

An Exhaustive Search Strategy for Detecting Persons in Beach Scenes using Digital Video Imagery and Neural Network-based Classification

Steve Green and Michael Blumenstein

Abstract— This paper presents an investigation of a neural-based technique for detecting and quantifying persons in beach imagery for the purpose of predicting trends of tourist activities at beach sites. The proposed system uses various pre-processing and segmentation techniques to initially isolate potential objects in cluttered scenes. A structural feature extraction technique is then used to represent objects of interest for training a neural classifier. An exhaustive search strategy, incorporating a neural network, is proposed to effectively scan beach images to determine whether objects are "person" or "non-person". Encouraging results are presented for person detection using video imagery collected from a beach site on the coast of Australia.

I. INTRODUCTION

The study of local beach behavior can provide valuable information for calculating advertising requirements, and providing information on whether a beach environment is sufficient to meet the current human demand. Specifically, detecting and quantifying 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. World Wide Web cameras (sometimes also called web cams) are currently posted to overlook most beaches around Australia (and the world) to provide local beach information accessible through the Internet. Coastalwatch Pty. Ltd. provides one such service.

The web cams that Coastalwatch provide offer 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. Therefore, aside from the inherent difficulties in dealing with variable outdoor imagery, 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

S. Green is with the School of Information and Communication Technology, Griffith University, PMB 50 Gold Coast Mail Centre, Bundall, Q 9726 AUSTRALIA; (e-mail: S.Green@griffith.edu.au).

M. Blumenstein is with the School of Information and Communication Technology, Griffith University, PMB 50 Gold Coast Mail Centre, Bundall, Q 9726 AUSTRALIA; (contact author; phone: +61 7 5552 8271; e-mail: M.Blumenstein@griffith.edu.au).

environment for people of all ages [1]. Currently, lifeguards provide assistance for swimmers who encounter trouble between the flags. Flagged areas are designated as patrolled areas on Australian beaches. 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.

II. OVERVIEW OF EXISTING TECHNIQUES FOR PERSON AND OBJECT DETECTION

The automated detection of persons and their behaviour in beach scenes is a novel application in the field of video surveillance. However, a number of techniques and systems have been proposed for automated analysis of humans in other indoor and outdoor situations. Some of these are reviewed and detailed in the paragraph that follows.

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 on and off a bus using image sequence analysis for allocating appropriate resources on bus lines. Pai et al. [7] present a system for pedestrian tracking using vision-based techniques to prevent traffic accidents.

Foresti et al. [8] present a motion-based people detection system. The system detects moving people in indoor scenes using a neural network to classify if the current frame in a video stream is static or contains mobile objects. Subsequently a finite state automaton is used to reduce detection errors. Finally, Haritaoglu et al. [9] propose a real time people detection system in outdoor scenes. The system uses shape analysis and tracking to locate people in a video frame. Foreground objects are extracted from an image scene, then each object is distinguished by their shape and periodic motion clues.

I. SYSTEM OVERVIEW

This research describes a model for the automatic detection of people on beaches using an exhaustive search strategy. The proposed system uses a classification-based approach, 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. An overview of the entire system is presented in Figure 1.

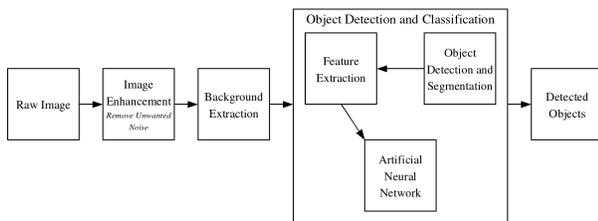


Fig. 1. System overview

A. 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. From one point of view, this is of great benefit when detecting people, as people in beach scenes contrast quite significantly with the sand and therefore are easier to detect, than say, in a forested scene. However, conversely, the presence of the ocean, constant wave motion and shadowing effects can make the process a challenging one.

Two different approaches to object segmentation have been carried out; the first approach used a background extraction technique to segment foreground objects, and the second approach utilized the benefits of the first technique but subsequently applied an exhaustive search to regions of interest (ROI).

