Towards automating underwater measurement of fish length: a comparison of semi-automatic and manual stereo–video measurements

Faisal Shafait[1]*, Euan S. Harvey[2], Mark R. Shortis[3], Ajmal Mian[4], Mehdi Ravanbakhsh[3], James W. Seager[5], Philip F. Culverhouse[6], Danelle E. Cline[7], and Duane R. Edgington[7]

[1]School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad H-12, Pakistan
[2]Department of Environment and Agriculture, Curtin University, Kent Street, Bentley, Perth, Western Australia 6102, Australia
[3]School of Science, RMIT University, GPO Box 2476, Melbourne, Victoria 3001, Australia
[4]School of Computer Science and Software Engineering, The University of Western Australia, 35 Stirling Hwy, Crawley, Perth, Western Australia 6009, Australia
[5]SeaGIS P/L, PO Box 1085, Bacchus Marsh, Victoria 3340, Australia
[6]Centre for Robotics and Neural Systems, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK
[7]Monterey Bay Aquarium Research Institute, 7700 Sandholt Road, Moss Landing, CA 95039, USA

*Corresponding author: tel: +92 (0)51 9085 2400; fax: +92 (0)51 9085 2002; e-mail: faisal.shafait@seecs.edu.pk

Underwater stereo–video systems are widely used for counting and measuring fish in aquaculture, fisheries, and conservation management. Length measurements are generated from stereo–video recordings by a software operator using a mouse to locate the head and tail of a fish in synchronized pairs of images. This data can be used to compare spatial and temporal changes in the mean length and biomass or frequency distributions of populations of fishes. Since the early 1990s stereo–video has also been used for measuring the lengths of fish in aquaculture for quota and farm management. However, the costs of the equipment, software, the time, and salary costs involved in post processing imagery manually and the subsequent delays in the availability of length information inhibit the adoption of this technology.

We present a semi-automatic method for capturing stereo–video measurements to estimate the lengths of fish. We compare the time taken to make measurements of the same fish measured manually from stereo–video imagery to that measured semi-automatically. Using imagery recorded during transfers of Southern Bluefin Tuna (SBT) from tow cages to grow out cages, we demonstrate that the semi-automatic algorithm developed can obtain fork length measurements with an error of less than 1% of the true length and with at least a sixfold reduction in operator time in comparison to manual measurements. Of the 22 138 SBT recorded we were able to measure 52.6% (11 647) manually and 11.8% (2614) semi-automatically. For seven of the eight cage transfers recorde,d there were no statistical differences in the mean length, weight, or length frequency between manual and semi-automatic measurements. When the data were pooled across the eight cage transfers, there was no statistical difference in mean length or weight between the stereo–video-based manual and semi-automated measurements. Hence, the presented semi-automatic system can be deployed to significantly reduce the cost involved in adoption of stereo–video technology.

Keywords: automated measurement, fish tracking, fisheries management, Southern Bluefin Tuna, stereo matching, stock assessment, underwater stereo–video, video sequences.

Introduction

The sustainability of wild fish stocks is of universal concern (Pauly *et al.*, 2002) because fish protein is a major contributor to the human diet in many regions of the world. Total catch effort is increasing world-wide to support the ever growing human population, with declines in the abundance and size of some targeted species. It has been demonstrated that the disturbance of keystone species in an ecosystem can have unpredictable and sometimes catastrophic outcomes (Pauly *et al.*, 2002). Accurate and reliable relative abundance and length information are important for monitoring fish populations and assemblages to facilitate the early detection of impacts from natural and anthropogenic activities such as fishing, climate change and pollution. Accurate and reliable fish length data are not only useful in fish stock assessment but also an important indicator of the health of wild fish stocks (Jennings and Polunin, 1997; Jennings and Kaiser, 1998).

With declining fish stocks, aquaculture is increasingly becoming an important source of dietary protein (Duarte *et al.*, 2009). The aquaculture industry has flourished world-wide, based on species such as tuna, salmon, and tilapia (Naylor *et al.*, 2000). In Australia, the most commonly farmed species are Southern Bluefin Tuna (SBT; *Thunnus maccoyii*) and Atlantic Salmon (*Salmo salar*). Data on the mean length and length frequency of aquaculture fish are a fundamental tool for farmers and managers. This generates data on fish growth and food conversion rates allowing farmers to optimize feeding and harvest strategies. Length data can also be used to estimate fish weight, enabling management agencies to assess the total biomass of fish that have been captured (Harvey *et al.*, 2003).

Invasive methods of sampling fish to measure length or biomass that involve capturing and handling fish before releasing them can adversely affect fish health, growth rates and the quality of the harvested product as a result of injury and stress (Ramsay *et al.*, 2009). This is especially true for SBT because capture instigates an instantaneous flight response that can cause injury to the fish or handlers.

Underwater stereo–video has been used to make accurate and precise, non-invasive measurements of fish length (Harvey and Shortis, 1995; Steeves *et al.*, 1998). Stereo–video systems are now used to monitor the relative abundance and length of shallow (Bornt *et al.*, 2015; Langlois *et al.*, 2015; McLaren *et al.*, 2015), deep-water demersal (Zintzen *et al.*, 2012; McLean *et al.*, 2015), and pelagic (Santana-Garcon *et al.*, 2014a, b, c) fishes. Underwater stereo–video systems have a longer history of use in aquaculture (Naiberg *et al.*, 1993; Petrell *et al.*, 1997; Shieh and Petrell, 1998; Harvey *et al.*, 2003; Costa *et al.*, 2006).

Most operational stereo–video measurement systems rely on a human operator using a mouse to locate and click on the snout and tail of a fish in both the left and right images on a computer screen. These locations are converted to stereo-image measurements in order to determine the 3D Euclidian length by intersection (Shortis *et al.*, 2009). Manual measurement is slow and delays the availability of length data. The initial cost of the equipment and software, combined with the ongoing costs of image processing and the delays in data availability, discourages the uptake of stereo–video systems by fisheries scientists, ecologists, managers and researchers involved in aquaculture and natural resource management. Of those who do use stereo–video systems, many choose to restrict the number of measurements to a percentage of the total possible in an effort to reduce the time and salary cost of operations (Phillips *et al.*, 2008).

While automatic identification, tracking, and measurement of objects such as vehicles and pedestrians are well established above water (Lipton *et al.*, 1998; Stauffer and Grimson, 2000; Bloisi and Iocchi, 2009; Chan and Vasconcelos, 2012), the loss of contrast from attenuation through the water and the dynamic environment makes the implementation of an automated solution for fish sizing problematic (Shortis *et al.*, 2013). Nevertheless, automation of some aspects of the process has been established in controlled environments for at least 15 years (Lines *et al.*, 2001).

Recent research in automated analysis of underwater videos has shown promising results for detection, counting, and species identification of fish of interest in partially or fully uncontrolled environments (Charalampidis *et al.*, 2012; Boom *et al.*, 2014; Ravanbakhsh *et al.*, 2015; Salman *et al.*, 2016; Shafait *et al.*, 2016; Shortis *et al.*, 2016). However, in general, these systems are not capable of fully automated operation or do not attempt length measurement, as they have been developed for single camera operations with the primary aim of identifying and counting species of interest. Further, the systems that do estimate the size of fish are based on a bounding region (Tillett *et al.*, 2000; Charalampidis *et al.*, 2012; Boom *et al.*, 2014) rather than identification of specific landmarks such as the tip of the snout and the fork of the tail.

