A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences

Mark R. Shortis[1], Mehdi Ravanbakskh[1], Faisal Shafait[2],
Euan S. Harvey[3], Ajmal Mian[2], James W. Seager[4],
Philip Culverhouse[5], Danelle Cline[6] and Duane Edgington[6]

[1] RMIT University, GPO Box 2476V, Melbourne, VIC 3001 Australia

[2] School of Computer Science and Software Engineering, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009 Australia

[3] School of Plant Biology, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009 Australia

[4] SeaGIS P/L, PO Box 1085, Bacchus Marsh, VIC 3340 Australia

[5] Centre for Robotics and Neural Systems, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA United Kingdom

[6] Monterey Bay Aquarium Research Institute, 7700 Sandholt Road, Moss Landing, CA 93933 USA

mark.shortis@rmit.edu.au

ABSTRACT

Underwater stereo-video measurement systems are used widely for counting and measuring fish in aquaculture, fisheries and conservation management. To determine population counts, spatial or temporal frequencies, and age or weight distributions, snout to fork length measurements are captured from the video sequences, most commonly using a point and click process by a human operator. Current research aims to automate the measurement and counting task in order to improve the efficiency of the process and expand the use of stereo-video systems within marine science. A fully automated process will require the detection and identification of candidates for measurement, followed by the snout to fork length measurement, as well as the counting and tracking of fish. This paper presents a review of the techniques used for the detection, identification, measurement, counting and tracking of fish in underwater stereo-video image sequences, including consideration of the changing body shape. The review will analyse the most commonly used approaches, leading to an evaluation of the techniques most likely to be a general solution to the complete process of detection, identification, measurement, counting and tracking.

Keywords: stereo-video, underwater, fish, counts, image matching and tracking, motion analysis and surface deformation, modelling of deformable surfaces, multi-view registration and integration

1. INTRODUCTION

The monitoring of fish for stock assessment in aquaculture, commercial fisheries and in the assessment of the effectiveness of biodiversity management strategies such as Marine Protected Areas and closed area management is essential for the economic and environmental management of fish populations. Video based

techniques for fishery independent and non-destructive sampling are now widely accepted. The advantages of using stereo-video for counting the numbers of fish, measuring their lengths and defining the sample area have been well demonstrated. However, the time lag and cost of processing video imagery decreases the cost effectiveness and uptake of this technology. Current research aims to minimise or completely eliminate the involvement of the human operator in the process of recognition and length measurement of fish recorded by underwater stereo-video surveys. The ultimate goal is to fully automate the recognition and measurement, in order to deal with the many thousands of hours of stereo-video footage that is routinely captured each year. Advances in automated techniques will substantially decrease the cost of processing and make the technology more accessible to a broad spectrum of end users.

Stereo-video systems have the advantages that the measurements are impartial and repeatable (Harvey et al., 2004), and calibration stability tests indicate very reliable levels of accuracy (Shortis et al., 2000). In recent years underwater stereo-video systems have been used in wild fish stock assessment (Watson et al., 2009) and in pilot studies to monitor length frequencies of fish in aquaculture cages (Harvey et al., 2003). Samples taken in aquaculture cages can approach 95% of the population and the measurement technique is non-invasive. Snout-to-tail and other body spans on the fish are measured from the video recordings and, using a length-weight regression (Pienaar and Thomson, 1969), the weights of the fish are estimated to an accuracy of a few per cent. Commercial systems such as VICASS (Shieh and Petrell, 1998) and the AQ1 AM100 (AQ1 Systems, 2013) are widely used in aquaculture to determine size distributions based on simple length and span measurements, and thereby deduce biomass from an estimated number of fish in the cage or tank.

The next significant advances in the technology of stereo-video monitoring of wild and aquaculture fish must be the automated candidate identification and body shape reconstruction of the fish in order to directly extract volumes, and potentially the identification of individual animals to validate sampling and monitor growth. Automated measurement will also enable monitoring of the condition of fish, at least to the extent of estimating the frequencies of superficial injuries and potentially identify infestation levels or secondary infections caused by parasites such as sea lice or skin and gill flukes.