1) Background Extraction

As a first step to 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 proposed for highlighting text in images [10] but can also be applied to beach image analysis. The moving average algorithm detects the mean grey level of the last n pixels. Subsequent processing enables the segmentation of objects from a binary image, whereby the boundaries of the objects are detected. A feature vector is then created, which is subsequently presented to a Multi-layer Perceptron (MLP) for classification [11].



Fig. 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 component containing the grey-scale image, and the edge detected image (not to scale).

2) Exhaustive Search

The background extraction technique described above provides a fast method for removing background information such as sand, sea and sky. Unfortunately, foreground objects (blobs) detected using this technique can be joined to other objects by shadows, by being too proximate to other objects, and occlusion. An exhaustive search technique is subsequently put forward to try and overcome the aforementioned problems. The technique described in this research applies a search window of increasing size over ROIs to detect person and non-person objects (See Figures 3 and 4).

As the name implies, the exhaustive search method is

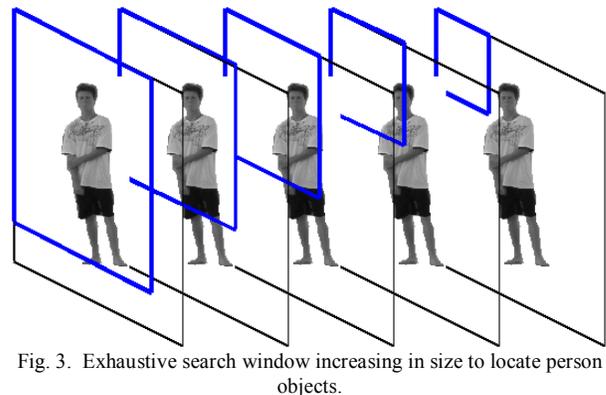


Fig. 3. Exhaustive search window increasing in size to locate person objects.

highly computationally inefficient. To reduce the search time, regions that were not of interest, needed to be excluded quickly. One method to exclude these regions quickly is to provide a hierarchy of classifiers, with the simplest and most

efficient classifier utilized first to remove non-objects. The proposed object detection system incorporates a single classifier (MLP).

Regions containing non-objects are removed from the search space by applying background extraction, and then building a Summed-Area Table (SAT) [12] of the binarised

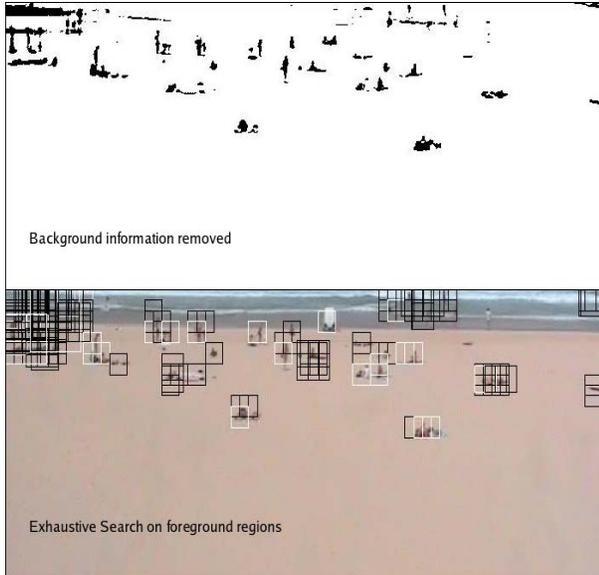


Fig. 4. Exhaustive search technique locating persons based on foreground objects, which are revealed by applying background extraction.

image. The SAT can then be used to quickly look at a ROI to see if it contains foreground pixels. In order to provoke a response, a ROI was required to contain between 20-80% of foreground pixels compared to the maximum possible number of foreground pixels. This was an empirical measurement obtained based on observation. If a ROI is classified as containing a person object, the region is removed from the search space by setting the foreground pixels to 'white', and then rebuilding the SAT.