While the full automation of the complete process of fish detection, tracking, species identification, and size measurement in uncontrolled environments may appear to be unrealistic at the time of publication, incremental approaches to automate individual components in isolation are desirable to demonstrate the feasibility of the concept and speed up the process of fish sampling. For example, Spampinato *et al.* (2008) performed automatic detection, tracking, and counting of fish achieving 85% accuracy. Salman *et al.* (2016) and Shafait *et al.* (2016) have performed fully automatic identification of fish species reporting over 95% accuracies on the ImageCLEF data set (http://groups. inf.ed.ac.uk/f4k/).

SBT is a highly migratory species inhabiting open oceans at mid-latitudes in the southern hemisphere. In the past, this species has supported pole and long-line commercial fisheries for Japan and Australia, and to a lesser extent Taiwan, New Zealand, Indonesia and Korea. In response to the imposition of a quota in 1988, the SBT industry in Australia shifted from wild catch to sea ranching in order to maximize the value of the harvested fish. In 2016, sea ranching operations based at Port Lincoln in South Australia accounted for more than 98% of the Australian total catch quota of 5665 tonnes.

In South Australian sea ranching operations, SBT are caught at sea in purse-seine nets and transferred to a tow cage. Fish are towed to waters near Port Lincoln where they are transferred to grow out cages. To determine the contribution toward quota, a sample of fish is caught and weighed to determine the mean fish weight. Fish are then transferred through a gate between the tow cage and grow out cage. Using an underwater video camera, the SBT are counted and the count multiplied by the mean weight to determine the quota caught.

Previous trials have demonstrated that an underwater stereo–video system can non-invasively and accurately measure the length of SBT being transferred into grow out cages (Philips *et al.*, 2008; Harvey *et al.*, 2011). The accuracy and precision of the mean of three or more measurements of the same fish have been shown to be significantly improved compared with a single

measurement, resulting in snout to tail measurements with an error of less than 1% of the real length (Harvey *et al.*, 2003). As discussed above, delays in the availability of length data are one of the negatives of using stereo–video and making manual measurements. Up to 1.5 min per fish is required to count and successfully measure a fish (Harvey *et al.*, 2011). This incorporates the time involved in counting every SBT and measuring a predetermined percentage, including measurements that are rejected. The total time involved depends on the number of fish in a tow cage, the proportion to be measured (e.g. 10 or 20%) and the number of analysts working simultaneously to measure fish that have been transferred. For example, a cage containing 3000 SBT with a requirement for 20% of the fish to be measured could take 15 operator hours before measurements were completed. There is a real world need to decrease image analysis costs and speed up the availability of length data making them an ideal focal species for incremental automation of stereo–video image analysis.

This research had three objectives: (1) to develop algorithms which, with minimal user input, can track an individual fish through sequential video frames and acquire multiple length measurements of the same fish; (2) to validate the algorithms by comparing the results of semi-automatic measurements to manual measurements made of the same fish; (3) to estimate the proportion of fish that can be measured using the proposed semiautomatic approach, as compared to the manual approach.

Material and methods
Imagery collection
This study is based on stereo–video imagery of SBT recorded in Port Lincoln, South Australia between January and April in 2011. This imagery was part of a commercial trial of stereo–video that aimed to record 10% of the catch for the 2011 season (Harvey *et al.*, 2011). Eight transfers from three tow cages were recorded with stereo–video (a total of 22 138 SBT). For the purpose of this study, the transfers have been labelled as A to H. The SBT were recorded as they swam from the tow cage to a grow out cage through a 3 × 3 m transfer gate connecting the two cages. The camera system was mounted 1.2 m beyond the gate with fish measurements ranging from 1.6 to 4.63 m.

Description of the camera system
Stereo–video imagery was recorded with an AM100 camera system (www.aq1systems.com.) The AM100 uses two Pulnix TMC 1327 Gigabit Ethernet (GigE) cameras, positioned approximately 700 mm apart and aligned with an inward convergence of 6° to optimize stereo overlap within the field of view. The cameras, power converter, and an Ethernet switch are mounted in a single integrated underwater housing. The housing was mounted on the wall of the grow out cage adjacent to the transfer gate. The AM100 cameras were connected to the logging computer on a vessel positioned close to the transfer gate. An umbilical cable supplied power to the cameras, transferred imagery to the logging computer and enabled control of the camera synchronisation and exposure. The image brightness was manually set (through gain and shutter speed adjustments) to produce a consistent exposure level for the stereo–video images. Images were recorded in compressed Audio Video Interleaved (AVI) file format directly onto the computer hard drive.

AM100 camera calibration
Prior to field recordings the stereo–video camera system was calibrated in a swimming pool following the technique described by Harvey and Shortis (1995, 1998) and Harvey *et al.* (2002, 2003). The calibration was performed by recording imagery of a 1 m by 1 m by 0.5 m purpose-built calibration frame and processing the images using the CAL software package (Version 1.33) (www.seagis.com.au/bundle.html.)

Camera calibration validation
Manual measurements (see section below) of three different length measurements (513.8, 750.7, and 1264.6 mm) on a scale bar were taken immediately following the calibration. These lengths encompass the range of lengths recorded for the SBT. Each length was measured twelve times over a range of distances (1.3–4.3 m from the camera system) using the centroiding functions in the measurement software. Measurements of the scale bar provide an independent validation of the calibration integrity of the system and give an indication of the best possible measurement accuracy and precision the AM100 can achieve.
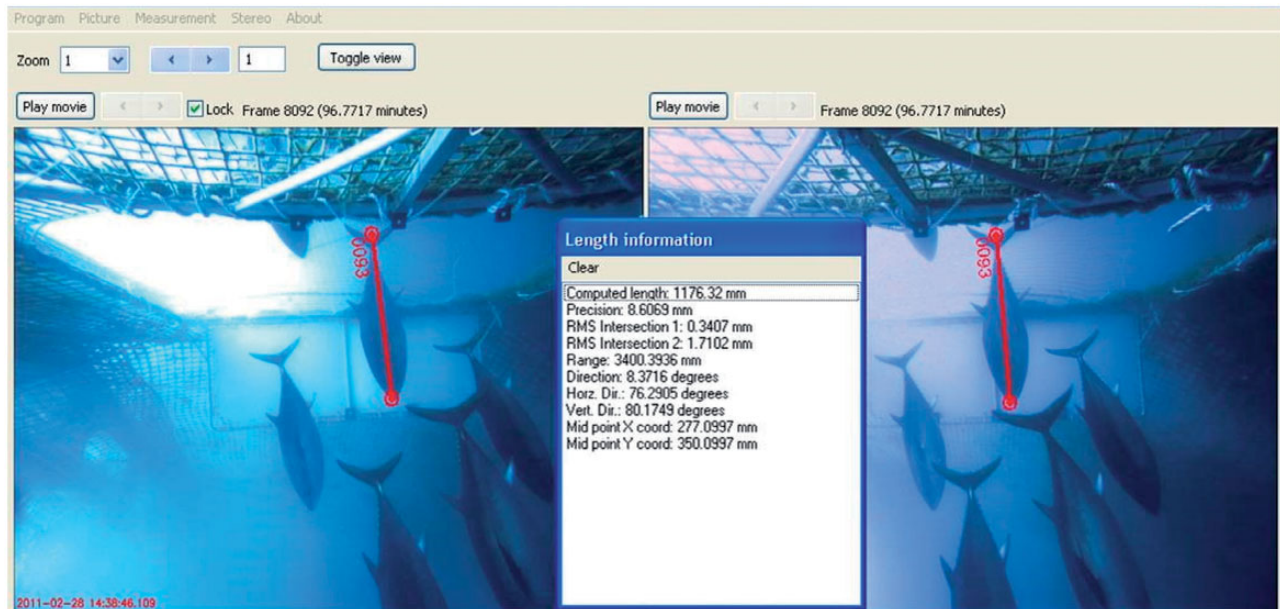
Camera calibration stability
To monitor the calibration of the camera system throughout the duration of the trials (both within and between recordings), a set of calibrated reference targets were positioned on tabs mounted at the back and front of the transfer gate and measured at regular time intervals during the recordings (Figure 1).