2. REVIEW OF PUBLISHED TECHNIQUES

Many published papers and reports describe video based systems for automated or semi-automated processes incorporating some or all of detection, identification, measurement, tracking and counting of fish. Few systems claim to be fully automated, capable of all functions mentioned and producing high levels of success or accuracy. The detailed reviews provided in this section are selected primarily on the basis of automation of some or all of the processes within an uncontrolled environment.

An approach to the automated measurement of fish in video sequences is described by Tillet et al. (2000) in what is one of the first published reports on a successful, operational, automated measurement system. The technique is based on 3D Point Distribution Models (PDM), which are composed of landmark locations on the outline of the fish, in this case Atlantic Salmon held in a small aquaculture tank. The PDM specific to the species is developed from a small sample of fish defined by manual measurement of stereo-images, leading to a mean shape and an estimate of the variation based on principal component analysis. The PDM is independent of the scale and orientation, but is limited only to the silhouette of the fish and does not model the full body shape.

The automated measurement process within the stereo-images commences with an edge detection based on the Sobel operator. Initial segmentation using a reference image is not required in this case because the tank wall in the background has relatively uniform intensity. A heuristic threshold is used to eliminate noise and reduce non-fish edges on the basis that higher magnitudes indicated stronger edges. The PDM is then fitted to the edge image based on an energy minimisation function that compares the proximity, the edge gradient strength and the gradient direction to the normal to the PDM. A two-step approach is used that first holds the PDM fixed and searches along the normal to the PDM, the candidate points are then re-projected to detect points on the same edge, and then the PDM is fitted to the best candidate edge points. This process is iterated and the RMS residual compared to a heuristic threshold.

The technique must be described as semi-automated, as there is no candidate region identification process and the initial location of the PDM is an operator selection. However, a success rate of over 90% is reported, with failures caused by overlapping fish, large orientation variations and large size variations. The fish lengths derived from the PDM realise a mean error of 5% and a precision of 2.8%, based on comparisons with manual image measurements. Comparisons of the fish depth and accuracy tests against calliper measurements were less favourable.

More recently, Spampinato et al. (2008) reports on a system for detection, tracking and counting fish that is in continuous operation on a sub-tropical reef in Taiwan waters. Developed by the University of Catania and the University of Edinburgh, the system is based on single video cameras and does not use stereo pairs, but is subject to cluttered scenes and poor visibility. The processing steps are as follows:

- Two algorithms are combined with an 'and' operation, specifically to reduce false positives in cluttered scenes, to create the background image:
  - A moving average algorithm, based on an adaptive (weighted) update for pixels classified into the background by comparison to a threshold; and
  - An adaptive Gaussian mixture model that classifies background pixels according to likelihood that the pixel is explained by the mixture model;
- Image areas classified as non-background fish candidates are then subject to morphological filters, erosion and closing, dilation and opening, and a median filter;
- The count of fish in the frame is then determined by a connected component labelling algorithm;
- The fish are then tracked using a combination of two algorithms:
  - a feature vector based on centroid of the image area, the motion vector, the area of the fish and the orientation (angle of the principal axis), comparing changes from frame to frame against heuristic thresholds; and
  - colour matching in HSV space using a comparison of pixel values against the probability that the hue belongs to the histogram of the target object.

The system achieves, on average, a detection rate of 85%, a tracking accuracy of 90%, and a counting success rate of 85%.

