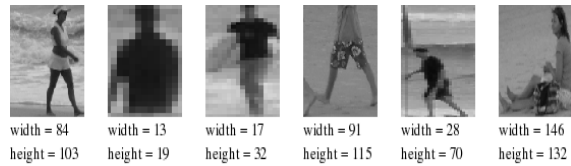


Figure 7: False person detection results (not to scale)

6. References

- [1] M. Browne, M. Blumenstein, R. Tomlinson and C. Lane, An intelligent system for remote monitoring and prediction of beach conditions, *Proceedings of the International Conference on Artificial Intelligence and Applications*, Innsbruck, 2005, pp. 533-537.
- [2] A. J. Schofield, T. J. Stonham and P. A. Mehta, Automated people counting to aid lift control, *Automation in Construction*, Vol. 6, No. 5-6, 1997, pp. 437-445.
- [3] A. J. Schofield, P. A. Mehta and T. J. Stonham, A System for Counting People in Video Images using Neural Networks to Identify the Background Scene, *Pattern Recognition*, Vol. 29, No. 8, 1996, pp. 1421-1428.
- [4] A. Iketani, A. Nagai, Y. Kuno and Y. Shirai, Real-Time Surveillance System Detecting Persons in Complex Scenes, *Real-Time Imaging*, Vol. 7, No. 5, 2001, pp. 433-446.
- [5] C. Sacchi, G. Gera, L. Marcenaro and C. S. Regazzoni, Advanced image-processing tools for counting people in tourist site-monitoring applications, *Signal Processing*, Vol. 81, No. 5, 2001, pp. 1017-1040.
- [6] F. Bartolini, V. Cappellini and A. Mecocci, Counting people getting in and out of a bus by real-time image-sequence processing, *Image and Vision Computing*, Vol. 12, No. 1, 1994, pp. 36-41.
- [7] C-J. Pai, H-R. Tyan, Y-M. Liang, H-Y. M. Liao and S-W. Chen, Pedestrian detection and tracking at crossroads, *Pattern Recognition*, Vol. 37, No. 5, 2004, pp. 1025-1034.
- [8] J-C. Tai, S-T. Tseng, C-P. Lin and K-T. Song, Real-time image tracking for automatic traffic monitoring and enforcement applications, *Image and Vision Computing*, Vol. 22, No. 6, 2004, pp. 485-501.
- [9] D. M. Ha, J.-M. Lee, Y.-D. Kim, Neural-edge-based vehicle detection and traffic parameter extraction, *Image and Vision Computing*, Vol. 22, 2004, pp. 899-907.
- [10] C. Wöhler and J. K. Anlauf, Real-time object recognition on image sequences with the adaptable time delay neural network algorithm — applications for autonomous vehicles, *Image Vision and Computing*, Vol. 19, No. 9-10, 2001, pp. 593-618.
- [11] J. Altmann, S. Linev and A. Weiß, Acoustic-seismic detection and classification of military vehicles--developing tools for disarmament and peace-keeping, *Applied Acoustics*, Vol. 63, No. 10, 2002, pp. 1085-1107.
- [12] J. R. Parker. Algorithms for Image Processing and Computer Vision. Wiley, 1997.
- [13] J. F. Canny, Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, 1986, pp. 679-698.
- [14] M. Blumenstein, X. Y. Liu and B. Verma, A Modified Direction Feature for Cursive Character Recognition, *Proceedings of the International Joint Conference on Neural Networks (IJCNN '04)*, Budapest, Hungary, 2004, pp. 2983-2987.