Manual measurements
Manual measurements of the lengths of SBT were made by a single operator using EventMeasure Stereo software (www.seagis.com.au/event; Version 3.08). Two AVI files containing image sequences from the left and right cameras were imported into the software and paired images were synchronized by frame number. Measurements of the fork length were made by an operator manually locating the tip of the snout and the fork of the tail of the target SBT within the left and right synchronized video streams using a mouse as shown in Figure 1. The two pairs of image coordinates are converted into coordinates in 3D object space using stereo intersection. To obtain length measurements, the Euclidean distance between the 3D measurements is computed automatically. An estimator of the quality of the measurements, the root mean square residual also known as RMS error or RMS residual parallax (Harvey and Shortis, 1998), and the precision of the length measurement is automatically logged. The distance and angle of the mid-point of the fish to the central point between the camera lenses are also automatically computed and logged (Figure 1).

Semi-automated measurement
The goal of the semi-automatic measurement strategy was to minimize the time required to make manual length measurements. It was also a goal to keep the measurement error below 1% of the true body length. This level of error was agreed between the Australian SBTIndustry Association and the Australian Fisheries Management Authority as a threshold that should be reached for the acceptance of any stereo–video-based measurement system. Additionally, it was agreed that each SBT had to be measured at least three times on different frames in the video

Figure 1. A screen capture of the software showing a measurement of a marked up fish and the resulting data. Targets for calibration stability testing are placed at the far end of the transfer gate and are visible in the images.

sequence and that the standard deviation of measurements across the three or more frames should not exceed 20 mm (see Harvey et al., 2011).

The first logical step in automation is a minimalistic input approach in which the operator marks the snout and tail fork of the fish to be measured in only one frame (for instance, the left image) of the stereo pair. The remainder of the measurements for that fish in subsequent frames of the video, as well as the corresponding measurements in the counterpart frame (i.e. the right image) of the stereo pair could be performed automatically using the computer vision algorithms described below. In theory, this would reduce the number of mouse clicks from four per image (the snout and fork of the fish in both the left and right image) or 12 for three repeated measurements of the same fish through three different frames, to a maximum of two clicks achieving a sixfold reduction in operator time and effort.

Template matching

Template matching is one of the most widely used methods to find an object of interest with known shape and texture in the given image. First, a template (usually a rectangular image region centred on the point of interest) representing the appearance of the target object is extracted from a sample image. Then, the similarity is computed between the template and different candidate locations in a new image where the object might be present. In this testing a normalized correlation coefficient was used as the similarity measure (Mahmood and Khan, 2012). The location in the image that yields the highest similarity score with the template is regarded as the best match for the template. The manually marked locations of the tip of the snout and tail fork of the fish in one frame serve as the centre points of the image templates that capture the appearance of the snout and tail of the fish to be measured. These templates are then used for matching, and hence locating, the snout and tail in the corresponding frame of the stereo pair, as well as in subsequent frames of the same video.
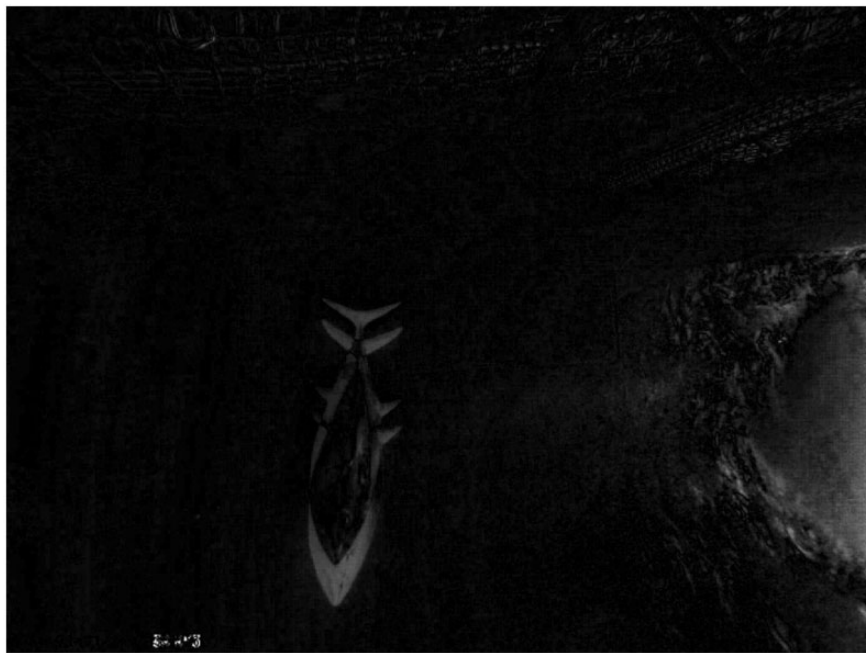
Joint shape-motion templates

The detection accuracy of template-based methods degrades in the presence of abrupt, spatially non-uniform changes in illumination, variability in the background and proximity of other similar-looking objects (Mahmood and Khan, 2010). To illustrate these challenges, an example of a stereo pair of frames is shown in Figure 2. Note that, due to the different viewing location and angles of the two cameras, the tails of the leftmost as well as the rightmost fish appear on substantially different backgrounds in the left camera and the right camera images. In this circumstance, the reliability of the match is affected and the success rate of the template matching declines. To circumvent these problems, the frame difference image as shown in Figure 3 is used as the basis for computing and matching templates. The frame difference image not only reduces the effect of variations in background and non-uniform illumination across the two frames in the stereo pair but also helps in encoding the swimming speed of the fish into the template. Hence, slow moving fish, or fish farther away from the camera can be distinguished from fast moving fish or those closer to the camera, resulting in more reliable matches. These joint shape–motion templates are also matched using normalized correlation coefficients.

