Detecting Marine Animals in Underwater Video: Let's Start with Salmon

R.N.Williams
School of Computing.
University of Tasmania
Sandy Bay, Tasmania, Australia
R.Williams@utas.edu.au

T.J.Lambert
School of Computing.
University of Tasmania
Sandy Bay, Tasmania, Australia
tristan@icsmultimedia.com.au

A.F.Kelsall
School of Computing.
University of Tasmania
Sandy Bay, Tasmania, Australia
akelsall@postoffice.utas.edu.au

T.Pauly
The Verdant Group Pty Ltd
Hobart, Tasmania
Australia
Tim.Pauly@sonardata.com

ABSTRACT

Environmental Decision Support Systems (EDSS) constitute an important emerging technology for the management of marine environments and resources world-wide. A requirement for such systems is the provision of accurate and comprehensive data on marine animal populations and habitats. Underwater video is increasing in importance as a technology for collecting this data, but manual analysis of the video is time-consuming and subjective, and so automated techniques need to be developed before video technology becomes effective as a major data collection method for EDSS. A central requirement in many applications is the automated detection of specific marine animals in the video footage. Our aim is to develop a range of techniques for detecting marine animals in underwater video scenes and, as a starting point, we have developed a segmentation technique to automatically detect salmon depicted in images taken from video cameras placed in fish farm cages.

Keywords (Required)

image analysis, underwater video, environmental decision support systems, marine environments, marine animals.

INTRODUCTION

Due to widespread concern about the state of the Earth's oceans, several large-scale scientific projects have begun to investigate the condition of our oceans on a global basis. These are very comprehensive projects involving extensive scientific contributions from many nations. Environmental Decision Support Systems (EDSS) constitute an important emerging technology for the management of marine environments and resources world-wide. Therefore EDSS are integral components in the information technology infrastructure being developed to manage and render useful the enormous quantities of data that will be required to solve the problems besetting marine environments world-wide. A fundamental requirement for marine EDSS is the provision of accurate and comprehensive base data on marine animal populations and habitats, at appropriate spatial and temporal scales, to support the models informing the decision support provided by these systems. Accurate base data is needed to ensure that the models employed by EDSS produce reliable conclusions.

UNDERWATER VIDEO ANALYSIS

Many traditional methods used to study marine organism populations (eg. dragging a net behind a boat and physically collecting samples) are essentially destructive in nature and only give an indication of the population characteristics in the area of interest because they combine the results over a large area (Wilson 2003), making it impossible to extrapolate the findings to surrounding areas.

With rapid improvements in video technology, underwater video is becoming an important method for collecting this data. It is a non-destructive means of data collection and provides much more detailed information on the fine-scale spatial and temporal variability of the data being collected; something that cannot be achieved using traditional drag-net techniques.

However, the move to using underwater video monitoring requires the use of labour-intensive manual processing techniques to analyse the video footage. Huge amounts of video data are now available for researchers to use, but analysis of the data requires a highly trained scientist to view the videos, making annotations of the animal populations and characteristics and then entering the annotations into a database. This imposes serious limitations in the amount of underwater data that can be processed. Manual processing of such large amounts of data has become impossible, leading to the need for an automated system capable of processing data at much greater speed.

DETECTING MARINE ANIMALS

A central requirement in many video analysis applications is the automated detection of specific marine animals depicted in the video footage. Being able to automatically locate, identify and track marine animals in video sequences would enable automated inventory of various marine species to be carried out (Dirk, Edgington and Koch 2004). The automated approach would also be able to identify and isolate certain footage of particular objects or events. This would greatly reduce the amount of footage that scientists have to analyse (Edgington et al 2003). In the future, automated tracking of individual animals by autonomous underwater vehicles will increase the amount of data that can be collected (Dirk, Edgington and Koch, 2004) and so event analysis will become more important. Given the diversity of animals of interest and the acknowledged limitations of current computer vision algorithms for detecting objects in natural settings, it is likely that many different approaches will be necessary to solve this problem for various marine applications.

