Automated measurements of fish within a trawl using stereo images from a Camera-Trawl device (CamTrawl)

Authors: Kresimir Williams[1*], Nathan Lauffenburger[1], Meng-Che Chuang[2], and Jenq-Neng Hwang[2], and Rick Towler[1]

1- Alaska Fisheries Science Center

National Marine Fisheries Service

7600 Sand Point Way NE

Seattle, WA 98115, USA

2- Electrical Engineering Department

University of Washington,

Box 352500,

Seattle, WA 98195, USA

* corresponding author

Keywords: underwater stereo-camera, automated image processing, camera based length measurement, background subtraction

Abstract

We present a method to automatically measure fish from images taken using a stereo-camera system installed in a large trawl (CamTrawl). Different visibility and fish density conditions were evaluated to establish accuracy and precision of image-based length estimates when compared with physical length measurements. The automated image-based length estimates compared well with the trawl catch values and were comparable with manual image processing in good visibility conditions. Greatest agreement with trawl catch occurred when fish were within 20° of fully lateral presentation to the cameras, and within 150 cm of the cameras. High turbidity caused substantial over- and underestimates of length composition, and a greater number of incompletely extracted fish outlines. Multiple estimates of individual fish lengths showed a mean coefficient of variation (CV) of 3% in good visibility conditions. The agreement between manual and automated fish measurement estimates were not correlated with fish length or range from the camera ($r^2 = 0 – 0.08$). Implementation of these methods can result in a large increase in survey efficiency, given the effort required to process the trawl catch.

1. Introduction

Cameras are an increasingly important tool for surveying marine living resources, and provide a non-extractive method of estimating fish abundance and demographic composition (Mallet and Pelletier., 2014). Image–based sampling provides many advantages compared to traditional catch sampling methods used on trawl surveys including higher spatial resolution, non-lethal observations, and removing the physical requirements for scientific sampling. A primary limitation of these methods, however, is the human and time resources required to extract accurate data from images. Two technological advancements increasingly applied to underwater visual surveys are making image-based sampling a viable method for survey work: 1) the use of stereo-imagery for precise measurement (Harvey et al., 2003; Gibson et al., 2009), and 2) the development of automated image processing techniques that can substantially reduce processing effort (Edgington et al., 2006; Shortis et al., 2013).

Automated image processing is a rapidly growing research field with wide ranging applications such as security monitoring, automated vehicle driving, and medical imaging (Sonka et al., 2014). A recent proliferation of computer algorithms for image processing have allowed for the development of automated software routines to reduce, and possibly remove, the cost of extracting scientific information from images (MacLeod et al, 2010). Underwater imagery has special challenges for processing, such as contrast loss with reduced water clarity, and greater absorption of higher frequencies in the visual spectrum reducing color information for color imaging Singh et al., 2015). The development of analytical methods is especially critical for practical implementation of camera systems to supplement or replace traditional survey methods, as typically there is substantial additional labor costs needed for image analysis.

The use of stereo cameras for precise manual length estimation has been well established in underwater image-based surveying with a variety of deployment platforms (Dunbrack, 2006; Watson et al., 2010; Rosen et al., 2013). Processing stereo camera images has been partially automated for determining fish lengths and mass estimations (Costa et al., 2006, Lines et al., 2001). Automated routines for underwater stereo-image processing can include a range of techniques including target location, segmentation (Tillett et al., 2000), shape and feature identification, and length measurements (Costa et al., 2006), but most automated algorithms that have been implemented are preliminary or experimental, and have not been incorporated into routine survey operations.

A critical component for the use of image-derived length data is quantifying the uncertainty associated with these estimates, and the conditions that influence that uncertainty (Harvey et al., 2010a; Williams et al., 2010b). Several factors that need more study to quantify uncertainty associated with length estimation are image resolution, influence of the camera to target range, stereo calibration accuracy and camera baseline separation, fish target density, and water clarity. It is important to understand the accuracy and reliability of the automated algorithm used in fish length estimation in the context of these factors.

An underwater stereo camera system (CamTrawl) has been developed by scientists at the National Oceanic and Atmospheric Administration's (NOAA) Alaska Fisheries Science Center (AFSC) for image-based fish sampling of midwater fish species as a compliment to standard trawl catch processing during acoustic-trawl surveys. The camera system is attached to the aft portion of a midwater trawl and captures images as the fishes pass through the net into the codend (Williams et al., 2010a). These data provide a unique opportunity to assess the precision and accuracy of image-based sampling, as image-based estimates can be directly compared with physical catch data on a trawl by trawl basis.

The routine use of CamTrawl during trawling operations results in millions of image-pairs. Time-consuming manual image-based length estimates do not provide sufficient efficiency advantages over direct physical measurements from the catch. While the development of an automated image analysis process also has substantial initial costs in terms of required personnel expertise and development time, the gains of automation can be expected to be realized in the long term, potentially as soon as a few sample collection seasons. Thus, the usefulness and practicality of CamTrawl is dependent on the successful development of automated image-processing software. This paper presents a field-ready automated technique with established precision and accuracy for the fish length estimates. It can be used to augment or replace current survey sampling practices, resulting in greater survey efficiency and abundance estimation accuracy.

2. Materials and Methods

2.1.CamTrawl hardware and image acquisition

The camera system consisted of a pair of solid state industrial grade high-resolution, high sensitivity machine vision cameras (JAI RM4200 GE) with an electronic global shutter. The cameras were mounted 28 cm apart on a rigid frame attached to the side of a midwater trawl near the codend, and captured lateral images of pollock as they passed though the trawl toward the codend. Images were recorded at depth using a small form factor computer. Illumination was achieved using light-emitting (LED) strobes. Further details of the camera system can be found in Williams et al. (2010a). Data format consisted of 2048 × 2048 pixel 8-bit monochrome images compressed using the jpeg standard. Images were collected at a rate of 4 Hz, with an exposure of 1.5 ms. The imaging chamber where the camera was attached had a square cross-section formed by rigid crossbars mounted to the outside of the trawl, with each side measuring approximately 1.5 m (Fig. 1). The trawl panel opposite of the camera was covered with a black fabric to provide a uniform background to aid in automated processing of the image data.