B. Modified direction feature (MDF)

MDF has been detailed elsewhere [13] 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 5). 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.

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.* [13]). Finally, a vector comprising the [LT, DT] values in each of the four possible traversal directions is created.

1) Rotationally Invariant MDF:

At present, MDF is not rotationally invariant. As people in beach scenes can be either standing up, or laying down, it is necessary to create an image database that has images of people in all possible positions to obtain a good person classification rate. Another way to approach this problem is to extend the current MDF to incorporate rotational invariance.

Currently, MDF examines both foreground transitions and direction values to build a feature vector. Foreground transitions are calculated by searching from all four sides of an image. To make this process rotationally invariant, it is hypothesized that foreground transition values can be calculated from the object's centroid in a 360 degree radius, starting at some maximum point on the object from the object's centroid. As there can be several maximum points, a method is required to find the same maximum point on an object regardless of the object's rotation. Currently, the Hotelling Transform along with eigenvectors for locating the principal axes of an object is being considered for this task [14]. Both foreground and direction values can then be calculated to create a feature vector that does not differ with respect to a particular object's rotation.

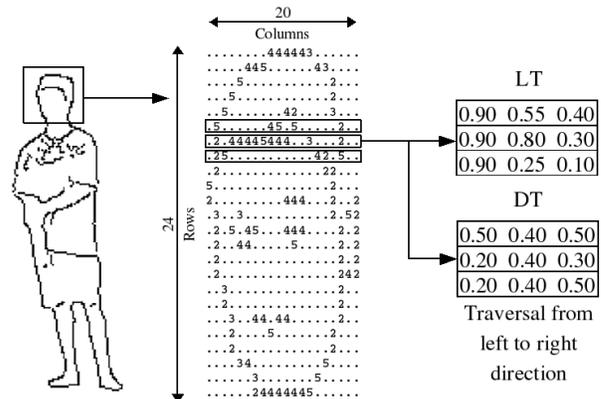


Fig. 5. In the above image the MDF extraction technique examines the person from the left-to-right direction. The head section shown only part of the process to find the LT and DT values.

C. Neural-based classification

In this research, an interconnected, feed-forward MLP 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.

II. RESULTS

A. Image database

To test the person detection system on beach scenes using the MDF feature extraction technique, a database was created including people and non-people objects. The image database consisted of 450 images of varying size. These images were extracted from DVD recordings taken from a Coastalwatch web camera at Surfers Paradise. Frames from the beach recordings were extracted, and image regions were

classified by a human operator into two categories; "person" and "non-person". The image database was then broken into three sets to facilitate three-fold cross validation, whereby each set was composed of 350 training images and 100 test images (see Figure 6).

B. Classifier settings

As previously mentioned, the image database was broken into separate sets to perform three-fold cross validation. On each set, training was conducted using 8, 12, 16, 20, 24, 28,

TABLE I
BP-MLP CLASSIFICATION RATES USING THE MDF EXTRACTION TECHNIQUE (TRUE POSITIVE TP, FALSE POSITIVE FP)

# Hidden Units	Test01 [%]		Test02 [%]		Test03 [%]		Average of tests 1-2-3 [%]	
	TP	FP	TP	FP	TP	FP	TP	FP
8	86	13	82	9	92	11	87	11
12	84	10	86	8	94	9	88	9
16	80	8	86	8	90	4	85	7
20	82	6	82	6	90	5	85	6
24	86	7	86	6	90	3	87	5
28	82	4	84	6	90	4	85	5
32	86	6	82	4	90	6	86	5
36	82	4	86	6	90	7	86	6

TABLE II
RESULTS FOR BOTH BACKGROUND EXTRACTION AND EXHAUSTIVE SEARCH METHODS, SHOWING THE NUMBER OF PEOPLE DETECTED (TRUE POSITIVE) IN EACH FRAME

Test01		
Frame#	True Positive	False Positive
Background Extraction		
10010	5	1
10080	8	1
10110	5	2
Exhaustive Search		
10010	19	4
10080	22	4
10110	20	10

Test02		
Frame#	True Positive	False Positive
Background Extraction		
10010	13	10
10080	6	13
10110	12	8
Exhaustive Search		
10010	32	16
10080	34	19
10110	30	20

32, and 36 hidden units, and the RMS error rate was recorded. An RMS error rate of 0.001 was used as a stopping criterion, but in some cases the MLP did not train to an RMS of 0.001. Table 1 shows the results for all three sets, and the average of the three sets.

