Nautical Scene Segmentation Using Variable Size Image Windows and Feature Space Reclustering

P. Voles, A.A.W. Smith, and M.K. Teal

Machine Vision Group, Bournemouth University, UK
{pvoles, asmith, mteal}@bournemouth.ac.uk

Abstract This paper describes the development of a system for the segmentation of small vessels and objects present in a maritime environment. The system assumes no a priori knowledge of the sea, but uses statistical analysis within variable size image windows to determine a characteristic vector that represents the current sea state. A space of characteristic vectors is searched and a main group of characteristic vectors and its centroid found automatically by using a new method of iterative reclustering. This method is an extension and improvement of the work described in [9]. A Mahalanobis distance measure from the centroid is calculated for each characteristic vector and is used to determine inhomogenities in the sea caused by the presence of a rigid object. The system has been tested using several input image sequences of static small objects such as buoys and small and large maritime vessels moving into and out of a harbour scene and the system successfully segmented these objects.

1 Introduction

Maritime vessels are today faced with the threat of piracy. Piracy is usually associated with the old swash buckling films and consequently we do not consider piracy in the modern age, however several incidents of piracy happen each day, particularly in the Mallaca straights and the South China Sea areas. Here fast RIB craft (Rigid Inflatable Boats) approach the stern of a large cargo ship, even super-tankers, and scale the ship using simple rope ladders. The small numbers of crew that these ships have on duty means pirate detection needs to be automated. Current Radar systems are of limited use in these situations as RIB craft are small almost non-metallic and consequently have poor radar returns and as such radar systems find them difficult to detect. To overcome this problem an image processing system is under development.

The maritime scene, however, has been found to be extremely complex to analyse [1], [2], producing large number of motion cues making identification and tracking in the visual environment complex. The system being developed here concentrates on the task of extracting the maritime vessels and other static nautical objects (buoys, mooring buoys, piers, etc.) from the sea to aid the recognition and tracking process. To accomplish this task three integrated algorithms have been developed, namely (i) variable size image window analysis, (ii) statistical

D. Vernon (Ed.): ECCV 2000, LNCS 1843, pp. 324–335, 2000.

analysis by reclustering and (iii) region segmentor. The variable window analysis determines a set of overlapping image windows and, for each image window, calcualtes the energy, entropy, homogeneity and contrast. This vector effectively forms a four-dimensional feature for each image window. The statistical analyser uses a new method of iterative reclustering of the feature space to determine the centroid of vectors representing the main feature in the scene (sea) [7]. The region segmentor calculates the Mahalanobis distance between the values of the feature centroid in each image window which identifies outliers from the mean. These outliers are potentially regions that contain inhomogeneities, effectively forming a feature map [3], which may indicate the presence of a rigid object. These extracted regions effectively form regions of interest (ROI) in the image, and the region segmentor identifies these ROI's in the original image sequence using white rectangular boxes.

Figure 1. Typical nautical scene.

2 Window Analysis

The segmentation of a maritime scene is complicated by the fact that waves cause noise (undesirable regions of interest) that does not have a Gaussian distribution, and consequently traditional ways of filtering are ineffective. The main properties of this noise are spatial dependent i.e. its appearance. Fig. 1 shows a typical maritime scene and we can see that the noise is not distributed uniformly in the image, a noise 'pattern' is formed which can clearly be seen in the bottom of the image.

Commonly used texture techniques (as described in many texts such as [10]) proved to be unsuccessful in describing the distribution of these noise patterns as they differ from scene to scene and frame to frame. However, from looking at a typical nautical scene, we observe a plane (sea level) that is almost parallel to the camera axis.

The bottom of the image contains that part of the sea that is closest to the camera, while the horizon is made up of points at infinity on the sea level plane [4]. Therefore, the resolution of observation is larger for any objects that are close to the bottom of image than for objects that are closer to the horizon. This also holds true for the noise patterns. Using this observation we can see that a variable size image window segmentation technique will require finer (smaller) image windows as we approach the horizon, but courser (larger) image windows could be used closer to the bottom of the image. The variable image analysis algorithm is passed the position of the image horizon and an initial window size.