To undertake this task using a computer, the analysis software will need to have some knowledge of the shape of the particular fish, and be able to match this shape to fish depicted within the image. Unlike humans, computers cannot easily separate objects within an image, or perceive objects as separate from the background. To make the task of identifying the fish easier for the computer to handle, image pre-processing techniques can be performed. Through the use of specific image enhancement techniques, the images may be taken to the stage where a computer can start to identify objects within that image. The ultimate goal of such a system is to autonomously detect an object, identify it and classify it within a database of known objects in real time. In order to achieve this, an accurate and efficient algorithm will need to be used.

Wilson (2003) addressed the problem of identifying very small mid-water organisms in video sequences and determining their species. They described a fully automated approach where the analysis was done in real time as the footage was recorded. This meant that identification of very small indistinguishable objects was extremely difficult, because it required significantly more processing than could be achieved in real-time.

Video analysis techniques for detecting marine animals are also being developed to provide automated animal tracking capabilities for Autonomous Underwater Vehicles (AUVs) and to assist scientists maneuvering Remotely Operated Vehicles (ROVs) to follow individual marine animals underwater (Rife and Rock 2001).

INPUT DATA

As a starting point for the development of new image analysis techniques for automatically detecting marine animals in video image sequences, we have constructed a prototype system for detecting and delineating fish shapes in video footage obtained from a local aquaculture facility. The images we used were provided by AQ1 Systems Pty Ltd. They are 8-bit grey scale Tagged Image File Format (TIFF) images with a size of 640 * 480 pixels and were produced by a dual-camera underwater video monitoring system, deployed within the fish cages of a local aquaculture firm. The system captured image sequences of the salmon in the cage as they swam past the cameras. Currently, these images are used to measure fish sizes using a manual analysis process. This process involves human input to locate fish within the image and select points at key locations on the fish from which the measurement of fish size can be made. One possible application for a fish detection system, such as the one we developed in this research project, would be to automate the fish sizing process. However, the work done here did not focus specifically on this application, but rather aimed to take the first step in the development of a marine animal detection system, capable of being used in a wide range of applications involving underwater video footage.

IMAGE PRE-PROCESSING

Contrast Enhancement

Due to the nature of the underwater environment, underwater images are of low contrast and so contrast enhancement is an important first step in processing these images. The contrast enhancement technique chosen was histogram equalization because it produces a higher contrast image without the need for user input and is thus capable of full automation. Histogram equalization was performed over all 256 grey levels of the image. The images before histogram equalization was applied contained mainly pixels with high value grey levels, resulting in a low contrasting bright image (Fig 1a and b). After

histogram equalization, these high value grey level pixels were expanded to more evenly cover the whole range of grey levels in the image (Fig 2a and b). The resulting images were of higher contrast containing more clearly defined fish.

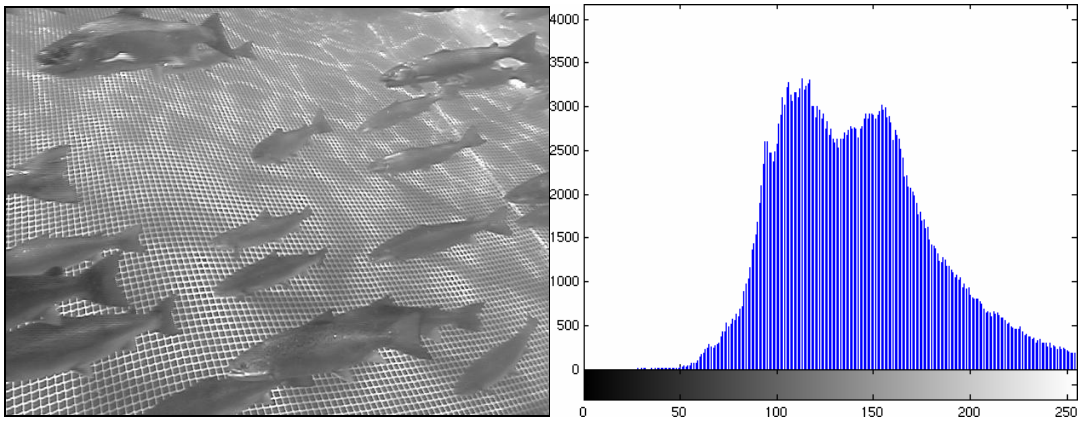


Image courtesy of AQ1 Systems Pty. Ltd,

Figure 1. Original Image and its Histogram.