Khanfar et al. (2010) and Charalampidis et al. (2012) describe a research project investigating recognition and tracking of fish in video sequences, developed at the University of New Orleans and tested at the Southeast Fisheries Science Centre of the NOAA National Marine Fisheries Service. The initial process applied to the

video broadly uses background subtraction, histogram analysis, segmentation, region growing, edge detection and boundary definition using an 'elastic' circle (Khanfar et al., 2010). More detailed descriptions of the essential steps in the process are as follows:

- create a background image by averaging 200 images without any fish present;
- on a pixel by pixel basis, divide the frame by the background;
- analyse the histogram of the resulting image on the basis that the largest peak is the background and smaller peaks are candidate fish;
- apply the identified thresholds and create a binary image, remove any regions less than a heuristic threshold, crop the original image to the rectangle containing the remaining candidate regions;
- extract the histogram of the cropped image, identify the largest peak as the background and the smaller peaks as candidate fish, set thresholds and extract the regions;
- apply dilation to remove noise, expand and merge adjacent regions, then apply erosion to restore the external boundaries of the regions;
- track, and count, the regions across a series of frames using the location, average intensity, length, width and area of each region;
- isolate each region and apply a Canny edge detector, remove any regions less than a heuristic threshold, identify and connect the closest end-points of the curves if the Euclidean distance is less than a heuristic threshold;
- use a decreasing radius 'elastic' circle to define the outer edges of the region; and
- update the background image based on the non-candidate regions.

The report indicates that the subsequent length measurement process is semi-automated, as an operator must identify the snout and tail fork of each fish on the left image from the stereo-pair, which is then automatically detected on the right image using image cross correlation (Thompson, 2013).

The approach to the image processing has been further refined to provide more sophistication and revised to place the emphasis on species recognition (Charalampidis et al, 2012). Some of the basic processing is retained in the four main steps of background subtraction, object detection, object tracking and feature extraction. The amended steps in the process are:

- compute the background image based on the median pixel intensity from multiple frames, subtract the background from the current image, identify non-background pixels using a sample variance from the median, update the background image;
- use morphological image processing (dilation, erosion, median filtering) to remove noise, eliminate small regions and join parts of candidate objects;
- candidate objects are tracked using the corner points of the bounding box and the velocity of the centre point, a Kalman filter is used to predict the location of each region in the next frame;
- Euclidean distances and a measure of the relative change in area are used to match regions from frame to frame, tested against heuristic thresholds;
- If a region matches to the distance but not the area, then it is assumed that a merge or a split of a candidate region has occurred;
- The outline of the fish object is extracted using the elastic circle approach;

- A nearest neighbour classifier (NNC) is used to match feature vectors between candidate fish objects and a training set of three species of fish based on a minimum, weighted Euclidean distance;
- Species specific features are extracted from the candidate region using vertical and horizontal Gabor filters (GF) to identify vertical and horizontal body stripe features associated with each species, if the NNC and GF classifications disagree then the candidate region is labelled as a non-fish;

Once a candidate fish is classified, the tracked region is labelled as the species or as a non-fish, except at the edges of the frame where the candidate fish enters or leaves the scene. Charalampidis et al. (2012) reports that, in more than 3000 frames used as a test sample, all regions were correctly identified.

3. ANALYSIS OF METHODOLOGIES

3.1 Detection

In context of this research, the detection methodologies comprise two steps: identification and subsequent delineation of the fish outline. Most of the existing work on fish detection from under-water image sequences employ either the differences between successive images (Lines et al. 2001; Spampinato et al. 2008) or histogram-thresholds (Khanfar et al. 2010) to segment a varying number of candidate regions in the frames. Whilst the former approaches appear to be autonomous, the latter method depends on the prior knowledge of the background portion of frames derived from a large number of frames with no fish present.

The identification step is followed by accurate delineation of the fish silhouette. Khanfar et al. (2010) used an edge detection algorithm to detect the initial fish outline followed by a shrinking circle to capture the outer boundaries more precisely. This method, however, can fail due to scene complexity of an uncontrolled marine environment and poor contrast of fish boundaries. Furthermore, the elastic curve employed can ultimately converge to the outer edge in which case the fish shape cannot be captured accurately. A model-based approach to fish detection can offer a solution to this problem.