Overlapping image windows are determined by growing the window size from an initial 16 by 16 pixels on the horizon line towards the bottom line of the image at a rate of 6% per window line. For our experiments we used rates from 5% to 10% depending on the camera angle under which the scenes are observed. The image windows are allowed to overlap by 33%. This effectively positions a grid on the sea plane as shown in Fig. 2. If we consider perspectivity then the correct shape of the projected grid tiles should be trapezoidal. This brings a complication to the process because we would have to use bilinear or other perspective transformation for each of the windows to transform it into a rectangle. These transformations are computationally intensive. However, it has been found that rectangles provide a good approximation of trapezoidal segments. The size of the windows and the amounts of overlays are stretched accordingly so the windows cover a whole region under observation and there are no uncovered 'blind spots' on the sides and at the bottom of the image.

Each window is then resized to the size of the smallest windows (a window near the horizon) by using either simple re-sampling or bilinear interpolation. Bilinear interpolation gives better results but is slower, while simple re-sampling gives poorer results but is much faster and for most applications is sufficient. The final task of the variable window analysis is to calculate the following statistical values[5] for each image window:

$$energy = \sum_{r=0}^{R} \sum_{c=0}^{C} P(r,c)^2 \ . \tag{1}$$

$$entropy = \sum_{r=0}^{R} \sum_{c=0}^{C} \log(P(r,c)) \cdot P(r,c) \ . \tag{2}$$

$$homogeneity = \sum_{r=0}^{R} \sum_{c=0}^{C} \frac{P(r,c)}{1+|r-c|} \ . \tag{3}$$

$$contrast = \sum_{r=0}^{R} \sum_{c=0}^{C} (r-c)^2 \cdot P(r,c) \ . \tag{4}$$

where r, c are row and column indexes, P(r,c) is the pixel value at position r,c and R, C are the image window boundaries. The calculated values are arranged to form a 4-element vector, giving N 4-element feature vectors, where N is the total number of windows in the segmentation.

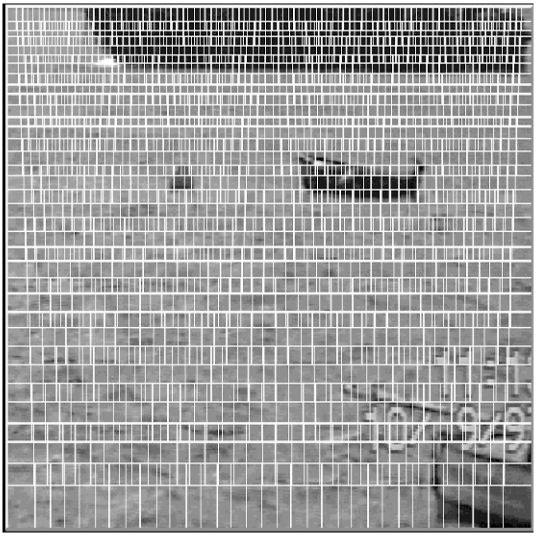


Figure 2. Variable Image Windows overlaid on the sea, minimum window size of 16 x 16 pixels with 33% overlap and an expansion rate of 6%.

3 Statistical Analyser

We can consider the vectors calculated from the variable window analysis as a population of points in a 4-dimensional feature space. The statistical analyser determines a set of characteristic features that could be used to describe the current sea state. This set is represented by a main cluster in the feature space.The previous algorithm used to find the main cluster is described in [9]. This algorithm uses histograms that are constructed for each of the four previously described characteristics. It divides the smoothed data histograms into subparts by local minima and assigns the largest subpart to the main cluster. This method does not perform well for smaller numbers of feature vectors. In these cases it becomes difficult to find the correct local minima because of the lack of data needed to create meaningful histograms. Another disadvantage is the presence of many thresholds whose settings influence the results significantly.