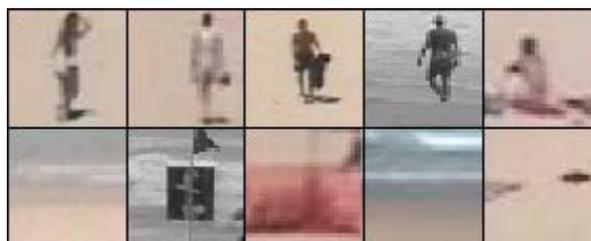


Fig. 6. Examples from the Person and Non-Person Image Database (not to scale)

C. 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 sets. 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 provided a good result considering the complexity of the beach scenes. A further test was undertaken on an independent test set of images containing fifty person

images and fifty images of ‘other beach objects’. The additional tests showed comparable results with those illustrated in Table 1, using corresponding classifier settings. The independent test set provided a slightly higher result for classifying persons (true positive), however provided a higher false positive rate for ‘other beach objects’.

D. Training results discussion

As may be seen from Table 1, the best result on average was a classification rate of 85% for people with a false positive rate of 5% using the MDF technique and an MLP with 28 hidden units. Figure 7 shows some results for cases of false positives and true positives. By observation of the data, people that were not identified correctly were either lying on the beach, occluded by other objects (non-person), or were obscured by the surf. In the case of false positives, objects that were classified as a person when the object was a non-person, tended to have an outline that was upright, causing the MLP to give an incorrect classification.

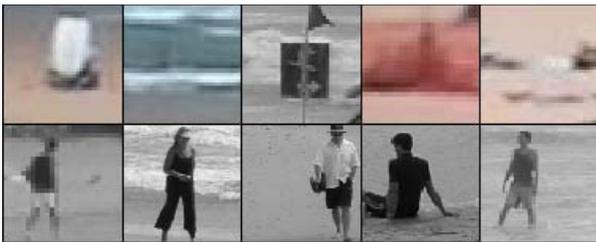


Fig. 7. Person detection results showing false positives (top row) and true positives (bottom row); both sets of images are not to scale.

E. Comparison between background extraction and exhaustive search results

Experiments were carried out to compare both background extraction and exhaustive search techniques using two separate sets of beach footage of ten minutes in length, taken at midday on different days. The beach footage was recorded from a Coastalwatch web camera located at Surfers Paradise beach. All frames in test01 contained approximately 31 people, and test02 contained approximately 42 people. Each frame in the two tests was recorded with white rectangles representing objects classified as a person (true positive), and black rectangles representing objects classified as non-person (false positive). In the case of the exhaustive search, black regions were not shown due to the large number of these generated by the exhaustive search method. For both tests, true positive and false positive counts were performed by a human operator to the best accuracy that could be achieved given the resolution of the images. False positive for the exhaustive search experiments were labeled as a single group when a large number of white rectangles appeared to be tightly packed in a given region (see Figure 8). In Table 2 it can be seen that the background extraction technique by itself did not deal with occlusion and shadowing effects very well, but did manage to locate most areas containing person objects. The

background extraction technique had a much lower false positive rate in that there were very few non-person objects classified incorrectly. The exhaustive search method gave a much better result for locating person objects that were occluded by other objects and were obscured by shadowing effects. The best results for persons correctly located were 22 in test01 and 34 in test02. The disadvantage of the exhaustive search method was the increase in false positive results. This can be clearly seen in Figure 8 (test02), where the fence posts were incorrectly classified as a person in some instances. The fence in test02 was not present in test01, and provided an extra level of difficulty that is not normally incurred on Surface Paradise beach. This illustrates the difficulty and variability inherent in detecting people in beach scenes.

