Automatic Bluefin Tuna (*Thunnus thynnus*) biomass estimation during transfers using acoustic and computer vision techniques

V. Puig-Pons[a,*], P. Muñoz-Benavent[b], V. Espinosa[a], G. Andreu-García[b],
José M. Valiente-González[b], V.D. Estruch[a], P. Ordóñez[a], I. Pérez-Arjona[a], V. Atienza[b], B. Mèlich[c],
F. de la Gándara[d], E. Santaella[d]

[a] *Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC), Universitat Politècnica de València (UPV), C/ Paraninmf 1, 46730, Gandia, Spain*
[b] *Institute of Control Systems and Industrial Computing (AI2), Universitat Politècnica de València (UPV), Camí de Vera (s/n), 46022, València, Spain*
[c] *Balfegó & Balfegó S.L. Plaça Polígon Industrial, 26-24, 43860, L'Ametlla de Mar, Tarragona, Spain*
[d] *Instituto Español de Oceanografía (IEO), C/Varadero, 1, 30740, San Pedro del Pinatar., Murcia, Spain*

ABSTRACT

In this work, acoustic and computer vision techniques are combined to develop an automatic procedure for biomass estimation of tuna during transfers. A side scan sonar working at 200 kHz and a stereo camera, positioned facing towards the surface to record the ventral aspect of fish, are set as acquisition equipment. Moreover, a floating structure has been devised to place the sensors between cages in transfers, creating a transfer canal that allows data acquisition while fish swim from donor to receiving cage. Biomass assessment is computed by counting transferred tuna and sizing a representative sample of the stock. The number of transferred tuna is automatically deduced from acoustic echograms by means of image processing techniques, whereas tuna size is computed from the stereo videos using our automatic computer vision procedure based on a deformable model of the fish ventral silhouette. The results show that the system achieves automatic tuna counting with error below 10%, achieving around 1% error in the best configuration, and automatic tuna sizing of more than 20% of the stock, with highly accurate Snout Fork Length estimation when compared to true data from harvests. These results fulfil the requirements imposed by International Commission for the Conservation of Atlantic Tunas for compliant transfer operations.

1. Introduction

Modern Atlantic Bluefin Tuna (ABFT) (*Thunnus thynnus*) (L. 1758) farming based on adult capture in breeding areas has recently been developed. In the Mediterranean Sea, ABFT is fished from May to July by a purse-seining fleet, fattened for four to six months in sea cages and finally, sold in Japan, Europe and America (De La Gándara et al., 2016). ABFT populations are considered severely overfished and tuna ranching was criticized, arguing that it can produce pressure on the stock and increase the depletion rate. To reverse the trend, a 15-year recovery plan was adopted by the International Commission for the Conservation of Atlantic Tunas (ICCAT). ICCAT is responsible for managing ABFT (and other species), and its recovery plan enforces three guidelines: establish a TAC system (Total Allowable Catch) or fishing quotas, impose closed fishing seasons and an obligatory minimum catch size of 30

kg (ICCAT, 2006).

To control fishing quotas, authorities perform a biomass assessment during fish transfers: counting transferred tuna and sizing at least 20% of the stock. A transfer operation is any transfer of live Bluefin tuna from catching vessel net to transport cage; any transfer of live Bluefin tuna from transport cage to another transport cage; from one farm to another; from trap to transport cage and from transport cage or trap to farming cage. Each cage has a hole in the net that is used as a gate for transfers, allowing the fish to pass from the donor to the receiving cage.

ICCAT directives impose a stereoscopic video system as a sizing and counting tool (ICCAT, 2015). Video systems record the side aspect of tuna during transfers, and tunas are counted and their average length estimated from recorded videos. Human operators must manually mark fish snout and fork in both stereo images to estimate fish length. This process is slow and laborious, and introduces the variability of manual

measuring in the biomass estimation. ICCAT requirements about video systems accuracy are detailed in ICCAT (2015), this recommendation establishes that error inherent in the technical specifications of the stereoscopic camera system, which does not will exceed a range of +/- 5%. A difference of 10% or more between the amount of Bluefin tuna reported caught by the fisher and the amount established by the control system (control camera) or the impossibility of sizing at least 20% of the fish constitutes a non-compliant transfer operation that must be repeated.

Underwater computer vision systems have been frequently used for monitoring wild fish stock and inspection in aquaculture, as reported in recent reviews (Boutros et al., 2015; Hao et al., 2015; Mallet and Pelletier, 2014; Saberioon et al., 2017; Shortis et al., 2013; Zion, 2012), because it is a very appropriate non-intrusive method that permits work even when the fish are alive. In the particular case of stereoscopic vision systems (two cameras in a side-by-side arrangement), the following applications have been achieved: fish sizing (Ruff et al., 1995; Tillett et al., 2000; Lines et al., 2001; Harvey et al., 2003; Costa et al., 2006; Dunbrack, 2006; Torisawa et al., 2011), (Williams and Lauffenburger, 2016); fish counting and sizing (Costa et al., 2009; Rosen et al., 2013); fish farm automation (Martinez-de Dios et al., 2003); wild fish stock assessment (Willis and Babcock, 2000; Watson et al., 2009; Harvey et al., 2012; Langlois et al., 2012; Seiler et al., 2012; Smale et al., 2012; Zintzen et al., 2012; Wakefield et al., 2013; Santana-Garcon et al., 2014; McLaren et al., 2015). Nevertheless, vision sensors and image processing methods have to overcome limited visibility, temporal and spatial variations in lighting, varying distances and relative orientations between cameras and objects, motion and density of the monitored targets, and even lack of physical stability. All these conditions represent a very demanding challenge, which has limited the development of automatic commercial solutions. Some of these limitations, especially those related to limited visibility due to water turbidity or temporal and spatial variations in lighting, can lead to non-compliant transfer operations. An approach to partially overcome some of these limitations was suggested in Puig et al. (2012) and developed in Atienza-Vanacloig et al. (2016) and Muñoz-Benavent et al. (2018a): video images were taken from below of swimming tuna looking towards the sea surface, enhancing image contrast and detecting tuna silhouettes to perform stereoscopic measurements or single image measurements with distance-to-fish information obtained from synchronised echosounding.

A proposal using an acoustic sensor is described here to replace the computer vision system currently used to count fish, with the future ultimate goal of using acoustic techniques for both fish counting and sizing. Acoustical methods were introduced early on for fish counting in rivers (Menin and Paulus, 1974), and became the standard for non-intrusive monitoring of fish migratory movements, spawning runs and escapement from commercial or recreational fisheries (De Rosny and Roux, 2001; Enzenhofer et al., 1998; Mesiar et al., 1990; Ransom et al., 1998; Romakkaniemi et al., 2000). A combination of side scan sonar transducers was proposed to cover the transverse dimension of salmon migration (Trevorrow, 1997). Another wide-aperture acoustic device, the DIDSON acoustic technique (Dual frequency IDentification SONar), has been used to monitor migrating adult river herring (Magowan et al., 2012) and to count farmed fish during transfers, but in this case a small transfer gate size (1 m × 1 m) was used to obtain accurate results (Han et al., 2009b, 2009a).