A new method is introduced that helps to approximate the distribution of the unlabelled feature vector data in feature space. The method takes feature vectors generated at the previous stage of the algorithm (variable size image windows analysis) as an input and it iteratively determines the centroid and the covariance matrix for the data in the main cluster. The problem here is that there is no useful knowledge about the data due to the nature of the problem (each scene segmented in the previous stage of the algorithm can be significantly different from the previous one in terms of sea appearance and presence of objects). The only usable knowledge is that there is a certain main cluster in the feature space which comprises the vectors corresponding to major features in the scene (presumably the sea). These vectors are relatively close to one another. Other vectors (outliers) represent regions where objects are in the scene and these vectors are relatively far from the main cluster and it's centroid. Unfortunately, due to the nature of the problem, we cannot use learning and classification algorithms (as described in Shalkoff [6]) as the feature data can change its values disobeying any rule at all. The distributions of feature data change from scene to scene and the only usable information is the presence of the main cluster and possible outliers.

We assume that the main cluster contains the majority of vectors and that these vectors are relatively close to one another. Other vectors or groups of vectors (representing the objects) are positioned relatively far from this main cluster. Therefore, if we calculate the centroid of all the vectors in the distribution by using the mean, or better, median then we can assume that this centroid of all vectors is not far from the centroid of only the vectors in main cluster. That is, because there are many vectors close together whose position will bias (or attract) the position of the centroid determined as the median of all vectors in the feature space. Experiments have proved that median performed better than mean because median is not influenced by a small number of outlying vectors. The next step after determining the centroid of the whole distribution in feature space is to choose which vectors actually fall into the main cluster. We assume that the main cluster lies within the boundary that corresponds to the mean distance of all the vectors from the determined centroid. Thus, the resulting group of vectors has a centroid corresponding to the median of all vectors in the distribution and includes vectors with distance's less than the mean distance of all the vectors in the distribution in the feature space.

The next step is similar to the one described above: once again, we determine the median centroid but now we use only the vectors lying within the mean distance from the previous centroid. We recalculate the mean distance from the newly calculated centroid for all the vectors in the group. The new main cluster consists of the vectors that lie within the new mean distance from the new centroid.

This process is repeated iteratively. The number of iterations is not significantly large as after each step the group of selected vectors shrinks significantly, especially if the main cluster is packed tightly together. Practical experiments proved that one to three iterations are sufficient.

We use the Mahalanobis measure to determine the distances among feature vectors:

$$k = (\overrightarrow{x} - \overrightarrow{\mu}) \cdot C^{-1} \cdot (\overrightarrow{x} - \overrightarrow{\mu})^T \quad . \tag{5}$$

where k is the distance, $\overrightarrow{x}$ is the feature vector, $\overrightarrow{\mu}$ is the centroid and C^{-1} is the inverse of covariance matrix. The reason for using the Mahalanobis distance is that the data is highly correlated. The Mahalanobis distance used in this method is slightly modified - the centroid used in the formula is not determined as a mean but as a median. The reason for that, as stated above, is the avoidance of outliers. This method only determines the main cluster and it's centroid approximately but as we haven't got any prior knowledge about the data it is sufficient to determine the outliers that represent the regions with objects in the scene. Experiments proved that the separation of outliers from the main cluster vectors is by means of orders (value of Mahalanobis distance of outliers from the centroid is by a few orders higher than the distance of vectors in the main cluster) even for highly scattered feature vectors. Another important property of the method is the fact that it does not shift the centroid of the feature vectors significantly if the data is relatively consistent and does not contain any outliers.

The main advantage of the algorithm is that there is no need for prior knowledge to approximate the distribution of the vectors in the main cluster. Another important advantage is the absence of any thresholds. The only value that is to be set is the number of iterations and as stated above, one to three iterations are sufficient. Figures 3a-3f show two iterations of the reclustering process in 2D projections of the feature space.

The statistical analyser applies the method described above onto the feature vectors determined by variable size image windows analysis and it determines the Mahalanobis distance from the main cluster centroid for each of the vectors.