Active contours (also called *snakes*) (Kass et al., 1987) are especially useful for delineating objects like fish bodies that are difficult to model with rigid geometric primitives. Moreover, active contours can be independent from edge gradients with flexibility in initialisation (Chan and Vese, 2001). The area-based active contour model (Chan and Vese, 2001) is based on the on techniques of curve evolution and level sets. While parametric active contours cannot handle automatic change of topology, level sets (Osher and Sethian 1988) allow for splitting and merging in a natural way and are thus more suited for detection of an unknown number of fish in a video image. Figure 1 shows the application of active contours on a low quality, low contrast image.

Furthermore, the incorporation of high-level prior knowledge about the shape of various fish species within the level set framework can provide a solution to such problems as poor contrast and occlusions caused by the uncontrolled and variable nature of marine environment. In the literature, there are some references of successful object extraction using shape information in the presence of image noise, clutter and occlusions (Cremers et al. 2006, Bailloeul et al. 2005, Cremers et al. 2003). The method has been used by Clausen et al. (2007) to segment overlapping fish in an aquaculture environment.
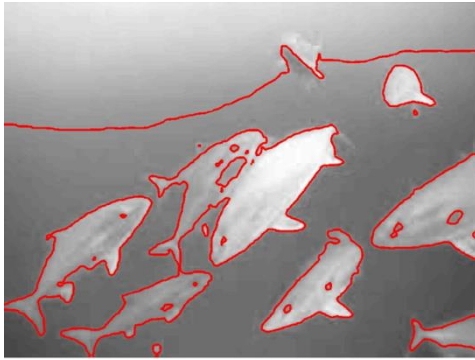
Figure 1. Active contours applied to a low quality, low contrast image.

Figure2. Haar classifier applied to Southern Bluefin Tuna tails.

3.2 Measurement

Underwater stereo systems are widely used to capture videos of swimming fish for subsequent measurement (Shortis et al. 2009). The simplest form of measurement is the fish snout to tail length which can be calculated if these two points can be identified in the stereo pair of images. This is done manually in most cases (Harvey et al. 2003), and a favourable orientation of the fish to the cameras and multiple measurements within the sequence of frames do improve the precision of the measurement (Harvey et al., 2002). One point to be noted is that fish are deformable and the Euclidean distance from snout to tail changes as the fish swims. If multiple measurements are taken as the fish swims, this will result in a sinusoid like signal and different methods must be used to estimate the length of the fish. To reduce the effect of swimming motion on length measurement, the operators manually inspect different frames in which a particular fish appears. Amongst these frames, measurements are made only on those frames in which the body of the fish appears to be straight.

Manual localization of fish snouts and tails, as well as selection of video frames in which these measurements are made, is labour intensive. Therefore, efforts are now directed towards automating this process. Once detected, the fish can be tracked using motion models and techniques such as Kalman filters (Charalampidis et al, 2012) that predict the location of the fish in the next frame and then matching within the limited neighbourhood of the predicted region corrects the prediction (Trucco and Plakas, 2006). By iteratively predicting and matching, a swimming fish can be tracked across a sequence of frames in a captured video. Accurate detection followed by tracking of swimming fish establishes correspondences between fish appearing in multiple frames, thereby allowing multiple measurements of the same fish to be taken. To automate an accurate measurement process, automatic detection of the fish snout and tail is required. Techniques such as bounding box dimensions and region extents (Charalampidis et al, 2012) may be limited by systematic errors in the distance estimation.

Template matching is one of the primitive methods that can be employed to accurately locate fish snout and tail in video frames. First, individual templates (usually rectangular image regions) centered on the snout and tail mid-points are extracted from sample videos. Then an efficient template matching strategy is employed to locate these templates in target videos. A certain degree of robustness against illumination changes can be achieved by using correlation between template and image regions of interest, instead of taking their absolute

differences (Mahmood and Khan, 2012). However, template based methods fail in the presence of perspective or affine transformations, requiring either use of multiple templates that capture appearance variations from different viewing angles, or using more sophisticated matching techniques that are invariant to affine or perspective transformations. These enhancements also significantly increase computational complexity of the template matching step.