Semi-automated single measurement

To robustly match the snout and tail templates across the stereo frames, a filtering mechanism is needed to automatically discard incorrect matches that occur due to the presence of other fish with a similar shape and texture as the target fish. In a stereo camera system, a point in one camera frame can only appear on a corresponding line, known as an epipolar line (Gruen and Baltsavias, 1988) in the frame of the other camera. For a calibrated stereo camera system, the epipolar lines can be accurately calculated and hence the search for corresponding points can be constrained to a narrow search space to efficiently achieve reliable results. This search space is defined by an envelope centred on the

Figure 2. A stereo pair of frames in which snout and tail of the fish to be measured is shown in the left frame with black squares, where the squares represent the image region used for template matching. Note that due to different perspectives of the camera, the background changes significantly between the two images. Two sample tails in each frame are marked with black ellipses to illustrate this effect.


Figure 3. A sample Frame Difference Image used as the basis for computing a joint shape-motion template.

epipolar line corresponding to the marked snout and tail points (see Figure 4). Once the correct correspondence is automatically established between the snout and tail points of the fish in the left and the right images, 3D locations of the snout and tail fork of the fish are computed using stereo intersection. The length measurement of the fish resolves to a simple 3D Euclidean distance between the snout and the tail fork. Compared to the manual method, this approach alone should result in a twofold reduction in measurement effort, since there is no need to manually locate the snout and tail fork in both frames of the image pair.

Automated repeat measurements

Once the locations of the snout and tail fork in both frames of the stereo pair have been determined, they are used to track the fish in the subsequent frames for repeated length measurements. Since the direction of motion of the fish is known, the region of interest (ROI) in the next frame can be estimated based on the

snout and tail fork location in the current frame. After ROI estimation, joint shape-motion templates are extracted from the current frame for both snout and tail. These templates are matched within the ROI to compute the new location of the snout and tail. A straightforward technique to compute these locations is to assign the location of the best matching scores as the snout and tail in the next frame. However, due to partial occlusion, sudden changes in illumination and the proximity of other fish, the best matching scores sometimes do not correspond to the true locations. Based on the observations that fish length and swimming direction does not abruptly change between two consecutive frames of a video, constraints on the length and orientation are used to prune the matching scores to find the most plausible new locations. Thresholds on the score are employed to discard the measurement if the similarity of the best match is too low. This case usually arises when the snout or tail is partially occluded, or a part of the snout goes out of the field of view of the camera. The threshold determines the sensitivity of the
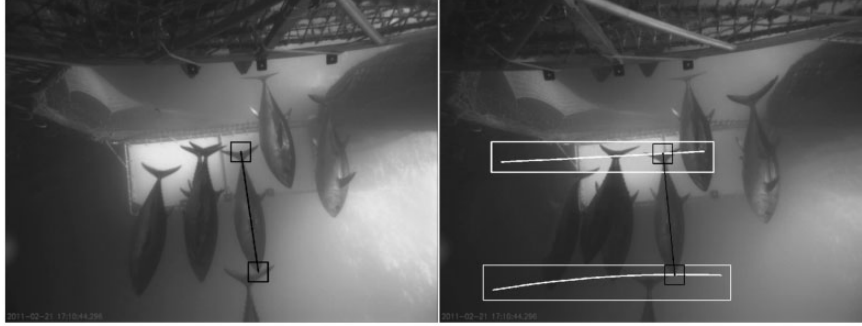
Figure 4. An illustration of search range restriction using epipolar geometry. The left image shows the marked points to be matched in the right stereo pair. White lines overlaid on the image depict epipolar lines corresponding to the marked snout and tail location in the left image. White rectangles show the regions of interest for locating the marked points based on the epipolar line. Note that lens distortions computed as part of the calibration of the stereo camera system results in curved epipolar lines.

proposed semi-automatic length measurement approach. If the threshold is low, a larger number of fish are measured, but with a relatively high mean error on length measurement. Setting the threshold to a higher value results in fewer, but more accurate measurements. For this case study on SBT, the threshold is set to a high value (0.9 on a scale from 0 to 1, where 1 indicates a perfect match) to meet the strict accuracy requirements of the end-users.

Quality control of length measurements

Harvey et al. (2003) describe how the error and precision of stereo–video measurements of SBT length stabilized after five repeated measurements from five sequential images of the same fish. Measurements were made as the fish was 'gliding' in a reasonably straight line rather than when the fish was turning or swimming strongly. This avoided length under-estimation due to sinusoidal changes in body shape associated with vigorous swimming. SBT can only be measured if both the snout and fork are visible in the left and right images of a stereo pair. The number of frames available for measurement is also dependent on the swimming speed of the fish. SBT can swim at burst speeds of up to 70 kmph (Blank et al., 2007). To avoid the elimination of many SBT from the measurement processes, but still retain the confidence of repeated measurements, a less restrictive minimum of three replicate measurements of each fish from different video frames was adopted. Where possible, larger numbers of repeated measures were made, especially for the semi-automatic measurement.

Apart from excluding fish which had less than three manual or semi-automated measurements of length, other quality control measures were implemented to reject average fish lengths if there were large differences in the lengths from the three (or more in the case of the semi-automated measurements) repeated measurements. SBT lengths were deemed unacceptable if the average of the three or more replicate measurements of any one fish had a standard deviation greater than 20 mm. Phillips et al. (2008) demonstrated that it was possible to obtain accuracies of over 99% of the real length of a fish when the standard deviations from 3 or more measurements were less than 20 mm. Measurements to the snout or fork that had standard deviations of 20 mm or greater, or RMS error values (calculated by the software) of 10 mm or greater were rejected. The RMS error provides an estimate of measurement quality, and a threshold of 10 mm

has been established from past experience of validation trials with objects of known lengths (Harvey and Shortis, 1995; Harvey et al., 2002, 2010).

Statistical analysis
Mean length and weight

A one-way analysis of variance (ANOVA) was used to test for differences in the mean length and mean weight between manual and semi-automated measurements for each transfer. Prior to the ANOVA, a Levene's test (Levene, 1960; Anderson et al., 2008) was conducted to check the homogeneity of variance of the length data. Where data were heterogeneous, the data were Log transformed. A two-way general linear model ANOVA was performed on the data pooled across transfers (Technique and Cage were fixed factors). All univariate analyses were conducted with Minitab (V16).

Length frequency

A Kolmorgrov-Smirnov test (Siegel, 1956) was used to find differences between the length frequency distributions ($p = 0.05$) between manual and semi-automated measurements for each transfer. This test calculates the maximum difference between a size class for the cumulative frequency distributions of two data sets (Bell et al., 1985). Length frequency has been expressed as a percentage of the total number of fish sampled.

Converting length to biomass

Differences in mean weight were also compared statistically. Weight estimates are usually based on species-specific regression analysis of length versus weight (Pienaar and Thomson, 1969; Kohler et al., 1995). In this case, the mean length of an individual SBT was converted to a weight using a regression equation provided by the Australian Department of Agriculture, Forestry and Fisheries (Harvey et al., 2011). The equation used was:

$$\text{Weight} = 0.000015588 \times \text{Length}^{3.0124}$$

The regression was applied to SBT lengths in units of centimetres (Harvey et al., 2011).

Table 1. Relative orientation parameters for the AM100 stereo–video system.