4 Region Segmentor

The statistical analyser has calculated the distances of the feature vectors from the centroid of the main cluster which represents the main feature in the image (presumably the sea), the region segmentor must now determine those image windows whose feature vectors have Mahalanobis distance above the set up threshold. The values of the Mahalanobis distance for each vector provide a measure of the likelihood of an image window being an object, the greater the distance value the more the likelihood of it being a vessel or other man-made object. Figure 5 shows the result of transforming the values of the Mahalanobis distance measure back into the image plane, the darker the image window, the greater the likelihood of that tile being a region of interest. The Mahalanobis distances are now scanned and the rate of change of the distance is calculated. If the rate of change is below a threshold value, the Mahalanobis distance is replaced with the minimum of that region. Finally Mahalanobis distances which have minimum values correspond to be the primary feature in the scene, namely

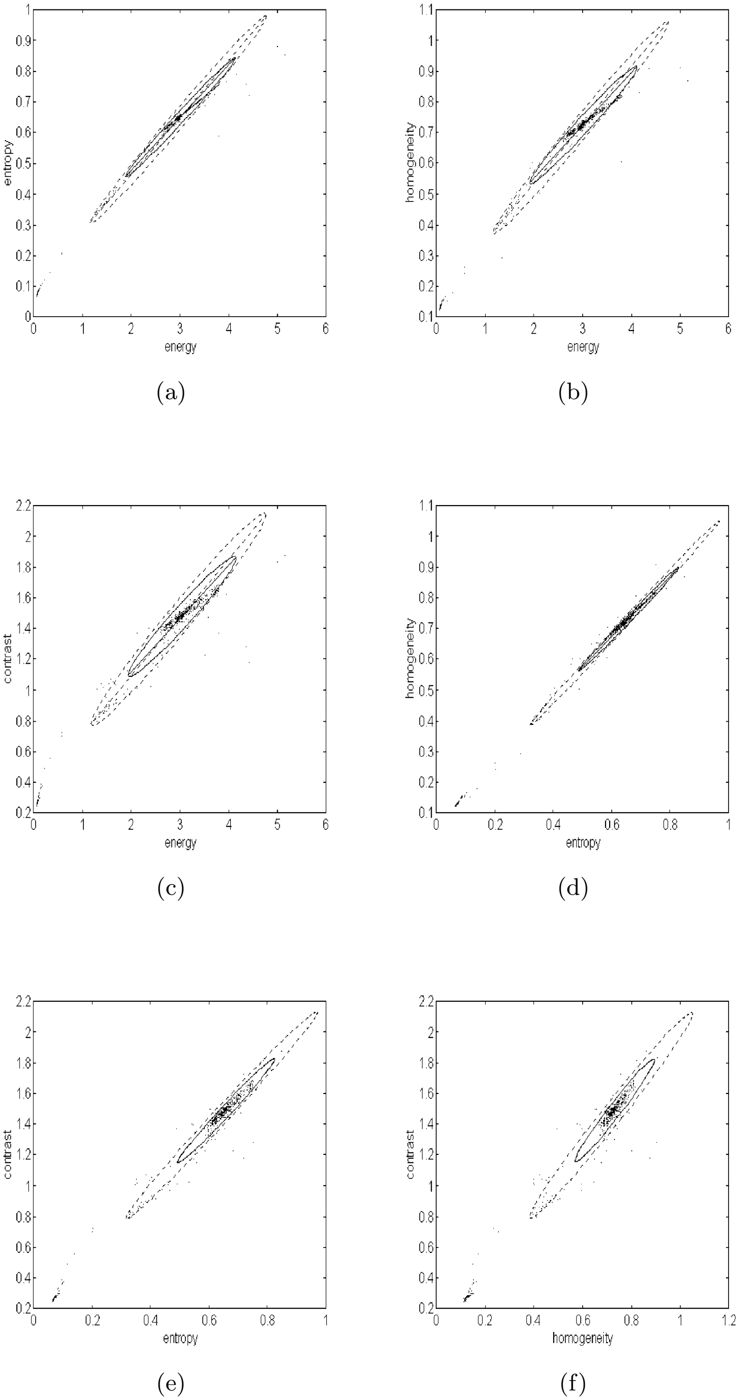


Figure 3. 2D projections of feature space showing the re-alignment of the elipsoid representing the main cluster (dashed line - 1st iteration, solid line 2nd iteration).

the sea (Fig. 4). The determination of the primary feature works even if the object covers the majority of the scene. The main feature then represents the object and outliers, determined by a large distance value, represent either sea or other smaller objects.