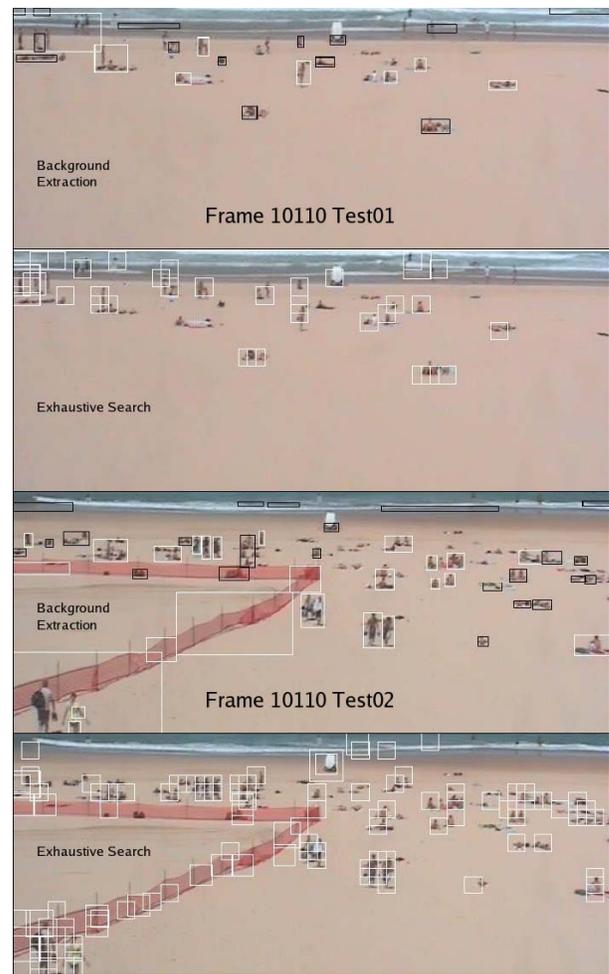


Fig. 8. Persons located in beach imagery using background extraction and exhaustive search methods

III. CONCLUSIONS AND FUTURE RESEARCH

In this paper, a person detection system for quantifying people in beach scenes has been described. A human operator segmented actual beach data from a Surfers

Paradise web camera into person and non-person categories. Tests were then carried out on images using two techniques (1) background extraction (2) and the exhaustive search technique on beach footage. The system incorporated a neural-based classification system for the detection of person and non-person objects. Encouraging results were presented for automated person detection and quantification, which can be applied to beach scene analysis for surveillance and coastal management applications.

Currently, the features described in this paper are not invariant to object rotation. Future research will be directed to address this point. Further work will also examine different object segmentation techniques, and determine ways to further deal with images that are not correctly classified. Finally, frames retrieved from beach media are at present processed individually. Future research will look at generating some confidence value for a person object based on temporal data from past frames. Also, motion detection may provide a more effective way of detecting person objects in beach imagery.

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] G.L. Foresti, C. Micheloni, and C. Piciarelli, "Detecting Moving People In Video Streams", *Pattern Recognition*, Vol. 26, No. 14, 2005, pp. 2232-2243.
- [9] I. Haritaoglu, D. Harwood, L.S. Davis, "W4: real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, 2000, pp. 809-830.
- [10] J. R. Parker. *Algorithms for Image Processing and Computer Vision*. Wiley, 1997.
- [11] S. Green, M. Blumenstein, M. Browne, R. Tomlinson, "The detection and quantification of persons in cluttered beach scenes using neural network-based classification," *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA '05)*, Las Vegas, USA, 2005, pp. 303-308.
- [12] F. Crow, "Summed-Area Tables for Texture Mapping," *Computer Graphics (SIGGRAPH '84 Proceedings)*, Vol. 18, No. 3, 1984, pp. 207-212.
- [13] 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.
- [14] R. C. Gonzales, R. E. Woods, *Digital image processing*, Addison-Wesley, 1993.