(a) Original image (courtesy of AQ1 Systems). (b) Histogram of the original image.

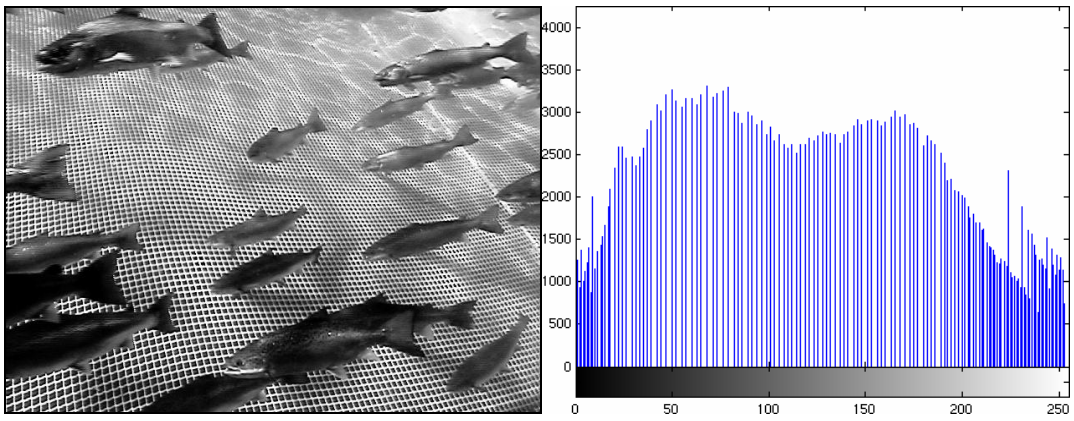


Image courtesy of AQ1 Systems Pty. Ltd,

Figure 2. Histogram Equalized Image and its Histogram.

(a) Image after histogram equalization. (b) Histogram of image after histogram equalization.

Background Removal

A major feature of the underwater images in this project is the netting used to enclose the fish cage. This netting represents the background of the image, with the fish being the foreground objects. Therefore background removal can be achieved by detecting the netting in the image and removing it, effectively distinguishing the fish from the background. The texture presented by the netting consists of a series of vertical and horizontal lines with high grey levels, in between which a square is formed with low value grey levels. Because edge detection operates by locating sharp transitions in grey levels, and this texture is a repeating pattern of transitions in grey levels, edge detection was used to detect the lines in the background netting. The edge detector chosen was the Sobel edge detector because it offered the ability to choose the direction of the edges being detected; either vertical, horizontal or both. Many fish within the images displayed some texture but, since most

of this texture consisted of horizontal lines, by using only the vertical edge detection capability of the Sobel detector, most of this texture was eliminated (Fig 3a).

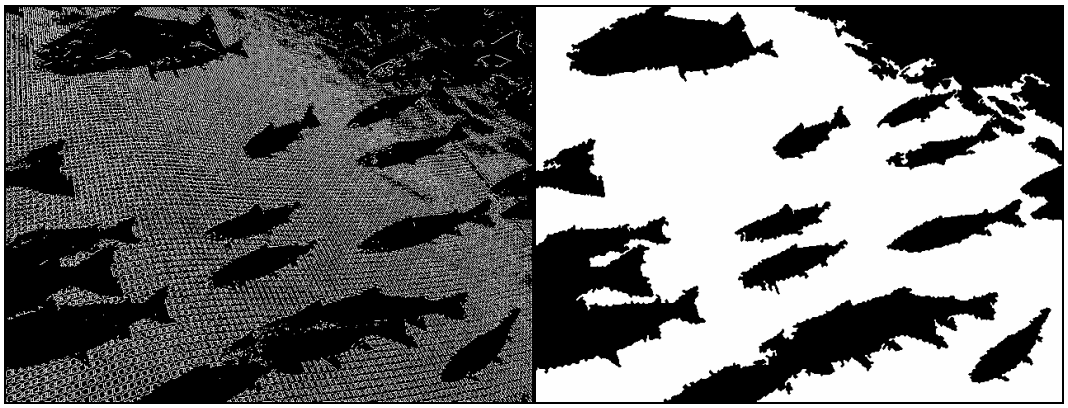


Figure 3. Image after Edge Detection and Image after Background Removal.