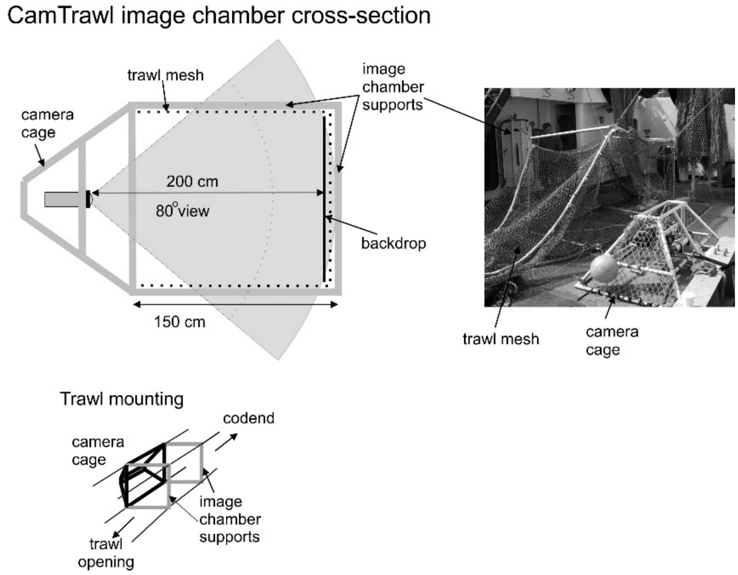


Figure 1. CamTrawl system description. Upper panel shows the cross section of the CamTrawl imaging chamber. The camera was attached to a 4-seam midwater trawl near the codend. The camera cage was attached to the port panel, with the trawl mesh removed from the panel to allow viewing of the passing fish. Rigid poles were mounted to the outside of the trawl to form a rectangular image chamber. A black tarp was mounted to the trawl panel opposite of the camera to provide a uniform background for image analysis.

2.2. Image data and analysis overview

Automated target detection and measurement from stereo optical images requires several steps. The initial step consists of target segmentation, where targets of interest are separated from the background. If fish targets are found in the left camera, the right image is segmented and corresponding individuals in the synchronous image pairs are matched using a process termed stereo-correspondence (Shapiro and Stockman, 2001). Fish lengths are subsequently estimated using stereo-triangulation, a process of reconstructing 3D positions of objects using the corresponding image points. All computations were performed using scripts written in the Matlab® computing language and were run on a standard desktop computer (Intel® Core i7 64-bit processor).

Three image data sets were analyzed from three trawl hauls taken during a single survey, each representing a unique set of challenges for automated length estimation (Fig. 2). The first set consisted of low densities of large pollock encountered in very clear water. The second set featured two distinct size modes of pollock encountered separately during the trawl haul, with the smaller fish occurring in high density and good visibility throughout the haul. The last set contained a high density of large fish imaged under poor visibility conditions due to high turbidity and a high density of krill in the images.

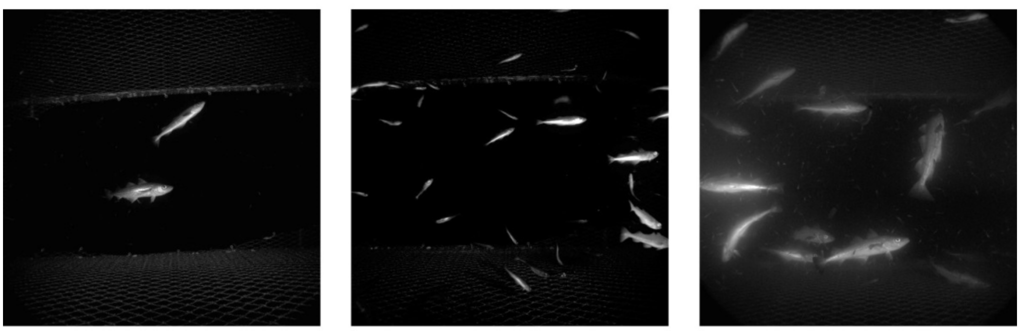


Figure 2. Data sets used for automated fish measurement. The data set 1 contained adult (30 – 40 cm) walleye pollock in low density and good visibility, the data set 2 contained a mixture of juvenile (<20 cm) and adult pollock with good visibility, and data set 3 contained high density of adult pollock in poor visibility with krill.

The catches consisted almost entirely of walleye pollock (*Gadus chalcogrammus*) by number in the first and third haul (99.0 and 99.5%, respectively). A significant proportion (60.6%) of the second haul consisted of eulachon (*Thaleichthys pacificus*). The eulachon were caught while the trawl was in deeper water based on the image data, so these data were excluded from the analysis, leaving only image data containing > 95% pollock targets.

2.3. Physical catch sampling

Trawl catches from the three trawl hauls used in this study were sorted to species and ~300 pollock from each haul were measured for length to the nearest 1.0 cm. When juvenile pollock co-occurred with adult pollock in the catch, juveniles were sampled separately and then merged with adult measurements to increase the precision of the length frequency estimate (Honkalehto and McCarthy, 2015).

2.4. Stereo calibration

Stereo analysis of image data requires that any distortions imposed by the camera optics be corrected and that the inter-camera geometry is known. To estimate these parameters, a set of 20 image pairs of a checkerboard pattern of known grid-cell dimensions was collected underwater and analyzed using a camera calibration software toolbox written in the Matlab® computing language (Bouguet, 2008) as described by in Williams et al., 2010b. First, the lens distortion coefficients are estimated for each camera, allowing the pixel coordinates for the checkerboard intersection points to be corrected to correspond to a rectilinear lens (undistorted) view. The corrected left and right image pixel coordinates for intersection points (with known physical inter-point distances on the checkerboard) are then used to iteratively solve for the translation (offset in space, or right camera relative to left) and rotation matrices (difference between left and right camera "aim" or central optical axes). The individual camera distortion coefficients and the translation and rotation matrices are then used to derive the 3D position coordinates of any fish targets simultaneously viewed by both cameras.

2.5. Segmentation

As images are collected inside the trawl during fishing, there are many images that do not contain fish and are relatively static from frame to frame, such as webbing and other trawl structural elements. These images are not needed in the subsequent analysis. These components are jointly considered the image background and are masked out of all images using a process called background subtraction. The components remaining after background masking are collectively referred to as the image foreground, containing the targets of interest. For this analysis step, images were down-sampled to a resolution of 512×512 pixels, greatly enhancing performance without adversely affecting the analysis outcome based on a comparison of results of analyses conducted at resolutions of 2048×2048 (original capture resolution), 1024×1024, and 512×512. At the latter resolution, most adult fish had pixel lengths of 80-100, while juveniles were between 15 and 30 pixels in length. Background subtraction was achieved using a median-based model (McFarlane and Schofield, 1995), which continually updates the background "image" from the static elements in the image as the analysis proceeds sequentially through the image frames (Fig 3b). A pixel intensity threshold was applied to the difference between the image being processed and the background image, resulting in a binary background "mask". This mask identifying foreground objects is analyzed for contiguous pixel regions using a four-sided connected components algorithm (Haralick, 1981), with each region receiving a separate label. Labeled areas are filtered to remove small objects that are not likely to be targets of interest, such as small organisms (< 5 cm) and trawl netting (Fig. 3c).