A better way of locating snout and tail is to use Haar-like features in a boosted classifier setup (Viola and Jones, 2001) that has shown high object detection accuracy, besides being able to operate in real-time. The method is in wide use for face detection, even employed by low-cost consumer cameras for real-time face-priority focus. To train the classifier, manually cropped images of the target object (snout or tail) are used so that the classifier can learn which features (among a set of possibly thousands of features) can locate the target with high accuracy. These features, once learned, are then used to construct the object classifier that can locate the presence of the object in cluttered scenes. Due to their high detection speed and ability to perform a scale-space search, Haar classifiers are a promising candidate for locating snout and tail of fish in underwater images. The results of independent detection of the snout and tail using Haar detectors can be further improved using relationships between the detected snouts and tails, for instance by constraining the search for tail detection based on the results of snout detection and vice versa. Figure 2 shows an example of a Haar classifier used to identify the tails of Southern Bluefin Tuna during a transfer between cages.

For obtaining accurate measurements, it is quite important to precisely locate the reference points (e.g. tip of the snout and the valley point of the tail). Once the Haar detector gives the location of snout and tail, pin-pointing reference points can be done by matching templates of snout and tail having the reference point explicitly marked.
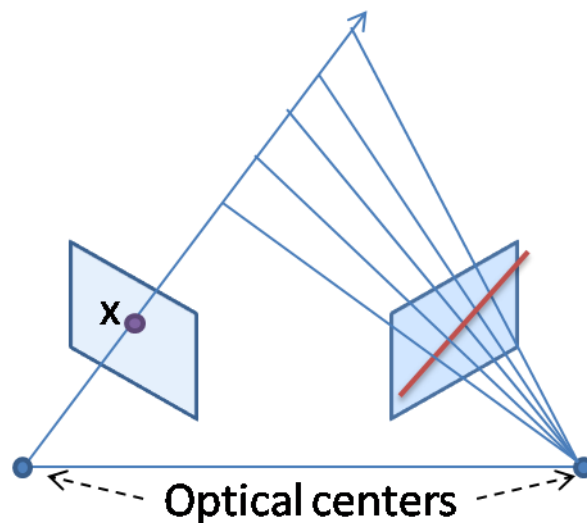

Figure 3. An illustration showing stereo-image geometry and the epipolar line in the right image of the stereo pair.

The last step is to establish correspondences between the stereo image pairs such that the snout and tail of each fish is correctly associated with the corresponding image points in the stereo pair. Corresponding points in stereo image pairs lie on epipolar lines (see Figure 3). Therefore, the search space for correspondences is

one-dimensional. Automatically establishing correspondences in stereo image pairs is a well-studied problem in photogrammetry and computer vision, both in a dense and a sparse fashion. Dense methods for correspondance establishment usually rely on block matching to compute disparity map (Scharstein and Szeliski, 2002). Sparse methods, on the other hand, first identify key points in the images and then capture texture around those locations in a feature descriptor such that correspondences can be established between key-points extracted from the stereo image pairs (Lowe, 2004). Use of a Haar classifier followed by template matching already provides the locations to be matched in the stereo pair. Hence, the search for the corresponding snout / tail reference point in one image can be restricted to the epipolar line corresponding to the snout / tail reference point in the second image.

Finally, using epipolar geometry and intersections, 3D locations of the snout/tail of a fish are computed. The length measurement of the fish resolves to a simple 3D Euclidean distance between the snout and tail. However, due to the swimming motion of fish, measurements taken from a single frame are not reliable. Hence, measurements should be made in all frames where the complete fish is visible. The Euclidean distances make a sinusoid-like pattern for swimming fish when plotted against time. The peak of the sinusoid is logically chosen as the final measurement of the fish length, as this corresponds to the straight body length.