This work presents a new automatic method for estimating tuna biomass (counting and sizing tuna) during transfers combining acoustic and computer vision systems from a ventral perspective of tuna, facing towards the sea surface from below the school. Both systems are previously calibrated, in the case of the vision sensor to estimate the intrinsic and extrinsic camera parameters with a checkerboard pattern, whereas in the case of the acoustic sensor, an on-axis calibration was carried out using a 38.1 tungsten carbide calibration sphere for a 200 kHz acoustic system (Simrad, 2008). Tuna counting is performed
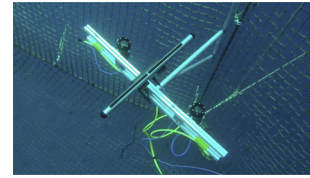

Fig. 1. Equipment for data acquisition composed of a 200 kHz side scan sonar and a stereo camera.

using acoustic techniques, whereas computer vision techniques based on a deformable model of the fish ventral silhouette are applied for tuna sizing. Moreover, a floating structure is proposed for the sensors. The transfer gate between cages has the maximum dimensions allowed by ICCAT (2015), that is a square of 10 m × 10 m, as preferred by ABFT ranchers to avoid possible entanglement of tuna during transfers. Our proposal can provide accurate measurements under real conditions and with minimal human intervention, as validated by comparing the results with ground truth data, fulfilling the requirements of ICCAT (2015). Ground truth data for sizing is obtained from measurements collected once the fish are harvested, whereas ground truth data for counting is provided by authorities' assessments during fish transfers.

2. Materials and methods

In real transfers, we only have one opportunity to obtain acoustic and visual data while tuna swim from donor to receiving cage. So having a controlled area, together with stability of the acquisition equipment and prevention of interferences of cage nets and auxiliary elements with the sensors, have been the main considerations to study. The equipment used to acquire acoustic and visual data is shown in Fig. 1 and explained below, in Section 2.1 and Section 2.2, respectively, and the different designs for a floating structure for the sensors are detailed in Section 2.3. Nine transfers were recorded with a side scan sonar and a stereo camera in Grup Balfegó facilities, an ABFT ranch off Spain's Mediterranean coast in Tarragona (4°052′11.7″ N and 0°48′15.2″ E).

2.1. Acoustic data acquisition

A 200 kHz side scan sonar operated by Simrad EK60 echosounder was employed using a transmitting power of 90 W, pulse length of 64 μs and 24 pings per second. These are the shortest pulse length and the maximum pulse repetition rate respectively, that EK60 can emit. Acoustic beam angle was 49 degrees perpendicular to the direction of swimming tuna and 0.5 degrees in swimming tuna direction. So, sonar generates an acoustic curtain that tunas must cross. The minimum pulse length is chosen to have an optimal spatial resolution to separate individual fish. The maximum ping rate would provide the maximum number of echoes from a single fish crossing the curtain. Time varied gain (TVG) is used in fisheries research in order to make the echo level independent of a target range. TVG function removes transmission loss effects from recorded data. Due to tuna is larger than the acoustic beam at the measuring range "20logR" was used as TVG function.

2.2. Stereo vision data acquisition

Video recordings were taken with a custom stereo camera, comprising two Gigabit Ethernet cameras, with 1720 × 1080 pixel resolution and framerate of 12 fps. The cameras are mounted in underwater housing, with a baseline of 85 cm and inward convergence of 5°. The system is rated to 40 m deep and has an umbilical cable that supplies Power over Ethernet to the cameras and transfers images to a logging computer, which encodes left and right videos. Camera synchronization is achieved using the IEEE 1588 Precision Time Protocol (PTP) (Eidson and Lee, 2002).

2.3. Floating structure for transfers

A floating structure was devised to place the sensors between cages in transfers, creating a transfer canal that allows data acquisition while fish swim from donor to receiving cage. Our floating structure was designed to meet the following requirements:

- Ensure stability of visual and acoustic sensors.
- Avoid handling the fish to avoid causing stress.
- Facilitate repeatability, by reducing the complexity of operations.
- Regarding gate size, achieve a compromise between ranchers' preferences and ICCAT requirements established in ICCAT (2015).
- Ensure full coverage of the gate, with both visual and acoustic sensors.
- Ensure physical stability of the structure and its assembly with the fish cages.
- Limit the maximum distance between gates to avoid creating a large tunnel and confusing the fish in their movement from donor to receiving cage.
- Minimize the presence of structure and cage elements in the cameras field of view to obtain a uniform background for the largest possible part of the image, that is, to maximize the uniform image background window.
- Minimize noise in acoustic data caused by the presence of structure and cage elements.

Three structure designs (S1, S2 and S3) has been proposed and tested in the Mediterranean Sea (see Fig. 2). A quadrangular prismatic structure is the base of S1 and S2. While in S1, sensors are centred at the bottom of the structure, S2 has an extensible mechanism to position the sensors inside the receiving cage. S3 is a pentagonal prismatic structure that allows modifications in the orientation of the visual sensors depending on requirements, to point towards the donor or receiving cage in order to minimise the background clutter. Each cage has a predefined

hole in the net that is used as a gate for the transfers. Fish farmers prefer gates to be as large as possible to avoid fish entanglement problems, so the maximum dimensions allowed by ICCAT (2015) are used, that is, a square of 10×10 m. The gates of the two cages in the transfers are attached to the rectangle in the upper half of the structure, creating a sort of tunnel between cages. Transducers and stereo camera are mounted 6 m deeper than the lowest side of the gate cage, facing towards the surface, to obtain a ventral silhouette of the fish and to cover the entire gate area with the acoustic beam (dark-blue volume in Fig. 2) and camera field of view (light-cyan volume in Fig. 2). From the authors' point of view, a smaller gate would allow the equipment to cover the entire gate area from closer to the lowest side of the gate, which would bring more accuracy in fish sizing and a less bulky structure, but farmers' preferences prevailed to avoid fish entanglement. This camera arrangement has three advantages: first, with this orientation, sunlight acts as a backlight system so objects are always darker than water; second, in this set up, body bending can be clearly appreciated and dealt with; third, the most reliable measurements are obtained when the fish are swimming in a plane orthogonal to the visual axis (Dunbrack, 2006). However, the problem of not seeing the fork of the tail from this perspective needs to be addressed. The solution adopted is detailed in Section 2.5. Ventral acoustic measurements are obtained when tuna cross the acoustic curtain.

2.4. Acoustic data processing for fish counting

The proposal developed for automatic tuna counting is based on the application of image processing techniques to acoustic echograms and is summarized in Fig. 3. An acoustic echogram (Fig. 4a) is generated for each transfer with the equipment and setup described in the previous sections. The acoustic beam insonifies the fish as they cross the acoustic curtain, generating traces in the echogram. The image processing algorithms must deal with different complex shapes of the traces, because they depend on multiple factors, such as fish swimming velocity,
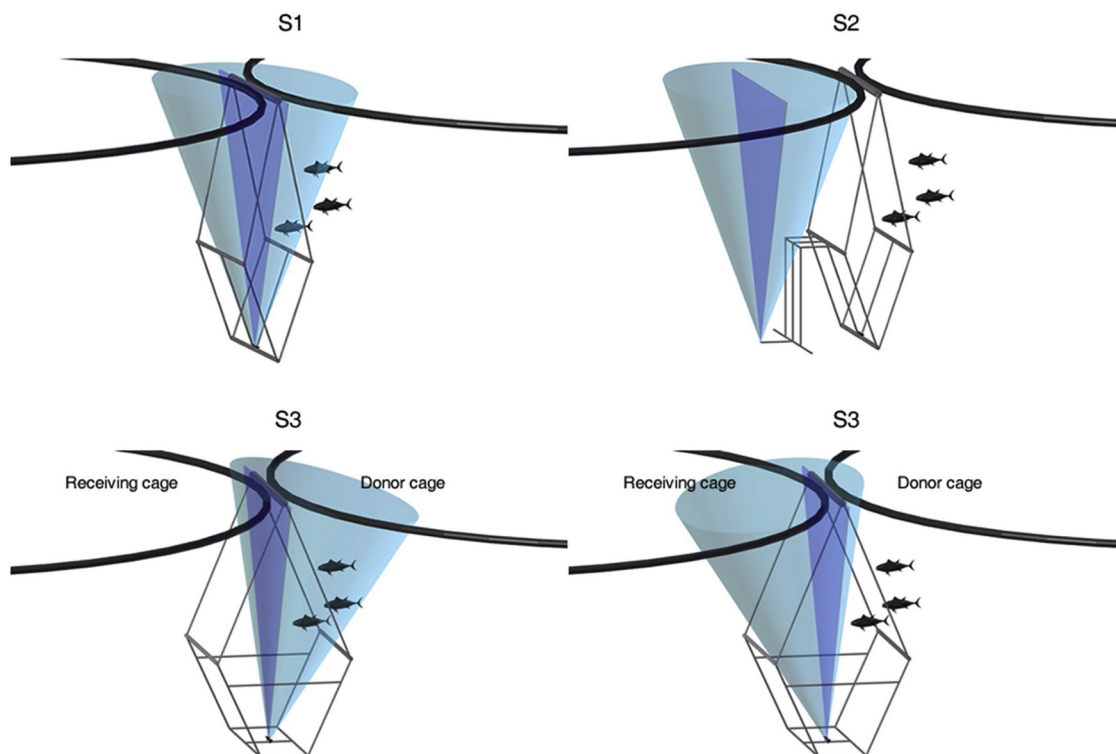