The basic dimensions of the remaining regions were estimated by conducting a regression analysis to rotate the data by the object major axis. Object length and height then corresponded to the pixel range along the horizontal and vertical dimensions. Targets whose aspect ratio (object length / object height) did not fall within a range of 3 and 7.5 were not used. In addition, objects with low occupancy ratio (< 35%), computed as the object area divided by the product of the length and height, were considered to be unlikely to be fish and removed from further analysis. This filtering effort effectively reduced many partially occluded targets, fish that were overlapping, and targets that were not fish (e.g., jellyfish). In addition, targets occurring in the upper and lower portions (~ 40 % of the vertical image extent, combined) of the images were difficult to fully separate from trawl netting, so targets from these areas were excluded from the analysis. The endpoints of each target were estimated by taking the extreme points along the major axis. Segmentation was performed on the left image only and is shown in Figure 3.

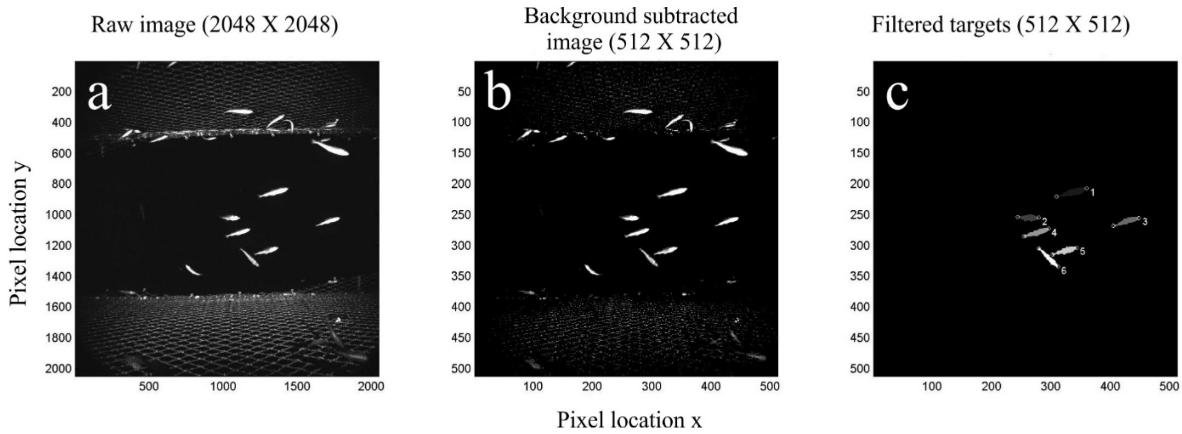


Figure 3. Example of the segmentation process for extracting fish targets from images. The raw image (a) is down sampled from 2048×2048 to 512×512 and the background is subtracted (b). Then a threshold is applied, contiguous regions identified and labeled, and candidate regions filtered for size and aspect ratio (c). Remaining regions then have the endpoints identified for length estimation.

2.6. Stereo correspondence (object matching)

A stereo correspondence technique has been implemented to match fish viewed in the left and right frames. Original resolution images (2048×2048) were used for this analysis step to provide maximum information content on matching targets. Epipolar geometry techniques (Hartley and Zisserman, 2003) were used to estimate the epipolar line, which constrains the location of an object seen in the left image to a line in the right image (Fig. 4). The exact position of the object along the epipolar line defines the range at which the object is located from the left camera. The endpoints of the target in the left image were used as starting points to locate the equivalent target in the right frame. Matching was done using a modified block match approach commonly used in stereo-image depth mapping (Lu and Liou, 1997). A block of pixels from the left image centered on the point of interest was compared to a candidate block of the same size from the right image, with candidate points being placed at regular intervals along the epipolar line, limited by the expected range limits within the image chamber (i.e., 50 – 190 cm from the camera). The size of the square block was matched to the target size, computed as 1/3 of the pixel distance between the target end points. The point with the highest Pearson correlation of pixel values from the test and candidate blocks would then represent the stereo-correspondence point, or equivalent object in the right image. By block matching both target endpoints simultaneously, computation time was reduced and the matches were made more robust by evaluating the combined correlation score for both points. Once a best match was found, a secondary, finer scale, localized block match with smaller inter-block intervals was performed independently to the fish snout and tail to enhance the correspondence. A minimum correlation score of 0.6 was required for a match to be accepted, reducing the probability of incorrect matches.

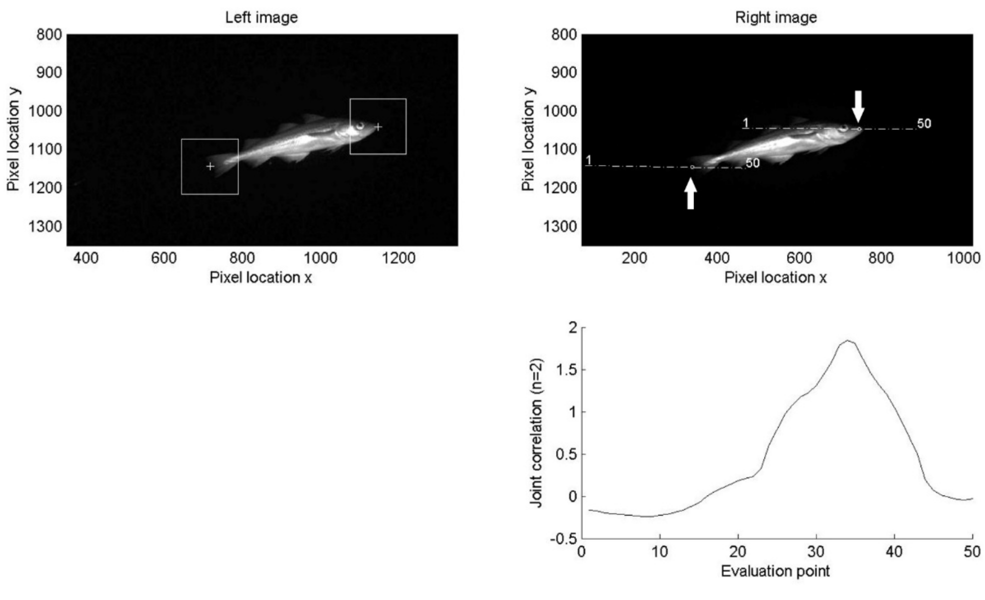


Figure 4. Stereo object correspondence using a modified block-match method. The left image endpoints are used to derive epipolar lines on the right image (dotted lines). The corresponding

points along the epipolar lines (arrows) are found by sequential correlation between reference sub-image blocks in the left image (white squares) and 50 equivalent evaluation blocks taken along the epipolar line. The blocks with the highest correlation score shown in the line chart are taken as the corresponding target endpoints in the right image.