(a) Image produced when Sobel edge detector is passed over the histogam equalized image.

(b) Image after background netting texture has been removed.

IMAGE SEGMENTATION

Initial Segmentation

Once the texture had been detected (Fig 3a) the next step was to remove it from the image. Several morphological operations were applied to the image, including dilation of vertical and horizontal lines to fill in the parts of the image depicting the netting, removal of small gaps within the textured and non-textured regions by merging unconnected segments smaller than 300 pixels in area with their surrounding regions and finally eroding the background segment, using a 3 x 3 diamond structuring element. This process provided a more accurate and smoother segmentation between the background and the fish (Fig 3b).

The next stage of the segmentation process was to remove the background as well as any segments connected to the border of the image and the final stage involved labeling and obtaining statistics for each separate segment. The labeling process involved identifying separate segments within the binary image and labeling each segment accordingly (Fig 4a). The labelled image was then used to obtain statistics about each segment in the image. The statistics calculated were the Area (number of pixels in the segment), Centroid (coordinates of the centre of the segment), Major and Minor Axis Lengths (the major and minor axis lengths of the segment), the Orientation (the angle between the major axis of the segment and the horizontal) and the PixelList (the coordinates of all the pixels in the segment).

Segment Analysis

Once the segmentation process had been performed, the segments and any information obtained about each segment would usually be passed on to a higher level process. This process might then use the information for tasks such as object tracking, shape matching and measuring. One problem is that all segments are not guaranteed to represent individual fish, so the analysis needs to determine whether a particular segment represents a single fish, multiple fish or some other artifact of the segmentation process. The shape of a salmon can be approximated as an ellipse so performing an ellipse matching process on each segment offered a quick way for determining how likely it was that a segment represented a single fish.

An ellipse, with the same Centroid, Major and Minor Axis Lengths and Orientation as the segment was plotted over the segment (Fig 4b). The segments shown here illustrate the fact that ellipse fitting provides an accurate match for the location, size and orientation of each segment. There is still no guarantee that the segment actually represents a fish within the image but fitting ellipses to the segments allows tests to be performed that will estimate the accuracy of the match. Segments not

showing a reasonable match are unlikely to represent a single fish within the image and should not be included in the final output of the system.

To calculate the confidence that the segment represented a single fish in the image, the number of pixels outside the ellipse that belong to the segment was added to the number of pixels inside the segment that do not belong to the segment (ie the logical XOR of the segment and the ellipse interior). This value was then divided by the total number of pixels belonging to the segment that were located inside the plotted ellipse (the logical AND of the segment and the ellipse interior), to provide a ratio that represented how well the ellipse matched the segment. These ratios are displayed, in red, on the image in Fig 4b.

Only segments with a confidence value above a specified threshold were retained for further analysis. However, those with a low confidence threshold were post-processed, using morphological techniques, to see if some segments could be sub-divided further in cases where fish outlines touched but did not occlude each other. In some cases this process was able to extract a few further segments, representing single fish, from the image.

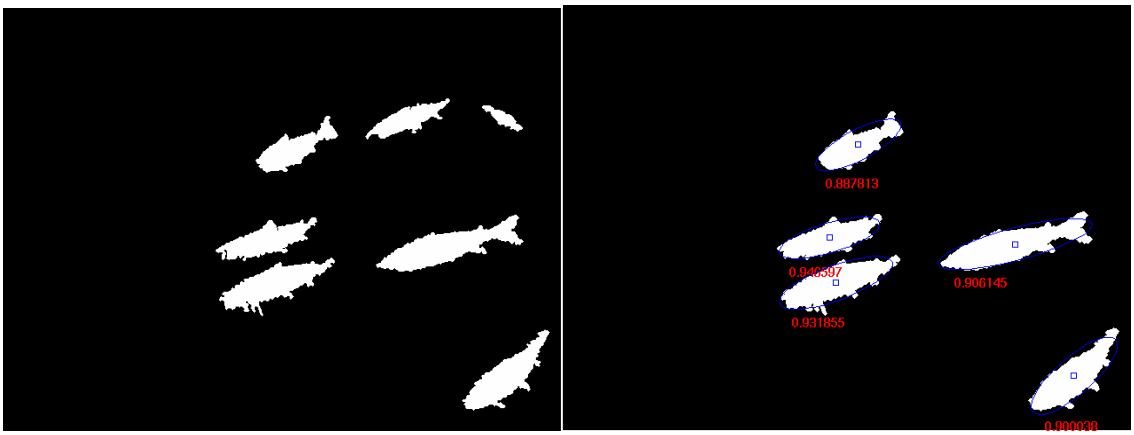


Figure 4. Image after Segmentation and Confidence Analysis.