Item	Value	Precision
Base separation	703.4 mm	1.5 mm
Delta ω	−0.271°	0.005°
Left π	−5.106°	0.047°
Left κ	−3.293°	0.099°
Right π	5.844°	0.037°
Right κ	−2.998°	0.095°

Results

System calibration

The stereo–video camera system was calibrated with a network precision of 1:16 000 and an RMS image residual of 0.12 pixels, which is commensurate with previous research (Boutros *et al.*, 2015). The relative orientation parameters derived from the calibration are shown in Table 1. The base separation is the distance between the left and right camera perspective centres and determines the scale of all subsequent measurements. The parameters ω, π, and κ are Euler angles that denote the relative orientation of the left and right cameras with respect to the base line. The left and right π values indicate the design for a 10° convergence of the camera optical axes in order to optimize the stereo coverage (Harvey *et al.*, 2010). The values of the parameters are in reasonable accord with the specifications for the AM100 system. The precision levels are commensurate with previous experience with the AM100 and similar systems (Harvey and Shortis, 1998).

Scale bar measurements

The three lengths (513.8 mm; 750.7 mm; 1264.6 mm) on the scale bar that were measured after the initial calibration had accuracies and precisions better than 1 mm (514.0 ± 0.6 mm; 751.6 ± 0.9 mm; 1265.0 ± 0.4 mm) when compared to the known lengths over the 12 measurements made at varying distances. These results are commensurate with previous experience with similar systems (Harvey and Shortis, 1995; Harvey *et al.*, 2002, 2003).

Camera calibration stability

Measurements of the targets mounted on the front and the back of the transfer gate were made at five different times during each transfer. The mean measurement in each case was taken from the average of three replicate measurements in the same manner as the criteria set for measurements of the SBT. The average error across all transfers and all times was 0.34%. The mean measurement error did not exceed 1% of the total length between targets (see Table 2).

Numbers of fish measured

Of the 22 138 SBT recorded during the trial, it was possible to make three or more repeat length measurements of 11,647 SBT (52.6%) manually and 2,614 SBT (11.8%) semi-automatically.

Time required to measure the lengths of fish

The time required to measure the fish associated with each transfer varied depending on the number of fish transferred and the number of operators. On average, slightly more than 1 min was required to measure a fish manually three or more times and make the appropriate annotation. To manually measure the average fraction of 56% of the SBT per transfer required a duration of 19–30 person hours depending on the number of fish in the transfer.

In comparison, the semi-automated measurement process required between 6 and 16 s to make three or more acceptable measurements of a fish, including the initial two point-and-click identifications by the operator. The variation in measurement time was a result of differences in imaging and transfer conditions. Overall, computations took longer to run when the scene was cluttered due to a number of fish entering the gate at the same time. The total time elapsed was less than 3 h to process each transfer.

Length estimates

Of the eight transfers, there was only one transfer (Transfer E) where the mean lengths were significantly statistically different for manual (113.38 cm) and semi-automated (111.79 cm) measurements. There was no statistical difference in the mean lengths when all the transfers were pooled (108.07 cm manual vs. 107.5 cm semi-automated) (see Table 3). The same pattern was observed with weight, with the same transfer (E) having significant differences in weight, but again there was an insignificant difference of 1.25% when all weights were pooled (21.59 kg manual compared with 21.20 kg semi-automated).

Length frequency

There were no significant differences in the length frequencies of measurements made manually or semi-automatically for seven of the eight transfers, with good agreement between manual and semi-automatic measurements in the percent allocated to each 5 cm length class bin (Figure 5).

Discussion

Underwater stereo–video is widely used in aquaculture, marine ecology and fisheries management. The delay in the availability of length information associated with the time required to manually measure the lengths of fish in video images is a major disadvantage and inhibits many potential users from adopting this technology (Harvey *et al.*, 2003, 2011; Phillips *et al.*, 2008; Shortis *et al.*, 2009, 2013).

This research has demonstrated that it is possible to semi-automate length measurements of SBT reducing the number of operator interactions. This is achieved using an interactive process in which an operator first locates the tip of the snout and fork of the tail of a fish in one image of a stereo pair. Using epipolar geometry, the algorithm matches the snout and tail of a fish in the other image of the stereo pair and then tracks the fish through the proceeding video sequence images. The average length measurement is then calculated from three or more measurements and accepted or rejected based on whether the standard deviation exceeded 20 mm. Measurements with an RMS value of 10 mm or greater were rejected. The time required to measure one fish in one frame decreases from over one minute for manual measurements to approximately 6–16 s for semi-automated measurements. Additionally, once a fish is marked, the algorithm tested can track the fish automatically through the subsequent video sequence and continue to make repeated measurements of the same individual. In some cases, up to 25 measurements were

Table 2. The absolute percentage error for scale bar measurements mounted on the transfer gate for each of the eight transfers.

	Transfer A (%)	Transfer B (%)	Transfer C (%)	Transfer D (%)	Transfer E (%)	Transfer F (%)	Transfer G (%)	Transfer H (%)
F0	0.00	0.50	0.68	0.60	0.43	0.21	0.07	0.05
F25	0.00	0.57	0.65	0.70	0.67	0.14	0.18	0.04
F50	0.03	0.39	0.66	0.49	0.71	0.13	0.15	0.08
F75	0.04	0.49	0.58	0.55	0.60	0.10	0.10	0.12
F100	0.06	0.40	0.72	0.54	0.67	0.11	0.12	0.05
B0	0.68	0.04	0.10	0.02	0.03	0.69	0.55	0.70
B25	0.50	0.01	0.02	0.14	0.00	0.48	0.67	0.40
B50	0.68	0.13	0.04	0.09	0.00	0.77	0.75	0.44
B75	0.77	0.12	0.07	0.12	0.01	0.73	0.78	0.29
B100	0.97	0.12	0.07	0.05	0.05	0.66	0.62	0.54

F = targets mounted on the front of the gate, B = targets mounted on the back of the gate (see Figure 1). 0 = measurements taken at the start of the transfer, 25 = measurements taken a quarter of the way through the transfer, 50 = measurements taken halfway through the transfer, 75 = measurements taken three quarters of the way through the transfer, 100 = measurements taken at the completion of the transfer.

Table 3. Comparison of semi-automatic length measurement accuracy with that of the manual measurement for each transfer.

Transfer	A	B	C	D	E	F	G	H	Mean
Mean length (cm) Manual	105.07	106.00	108.10	113.12	113.40	106.52	104.91	106.59	108.07
Mean length (cm) Semi-automatic	105.54	106.31	107.92	112.72	111.81	105.26	105.53	107.02	107.50
One Way ANOVA length	No diff	No diff	No diff	No diff	Sig. diff	No diff	No diff	No diff	No diff
Maximum length (cm) Manual	139.93	134.62	135.97	154.32	145.45	144.97	142.99	145.95	
Maximum length (cm) Semi-automatic	135.53	130.89	137.18	189.22	138.53	133.75	139.72	144.56	
Minimum length (cm) Manual	56.43	55.44	76.84	74.53	84.54	63.01	76.05	75.30	
Minimum length (cm) Semi-automatic	76.80	79.79	84.52	80.87	90.27	71.21	76.14	76.86	
Kolmorgrov-Smirnov length frequency test	No diff	No diff	No diff	No diff	No diff	No diff	No diff	No diff	No diff
Mean weight (kg) Manual	19.96	20.32	21.45	24.88	24.72	20.71	19.64	20.51	21.59
Mean weight (kg) Semi-automatic	20.13	20.38	21.69	24.84	23.68	19.99	19.99	20.72	21.20
One Way ANOVA Weight	No diff	No diff	No diff	No diff	Sig. diff	No diff	No diff	No diff	
# SBT measured Manual	1459 (65.7%)	1118 (47.7%)	1386 (81.3%)	1616 (55.4%)	1430 (71.8%)	1757 (52%)	1291 (35.8%)	1590 (40%)	1455.8 (52.6%)
# SBT measured Semi-automatic	323 (14.5%)	191 (8.1%)	82 (4.8%)	202 (6.9%)	450 (22.6%)	416 (12.3%)	543 (15.1%)	407 (10.2%)	326.8 (11.8%)
Number of SBT counted	2222	2344	1705	2915	1993	3376	3607	3976	2767.2
Time Manual Measurement (min)	1380	1034	1362	1438	1260	1943	1434	1488	1417.4
Computation Time Semi-automatic measurement (min)	58	124	31	46	114	138	165	98	96.8