2.7. Length estimation

To estimate the length of each target, the matched fish endpoint pixel coordinates (e.g., tip of the snout and end of the centerline of the tail) from both images were transformed into 3D points using a stereo triangulation function (Bouguet, 2008). Two length estimates were made: the first was a straight line Cartesian distance between the target endpoints, and the second was a sum of three contiguous linear segments defined by two additional points along the fish body center line to account for body curvature (Fig. 5). The angle of the fish body relative to the camera in the planar view (y axis in 3-D projection), or the deviation from orthogonal position, was computed by estimating the angle defined by the points (x_t,y_t), (x_s,y_s), and (x_t, y_s), where x_t and y_t are the x and y position of the fish tail in 3D coordinates (z is not used for this computation), and x_s, and y_s are the same for the snout.

To distinguish between errors inherent in the stereoscopic method and errors in automatic estimates of fish end points and matching between left and right views, manual image-based lengths were taken for a randomly selected subset (*n* = 300) of targets in each dataset. This step involved manually estimating the pixel coordinates of the snout and tail of corresponding fish targets in the left and right views using a manual stereo- analysis software package (Williams et al., 2016). Manual analysis package allowed for a close-up view of the snout and tail to minimize errors in user inputs.

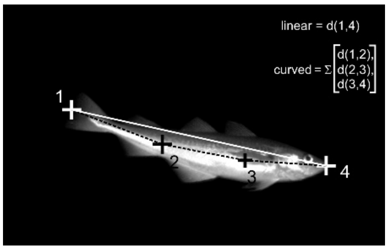


Figure 5. Length estimation was conducted as a 3D straight line-distance between fish endpoints (1 and 4) and alternatively as a sum of the three linear segments representing the centerline length of the fish to account for curvature.

2.8. Target tracking

Individual fish were often encountered over several frames as they passed through the imaging chamber toward the codend. To estimate the true fish passage rates and the precision of multiple measurements on the same individual, several hundred sequential images were manually analyzed by tracking individuals across frames. Fish were tracked only in the left camera images using a purpose-made program that allowed previous frame fish positions to be overlaid onto the image data to aid visual tracking.

3. Results

3.1. General results

Basic descriptions of each data set and general results from the analysis are given in Table 1. The first set with a lower density of large pollock and good water clarity, resulted in a high degree of agreement in mean fish length estimates. Only 15% of automatically acquired targets were used for length estimation in this data set (targets in analysis / total targets). This low value resulted from the restriction of the analysis area to the central portion of the image containing the backdrop, filtering for favorable fish orientation, as well as filtering out occluded targets and false detections (non-fish objects). A higher level of target filtering was required in the second Data set, with only 10% of targets used for lengthing. The last data set resulted in poor performance with a > 6 cm difference in mean length, and a comparable level of target retention as the previous sets (11%). Mean processing time per frame was 0.4 s, meaning it took approximately 60% longer to process than to collect the data. The third data set required substantially longer per frame to process due to the higher density.

3.2. Length Frequency Comparison

A comparison of image and catch-based length estimates shows good general agreement in the first two datasets, while the third data set shows substantial over- and underestimates of size from images (Fig. 6). Manual image-based measurements did not substantially differ from automated methods in the first data set. However, in the second data set, the juvenile size mode

Table 1. Summary of characteristics and analysis results for the three image and catch data sets compared for performance of automated image-based length estimation. Manual count of fish in image frames included the entire image area, whereas the automated analysis was restricted to the central portion (60%) of the image.

Data set	Pollock caught by trawl	Mean fish size (catch)	Mean fish size (images)	Frames analyzed	Analysis time (min)*	Raw targets (left camera)	Targets in analysis (both cameras)	Mean targets / frame (manual)	Measured targets / frame (automated)	Number of fish / minute	Mean track length
1	1348	45.92	45.91	4800	25.53	2605	389	1.93	0.17	76.7	6.01
2	1856	23.59	19.61	4500	22.76	17221	1754	4.22	0.55	187.59	5.38
3	4128	37.81	31.43	5000	45.82	26788	2936	4.20	0.10	148.85	6.76

* based on a computer with a Intel Core i7 processor

was shifted by 1-2 cm in automated measurements (i.e., see FL ~18-21 cm), indicating a tendency of this method to overestimate length for this size group. While manual image-based measurements were more consistent with the catch estimates in the third dataset, a substantial number of fishes > 45 cm encountered in the catch were not proportionally represented in the image-based lengths. The linear distance estimate was virtually identical to the curved approach, indicating that fish curvature does not play a large role in potential image-based length estimation errors in these data.