(a) Labelled image after the segmentation has been performed.

(b) Labelled image after further analysis and confidence estimation has been performed.

SHAPE MATCHING

The Active Shape Model Technique

The first type of adaptable model used in image analysis was the active contour model or "snake", which comprised a spline curve that aligns itself with the edges found in an image. The curve parameters are weighted in such a way that certain bends are not possible, giving the spline characteristics such as stiffness and elasticity (Kass, Witzin and Terzopolous 1987). These work well for ideal objects but are unable to utilize any high-level knowledge about the shape of the object to be detected within the detection process.

Active shape models (ASM) are based on the similar principles to contour models, providing a flexible means to identify objects within an image. Each model is a collection of labelled points defining the boundaries of a specified shape. Using a training set of images, the computer extends the model by obtaining statistics of variations between the points. Using the mean values of each point, a model representing the average appearance in the training set is obtained. The main object deformations are also known, giving a number of modes of variation describing the ways the object in the training set tends to deform from the mean. This produces a point distribution model with a number of parameters that can be altered during a search to identify the object when it is deformed (Cootes, Taylor, Cooper and Graham 1992).

To improve the reliability and accuracy of active shape models, the use of statistical grey-level models may be incorporated into the ASM process. When defining a point, the assumption can be made that the grey levels around the point are going to be similar across various images. Therefore the grey-levels around the point can be modeled with a mean, a degree of variance and a number of modes of variation (Cootes and Taylor 1993). Incorporating statistical grey-level searching into the ASM technique increases the reliability and accuracy of the technique.

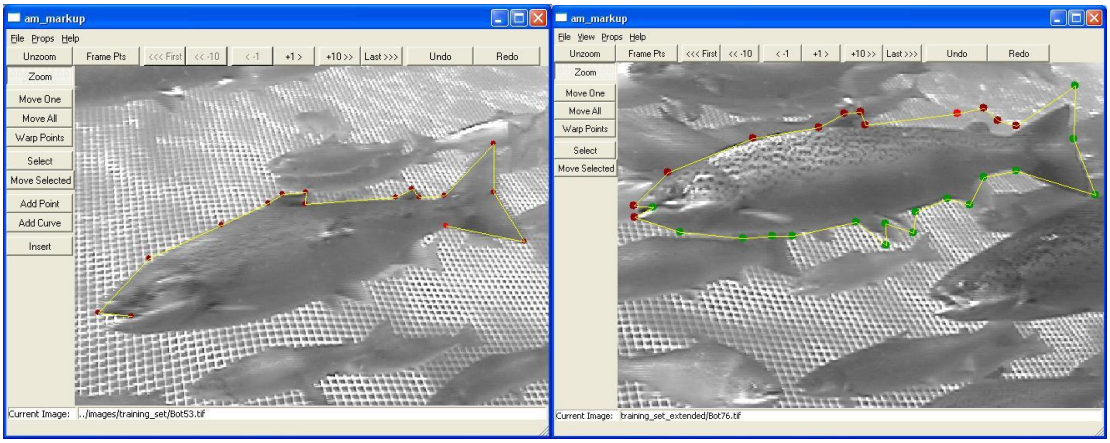
Active shape models have been developed mainly for face recognition (Cootes, Edwards and Taylor 1998) and medical imaging (Cootes, Taylor and Haslam 1994) but are potentially useful in a wide variety of applications. There has been a study into the suitability of active shape models for use with marine starfish (Kuksin 2001) but without clear-cut results.

The ASM Toolkit is a set of Active Shape Model tools developed by Cootes. It is a stand-alone application and includes features such as the ability to manually input a point distribution model and a range of grey-level models to help identify points in new images. Cootes developed the ASM toolkit for research purposes and it is freely available to download.