Fig. 2. Design of the proposed and tested floating structures. The top of the cages is represented in black, camera field of view is represented in light-cyan and acoustic beam angle in dark-blue. S1: quadrangular prismatic structure with the sensors centred at the bottom. S2: quadrangular prismatic structure with an extensible mechanism to position the sensors inside the receiving cage. S3: pentagonal prismatic structure allowing modifications in the orientation of the sensors.
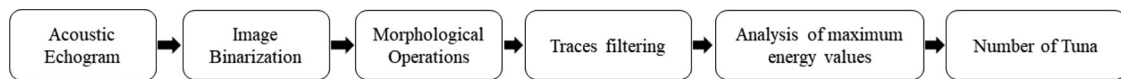
Fig. 3. Sequence of echogram processing algorithms involved in the process of ABFT counting.

insonification angle and distance, among others.

In the first step, the echogram is converted to a binary image using the threshold level defined by Otsu's method (Otsu, 1979). Then, a sequence of morphological operations is applied: thickening to provide more compact traces, open to remove protrusions (noise) and break weak connections and close to smooth out contours and fill small holes (Fig. 4d). Afterwards, traces are geometrically characterized and filtered using two criteria: acoustic backscattering (traces with a maximum backscattering value below a predefined threshold are discarded) and region properties (traces out of predefined ranges for area, convex Hull area, Euler number and solidity (Solomon and Breckon, 2011) are discarded (Fig. 4e and f).

Finally, an analysis of the maximum backscattering values distribution is applied to identify the number of fish in each of the remaining traces. For each trace, the maximum backscattering values set is built from the maximum backscattering value of each ping, that is, of each column, and a Savitzky-Golay filter (Savitzky and Golay, 1964) is applied before evaluating the number of local maxima. The Savitzky-Golay method is based on calculating a local polynomial regression to generate a smoothed function of the input data. The main advantage of this approach is that it tends to preserve characteristics of the initial distribution such as relative maxima and minima, as well as the width of the peaks, which normally disappear with other averaging techniques. Each local maximum obtained after filtering is identified as a fish passing through the acoustic curtain. As an example, Fig. 4g shows the maximum backscattering values distribution for the trace in Fig. 4b, which is generated when the fish crosses the acoustic curtain (yellow rectangle in Fig. 4h). As only one local maximum is found, only one fish is counted. However, if two targets inside the beam are at the same transducer range, echo amplitudes from each of them are added and the echo value received is increased when both acoustic signals are in phase. To consider this phenomenon, an acoustic backscattering threshold is established to decide when high backscattering values correspond to two fish. The appropriate threshold has been established from observation of the different transfers to the 95th percentile of maximum backscattering values distribution.

2.5. Stereo vision data processing for fish sizing

The computer vision algorithms involved in the process of fish sizing are summarized in Fig. 5. Image segmentation was implemented using local thresholding (Petrou and Petrou, 2011), a region-based technique for extracting compact regions (blobs) on each video frame, and morphological operations. The segmented blobs are geometrically characterized and sifted using shape (aspect ratio), pixel density and dimensional filters. An edge detection algorithm is then applied and a minimization algorithm is used to fit a deformable tuna model. A Fitting Error Index (FEI), based on the quadratic distance between the model points and target edges points, is computed to analyse the goodness of the fitting. FEI takes values between 0 and 10, where FEI = 0 denotes a perfect fit between the segmented blob and the geometric model. Fittings with high values (FEI > 3) are discarded. The results for left and right videos, obtained separately, are merged to calculate SFL of the fish. The image plane information is transformed to 3D measurements using the calibration parameters of the stereoscopic vision system and 3D triangulation. Samples are discarded if the stereo correspondence is not met for the first and last model vertebrae, that is, if the distance from the points to the epipolar lines is greater than a threshold. The problem of not seeing the fork of the tail from the ventral perspective is addressed by excluding the caudal fin from the

deformable tuna model and applying the relation SFL= 1.0312 ML+ 0.065641, deduced from experimental samples in (Muñoz-Benavent et al., 2018a), where ML is the Model Length. See Atienza-Vanacloig et al. (2016) and (Muñoz-Benavent et al. (2018a) for further details on the computer vision algorithms.

2.6. System validation

Nine transfers were recorded in Grup Balfegó facilities and the automatic procedures for fish counting and sizing presented in the previous sections were applied. Table 1 shows a summary of measurements of the nine transfers, and for each transfer, the floating structure design used, the automatic counting and sizing results and the ground truth data. To reduce tuna stress levels and optimise costs of in-situ tests, it was decided to make two-way transfers whenever possible, in which the donor and receiving cages were exchanged. For the first transfer, T1.1 refers to the one-way and T1.2 to the return situation, and similarly for the other three two-way transfers. T5 is a one-way transfer, because it was not an ad-hoc operation, but a routine operation of the ranchers. Fig. 6 shows, on the left, the S3 floating structure on the quay of the port ready to be transported to the cages and, on the right, a surface view of its position between two cages during transfers.

True data is composed of the number of transferred tuna and SFL statistical information (SFL mean, S$\bar{F}$L, and SFL frequency distribution). On the one hand, ground truth data for sizing is obtained from measurements collected by Grup Balfegó, which measures the fish on deck with callipers after harvests. Although the entire fish stock could not be measured at harvesting, the sample is considered statistically significant. On the other hand, ground truth data for counting is provided by authorities' assessments during fish transfers, as all transfer operation must be controlled by authorities. The same cage is always used for transfers, but the number of transferred tuna decreases because some of them are harvested to satisfy client demand.

The results of automatic acoustic counting are compared with the number of transferred tuna. During the counting process, different sizes of the *M* x *N* rectangular structuring element for morphological operations are used. The criteria for selecting the matrix size is based on tuna behaviour during transfer: swimming velocity and swimming tilt. High tuna swimming velocities entails the use of a matrix with smaller *N* values (which means fewer columns). However, when swimming tilt increases, tuna traces size increases too and the number of rows (*M*) and columns (*N*) has to be larger. To automate the counting procedure, a compromise solution is employed considering both situations.

Regarding the automatic visual sizing results, differences in S$\bar{F}$L between harvests and automatic sizing are examined with analysis of variance tests. Since the two groups have unequal sample sizes and homoscedasticity (homogeneity of variance) cannot be ensured, Welch's ANOVA test (Welch, 1951) is used, as recommended in Rasch et al. (2011) and McDonald (2014). Differences in SFL frequency distributions are analysed with the Kolmogorov-Smirnov test (Massey, 1951).