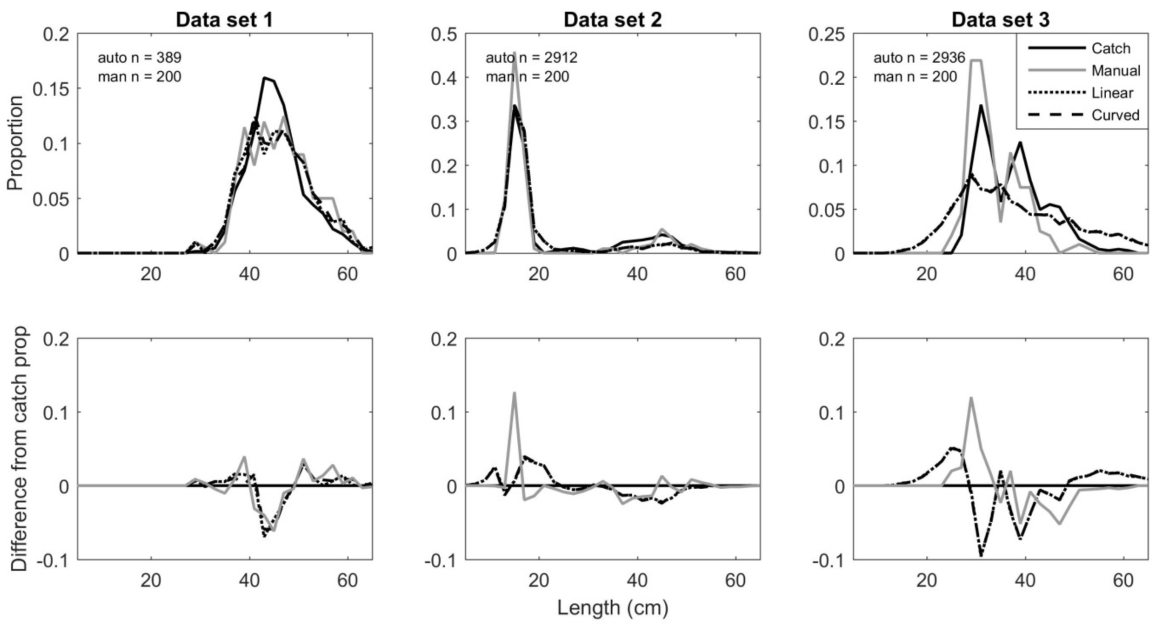


Figure 6. Results comparing the catch-derived length composition of walleye pollock with manual stereo-image based measurements and linear and curved methods of automated length estimation. The lower panel shows the signed difference between catch-based length and the image-based methods for each 2 cm length class.

A quantitative comparison of the length frequencies was made by looking at the mean absolute error (MAE) between frequencies for each 2 cm length class (Fig. 7). The MAE values clearly show the small differences in linear and curved automated methods. For the first two sets, manual image-based measurements do not appear to outperform the automated method. As expected, higher errors were observed in set 3 with both manual and automated image-based methods.

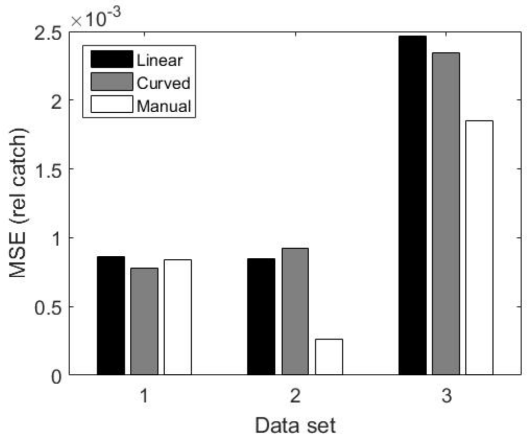


Figure 7. The mean absolute error (MAE) from manual stereo-image based measurements and linear and curved methods of automated length estimation relative to catch based length composition.

3.3. Range and angle dependency

The accuracy of the automated image-based length estimates, compared to catch-based lengths, was examined as a function of target orientation and range relative to the camera. In the first two data sets, automated length estimates were more consistent with the catch-based lengths for targets that were oriented within 20 degrees of orthogonal to the camera axis (those approximating a lateral view in the images), especially in data set 2 (Fig. 8a). Catch- and image-based estimates exhibited poorer agreement for data set 3. This set also exhibited a much wider distribution of horizontal angle estimates. This increased variability likely results from errors in locating corresponding end points for targets due to the low visibility, as well as actual increased variability in horizontal position likely due to reduced ability of the fish to see (Olla et al., 2000) the trawl in water that had reduced clarity compared to the other data sets. Greater agreement with catch data was observed with all data sets (Fig. 8b) when only targets closer than 150 cm to the camera were used, with substantial improvements to data sets 1 and 3.

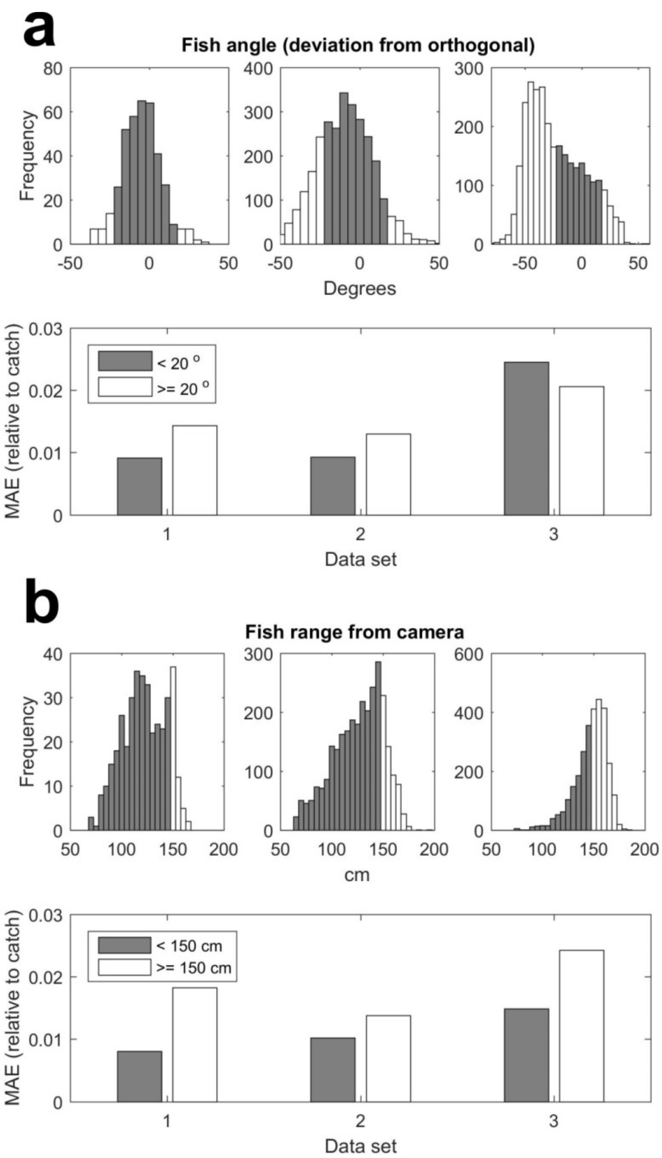


Figure 8. The upper panel (a) represents the frequency distribution of fish horizontal angles (yaw) relative to the camera, with the shaded bars showing the central +/- 20 degree band. The

lower panel (b) shows the distribution of estimated individual fish distances from the camera, with the shaded bars highlighting the portion of the distribution < 150 cm. The maximum extent of the range possible in the image chamber was 190 cm. The differences in MAE between these two categories is given below the histograms.

3.4. Automated versus manual stereoscopic measurements

A more detailed analysis compared target-specific differences between the manual and automated image-based measurements. While data sets 1 and 2 indicated approximately normal errors (standard deviation ~ 2 cm) and little bias in automated methods, the third set shows greater error and a tendency for automated methods to over-estimate fish sizes (Fig. 9). Errors were not strongly related to the size of fish, or with range, although the tendency for large positive overestimates (>10 cm) of the automated method appeared more prevalent with range in data set 3.