Creating a Salmon Shape Model

The aim of the video system being developed is to extract segments from the video images which are likely to contain single isolated fish and then attempt to match an active shape model, constructed specifically for the salmon species, to the segment. If a close match is obtained, this confirms that the segment represents a salmon and enables various characteristics of the individual fish to be measured, using the coordinates of the points representing the matching shape model.

To train the model, 20 example salmon shapes were chosen from the images provided by AQ1 Systems Pty. Ltd. They were selected so that the model would incorporate the typical variations in shape that a fish may present to the camera as it swam. In order to build a model from the training set, an ideal image needed to be traced to initially locate the key points on the fish outline (Fig 6a). Each defined point identifies the same feature of the fish. Each image in the training set is loaded and the points are moved into position manually for each fish sample (Fig 6b). Once the points around the fish have been set, a points file is generated for that instance, saving the coordinates of each point.



Images courtesy of AQ1 Systems Pty. Ltd.

Figure 6 Creating the Shape Model.

(a) Defining the point locations from the first sample.

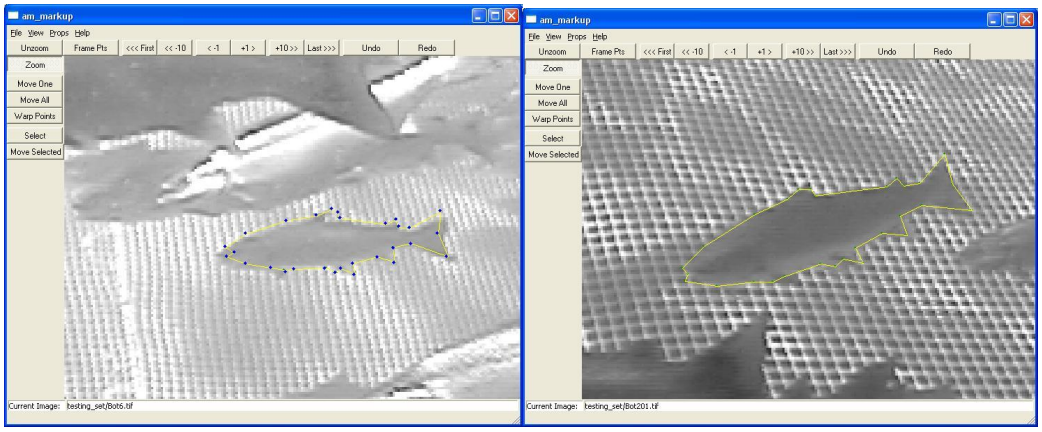
(b) Transforming the points around the sample in the training set.

Once the training was complete, the ASM toolkit read the data collected and built two models from the points and images in the training-set. The first was a statistical shape model, a multi-resolution model defining modes of variation between points, and the second was a statistical texture model, defining how the textures sampled at various points within the shape defined by the shape model can vary.

Matching Salmon Shapes to Segments

To search for fish within an image, the image enhancement, segmentation and ellipse analysis steps were undertaken first, producing a text file for each image, containing a list of candidate segments within the image together with details of the matching ellipse for each segment. This file is provided to an initial positioning program which takes the ASM model of salmon learned from the training data and positions a version of this model as closely as possible over the ellipse representing

each of the candidate segments in the image. The coordinates of the points in the model are transformed so that the centroid, major axis length and orientation of the model shape matches those of the ellipse representing the segment. (Fig 6a).



Images courtesy of AQ1 Systems Pty. Ltd.

Figure 6 Initial Positioning of a Shape Model and a Successful Match.

(a) Example of the model being automatically positioned.

(b) Example of a successfully matched shape model.

The ASM Toolkit then searched the space of possible deformations of the model (provided as a result of the learning process) to find a specific deformation which most closely matches the detailed shape of the segment and its detailed grey level pattern. The text file produced by this process contains model points that should closely outline the identified fish (Fig 6b). This file can then be loaded into another analysis program that would be able to make appropriate calculations relating to the fish depicted in the image, such as the length of the fish, its width, the direction in which it is swimming and other details of its shape.