3. Results

T1 transfers were carried out using the S1 floating structure. Automatic acoustic counting errors below 10% were obtained, but automatic visual sizing was impossible due to poor water visibility, as can be observed in Fig. 7a. In T1.2, error is around 10% due to a line going through fish groups in the echogram (see Fig. 8), which hindered the counting process and increased counting error. That line is the result of
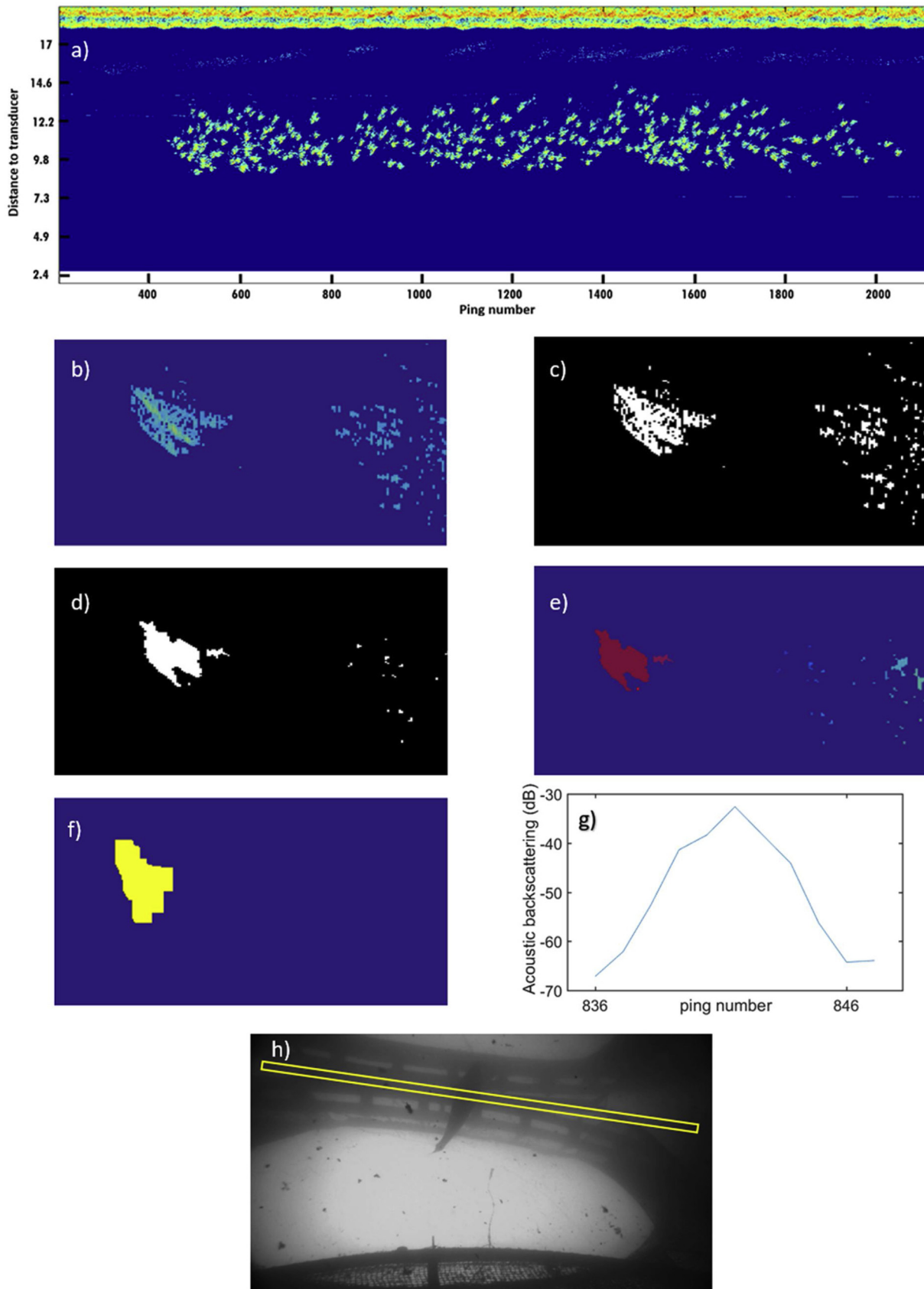
Fig. 4. a) Transfer echogram. b) Trace extracted from transfer echogram. c) Binary image obtained from echogram. d) Result to apply morphological operations (removing noise and filling small holes). e) Blobs labelling and characterization. f) Segmentation process result (e and f correspond to traces filtering process). g) Maximum backscattering values distribution for the trace. h) Video frame when tuna crosses the acoustic curtain (yellow rectangle in the image).
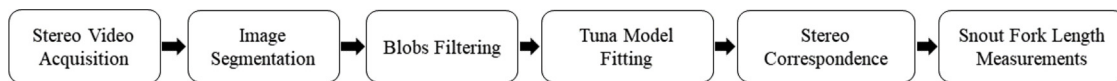
Fig. 5. Sequence of stereo video processing algorithms involved in the process of ABFT sizing.

Table 1
Summary of measurements with the nine transfers (four two-way transfers and one one-way transfer). For each transfer: floating structure design used, ground truth data (transferred tuna and Snout Fork Length mean, $S\bar{F}L$, from harvests after transfers date), automatic acoustic fish counting and automatic visual fish sizing results.

Transfers	Floating structure	Ground truth			Automatic acoustic fish counting (error %)	Automatic visual fish sizing			
		Transferred tuna	harvests (% from total)	$S\bar{F}L$		# measurements (% from total)	$S\bar{F}L$(error %)	Welch's ANOVA p-value	Kolmogorov-Smirnov p-value
T1.1	S1	608	–	–	574 (5.6%)	Poor water visibility	–	–	–
T1.2		608			549 (9.7%)				
T2.1	S1	577	–	–	Bad sea conditions	25 (4.3%)	–	–	–
T2.2		577				13 (2.3%)	–		
T3.1	S2	577	–	–	554 (4.0%)	Cameras damaged during operation	–	–	–
T3.2		577			535 (7.3%)				
T4.1	S2	577	304 (58.9%)	2.22	575 (1.0%)	212 (36.9%)	2.38 (7.2%)	0	0
T4.2		577			570 (1.2%)	378 (65.7%)	2.39 (7.7%)	0	0
T5	S3	503	230 (45.7%)	2.23	493 (2.0%)	115 (22.9%)	2.24 (0.4%)	0.632	0.587

Fig. 6. Left, the S3 floating structure on the quay of the port ready to be transported to the cages. Right, surface view of its position between two cages during transfers.

returned echo from reinforcement bars on the floating structure, which were removed for T2 transfers.

In T2 transfers, sea conditions made automatic acoustic fish counting impossible and a very problematic image segmentation was revealed with S1. The top of the floating structure was visible in the image and covered a large part of the image background, as can be seen in Fig. 7b, making it very difficult to segment tunas in the foreground from the floating structure in the background. Only 4.3% and 2.3% of the fish could be automatically measured. Other segmentation processes rather than local thresholding could be studied, but it is not straightforward, so we decided to redesign the floating structure. To avoid having parts of the floating structure in the camera field of view, and thus achieve a uniform image background, an extensible mechanism was added to place the sensors inside the receiving cage (S2).