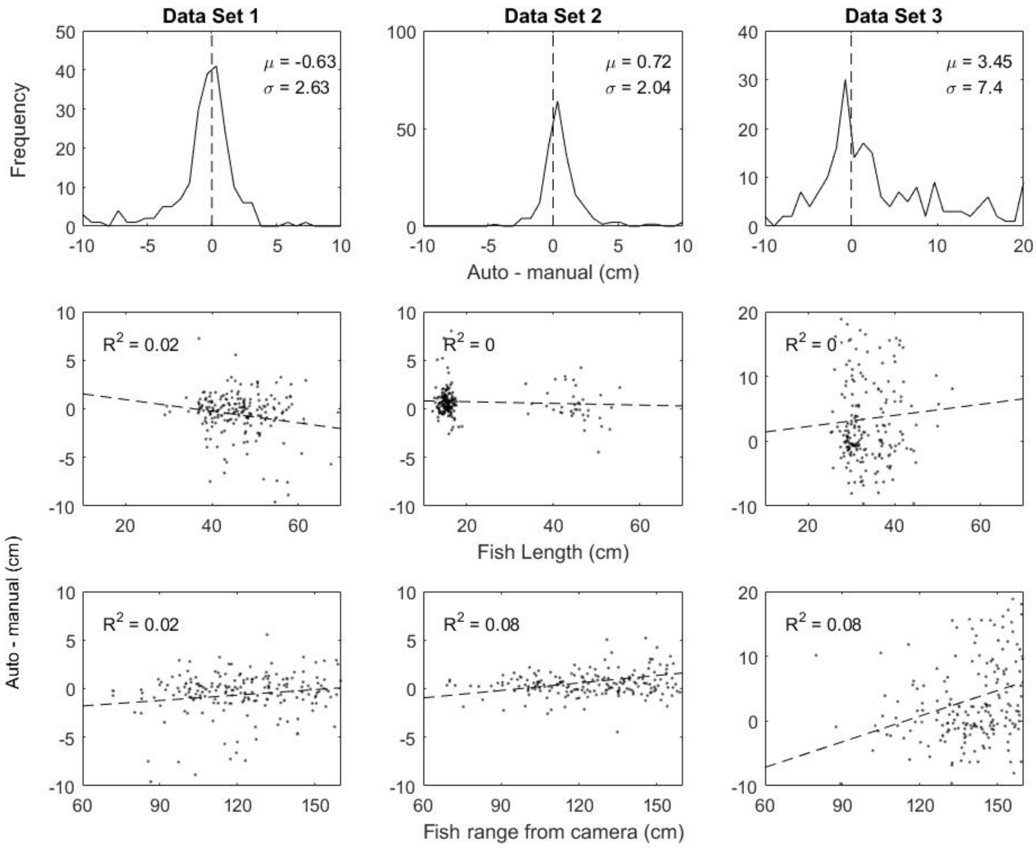


Figure 9. Comparison of manual image-based measurements derived by clicking on the head and tail of corresponding fish in stereo-image pairs and the automated measurements derived for individual fish (n= 300 per data set). The upper panel shows the error distribution (manual-automated), and the lower panels show the correlation of error with fish length, and range from the camera, respectively.

3.5. Automated target assessment

A visual review of the automatically detected targets (n = 200) was conducted to determine the frequency of specific errors in detection across the different data sets. The majority of targets in data sets 1 and 2 appeared to be valid; that is, the targets consisted of a single individual, with straight bodies, and were fully segmented in both left and right views, and thus suitable for making accurate length estimates (Fig. 10). Most of the automatically detected targets in data set 3 were classified as incomplete, meaning that segmentation failed to capture the entire fish body in one or both views segment the entire fish, with the fish tail missing in most cases. A higher number of segmentation errors involving multiple targets overlapping due to higher target densities in the frame occurred in data sets 2 and 3. However, less than 7% of targets were seen in a curved position, and less than 3% were non-pollock fish or trawl objects in all three data sets.

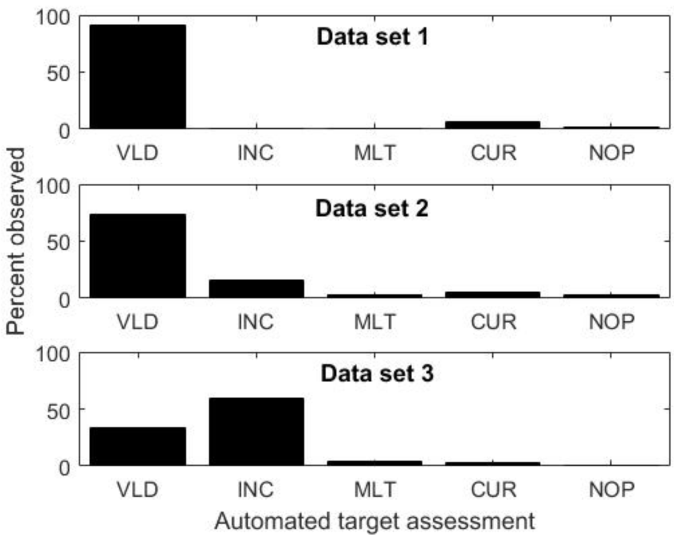
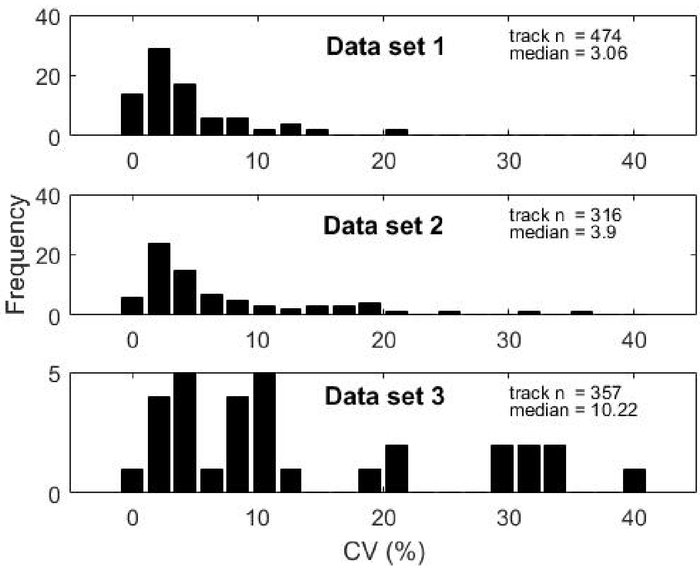


Figure 10. Frequency of observation of different categorizations of walleye pollock automatically detected targets across three image data sets. The categories are abbreviated accordingly: VLD = valid, fully segmented individual pollock targets presented with little curvature, INC = incompletely segmented pollock targets, MLT = multiple targets detected as one, CUR = pollock showing substantial body curvature, and NOP = non pollock targets.