TECHNIQUE EVALUATION

Test Images

Before the evaluation began, a test set of 60 images was extracted from the image set provide by AQ1 Systems Pty. Ltd. This was done by first removing all images used in the training process, then dividing the rest into three classes: good, medium and bad, based on criteria such as image contrast and the number of non-occluded fish depicted in the image. From these, a random set of images was chosen with 20 images from each class, giving a reasonable representation of real world performance. A manual analysis of all 60 images was then undertaken, during which 1122 individual salmon were identified. Of these, 125 were classified as "isolated fish", meaning that their shapes within the image did not occlude and were not occluded by any other fish shape. These included fish whose shapes just touched each other but did not overlap.

Performance Analysis

The testing process involved first applying just the initial segmentation and ellipse analysis stages of the process to the test set of images. Results from this evaluation (Table 1) show that 95 segments representing single fish were found. An attempt was then made to match each of these 95 segments to the active shape model. The result of each attempt was assessed by eye using a consistent evaluation criterion. Generally it was clear if a match occurred. However, there were some borderline cases. Two evaluation criteria were used to decide whether the model matched a sample or not. The first was how closely the model matched the visible features of the sample, such as the top and bottom fins and the body. In some cases the minor fins above and below did not match but this was not considered important as long as the body was a close match. The second criterion was how well the model matched the length of the sample. If the model did not accurately match the nose of the fish or the tail it was considered a unsuccessful match.

Number of images processed in the trial.		60
Total number of fish identified manually in these images.		1122
Number of isolated fish identified manually in these images.		125
Number of candidate segments produced by the segmentation stage.		95
Number of these candidate segments correctly matched by the shape model		54
Number of extra segments produced by morphological post-processing.		35
Number of these candidate segments correctly matched by the shape model.		11

Table 1. Performance Evaluation Results for Segmentation and Shape Matching.

When the shape matching process was applied, 54 of these segments were successfully matched to a shape model (Table 1). This was increased by another 11 as a result of morphological post-processing being applied to the rejected segments, revealing some further segments with a reasonable likelihood of representing single fish. Overall this meant that, of the 125 isolated fish identified manually within the 60 test images, 65 (ie 52%) were successfully matched using the salmon shape model. Unsuccessful matches were the result of a number of problems, including inaccurate initial segmentation, segments representing fish swimming towards or away from the camera and segments representing fish shadows. Also, another problem is introduced when segments representing two or more fish are shaped in such a way that it appears as though it is representing one fish. However, it is likely that substantial improvements can still be made by improving the fitness measures used to filter the output of the initial segmentation.

CONCLUSIONS

As can be seen from Table 1, only a small proportion of the total number of fish depicted in the 60 images were "isolated", and it is only these fish which can be potentially detected by this system. Of approximately 1122 fish identified by eye, only 125 (11.1%) were amenable to detection with the current system. This means that any system needing to comprehensively detect and measure or count all of the fish depicted within the image would need to use much more sophisticated image analysis techniques. This system may, however, constitute a useful starting point for such a system. On the other hand, given the fact that large numbers of images would be available, and the system in its final form would be fully automated, then sufficient numbers of images could be analysed so that even a 11% sample of the fish depicted in the images would be of sufficient size to provide statistically significant measurements, such as average length, of the whole fish population, provided that any statistical bias in the sampling process was carefully considered and accounted for.

Of the 125 isolated fish depicted within the 60 images, the overall analysis process, consisting of the enhancement, segmentation, ellipse analysis and shape model matching stages, successfully detected and shape-matched 65 segments, giving a successful detection/matching rate of 52 %. If we consider all of the 1122 fish depicted in the 60 images, the successful detection rate of the method is only 5.8%. This is a low rate but given the task being performed, running the process for an extended length of time should eventually obtain the statistics required. Also, it is important to note that some fish, such as tuna, do not swim in as close proximity as do the salmon, so this issue may be less significant for other marine species as it is for salmon. It may also be less significant in research aquaculture facilities and in natural marine environments, where the density of salmon (and other fish species) is generally less than that in commercial fish cages.

We will be extending the work reported in this paper in a number of ways. Firstly, the model training process will be studied further. Training will be carried out with a larger number and wider range of images in the training set, to see if performance can be improved. The current training set consisted of 20 samples each representing clearly isolated fish, and a preliminary investigation attempted firstly to decrease the size of the set to 15, resulting in a decrease in performance, and then to increase the size of the set to 40 with a greater diversity of samples, also resulting in a performance decrease. However, a more thorough and systematic investigation of training set size and composition may lead to an improved flexible shape model to use in future. Secondly, further improvements in the preprocessing of the images are likely to lead to an overall improvement in performance, given the low contrast and difficult background condition experienced in underwater imagery.