With the S2 floating structure, two two-way transfers were conducted, T3 and T4. In T3, automatic acoustic counting errors of 4% and 7.3% were obtained, being greater in T3.2 due to two factors. First, divers' position during transfer affected data recording and second, a few tunas swam back to the donor cage and the counting algorithm was not prepared for such an eventuality. An improvement was made to the transfers' measurement procedure, fixing divers' position during the recordings. The algorithm was also modified to increase the degree of automation. As a result of these actions, on T4 transfers automatic acoustic counting error decreased to around 1%. Regarding the visual procedure, the stereo video could not be recorded in T3 because cameras were damaged during the operation, but recording was possible in T4, so automatic visual sizing measurements were extracted. In particular, a large amount of automatic visual sizing measurements was obtained, 36.9% of the fish (212 out of 575) and 65.7% of the fish (378 out of 575) for T4.1 and T4.2, respectively. However, statistical comparison between automatic visual sizing measurements and ground truth data from harvests gives poor results: there is a difference in $S\bar{F}L$
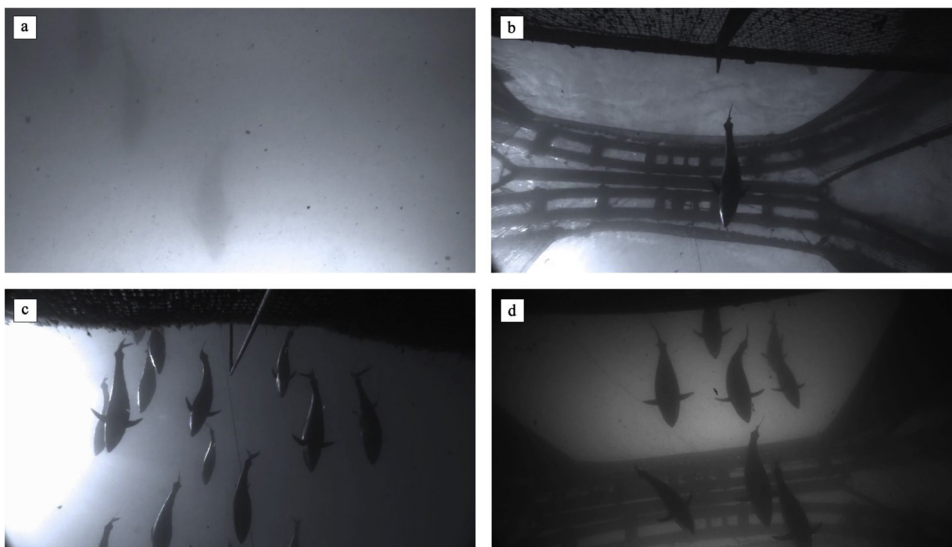

Fig. 7. (a) Video capture showing the poor water visibility that made it impossible to obtain automatic visual sizing measurements in T1 transfers. (b) Video capture showing the S1 floating structure visible in the image background, which strongly hinders automatic visual measurements in T2 transfers. (c) Video capture with S2 floating structure, which places the sensors inside the receiving cage and allows a large uniform image background window in T4 transfers. (d) Video capture with S3 floating structure, which enlarges the uniform image background window in T5 transfer, compared to S1 (Fig. 7b).
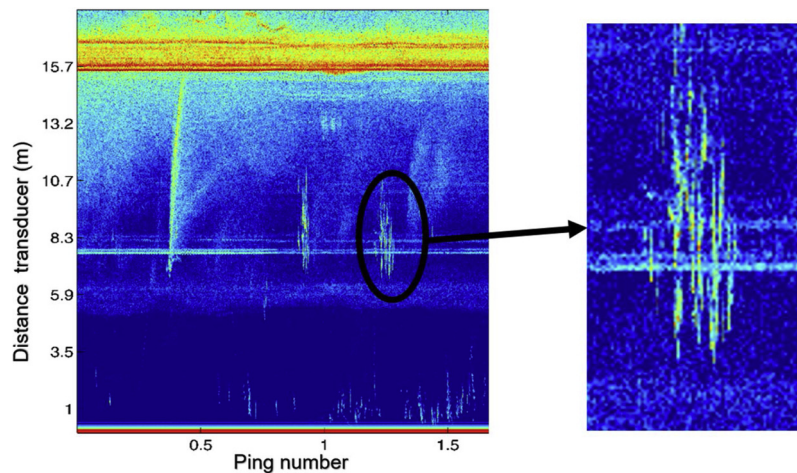
Fig. 8. Left: Echogram of T1.2 transfer (left). Right: Zoom of a fish group, where it can be observed that echoes from the reinforcement bar of the structure generate a line in the echogram that hinders the automatic acoustic counting process.

of 6 and 7 cm, for T4.1 and T4.2, respectively, and the statistical tests for $S\bar{F}L$ and SFL distribution frequency give null p-values, indicating a statistically significant difference between ground truth and automatic measurements. Analysis of the video frame timestamps revealed a non-negligible time desynchronization between left and right cameras, which was solved with an update of the camera firmware and acquisition software for the last T5 transfer.

With the S2 floating structure, good results in terms of automatic acoustic counting and amount of automatic visual sizing measurements were achieved, because the design avoids any source of noise that could disturb the acoustic sensor and a uniform background is accomplished in most of the image, facilitating the image segmentation process. However, its use was strongly discouraged mainly for two reasons: firstly, repeatability in the position and orientation of the sensors cannot be ensured with the extensible mechanism, and secondly, the structure is not self-contained, which hinders the divers' setup operations and can cause damage to the fish stock. The extensible mechanism was removed in S3 and the floating structure was redesigned to enlarge the uniform background window in the image. This is accomplished by tilting the cage doors and orientating the cameras to point to the donor cage, but only by a few degrees to keep a ventral view of the fish. A larger uniform background window is achieved in S3 (Fig. 7d) compared to S1 (Fig. 7b).

Structure design (T1 and T2 transfers), divers' position (T3 transfers) and faulty camera synchronisation (T4 transfers) affected measurements by decreasing the automation possibilities. These problems were solved for the last transfer (T5), where the automatic acoustic counting procedure obtained 2% error and the automatic visual sizing procedure obtained 0.4% error (1 cm) in $S\bar{F}L$ with 22.9% of the transferred fish automatically measured, fulfilling the requirements of ICCAT (2015). As Table 1 shows, the tests for $S\bar{F}L$ and SFL distribution frequency give p-values higher than the 5% significance level. In conclusion, there is no statistically significant difference between ground truth and automatic measurements, thereby validating the measurements obtained with the proposed automatic system. Fig. 9 shows the normalized SFL frequency histograms of the automatic visual sizing measurements and ground truth data for T4 and T5 transfers.

4. Discussion

ICCAT established in ICCAT (2015) that at least 20% of transferred tunas must be sized and counted using a stereoscopic system. The most widely used commercial systems are AQ1 AM100 (Phillips et al., 2009) and AKVAsmart, formerly VICASS (Shieh and Petrell, 1998), but both have a significant limitation, they require human interaction. In both

systems, human operators must inspect the videos and select frames in which the fish is isolated and straight, then fish snout and fork have to be manually marked with a mouse click on both stereo images to estimate fish Snout Fork Length (SFL). This process is slow and laborious, and introduces the variability of manual measuring in the biomass estimation. Instead, in this work we propose an automatic sizing procedure based on computer vision techniques. The automatic process is the main difference of this work with respect to other studies with similar goals, such as Lines et al. (2001), (Harvey et al. (2003) and Shafait et al. (2017). These authors obtained good results measuring fish lengths with stereovision systems, but their proposals have one or several of the following common limitations: measurements are not extracted automatically, measurements are taken in a narrow range, or the ground truth comprises only a few samples. In fact, Lines et al. (2001) work with only 60 images of 17 fish and measure in a range from 1 to 2 m. (Harvey et al. (2003) compare their results with harvests of 54 Southern Bluefin Tuna (SBT) from thousands in the cage and measuring in a range of about 1 m. Shafait et al. (2017) use a semiautomatic method to measure SBT that requires human intervention to locate the tip of the snout and fork of the tail of the fish.

In our case, minor human intervention is required, to discard repetitive measurements and to setup morphological operations parameters. The resulting SFL measurements are supervised to discard repetitive measurements of the same fish, so the number of measured fish can be compared with ICCAT requirements. For the final automation procedure, tracking algorithms are being developed to automate this decision. In the same way, intervention is required to decide M and N values of structure element used in morphological operations during counting. Study and characterisation of tuna behaviour during transfer process would allow to improve counting automation degree. The full automation of the process will be developed in a new version.