3.6. Error in repeat measurements

Manual tracking of individuals allowed for an evaluation in the consistency in repeated length estimates for the same individual seen in several image frames. Most of the repeat measurements fell within 10% of the mean length, with increased variability observed in data set 3 (Fig. 11). Specifically, data set 3 had a median CV of 10 %, with a substantial number of individuals varying by >30% between measurements. Mean track lengths were greatest for this data set (Table 1), meaning that the individual target length CVs were based on larger sample sizes.



one, CUR = pollock showing substantial body curvature, and NOP = non pollock targets.

Figure 11. Frequency histogram of the coefficient of variance (CV) derived from multiple automated measurements of individual pollock in three image data sets. Measurements were aggregated using manual tracking of individuals across frames.

4. Discussion

This study demonstrates that length frequencies automatically extracted from stereo-camera image pairs can provide sufficiently accurate results when compared with traditional catch sampling under certain conditions. Image-based length composition data, while containing more errors than physical measurements, did not deviate substantially from the catch length frequency modal lengths or in estimates of presence and relative proportion of length classes when optical conditions were good. In addition, the automated method compared very well with manual stereo-image processing, the latter which is used extensively in fisheries image-based surveys for fish sizing (e.g., Cappo et al, 2006; Williams et al., 2010b). Manual processing outperformed the automated method in poor visibility/high density situations, suggesting that human visual acuity still outperforms the ability of automated processing in challenging conditions. Further algorithm development is needed to close this gap.

The component algorithms used in this study are standard, well-established methods slightly customized for specific features of the image data used in this study. The primary contribution of this work is not in the technical sophistication of these components, but in the synthesis of an operational pathway for field use, which includes attention to quantification of the different sources of uncertainty in the process. The general framework of the image-processing operations could be applied to other stereo-camera platforms, with additional tuning or substitution of component algorithms where necessary. For example, segmentation of targets in CamTrawl data was fairly simplistic, as the fish pass by a uniform mostly static background that presents sufficient contrast to easily isolate individual fish outlines. To achieve quality segmentation in situations where the background consists of natural benthic habitat that is not static and may be heavily patterned, a more sophisticated algorithm will be required to extract the targets. Once that step is achieved, however, many of the subsequent steps presented in this study, including stereo-matching and length estimation, could be expected to perform well.

The CamTrawl dataset provides a unique opportunity compared with most other image-based assessment field work, as catch data are readily available to directly validate the image-based size and species compositions. These comparisons have to be made on an aggregate basis as opposed to a one-to-one comparison with physical measurements that have been conducted in the aquaculture setting (Harvey et al., 2003). However, the assessment of the accuracy and precision of length measurements are critical for the successful implementation of this approach into abundance estimation surveys. As expected, repeated image-based measurements from tracked targets revealed higher variability under good measurement conditions (CV ~ 3%) than would be expected from physical measurements using an electronic fish length board (CV ~ 0.5%, personal comm., Rick Towler, AFSC), indicating relatively lower precision for this method. Although repeated measurement results for this study are more variable than similar manual stereo-video measurements made on captive bluefin tuna in a controlled environment (CV = 0.21%, Harvey et al., 2003), results from the present study compare well with *in situ* rockfish measurements (CV = 5.85 %, Williams at al., 2010b).

Stereo-camera data allows independent ranging of both ends of a fish target, removing the need for assumptions about fish orientation such as are required when using parallel laser systems (Dunlop et al., 2015). Despite the potential for getting length from fish in a variety of angular positions relative to the camera, best results for the present study were achieved when the data were restricted to fish orientations +\- 20 degrees from normal (Fig 8a). Similarly, fish on the far end of the sampling space (>150 cm) introduced more error into the estimates, showing the limitations of accuracy at increasing ranges. Both of these factors present challenges for towed or ROV/AUV based camera platforms where fish behavior can affect both orientation and range (Stoner et al., 2008).

This study revealed the limitations of using automated image processing in certain situations, such as encountered in data set 3, namely high turbidity/reduced water clarity and high density of fish. Of these two challenges, the former presents a more difficult situation, as seen by the number of incompletely imaged animals in the target review (Fig. 10). Additional image pre-processing may be able to greatly improve performance in these conditions however, as similar challenges occurring in terrestrial imagery have been studies (Watkins et al., 2000) High density has the primary effect of limiting the number of acceptable targets; for example, fish that are not overlapping with or occluded by others in both images. The results of this study were heavily influenced by the selection of targets for length estimation. While the manual image-based measurement approach would be better suited to extracting lengths for all fish encountered, subsampling of targets may be essential for achieving reliable and usable results using automated methods. However, automated methods are not well poised to properly analyze occluded and overlapping fish targets at this time.

Future additions to the automated processing workflow include implementation of automated target tracking specifically developed for low frame rate situations (Chuang et al., 2015), which will allow for higher precision to be achieved by averaging multiple measurements (Harvey et al., 2003). Automated species classification (Chuang et al., 2014) will also enable species-specific length compositions to be determined, expanding the usefulness of the CamTrawl automated lengthing to catches with greater catch diversity.

Implementation of image-based length data into survey abundance analyses will also require a thorough understanding of how measurement errors will affect uncertainty in population abundance and size structure. As with all new approaches to data collection, it is critical to reduce the risks of methodological biases, potentially derived by non-random effects of the stereo analysis or automated image processing methods. Even randomly distributed length measurement errors can have non-random effects on quantities derived from length estimates, such as biomass estimates using non-linear length-weight relationships. In the case of the CamTrawl system, these effects can be directly estimated by comparing survey abundance estimates derived from catch- and image-based length measurements.

In conclusion, this study shows the potential benefits of using automated methods for measuring fish from stereo methods, which can yield substantially greater efficiency compared with traditional physical or manual image-based sampling. Satisfactory results can often be achieved using basic image analysis algorithms, especially where the image data are collected in controlled or semi-controlled environments. In our case, this was within a large trawl. In some situations, some level of human intervention may be necessary, such as poor visibility conditions or less controlled environments. The analytical process here represents an operational system that is field ready, and continuing developments will help further improve performance and extract additional data from the images.