3.3 Mathematical Modelling

If a swimming fish is tracked across a sequence of images then the change in shape of the body must be described by a deformable model. The aim of the deformable model of the fish is to produce a time-dependent 3D description of the surface model. The length of the fish can then be derived from all instances of the model using the curved profile of the body shape.

However, the aim of extracting the fish length is to use the correlation with weight to determine biomass. If an accurate surface model can be created and refined across many images in a video sequence, the biomass can be estimated directly from the volume contained within the surface. The surface model at any instance can be created using the delineated outline of the fish from active contours, followed by epipolar searching and image matching (Gruen and Baltsavias, 1988), or a semi-global matching (Hirschmüller, 2005) approach within the defined region of the image.

Development of a suitable deformable model is a major challenge because there is relatively little prior research in marine science that can be used directly as a basis for the development. However there are several possible approaches from other disciplines that require further investigation. Common to any approach will be a species-specific base fish model. When a fish that is not currently being tracked is identified, the base fish model is modified to suit the dimensions of the new individual. From this point the fish model is refined as the fish is tracked across multiple frames in the video sequence.

One approach to be considered is the incremental construction of 3D meshes (Guibas et al., 1992), with deformation modelled by least squares fitting of registration points followed by re-sampling over the 3D mesh. A similar approach could use matching over registration points with a thin-plate spline model (Bookstein, 1989) allowing interpolation via a parameterised surface. An extension of this approach using both registration points and intensity-based matching (Johnson and Christensen, 2002) may prove useful. Other 3D deformable models originating in medical imaging need to be considered. Approaches that model 3D surfaces using decomposition over a set of basis functions (Noblet et al., 2005) are of particular interest.

Successful development of a deformable template model is a key to the success of this approach. The deformable model is required to build an accurate representation of the fish over multiple video frames as the fish swims, and it is then required to successfully recognise the same individual presenting to the camera system. This latter requirement is a highly desirable but challenging requirement necessary to identify fish in either the same camera system deployment to avoid re-counting, or in camera deployments over time to monitor growth. The identification of individual fish will be based on the deformable model, draped with the surface images to identify individuals through size, distinctive markings and parasites. The automated measurement process will also enable monitoring of the condition of fish which is important in both wild environments and aquaculture.

4. PROPOSED APPROACH

A synthesis of the current approaches to the recognition and measurement of fish in video sequences and the analysis of the methodologies leads to a likely broad scenario of the following steps:

- establish a training set of images of species of interest;
- capture many images without fish in the scene and compute an average or median to create the background image;
- use differences between the current frame and the background image, and either histogram thresholds or variance from the median, to identify candidate regions;
- apply morphological operators, connected component labelling and heuristic geometric tests to remove noise and refine the candidate regions;
- employ edge detection and geometric algorithms or active contours to delineate the fish in each of the candidate regions;
- simultaneously validate the candidate regions as fish or 'not fish of interest', and for the fish classify the species using a combination of multiple techniques such as:
  - o a Haar detector, based on the training set, to identify the snout, tail and potentially other distinctive features of the fish,
  - o a Nearest Neighbour Classifier to match feature vectors to the training set, and/or
  - o Gabor filters to identify body markings on the fish;
- count the fish based on the validated regions;
- dependent on image quality, establish a surface shape of the body of the fish using an epipolar search and stereo-matching or a semi-global matching approach;
- compute the biomass based on the Haar detector length, or the volume within the surface model, of the fish;
- predict the trajectory using a Kalman filter and track the fish using a combination of feature vectors and colour space matching;
- identify merged and split regions using frame to frame geometric differences for the candidate regions, retaining the non-fish candidate merged regions in the tracking set;
- repeat the surface shape measurement and reconcile with a deformable template model to refine the accuracy and reliability of the surface model generation;
- update the background image using the non-fish areas of the image, repeat from the beginning.

