Artificial Intelligence Techniques for Marine Video Analysis

Rob Fearn, Ray Williams, Mike Cameron-Jones - School of Computing, University of Tasmania.
Julian Harrington - Tasmanian Aquaculture and Fisheries Institute

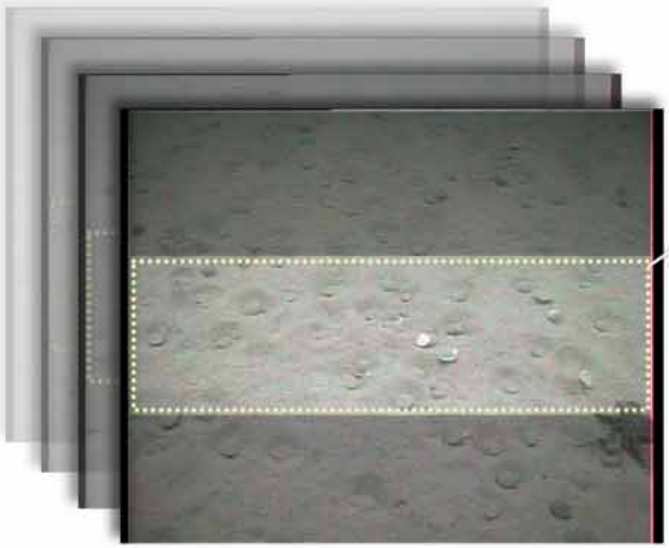
Research Overview

In recent years, the use of underwater video cameras has become a popular, low impact alternative to traditional techniques (such as drag netting) for gathering quantitative data sets suitable for statistical analysis of marine animals and habitat. Video offers a number of benefits, including the ability to measure spatial and temporal variation across transects, and is likely to increase in importance with the future deployment of a variety of Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs) for ocean observation.

The Tasmanian Aquaculture and Fisheries Institute (TAFI) have been aware of the potential benefits of capturing underwater video for more than five years. As a result they have made a concerted effort to capture underwater footage of commercial and non-commercial scallop beds during regular dredging and diving research expeditions. The potential benefits of this footage have yet to be realised due mainly to the time required to manually annotate the footage for statistical purposes

Our research focusses on the process of automatically annotating TAFI's scallop bed footage using artificial intelligence techniques. We propose a multi-layered approach that begins with an attention selection procedure to perform an initial sweep of each video frame in order to identify regions of possible interest. This process is performed using the University of Southern California's (USC's) "iLab Neuromorphic Visual Toolkit" (iNVT).

Each region of interest is then extracted as a separate sub-image and basic image processing techniques such as thresholding and normalisation are applied to enhance this sub-image. Candidate regions are then extracted from the sub-image and various features of each region measured. These features will enable the region to be classified using traditional Artificial Intelligence techniques, so that statistical data can be derived on the number and sizes of commercial scallops observed in the video footage.



The Process

Video frames are extracted one by one and analysed. The middle third of each image is the most useful as it offers a good compromise between clarity and maximum field of view.

iNVT is used as a coarse grained search to determine regions of possible interest, and these regions are then extracted as sub-images. Rules are applied to the sub-images in order to clean them up, leaving only clearly visible regions with specific characteristics.

The newly found regions can be used to generate a bounding ellipse within the original sub-image, enabling us to extract features that are appropriate for classification. Classification may also rely on the processed sub-images as well.