Standards and procedures for stereoscopic cameras systems imposed by ICCAT do not make any explicit reference to the camera operator position during transfer recording, what hinders procedure repeatability. Instead, we designed a floating structure to place the sensors between cages, which allowed installation of the sensors in a stable position and enhanced the repeatability of measurement conditions. Moreover, the structure minimizes the effects of adverse weather conditions, serving as an anchor between donor and receiving cages.

The proposed system has been evaluated performing nine transfers under controlled conditions, which propitiated the improvement of the system from the experience. In the last transfer, the automatic procedure was validated, since it achieved accurate automatic ABFT counting and sizing, fulfilling the requirements imposed by ICCAT. Regarding the acoustic sensor, the ventral aspect is related to great backscattering
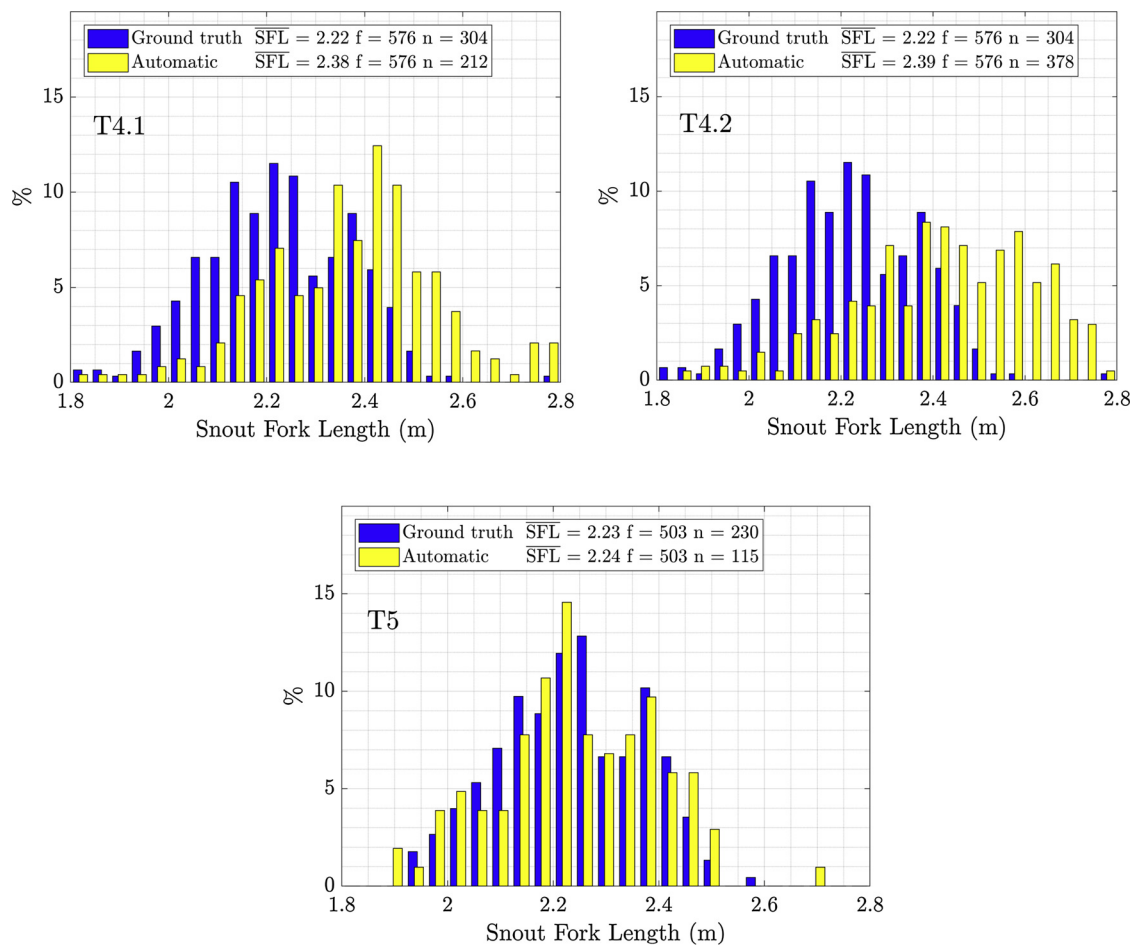
Fig. 9. Normalized SFL frequency histogram. Ground truth in dark-blue and automatic measurements in light-yellow. $\overline{\text{SFL}}$, mean SFL ; f, number of fish; n, number of samples.

level due to fish swimbladder (Foote, 1985). Moreover, the short pulse used provides good range resolution, which implies that two tuna can be detected and counted when they are separated more than 10 cm. This low value allows isolating the great part of traces, as validated with respect to ground truth data. Regarding the vision sensor, the accuracy of the procedure is validated with respect to ground truth from harvests, as it was previously proved also in Muñoz-Benavent et al. (2018a) and Muñoz-Benavent et al. (2018b) with respect to manual measurements and ground truth data from harvests.

Future development of a combined acoustical and vision system should offer complementary or even redundant information of both counting and sizing, fulfilling the desired accuracy requirements in any operational situation or environmental circumstance. Note that acoustic sensors have an advantage over vision sensors, since they are not affected by water turbidity. In fact, fish sizing with the current vision system was impossible to perform in T1 transfer (Fig. 7a) due to poor visibility. Acoustical sizing could be achieved with different strategies based either on the measurement of individual backscattered energy (Burwen et al., 2003; Simmonds and MacLennan, 2007) or of vertical dimensions of fish (Soliveres et al., 2017), both improved by using broadband echosounding and pulse compression techniques (Chu and Stanton, 1998).

5. Conclusions and further work

An automatic method for biomass estimation using acoustic and computer vision techniques has been developed. A new setup with a 200 kHz Side Scan Sonar transducer and a stereo camera was used to record transfers from a tuna ventral view.

Regarding acoustic information, the Side Scan Sonar generated an acoustic curtain that tuna crossed during transfers and a counting algorithm was implemented to obtain the number of transferred tuna number with an automatic process, without operator intervention. However, the size of the structuring element used in morphological operations is conditioned by tuna behaviour and the selection of an appropriate value should be improved by the use of decision algorithms (machine learning), which would allow more automation of the counting process.

Regarding visual information, an automatic computer vision procedure based on a deformable model of the tuna body was used to size a statistically representative number of fish. Biomass is commonly estimated using length-weight relations, Lines et al. (2001) and (Martinez-de Dios et al. (2003). The mathematical model between fish length (L) and mass (W) is $W = aL^b$, where a and b are empirically characterized species and strain-dependent parameters (Zion, 2012). Recent studies attempt to show that biomass can be better estimated if fish measurements in dimensions other than length, like width and depth, are available (Aguado-Gimenez and Garcia-Garcia, 2005; Harvey et al., 2003). Nevertheless, as stated in Harvey et al. (2003), measuring the width of a fish is relatively subjective due to the lack of defined points in the fish silhouette. We are currently working on finding the best way of estimating fish width and generating a ground truth dataset that includes width measurements at harvesting.

Acknowledgements

References

Aguado-Gimenez, F., Garcia-Garcia, B., 2005. Growth, food intake and feed conversion rates in captive Atlantic bluefin tuna (Thunnus thynnus Linnaeus, 1758) under fattening conditions. Aquacult. Res. 36, 610–614. https://doi.org/10.1111/j.1365-2109.2005.01210.x.

Atienza-Vanacloig, V., Andreu-García, G., López-García, F., Valiente-Gonzalez, J.M., Puig-Pons, V., 2016. Vision-based discrimination of tuna individuals in grow-out cages through a fish bending model. Comput. Electron. Agric. 130, 142–150. https://doi.org/10.1016/j.compag.2016.10.009.

Boutros, N., Shortis, M.R., Harvey, E.S., 2015. A comparison of calibration methods and system configurations of underwater stereo-video systems for applications in marine ecology. Limnol. Oceanogr. Methods 13, 224–236. https://doi.org/10.1002/lom3.10020.