captured, depending on the number of video frames in which the fish was recorded, accounting for the range in measurement time per fish. Taking the average of a greater number of measurements should result in greater measurement accuracy and precision, however there was no pattern in the standard deviation. Recording a greater number of measurements increases the probability that a fish is not gliding with a straight body, but a greater number of repeat measurements is associated with slower, less vigorous swimming.

With the exception of one transfer, the mean length measurements did not differ between the manual and semi-automated measurements. When measurements were pooled across all transfers, there was no statistically significant difference in the mean lengths or weight, or the length frequencies.

Overall, 52.6% of the total SBT were measured manually, although many fish were rejected because the imagery or measurements did not meet the quality control criteria that have been set in place. It was reported by the image analyst that rejected fish were often not straight enough to be measured three times with a standard deviation less than 20 mm. Approximately, one third of the missed fish had a snout or a tail obscured by another fish.

Another 20% were swimming too rapidly for the video camera, preventing measurements in the minimum of three different video frames.

For the semi-automatic approach, the number of fish successfully measured was substantially lower than that of the manual approach due to additional rejections, primarily caused by low image matching confidence as a result of occlusion, spatially non-uniform illumination and other image disturbances. While the effects of spatially non-uniform illumination and other image disturbances could possibly have been decreased by having closer camera separations and having the stereo cameras parallel rather than inwardly converged, an engineering change of this nature would have an adverse impact on the area of stereo coverage for the camera system and would compromise measurement accuracy and precision (Boutros *et al.*, 2015), potentially to a level where it would not meet the measurement accuracy and precision thresholds set by the Australian industry and government. When a large number of fish simultaneously swim through the camera's field of view, partial occlusion of the fish as well as errors in matching snout/tail through image analysis substantially increase. These result in an increased number of automatic rejections of
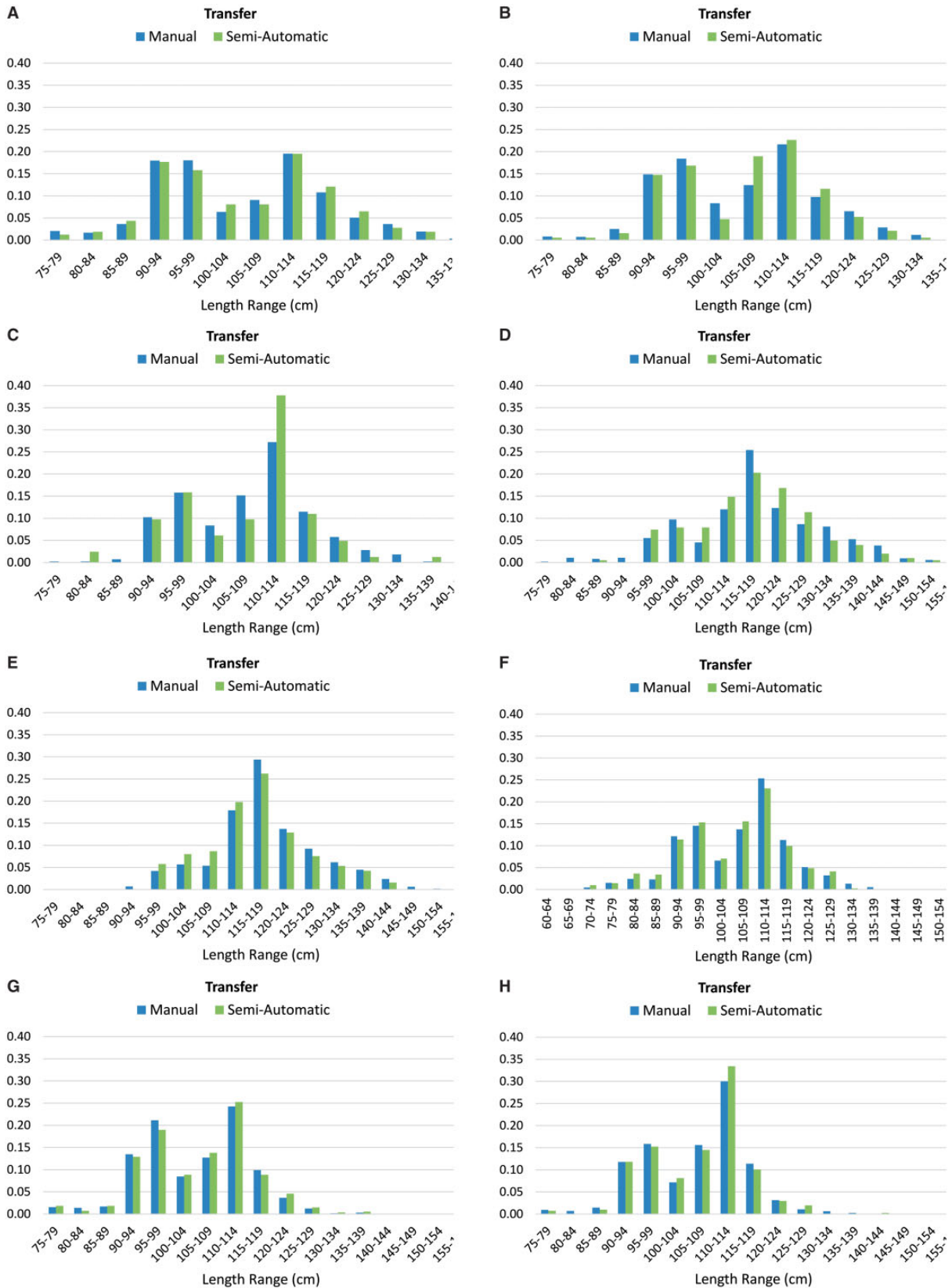
Figure 5. Length frequency histograms for manual and semi-automated measurements from eight transfers.

the length measurement based on the filtering criteria discussed in Section 2. Hence, the percentage of fish measured by the semi-automatic system decreases in this scenario. Inspection of the results revealed that Transfer E (see Table 3) had the highest percentage of fish measured by our system since a major proportion of the fish passed through the gate in isolation. Hence, measures taken to reduce the number of fish passing simultaneously through the gate (e.g. reducing gate size) will increase the percentage of fish measured by the system.