The steps outlined here are presented in a simplistic linear fashion, however in reality the detection, identification and measurement will be an iterative process. As fish and fish candidates are tracked through a sequence, classification as a species of fish at any point in the trajectory will result in re-processing of the images in the sequence to extract a revised estimate of the fish count as well as further refinement of the deformable template model of each fish in the sequence. The optimisation of the model will provide the maximum confidence in the biomass calculation and the evaluated condition of each individual.

5. CONCLUSIONS

This paper has presented a review of the techniques used for the detection, identification, measurement, tracking and counting of fish in underwater stereo-video image sequences, including consideration of the changing body shape. The review has identified the common approaches and their shortcomings, leading to an evaluation of the techniques most likely to be a general solution to the identification and measurement task. Finally, the paper has proposed an approach that is likely to provide the highest possible success rate in automation of the process and will be the subject of future research and development.

REFERENCES

AQ1 Systems, 2013. http://www.aq1systems.com/products (accessed March 14, 2013).

Bailloeul, T., Prinet, V., Serra, B. and Marthon, P., 2005. Spatio-temporal prior shape constraint for level set segmentation. *Energy Minimization Methods in Computer Vision and Pattern Recognition* (EMMCVPR), LNCS (3757), pp. 503-519.

Bookstein, F. L., 1989. Principle warps: thin-plate splines and decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(6): 567-585.

Chan, T. F. and Vese, L. A., 2001. Active contours without edges. *IEEE Trans. on Image Processing*, 10(2), pp. 266-277.

Charalampidis, D., Gundam, M., Joginipelly, A., Quinteros, M., Ioup, G., Ioup, J., Yoerger, E. J., and Thompson, C. H., 2012. Feature analysis for classification of fish in underwater video. Final Report, LA Board of Regents Contract NASA(2011)-STENNIS-02, December 2012. 19 pages.

Clausen, S., Greiner, K., Andersen, O., Lie, K., Schulerud, H., and Kavli, T., 2007. Automatic segmentation of overlapping fish using shape priors. *15th Scandinavian Conference on Image Analysis*, Aalborg, Denmark, LNCS (4522), pp. 11-20.

Cremers, D., Sochen, N. and Schnoerr, C., 2003: *Towards recognition*-based variational segmentation using shape priors and dynamic labeling. *4th International Conference on Scale Space Theories in Computer Vision*, LNCS (2695), pp. 388-400.

Cremers, D., Osher, S. and Soatto, S., 2006. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International Journal of Computer Vision*, 63(3), pp. 335-351.

Gruen, A. and Baltsavias, E. P. 1988. Geometrically constrained multiphoto matching. *Photogrammetric Engineering and Remote Sensing*, 54(5): 633 - 641.

Guibas, L.J., Knuth, D.E., and Sharir, M., 1992. Randomized incremental construction of Delaunay and Voronoi diagrams. *Algorithmica* 7: 381-413.

Harvey, E. S., Shortis, M. R., Stadler, M. and Cappo, M., 2002. A comparison of the accuracy of measurements from single and stereo-video systems. Marine Technology Society Journal, 36(2): 38-49.

Harvey, E. S., Cappo, M., Shortis, M. R., Robson, S., Buchanan, J. and Speare, P., 2003. The accuracy and precision of underwater measurements of length and maximum body depth of southern bluefin tuna (Thunnus maccoyii) with a stereo-video camera system. *Fisheries Research*, 63: 315-326.

Harvey, E. S., Fletcher, D., Shortis M. R. and Kendrick, G., 2004. A comparison of underwater visual distance estimates made by SCUBA divers and a stereo-video system: Implications for underwater visual census of reef fish abundance. *Marine and Freshwater Research*, 55(6): 573-580.

Hirschmüller, H., 2005. Accurate and efficient stereo processing by Semi Global Matching and Mutual Information. *IEEE Conference for Computer Vision and Pattern Recognition*, 2, 807–814. San Diego, CA, USA.