Burwen, D.L., Fleischman, S.J., Miller, J.D., Jensen, M.E., 2003. Time-based signal characteristics as predictors of fish size and species for a side-looking hydroacoustic application in a river. ICES J. Mar. Sci. 60, 662–668.

Chu, D., Stanton, T.K., 1998. Application of pulse compression techniques to broadband acoustic scattering by live individual zooplankton. J. Acoust. Soc. Am. 104, 39. https://doi.org/10.1121/1.424056.

Costa, C., Loy, A., Cataudella, S., Davis, D., Scardi, M., 2006. Extracting fish size using dual underwater cameras. Aquacult. Eng. 35, 218–227. https://doi.org/10.1016/j.aquaeng.2006.02.003.

Costa, C., Scardi, M., Vitalini, V., Cataudella, S., 2009. A dual camera system for counting and sizing Northern Bluefin Tuna (Thunnus thynnus; Linnaeus, 1758) stock, during transfer to aquaculture cages, with a semi automatic artificial neural network tool. Aquaculture 291, 161–167. https://doi.org/10.1016/j.aquaculture.2009.02.013.

De La Gándara, F., Ortega, A., Buentello, A., 2016. Tuna aquaculture in Europe. In: Benetti, D.D., Partridge, G.J., Buentello, A. (Eds.), Advances in Tuna Aquaculture. Elsevier, San Diego, pp. 115–157. https://doi.org/10.1016/B978-0-12-411459-3.00005-9.

De Rosny, J., Roux, P., 2001. Multiple scattering in a reflecting cavity: application to fish counting in a tank. J. Acoust. Soc. Am. 109, 2587–2597.

Dunbrack, R.L., 2006. In situ measurement of fish body length using perspective-based remote stereo-video. Fish. Res. 82, 327–331. https://doi.org/10.1016/j.fishres.2006.08.017.

Eidson, J., Lee, K., 2002. IEEE 1588 standard for a precision clock synchronization protocol for networked measurement and control systems. Sensors for Industry Conference, 2002 98–105.

Enzenhofer, H.J., Olsen, N., Mulligan, T.J., 1998. Fixed-location riverine hydroacoustics as a method of enumerating migrating adult Pacific salmon: comparison of split-beam acoustics vs. visual counting. Aquatic. Living Resour. 11, 61–74.

Foote, K.G., 1985. Rather-high-frequency sound scattering by swimbladdered fish. J. Acoust. Soc. Am. 78, 688–700.

Han, J., Asada, A., Mizoguchi, M., 2009a. DIDSON-based acoustic counting method for juvenile Ayu Plecoglossus altivelis migrating upstream. J. Marine Acoust. Soc. Japan 36, 250–257. https://doi.org/10.3135/jmasj.36.250.

Han, J., Honda, N., Asada, A., Shibata, K., 2009b. Automated acoustic method for farmed Fish counting and sizing during its transfer using DIDSON. Sixth InErnational Symposium on Underwater Technology. pp. 363–368.

Hao, M., Yu, H., Li, D., 2015. The measurement of Fish size by machine vision - a review. Li, D., Li, Z. (Eds.), Computer and Computing Technologies in Agriculture IX - 9th IFIP WG 5.14 International Conference, CCTA 2015 15–32. https://doi.org/10.1007/978-3-319-48354-2_2.

Harvey, E., Cappo, M., Shortis, M., Robson, S., Buchanan, J., 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. Fish. Res. 63, 315–326. https://doi.org/10.1016/S0165-7836(03)00080-8.

Harvey, E., Butler, J.J., McLean, D.L., Shand, J., 2012. Contrasting habitat use of diurnal and nocturnal fish assemblages in temperate Western Australia. J. Exp. Mar. Biol. Ecol. 426, 78–86. https://doi.org/10.1016/j.jembe.2012.05.019.

ICCAT, 2006. Recommendation by ICCAT to Establish a Multi-Annual Recovery Plan for Bluefin tuna in Eastern Atlantic and Mediterranean. Rec.[06-05]. .

ICCAT, 2015. Recommendation by ICCAT amending the recommendation 13-07 by ICCAT to establish a multi-annual recovery plan for Bluefin Tuna in the eastern Atlantic and Mediterranean. Rec [14-04]. 2015 Compendium Management Recommendations and Resolutions Adopted by ICCAT for Conservation of Atlantic Tunas and Tuna-Like Species. pp. 47–82.

Langlois, T.J., Harvey, E.S., Meeuwig, J.J., 2012. Strong direct and inconsistent indirect effects of fishing found using stereo-video: testing indicators from fisheries closures. Ecol. Indic. 23, 524–534. https://doi.org/10.1016/j.ecolind.2012.04.030.

Lines, J.A., Tillett, R.D., Ross, L.G., Chan, D., Hockaday, S., McFarlane, N.J.B., 2001. An automatic image-based system for estimating the mass of free-swimming fish. Comput. Electron. Agric. 31, 151–168. https://doi.org/10.1016/S0168-1699(00)00181-2.

Magowan, K., Reitsma, J., Murphy, D., 2012. Use of dual-frequency identification sonar to monitor adult River herring in a small coastal stream. Marine Coastal Fish. 4, 651–659. https://doi.org/10.1080/19425120.2012.730916.

Mallet, D., Pelletier, D., 2014. Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012). Fish. Res. 154, 44–62. https://doi.org/10.1016/j.fishres.2014.01.019.

Martinez-de Dios, J.R., Serna, C., Ollero, A., 2003. Computer vision and robotics techniques in fish farms. Robotica 21, 233–243. https://doi.org/10.1017/S0263574702004733.

Massey, F.J., 1951. The Kolmogorov-Smirnov test for goodness of fit. J. Am. Stat. Assoc. 46, 68–78.

McDonald, J.H., 2014. Handbook of Biological Statistics, 3rd ed. Sparky House Publishing, Baltimore, Maryland.

McLaren, B.W., Langlois, T.J., Harvey, E.S., Shortland-Jones, H., Stevens, R., 2015. A small no-take marine sanctuary provides consistent protection for small-bodied bycatch species, but not for large-bodied, high-risk species. J. Exp. Mar. Biol. Ecol. 471, 153–163. https://doi.org/10.1016/j.jembe.2015.06.002.

Menin, A., Paulus, R., 1974. Fish counting by acoustic means. Engineering in the Ocean Environment, Ocean'74-IEEE International Conference On 166–168.

Mesiar, D.C., Eggers, D.M., Gaudet, D.M., 1990. Development of techniques for the application of hydroacoustics to counting migratory fish in large rivers. Rapports et Proces-Verbaux des Reunions, Conseil International pour l'Exploration de la Mer 189, 223–232.

Muñoz-Benavent, P., Andreu-García, G., Valiente-González, J.M., Atienza-Vanacloig, V., Puig-Pons, V., Espinosa, V., 2018a. Automatic Bluefin Tuna sizing using a stereoscopic vision system. ICES J. Mar. Sci. 75, 390–401. https://doi.org/10.1093/icesjms/fsx151.

Muñoz-Benavent, P., Andreu-García, G., Valiente-González, J.M., Atienza-Vanacloig, V., Puig-Pons, V., Espinosa, V., 2018b. Enhanced fish bending model for automatic tuna sizing using computer vision. Comput. Electron. Agric. 150, 52–61. https://doi.org/10.1016/j.compag.2018.04.005.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66. https://doi.org/10.1109/TSMC.1979.4310076.

Petrou, M., Petrou, C., 2011. Image segmentation and Edge detection. Image Processing: The Fundamentals. John Wiley & Sons, Ltd, Chichester, UK, pp. 527–668. https://doi.org/10.1002/9781119994398. ch6.