It has been stated that the number of isolated fish locations found should provide enough statistical data over a large quantity of images. However, to truly test this significance, further testing would be needed on a much larger number of images than the 60 used here, before we can provide conclusive results on the ability of an automated system to provide enough statistical data for meaningful measurements to be obtained.

The ultimate goal of this research is to develop a fully automated system for detecting and identifying fish and other marine animals in video image sequences. This work represents an initial stage in that process. Such a system, or more probably a number of different systems each one trained to detect a specific marine species from video footage, would be a very important contributor to the process of gathering the enormous quantities of data, at sufficiently small spatial and temporal scales, to support the large environmental decision support systems (EDSS) being deployed to assist in future management of our vital but increasingly threatened marine resources. For EDSS to have an impact on our decision-making capacity for managing our oceans and other marine environments effectively, accurate base data will be required. Underwater video is an essential technique for gathering a significant component of this base data and the technique described here represents an initial step in the automation of the analysis process for underwater video. Without substantial automation, the time-consuming nature of manual video analysis will prevent it from fulfilling its potential to supply much of the base data needed to support the models used by environmental decision support systems.

ACKNOWLEDGMENTS

We wish to thank The Verdant Group Pty. Ltd. and AQ1 Systems Pty. Ltd. for providing the images that have been used in this study. We also wish the thank Tony Gray, our technical manager, for the excellent technical support he has provided for this research, both in the provision of hardware and for the installation and maintenance of the MATLAB software package and its Image Processing Toolbox, with which some of the algorithm development work was undertaken.

REFERENCES

1. Ballard, D.H. and Brown, C.M. (1982) *Computer Vision*, Prentice Hall, New York.
2. Cootes, T.F., Taylor, C.J., Cooper, D.H. and Graham, J. (1992) Training Models of Shape from Sets of Examples, *Proceedings of the 1992 British Machine Vision Conference*.
3. Cootes, T.F. and Taylor, C.J. (1993) Active Shape Model Search using Local Grey-level Models: A Quantitative Evaluation *Proceedings of the 1993 British Machine Vision Conference*.
4. Cootes, T.F., Hill, A., Taylor, C.J. and Haslam, J. (1994) The Use of Active Shape Models for Locating Structures in Medical Images, *Image and Vision Computing*, 12, 6, 355-366.
5. Cootes, T.F., Taylor, C.J. and Lanitis, A. (1994) Active Shape Models: Evaluation of a Multi-resolution Method for Improving Image Search, *Proceedings of the 1994 British Machine Vision Conference.*
6. Cootes, T.F., Edwards, G.J. and Taylor, C.J. (1998) Active Appearance Models, *Proceedings of the Fifth European Conference on Machine Vision*, Freiburg, Germany.
7. Dirk, I., Edgington, D.R and Koch, C. (2004) Detection and Tracking of Objects in Underwater Video, *Proceedings of the IEEE International Conference of Computer Vision and Pattern Matching* (CVPR), Washington DC.
8. Edgington, D.R., Salamy, K.A., Risi, M., Sherlock, R.R., Walther, D. and Koch, C. (2003) Automated Event Detection in Underwater Video, *Proceedings of the MTS/IEEE Oceans Conference*, San Diego CA.
9. Kass, M., Witkin, A. and Terzopolous, D. (1987) Snakes: Active Contour Models, *International Journal of Computer Vision*, 1, 4, 321-331.
10. Kuksin, J. (2004) Detecting Starfish in Subsea Videos using Flexible Shape Models, Internal Report, University of Manchester, Manchester.
11. Rife, I. and Rock, S.M. (2001) A Pilot-Aid for ROV Based Tracking of Gelatinous Animals in the Midwater, *Proceedings of the Oceans 2001 Conference*, Honolulu.
12. Wilson, A. (2003) First steps towards autonomous recognition of Monterey Bay's most common mid-water organisms: Mining the ROV video database on behalf of the Automated Visual Event Detection (AVED) system. *Technical Report, Monterey Bay Aquarium Research Institute).*