Acknowledgements

This project would have not been possible without the advice and assistance of Chris Wilson, Scott McEntire, Craig Rose, and David King from the AFSC. We would also like to thank the crew of the Oscar Dyson for their efforts in supporting all of the field work, including midwater trawl modifications. This project was funded by NOAA's advanced sampling technology working group and the NOAA's Automated Image Analysis Strategic Initiative (AIASI).

References

Bouguet, J. Y. 2008. Camera calibration toolbox for Matlab [online].[Available from http://vision.caltech.edu/bouguetj/calib_doc/index.html (accessed September 2008)].

Cappo, M., Harvey, E. and Shortis, M., 2006. Counting and measuring fish with baited video techniques-an overview. *In* Australian Society for Fish Biology Workshop Proceedings (pp. 101-114).

Chuang, M. C., Hwang, J. N., Williams, K., and Towler, R. 2015. Tracking Live Fish from Low-Contrast and Low-Frame-Rate Stereo Videos. IEEE Trans. on Circuits and Systems for Video Technologies (CSVT) 25(1): 167-179.

Chuang, M. C., Hwang, J.N., Kuo, F. F., Shan, M. K. and Williams, K., 2014, October. Recognizing live fish species by hierarchical partial classification based on the exponential benefit. In Image Processing (ICIP), 2014 IEEE International Conference on (pp. 5232-5236). IEEE.

Costa, C., Loy, A., Cataudella, S., Davis, D., and Scardi, M. 2006. Extracting fish size using dual underwater cameras. Aquacultural Engineering 35(3): 218-227.

Dunbrack, R. L. 2006. In situ measurement of fish body length using perspective-based remote stereo-video. Fisheries Research 82(1): 327-331.

Dunlop, K. M., Kuhnz, L. A., Ruhl, H. A., Huffard, C. L., Caress, D. W., Henthorn, R.G., Hobson, B.W., McGill, P. and Smith, K. L. 2015. An evaluation of deep-sea benthic megafauna length measurements obtained with laser and stereo camera methods. Deep-Sea Research Part I: Oceanographic Research Papers 96: 38-48.

Edgington, D. R., Cline, D. E., Davis, D., Kerkez, I. and Mariette, J. 2006. Detecting, tracking and classifying animals in underwater video. In OCEANS 2006, IEEE International Conference on (pp. 1-5). IEEE.

Gibson, R. N., Atkinson, R. J. A. and Gordon, J. D. M., 2009. A review of underwater stereo-image measurement for marine biology and ecology applications. Oceanography and Marine Biology: an Annual Review 47: 257-292.

Haralick, R. M. 1981. Some neighborhood operators. pp. 11-35. *In* Real-Time Parallel Computing. Springer US.

Hartley, R., and Zisserman, A. 2003. Multiple View geometry in computer vision. Cambridge University Press.

Harvey, E. S., Goetze, J., McLaren, B., Langlois, T. and Shortis, M. R. 2010. Influence of range, angle of view, image resolution and image compression on underwater stereo-video measurements: high-definition and broadcast-resolution video cameras compared. Marine Technology Society Journal 44(1): 75-85.

Harvey, E., Cappo, M., Shortis, M., 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(3): 315-326.

Honkalehto, T., and McCarthy, A. 2015. Results of the acoustic-trawl survey of walleye pollock (*Gadus chalcogrammus*) on the U.S. and Russian Bering Sea Shelf in June - August 2014

(DY1407). AFSC Processed Rep. 2015-07, 63 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

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

Lu, J. and Liou, M. L. 1997. A simple and efficient search algorithm for block-matching motion estimation. Circuits and Systems for Video Technology, IEEE Transactions on, 7(2): 429-433.

Mallet, D., and Pelletier, D. 2014. Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012). Fisheries Research 154: 44-62.

McFarlane, N. J., and Schofield, C. P. 1995. Segmentation and tracking of piglets in images. Machine Vision and Applications 8(3): 187-193.

MacLeod, N., Benfield, M. and Culverhouse, P., 2010. Time to automate identification. Nature, 467(7312), pp.154-155.

Olla, B. L., Davis, M. W. and Rose, C. 2000. Differences in orientation and swimming of walleye pollock *Theragra chalcogramma* in a trawl net under light and dark conditions: concordance between field and laboratory observations. Fisheries Research, 44(3), pp.261-266.

Rosen, S., Jörgensen, T., Hammersland-White, D. and Holst, J. C. 2013. DeepVision: a stereo camera system provides highly accurate counts and lengths of fish passing inside a trawl. Canadian Journal of Fisheries and Aquatic Sciences 70(10): 1456-1467.

Shapiro, L. G., and Stockman, G. C. 2001. Computer Vision. Prentice Hall, pp. 279-325.

Shortis, M. R., Ravanbakskh, M., Shaifat, F., Harvey, E.S., Mian, A., Seager, J. W., Culverhouse, P. F., Cline, D. E. and Edgington, D. R. 2013. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. (pp. 87910G-87910G) *In* SPIE Optical Metrology 2013: International Society for Optics and Photonics.

Singh, H., Roman, C., Pizarro, O., Eustice, R. and Can, A., 2007. Towards high-resolution imaging from underwater vehicles. The International journal of robotics research, 26(1), pp.55-74.

Sonka, M., Hlavac, V. and Boyle, R., 2014. Image processing, analysis, and machine vision. Cengage Learning.

Stoner, A. W., Ryer, C. H., Parker, S. J., Auster, P. J. and Wakefield, W. W. 2008. Evaluating the role of fish behavior in surveys conducted with underwater vehicles. Canadian Journal of Fisheries and Aquatic Sciences 65(6): 1230-1243.

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(1): 123-141.

Watkins, W.R., Tofsted, D.H., CuQlock-Knopp, V.G., Jordan, J.B. and Merritt, J.O., 2000, June. Navigation through fog using stereoscopic active imaging. In AeroSense 2000 (pp. 20-28). International Society for Optics and Photonics.

Watson, D. L., Harvey, E. S., Fitzpatrick, B. M., Langlois, T. J. and Shedrawi, G. 2010. Assessing reef fish assemblage structure: how do different stereo-video techniques compare? Marine Biology 157(6): 1237-1250.

Williams, K., Towler, R. and Wilson, C., 2010. Cam-trawl: a combination trawl and stereo-camera system. Sea Technology 51(12): 45-50.

Williams, K., Rooper, C. N. and Towler, R., 2010. Use of stereo camera systems for assessment of rockfish abundance in untrawlable areas and for recording pollock behavior during midwater trawls. Fishery Bulletin 108(3): 352-362.

Williams, K., R. Towler, P. Goddard, R. Wilborn, and C. Rooper. 2016. Sebastes stereo image analysis software. AFSC Processed Rep. 2016-03, 42 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.