To be able to determine the total biomass of fish caught in a cage requires not only lengths of a statistically valid number of fish but also requires counts of the number of fish in a cage. Presently, counting is done manually. Phillips *et al.* (2008) suggest that at least 10% of a population or cage should be sampled. Counts are initially done in real time at sea to gain an approximate estimate of the number of fish being transferred into a cage to prevent overstocking as part of efficient farm management. Counts are done on land in a rigorous manner by representatives of the Australian government and company to enumerate the quota caught by a particular operator. Developing a fully automated system for counting SBT is challenging due to occlusions of one fish behind another, the rapid speed at which some fish swim through the transfer gate and changing light during a transfer, caused by changes in the position of the sun, boat, or changes in cloud cover. It is possible to reduce the speed at which fish swim and the number of fish coming through the gate simultaneously by not forcing them out of the tow cage by 'drying' the net out. It may be possible to gain greater control over the lighting by modifying the design of the gate and by implementing guidelines on camera location relative to the orientation of the sun. It may also be possible to develop an interactive fish counting system, where the image processing software identifies areas in a recording where occlusion may be occurring and an observer must count the number of fish manually.

A key question is whether this semi-automated measurement approach which has been developed for SBT in sea ranching could be adapted for surveys of wild fishes? Both diver operated stereo–video and baited remote (McLaren *et al.*, 2015) systems are becoming increasingly popular techniques for sampling demersal (Bornt *et al.*, 2015) and pelagic fishes (Santana-Garcon *et al.*, 2014a). It is possible that this approach could be used to decrease the measurement time, but for wild fish surveys the major component of analysis time is spent identifying species. Variable lighting, moving camera (for diver operated, drift and ROV stereo–video systems) and moving background (for example, kelp or seagrass in a temperate shallow water benthic habitat) will make the tracking and template matching particularly challenging. This is an area that needs to be prioritized for future research (Shortis *et al.*, 2016).

Conclusions

This paper has presented a method for, and the results of experiments in, the semi-automatic measurement of the lengths of fish. Based on imagery recorded using an underwater stereo–video system during transfers of SBT from tow cages to grow out cages, it has been demonstrated that the semi-automatic algorithm produces fork length measurements with an error of less than 1% of the true length and with at least a sixfold reduction in operator time when compared to manual measurements. For the majority of the transfers recorded, there were no statistical differences in the mean length, weight or length frequency between manual and semi-automatic measurements. Pooled data across all eight transfers show no statistical differences between the mean length and total biomass estimates from the two methods.

However, there are three limitations of the existing system: (i) the camera frame rate is not high enough to capture fast swimming fish, (ii) the lighting conditions pose difficulties for low dynamic range of the cameras, and (iii) some fish are swimming so strongly that their body deforms from a linear shape as they flex their muscles to drive themselves forwards quickly. The first two limitations can be addressed by using newer, faster, 12 bit cameras. The latter requires better models of the three dimensional deformations of the fish. All are tractable and will ensure that semi-automatic SBT surveys will provide improved results and replace manual methods in the future.

Notwithstanding these issues and the limitations of current technology and algorithms, fully automated systems, including the identification of species in the wild and counts of fish in aquaculture, remains a desirable and achievable aim (MacLeod *et al.*, 2010; Shortis *et al.*, 2016). Future systems for underwater fish measurement and biomass estimation will see advancement in the complete processing pipeline, starting from improved video cameras that are optimized for the task in terms of signal to noise ratio and frame rate, more accurate detection algorithms to initialize the tracking templates, and tracking algorithms that can more accurately predict the location of the fish snout and tail based on swimming action. The latter is also expected to improve stereo matches and the complete system will eventually provide a greater number of measurements per fish and thereby more accurate length estimates. When combined with species identification, in the future fully automated systems will be integrated with a variety of stereo–video systems in a variety of different environments and will be rapidly adopted for population monitoring in fisheries and for conservation management in the wild.

Acknowledgements

References

Anderson, M., Gorley, R., and Clarke, K. 2008. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. Primer-E Ltd., Plymouth, UK.

Bell, J. D., Craik, G. J. S., Pollard, D. A., and Russell, B. C. 1985. Estimating length frequency-distributions of large reef fish underwater. Coral Reefs, 4: 41–44.

Blank, J. M., Farwell, C. J., Morrissette, J. M., Schallert, R. J., and Block, B. A. 2007. Influence of swimming speed on metabolic rates of juvenile Pacific bluefin tuna and yellowfin tuna. Physiological and Biochemical Zoology, 80: 167–177.

Bloisi, D., and Iocchi, L. 2009. Argos – a video surveillance system for boat traffic monitoring in Venice. International Journal of Pattern Recognition and Artificial Intelligence, 23: 1477–1502.

Boom, B. J., He, J., Palazzo, S., Huang, P. X., Beyan, C., Chou, H. M., Lin, F. P. *et al.* 2014. A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage. Ecological Informatics, 23: 83–97.

Bornt, K. R., McLean, D. L., Langlois, T. J., Harvey, E. S., Bellchambers, L. M., Evans, S. N., and Newman, S. J. 2015. Targeted demersal fish species exhibit variable responses to long-term protection from fishing at the Houtman Abrolhos Islands. Coral Reefs, 34: 1297–1312.

Boutros, N., Shortis, M. R., and Harvey, E. S. 2015. A comparison of calibration methods and system configurations of underwater stereo-video systems for applications in marine ecology. Limnology and Oceanography: Methods 13: 224–236.

Chan, A. B., and Vasconcelos, N. 2012. Counting people with low-level features and Bayesian regression. IEEE Transactions on Image Processing, 21: 2160–2177.

Charalampidis, D., Gundam, M., Joginipelly, A., Quinteros, M., Ioup, G., Ioup, J., Yoerger, E. J. *et al.* 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

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

Duarte, C. M., Holmer, M., Olsen, Y., Soto, D., Marbà, N., Guiu, J., Black, K. *et al.* 2009. Will the oceans help feed humanity? BioScience, 59: 967–976.

Gruen, A., and Baltsavias, E. P. 1988. Geometrically constrained multiphoto matching. Photogrammetric Engineering and Remote Sensing, 54: 633–641.

Harvey, E. S., Cappo, M., Shortis, M. R., Buchannan, J., and Robson, S. 2003. The use of an underwater stereo-video system for measuring the lengths of caged Southern Blue Fin Tuna. Fisheries Research, 63: 315–326.

Harvey, E. S., Goetze, J., McLaren, B., Langlois, T., and Shortis, M. R. 2010. The 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: 75–85.

Harvey, E. S., Seager, J., Dodd, R., and Jones, T. 2011. Proof of the use of stereo-video for counting and measuring the lengths of Southern Bluefin Tuna in commercial trials. Australian Fisheries Management Authority RA/1/1451/2. Final Report, 46 pages.

Harvey, E. S., and Shortis, M. R. 1995. A system for stereo-video measurement of sub-tidal organisms. Marine Technology Society Journal, 29: 10–22.

Harvey, E. S., and Shortis, M. R. 1998. Calibration stability of an underwater stereo-video system: Implications for measurement accuracy and precision. Marine Technology Society Journal, 32: 3–17.