Johnson, H. J., and Christensen, G. E., 2002. Consistent landmark and intensity-based image registration. *IEEE Transactions on Medical Imaging*. 20(5): 450-461.

Kass, M., Witkin, A., and Terzopoulos, D., 1987. Snakes: Active contour models. *International Journal of Computer Vision*. 1(4): 321-331.

Khanfar, H., Charalampidis, D., Ioup, G., Ioup, J. and Thompson, C. H., 2010. Automated recognition and tracking of fish in underwater video. Final Report, LA Board of Regents Contract NASA(2008)-STENNIS-08, 4 June 2010. 40 pages.

Lines, J.A., Tillett, R.D., Ross, L.G., Chan, D., Hockaday, S. and McFarlane, N.J.B., 2001. An automated image-based system for estimating the mass of free-swimming fish. *Journal of Computers and Electronics in Agriculture*, 31(2): 151–168.

Lowe, D. G., 2004. Distinctive image features from scale-invariant key points. *International Journal of Computer Vision* 60(2): 91-110.

Mahmood, A., Khan, S., 2012. Correlation-coefficient based fast template matching through partial elimination. *IEEE Transactions on Image Processing* 21(4): 2099-2108.

Noblet, V., Heinrich, C., Heitz, F., and Armspach, J., 2005. 3-D deformable image registration: A topology preservation scheme based on hierarchical deformation models and interval analysis optimisation. *Transactions on Image Processing*. 14(5): 553-566.

Osher, S. and Sethian, J.A., 1988. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1): 12-49.

Pienaar, L.V. and Thomson, J. A., 1969. Allometric weight-length regression model. *Journal of the Fisheries Research Board of Canada,* 26:123-131.

Scharstein, D., Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision* 47(1-3): 7-42.

Shieh, A. C. R. and Petrell, R. J. 1998. Measurement of fish size in Atlantic salmon (salmo salar l.) cages using stereographic video techniques. *Aquacultural Engineering*, 17(1): 29-43.

Shortis, M. R., Miller, S., Harvey, E. S. and Robson, S. 2000. An analysis of the calibration stability and measurement accuracy of an underwater stereo-video system used for shellfish surveys. *Geomatics Research Australasia*, 73: 1-24.

Shortis, M. R., Harvey, E. S. and Abdo, D. A., 2009. A review of underwater stereo-image measurement for marine biology and ecology applications. In *Oceanography and Marine Biology: An Annual Review*, Volume 47, Gibson, R. N., Atkinson, R. J. A. and Gordon, J. D. M. (Editors). CRC Press, Boca Raton FL, USA. ISBN 978-1-4200-9421-3. 342 pages.

Spampinato, C., Chen-Burger, Y.-H. Nadarajan, G. and Fisher, B., 2008. Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos. Proc. of *3rd Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, Vol. 2, pp. 514-519.

Thompson, C. H., 2013. Personal communication, February 2013.

Tillett, R., McFarlane, N. and Lines, J. 2000. Estimating dimensions of free-swimming fish using 3D point distribution models. *Computer Vision and Image Understanding*, 79, 123-141.

Trucco, E. and Plakas, K., 2006. Video Tracking: A Concise Survey, *IEEE Journal of Oceanic Engineering*, 31(2): 520–529.

Viola, P. and Jones, M., 2001. Rapid Object Detection using a Boosted Cascade of Simple Features, IEEE Int. Conf. on *Computer Vision and Pattern Recognition* (CVPR), Kauai, HI, USA, pp. 511-518.

Watson, D.L., Anderson, M.J., Kendrick, G.A., Nardi. K. and Harvey, E.S., 2009. Effects of protection from fishing on the lengths of targeted and non targeted fish species at the Houtman Abrolhos Islands, Western Australia. *Marine Ecology Progress Series*, 384: 241-249.

Willis, T. J. and Babcock, R. C. , 2000. A baited underwater video system for the determination of relative density of carnivorous reef fish. *Marine and Freshwater Research*, 51, 755-763.