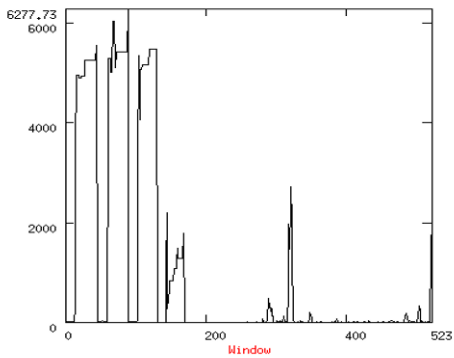


Figure 4. Mahalanobis distance of feature vectors from the centroid of the main cluster after process of homogenizing (distance values are substituted by local minima).

5 Discussion

A static camcorder was set up at the entrance to Portsmouth harbour and an image sequence showing small motor vessels and in particular RIBs moving out of the harbour was filmed. From this sequence a 1500 frame clip was digitised to disk at a rate of 10 frames per second. A second sequence was filmed at Poole harbour showing yachts and buoys moving in the scene and a third showing a medium sized vessel approaching a pier.

The error rate of the segmentation was determined as a ratio between number of frames where the segmentation was incorrect (i.e., rigid objects present in scene were not found or false regions without any objects were marked) and total number of frames in each sequence. This ratio is stated in percentage terms.

Figures 6a and 6b show the Portsmouth scene where a larger motor vessel led a procession of five smaller motor vessel out of the harbour. The algorithm correctly segmented the motor vessels 91% of the time, however, as the vessels moved across the scene several segmented regions were merged. This particular sequence included a number of RIBs.

Figures 7a and 7b show the Poole scene where small and large yachts were moving into and out of the harbour entrance together with a small buoy. The

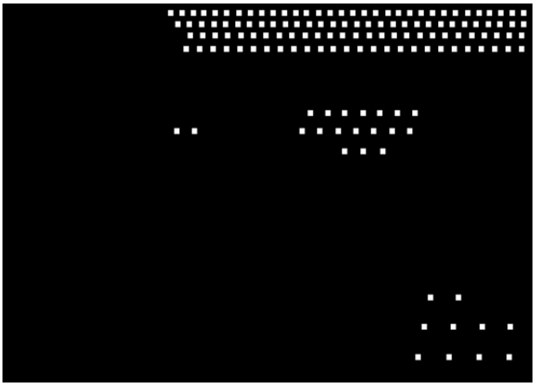


Figure 5. Mahalanobis distance transformed back to image plane. Blobs are positioned at the centers of windows used in segmentation. Brightness of the blob corresponds to the likelihood of object beeing present in the window.

system segmented out the yachts 95% of the time, however, again as vessels crossed, the segmented regions were merged. The system did however, incorrectly segment the buoy 15% of the time but this is still an improvement on the results shown in [9].

Figure 7a shows that algorithm has found only the bottom of the large yacht. The reason for this is, that the algorithm is segmenting only the sea. It ignores everything above the shore. Thus, this algorithm serves only as a partial solution of maritime scene segmentation task.

Figures 8a and 8b show the ability of the system to correctly segment either static and moving objects in the scene even if these cover large areas of the image.

6 Conclusion

A method for segmenting static man-made objects and small vessels moving in a maritime scene has been developed and has been shown to provide reliable segmentation results for a number of maritime scenes. The algorithm uses simple mathematical operators to build a statistical character of the sea. A new method of feature space re-clustering is has been introduced for statistical analysis, based on the work first described in [9].