Harvey, E. S., Shortis, M. R., Stadler, M., and Cappo, M. 2002. A comparison of the accuracy and precision of digital and analogue stereo-video systems. Marine Technology Society Journal, 36: 38–49.

Jennings, S., and Kaiser, M. 1998. The effects of fishing on marine ecosystems. *In* Advances in Marine Biology, pp. 201–352. Ed. by J. H. S. Blaxter, A. J. Southward and P. A. TylerAcademic Press, London.

Jennings, S., and Polunin, N. V. C. 1997. Impacts of predator depletion by fishing on the biomass and diversity of non-target reef fish communities. Coral Reefs, 16: 71–82.

Kohler, N. E., Casey, J. G., and Turner, P. A. 1995. Length-weight relationships for 13 species of sharks from the western North Atlantic. Fishery Bulletin, 93: 412–418.

Langlois, T. J., Newman, S. J., Cappo, M., Harvey, E. S., Rome, B. M., Skepper, C. L., and Wakefield, C. B. 2015. Length selectivity of commercial fish traps by *in situ* comparisons with stereo-video surveys: is there any evidence of bias. Fisheries Research, 161: 145–155.

Levene, H. 1960. Robust tests for equality of variances. Contributions to probability and statistics: Essays in honor of Harold Hotelling, 1: 278–292.

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: 151–168.

Lipton, A. J., Fujiyoshi, H., and Patil, R. S. 1998. Moving target classification and tracking from real-time video. *In Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision*, WACV '98, pp. 8–14.

MacLeod, N., Benfield, M., and Culverhouse, P. 2010. Time to automate identification. Nature, 467: 156–155.

Mahmood, A., and Khan, S. 2010. Exploiting transitivity of correlation for fast template matching. IEEE Transactions on Image Processing, 19: 2190–2200.

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

McLaren, B. W., Langlois, T. J., Harvey, E. S., Shortland-Jones, H., and Stevens, R. 2015. A small no-take marine sanctuary provides consistent protection for small-bodied by-catch species, but not for large-bodied, high-risk species. Journal of Experimental Marine Biology and Ecology, 471: 153–163.

McLean, D., Green, M., Harvey, E. S., Williams, A., Daley, R., and Graham, K. J. 2015. Comparison of baited longlines and baited underwater cameras for assessing the composition of continental slope deepwater fish assemblages off southeast Australia. Deep Sea Research Part I: Oceanographic Research Papers, 98: 10–20.

Naiberg, A., Petrell, R. J., Savage, C. R., and Neufeld, T. 1993. Non-invasive fish size assessment method for tanks and sea cages using stereo–video. *In* Techniques for Modern Aquaculture. Ed. by J. K. Wang. American Society for Agricultural Engineering, MI, USA, pp. 372–381.

Naylor, R. L., Goldberg, R. J., Primavera, J. H., Kautsky, N., Beveridge, M. C., Clay, J., Folk, C. *et al.* 2000. Effect of aquaculture on world fish supplies. Nature, 405: 1017–1024.

Pauly, D., Christensen, V., Guenette, S., Pitcher, T. J., Sumaila, U. R., Walters, C. J., and Watson, D. R. 2002. Towards sustainability in world fisheries. Nature, 418: 689–695.

Petrell, R. J., Shi, X., Ward, R. K., Naiberg, A., and Savage, C. R. 1997. Determining fish size and swimming speed in cages and tanks using simple video techniques. Aquaculture Engineering, 16: 63–84.

Phillips, K., Boero Rodriguez, V., Harvey, E., Ellis, D., Seager, J., Begg, G., and Hender, J. 2008. Assessing the operational feasibility of stereo-video and evaluating monitoring options for the Southern Bluefin Tuna Fishery ranch sector. FRDC Project No. 2008/044: 42 pages.

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

Ramsay, J. M., Watral, V., Schreck, C. B., and Kent, M. L. 2009. Husbandry stress exacerbates mycobacterial infections in adult zebrafish, Danio rerio (Hamilton). Journal of Fish Diseases, 32: 931–941.

Ravanbakhsh, M., Shortis, M. R., Shafait, F., Mian, A., Harvey, E. S., and Seager, J. W. 2015. Automated fish detection in underwater images using shape-based level sets. The Photogrammetric Record, 30: 46–62.

Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., and Harvey, E. 2016. Fish species classification in unconstrained underwater environments based on deep learning. Limnology and Oceanography: Methods, 14: 570–585.

Santana-Garcon, J., Braccini, M. J., Langlois, T. J., Newman, S. J., McAuley, R. B., and Harvey, E. S. 2014a. Calibration of pelagic

stereo-BRUVs and scientific longline surveys for sampling sharks. Methods in Ecology and Evolution, 5: 824–833.

Santana-Garcon, J., Newman, S. J., Langlois, T. J., and Harvey, E. S. 2014b. Effects of a spatial closure on highly mobile fish species: an assessment using pelagic stereo-BRUVs. Journal of Experimental Marine Biology and Ecology, 460: 153–161.

Santana-Garcon, J., Newman, S. J., and Harvey, E. S. 2014c. Development and validation of a mid-water baited stereo-video technique for investigating pelagic fish assemblages. Journal of Experimental Marine Biology and Ecology, 452: 82–90.

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: 29–43.

Shafait, F., Mian, A., Shortis, M., Ghanem, B., Culverhouse, P., Edgington, D., Cline, D. *et al.* 2016. Fish identification from videos captured in uncontrolled underwater environments. ICES Journal of Marine Science, 73: 2737–2746.

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, 47, pp. 257–292. Ed. by R. N. Gibson, R. J. A. Atkinson and J. D. M. Gordon. CRC Press, Boca Raton, FL. 342 pp.

Shortis, M. R., Ravanbakhsh, M., Shafait, F., Harvey, E. S., Mian, A., Seager, J. W., Culverhouse, P. F. *et al.* 2013. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. *In* Proceedings of the SPIE 8791. Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection. Munich. Germany, pp. 87910G.

Shortis, M. R., Ravanbakhsh, M., Shafait, F., and Mian, A. 2016. Progress in the automated identification, measurement and counting of fish in underwater image sequences. Marine Technology Society Journal, 50: 4–16.

Siegel, S. 1956. Nonparametric Statistics for the Behavioral Sciences, 2nd edn. McGraw-Hill Book Company, Inc., New York.

Spampinato, C., Chen-Burger, Y., Nadarajan, G., and Fisher, R. 2008. Detecting, tracking and counting fish in low quality unconstrained underwater videos. *In* Proceedings of the third International Conference on Computer Vision Theory and Applications (VISAPP), 2, pp. 514–519.

Stauffer, C., and Grimson, W. E. L. 2000. Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22: 747–757.

Steeves, G., Peterson, R., and Clark, L. 1998. A quantitative stereoscopic video system for visually measuring the linear dimensions of free-swimming fish. *In* Proceedings of the Oceans'98 Engineering for Sustainable use of the Oceans, Nice, France, pp. 1405–1408.

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.

Zintzen, V., Anderson, M. J., Roberts, C. D., Harvey, E. S., Stewart, A. L., and Struthers, C. D. 2012. Diversity and composition of demersal fishes along a depth gradient assessed by baited remote underwater stereo-video. PLoS One, 7: e48522.

Handling editor: Howard Browman