Phillips, K., Rodriguez, V.B., Harvey, E., Ellis, D., Seager, J., Begg, G., Hender, J., 2009. Assessing the Operational Feasibility of Stereo-Video and Evaluating Monitoring Options for the Southern Bluefin Tuna Fishery Ranch Sector. Fisheries Research and Development Corporation and Bureau of Rural Sciences, Australia.

Puig, V., Espinosa, V., Soliveres, E., Ortega, A., Belmonte, A., de la Gándara, F., 2012. Biomass estimation of bluefin tuna in sea cages by the combined use of acoustic and optical techniques. Collect. Vol. Sci. Pap. ICCAT 68, 284–290.

Ransom, B.H., Johnston, S.V., Steig, T.W., 1998. Review on monitoring adult Salmonid (Oncorhynchus and Salmo spp.) escapement using fixed-location split-beam hydroacoustics. Fish. Res. 35, 33–42. https://doi.org/10.1016/S0165-7836(98)00057-5.

Rasch, D., Kubinger, K.D., Moder, K., 2011. The two-sample t test: pre-testing its assumptions does not pay off. Statistical Papers 52, 219–231. https://doi.org/10.1007/s00362-009-0224-x.

Romakkaniemi, A., Lilja, J., Nykänen, M., Marjomäki, T.J., Jurvelius, J., 2000. Spawning run of Atlantic Salmon (Salmo salar) in the River Tornionjoki monitored by horizontal split-beam echosounding. Aquatic Living Resour. 13, 349–354. https://doi.org/10.1016/S0990-7440(00)01083-4.

Rosen, S., Jörgensen, T., Hammersland-White, D., Holst, J.C., Grant, J., 2013. DeepVision: a stereo camera system provides highly accurate counts and lengths of fish passing inside a trawl. Can. J. Fish. Aquat. Sci. 70, 1456–1467. https://doi.org/10.1139/cjfas-2013-0124.

Ruff, B.P., Marchant, J.A., Frost, A.R., 1995. Fish sizing and monitoring using a stereo image analysis system applied to fish farming. Aquacult. Eng. 14, 155–173. https://doi.org/10.1016/0144-8609(94)P4433-C.

Saberioon, M., Gholizadeh, A., Cisar, P., Pautsina, A., Urban, J., 2017. Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues. Rev. Aquacult. 9, 369–397. https://doi.org/10.1111/raq.12143.

Santana-Garcon, J., Newman, S.J., Harvey, E.S., 2014. Development and validation of a mid-water baited stereo-video technique for investigating pelagic fish assemblages. J. Exp. Mar. Biol. Ecol. 452, 82–90. https://doi.org/10.1016/j.jembe.2013.12.009.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.

Seiler, J., Williams, A., Barrett, N., 2012. Assessing size, abundance and habitat preferences of the Ocean Perch Helicolenus percoides using a AUV-borne stereo camera system. Fish. Res. 129, 64–72. https://doi.org/10.1016/j.fishres.2012.06.011.

Shafait, F., Harvey, E.S., Shortis, M.R., Mian, A., Ravanbakhsh, M., Seager, J.W., Culverhouse, P.F., Cline, D.E., Edgington, D.R., 2017. Towards automating underwater measurement of fish length: a comparison of semi-automatic and manual stereo– video measurements. ICES J. Mar. Sci. 10–1093. https://doi.org/10.1093/icesjms/fsx007.

Shieh, A.C.R., Petrell, R.J., 1998. Measurement of fish size in atlantic salmon (salmo salar l.) Cages using stereographic video techniques. Aquacult. Eng. 17, 29–43. https://doi.

org/10.1016/S0144-8609(97)00012-5.

Shortis, M., Ravanbakskh, M., Shaifat, F., Harvey, E.S., Mian, A., Seager, J.W., Culverhouse, P.F., Cline, D.E., Edgington, D.R., 2013. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. Proc. SPIE 8791, Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection, 87910G. https://doi.org/10.1117/12.2020941.

Simmonds, J., MacLennan, D., 2007. Fisheries Acoustics: Theory and Practice, second edition. Fisheries Acoustics: Theory and Practice: Second Edition. Blackwell Publishing Ltd https://doi.org/10.1002/9780470995303.

Simrad, 2008. Simrad ER60, Scientific Echo Sounder. Reference Manual.

Smale, D.A., Kendrick, G.A., Harvey, E.S., Langlois, T.J., Hovey, R.K., Van Niel, K.P., Waddington, K.I., Bellchambers, L.M., Pember, M.B., Babcock, R.C., Vanderklift, M.A., Thomson, D.P., Jakuba, M.V., Pizarro, O., Williams, S.B., Niel, V., 2012. Regional-scale benthic monitoring for ecosystem-based fisheries management (EBFM) using an autonomous underwater vehicle (AUV). ICES J. Mar. Sci. 69, 1108–1118. https://doi.org/10.1093/icesjms/fss082.

Soliveres, E., Poveda, P., Estruch, V.D., Pérez-Arjona, I., Puig, V., Ordóñez, P., Ramis, J., Espinosa, V., 2017. Monitoring fish weight using pulse-echo waveform metrics. Aquacult. Eng. 77, 125–131. https://doi.org/10.1016/J.AQUAENG.2017.04.002.

Solomon, C., Breckon, T., 2011. Fundamentals of Digital Image Processing: A Practical Approach With Examples in Matlab. John Wiley & Sons.

Tillett, R., Mcfarlane, N., Lines, J., 2000. Estimating dimensions of Free-swimming Fish using 3D Point distribution models. Comput. Vision Image Understanding 79, 123–141. https://doi.org/10.1006/cviu.2000.0847.

Torisawa, S., Kadota, M., Komeyama, K., Suzuki, K., Takagi, T., 2011. A digital stereo-video camera system for three-dimensional monitoring of free-swimming Pacific bluefin tuna, Thunnus orientalis, cultured in a net cage. Aquat. Living Resour. 24,

107–112. https://doi.org/10.1051/alr/2011133.

Trevorrow, M.V., 1997. Detection of migrating salmon in the Fraser River using 100-kHz sidescan sonars. Can. J. Fish. Aquat.Sci. 54, 1619–1629. https://doi.org/10.1139/f97-069.

Wakefield, C.B., Lewis, P.D., Coutts, T.B., Fairclough, D.V., Langlois, T.J., 2013. Fish assemblages associated with natural and anthropogenically-modified habitats in a Marine embayment: comparison of baited videos and Opera-house traps. PLoS One 8, e59959. https://doi.org/10.1371/journal.pone.0059959.

Watson, D., Anderson, M., Kendrick, G., Nardi, K., Harvey, E., 2009. Effects of protection from fishing on the lengths of targeted and non-targeted fish species at the Houtman Abrolhos Islands, Western Australia. Mar. Ecol. Prog. Ser. 384, 241–249. https://doi.org/10.3354/meps08009.

Welch, B.L., 1951. On the comparison of several mean values: an alternative approach. Biometrika 38, 330–336. https://doi.org/10.1093/biomet/38.3-4.330.

Williams, K., Lauffenburger, N., 2016. Automated measurements of fish within a Trawl using stereo images from a camera-trawl device (CamTrawl). Methods Oceanogr. 17, 138–152. https://doi.org/10.1016/j.mio.2016.09.008.

Willis, T.J., Babcock, R.C., 2000. A baited underwater video system for the determination of relative density of carnivorous reef fish. Mar. Freshwater Res. 51, 755. https://doi.org/10.1071/MF00010.

Zintzen, V., Anderson, M.J., Roberts, C.D., Harvey, E.S., Stewart, A.L., 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. https://doi.org/10.1371/journal.pone.0048522.

Zion, B., 2012. The use of computer vision technologies in aquaculture - a review. Comput. Electron. Agric. 88, 125–132. https://doi.org/10.1016/j.compag.2012.07.010.