One advantage of the algorithm is the use of only the current image in the segmentation process, the algorithm does not rely on any change between consecutive images to provide the regions of interest. It does not rely on any prior knowledge about the characteristics representing the sea. It efficiently eliminates the noise caused by the motion of the sea, and has demonstrated within the constraints of the project that this is both scene and time independent.

However, the algorithm as it stands requires initial start positions for the horizon and the minimum window size which must be passed to the algorithm at the start of processing. It also requires the threshold value for separating the main feature from outliers which is the main drawback at the moment. It does not give an exact and final identification about any objects in the scene, it provides only a measure of the objects presence.

Future enhancements to the algorithm are aimed at addressing automating the horizon identification, determining a function for homogenising the Mahalanobis distance measure to preserve outliers and using connectivity analysis to produce improved object detection. The future development is also oriented to find and process the temporal correspondence of the detected regions in the sequence.

Another important enhancement to the algorithm is aimed at substituting the final thresholding of the Mahalanobis distances with a clustering algorithm that connects the regions with similar Mahalanobis distances. A good description of such a clustering algorithm is given in [8].

References

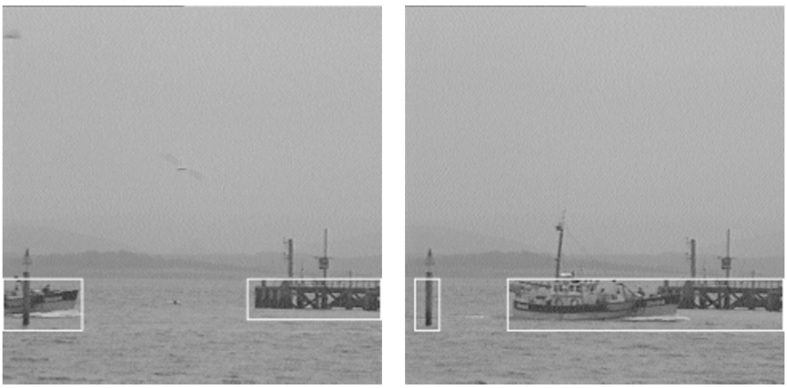
1. Sanderson, J.G., Teal, M.K., Ellis, T.J.: Identification and Tracking in Maritime Scenes. IEE Int. Conference on Image Processing and its applications. (1997) Vol. 2 463–467
2. Smith, A.A.W., Teal, M.K.: Identification and Tracking of Maritime Objects in Near-Infrared Image Sequences for Collision Avoidance. IEE 7th Int. Conference on Image Processing and its applications. (1999) Vol. 1 250–254
3. Campbell, .N.W., Thomas, B.T.: Segmentation of natural images using self organising feature maps. British Machine Vision Conference Proceedings. (1996) 223–232
4. Mohr, R., Triggs, B.: Projective geometry for image analysis. A tutorial given at ISPRS in Vienna. (1996)
5. Jain, R., Kasturi, R., Schunck, B.G.: Machine Vision. (1995) 234–241
6. Schalkoff, R.: Pattern Recognition, Statistical, Structural and Neural Approaches. John Wiley & Sons Inc. (1992).
7. Sanderson, J.G., Teal, M.K., Ellis, T.J.: Characterisation of a Complex Maritime Scene using Fourier Space Analysis to Identify Small Craft. 7th IEE Int. Conference on Image Processing and its applications (1999) Vol. 2 803–807.
8. Shapiro, L.S.: Affine analysis of image sequences. Cambridge University Press (1995).
9. Voles, P., Smith, A.A.W., Teal, M.K.: Segmentation of Nautical Scenes Using the Statistical Characteristics of Variable Size Image Windows. accepted for CVPRIP 2000.
10. Sonka, M., et al.: Image Processing, Analysis and Machine Vision. Thomson Computer Press (1993)

(a) (b)

Figure 6. Procession of small motor vessels (a) frame 200, (b) frame 1100.

(a) (b)

Figure 7. Large and small yachts and a buoy, (a) frame 300, (b) frame 900.



(a) (b)

Figure 8. Medium sized vessel approaching a pier, (a) frame 100, (b) frame 200.