Atienza-Vanacloig, V.; Andreu García, G.; López García, F.; Valiente González, JM.; Puig Pons, V. (2016). Vision-based discrimination of tuna individuals in grow-out cages through a fish bending model. Computers and Electronics in Agriculture. 130:142-150. doi:10.1016/j.compag.2016.10.009.

Vision-based discrimination of tuna individuals in grow-out cages through a fish bending model

Vicente Atienza-Vanacloig[1,*], Gabriela Andreu-García[1], Fernando López-García[1], José M. Valiente-González[1], Vicente Puig-Pons[2]

[1]*Institute of Control Systems and Industrial Computing (AI2)*
[2]*Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC)*
Universitat Politècnica de València (UPV)
Camino de Vera (s/n), 46022 Valencia (Spain)

Email: {vatienza,gandreu,flopez,jvalient}@disca.upv.es*
* Corresponding author. Tel : +34-963-877-000 ; Fax +34-963-879-009

Abstract

This paper proposes a robust deformable adaptive 2D model, based on computer vision methods, that automatically fits the body (ventral silhouette) of Bluefin tuna while swimming. Our model (without human intervention) adjusts to fish shape and size, obtaining fish orientation, bending to fit their flexion motion and has proved robust enough to overcome possible segmentation inaccuracies. Once the model has been successfully fitted to the fish it can ensure that the detected object is a tuna and not parts of fish or other objects. Automatic requirements of the fishing industry like biometric measurement, specimen counting or catch biomass estimation could then be addressed using a stereoscopic system and meaningful information extracted from our model. We also introduce a fitting procedure based on a fitting parameter --Fitting Error Index (FEI)-- which permits us to know the quality of the results. In the experiments our model has achieved very high success rates (up to 90%) discriminating individuals in highly complex images acquired for us in real conditions in the Mediterranean Sea. Conclusions and future improvements to the proposed model are also discussed.

Keywords: Shape Modelling, Fish detection, Underwater Video Processing, Computer Vision, Image Segmentation, Automatic Biomass Estimation.

1. Introduction

In recent years, great progress has been achieved in all underwater applications (Zion 2012). However, most of them currently require human intervention in some of their stages which is critical for obtaining valid results. Applications and techniques which need human intervention are described in the literature as semi-automatic. But some authors like (Lines et al. 2001), (Shortis et al.2013) and (Zion, 2012), remark that further progress in fisheries management and research into aquatic biodiversity requires fully automatic processing of underwater video recordings to extract the most meaningful information for an application proposal.

A real challenge for this kind of application is the automatic discrimination of isolated fish in the image, ensuring that the object identified is a whole fish (hereinafter "good-fish") rather than a portion of it, or two or more overlapped fish (hereinafter "bad-fish") (Costa 2006 et al.). The characterisation of a single fish is an essential processing step in the most significant applications of underwater video, such as fish detection, species identification (Spampiato et al., 2010) (Zion et al., 2007), biometric measurements (Tillett et al. 2000) (Harvey et al. 2003) (Costa et al. 2006), biomass estimation in fish cages or tanks (Lines et al. 2001) (Martinez et al. 2003), tracking and counting fish (Lee et al. 2004).

Our goal is to develop a vision-based application to automatically discriminate individuals or whole tuna in underwater images acquired under real conditions. This application has to overcome the fact that real underwater fish images are generally poor quality because they suffer from limited range, non-uniform lighting, low contrast, colour attenuation and blurring (Shortis et al.2007) which represent a challenge for researchers. Figure 1 shows some colour video frames used in this work which illustrate some of these difficulties. We need to be able to assure that the object detected is a whole fish because, once the fish has been discriminated, the process can be continued performing biometric measurements for the purpose of species identification, biomass estimation or fish counting. Image processing and computer vision methods can be used for these purposes.

Figure 1: Some examples of video frames (left VideoA, right VideoB) used in this work

Commercial biomass estimation systems most widely used in aquaculture are VICASS (AKVA group, 2014) and AQ1 (AQ1 Systems 2013) which belong to the above mentioned semi-automatic category. These systems need human operators to manually inspect different frames in which a particular isolated fish appears (Harvey et al. 2003). Then, they mark the fish snout and tail, and the fish length and span are automatically computed. To reduce the effect of swimming motion on length measurements only frames in which the body of the fish appears to be straight are considered. If the system works with stereo vision, the marking process is made on corresponding points in the image pair. These systems determine size distributions based on simple length and span measurements, and thereby deduce biomass from an estimated number of fish in the cage or tank.

Currently, Bluefin tuna catch quotas are monitored to compute two statistical factors: the number of fish caught and the catch weight. The number of fish is obtained by counting all the individuals transferred from tow cages to grow-out cages. Bluefin tuna transfers are usually made by joining tow and grow-out cages through gates that allow fish to pass from one cage to another, while experienced divers equipped with video cameras monitor these underwater tasks. Subsequently, these films are watched by human inspectors in order to manually count the number of fishes transferred. The average weight of these live samples is usually estimated by collecting a given number of fish from the tow cage (Harvey et al. 2003). The individuals counted during a transfer are multiplied by the average weight to derive total biomass per tow cage.

Nevertheless, we consider that video cameras could be attached to gate sides given that it is mandatory to record the fish swimming through during the transfer. These films could be analysed automatically by computer vision techniques. These techniques have the advantage of not stressing the fish (stress can cause death) and provide continuous, objective and reproducible results.

Another interesting scenario that benefits from non-intrusive vision-based weight estimation is fish fattening monitoring. It can be used to control the feeding process without the need to stress or sacrifice specimens.

Tuna monitoring does not require precise counting of individuals because the objective is to obtain statistical estimations of fish weight. Espinosa et al. (2011) present real values for obtaining the relationship between Bluefin tuna length (L) and its mass (W). This relationship has been investigated for many years (Zion, 2012) and the most common mathematical model is $W = aL^b$, where the values of coefficients a and b depend on the fish species.

The first step in automating any process is the detection of candidates and be able to ensure that each one corresponds with a whole individual. Furthermore, body bending while free-swimming means that the same individual can be observed with very different shapes and fish size and orientation can vary in relation to the visualized frame. So, robust fish detection methods which cannot be affected by these variations are required.

The influence of swimming motion on fish shape can be minimised by designing a robust deformable fish model (Lines et al. 2001) to fit fish size and gesture. When the model successfully fits the object detected in the image, it can help to accurately locate its different parts and deduce useful information including, for example, whether the detected object is a whole fish or not, if the fish is straight or not and the angle of curvature of its body. With an estimation of the exact curvature of the body, biometric measures like fish length could be robustly obtained. Other advantages of the model would be to correct segmentation errors caused by noise or variable lighting and to successfully detect the silhouette of foreground fish in crowded images.

This paper proposes a deformable and adaptive robust model that automatically fits the ventral silhouette of Bluefin tuna in images acquired in natural conditions. The differences of the present work with regard to other works in literature are: i) video images are taken in the natural environment without artificial illumination and without background screens, ii) the image can contain fish clusters with semi-crowded situations and overlapping fish, iii) the fish is extracted from images by a fully automatic process, iv) all fish edges and contours considered in our process are outlined without human operators, v) fish direction -- which is unknown -- is obtained automatically.

In this paper materials and methods are described in section 2. Section 3 describes experiments and results which show that our model is able to identify Bluefin tuna fish.

We discuss the results in section 4 and present our conclusions and future work in section 5.

2. Materials and methods

The automatic identification of individual fish in an underwater image is a complex issue. Important aspects like overlapped individuals in the image and sunlight effects that cause many segmentation problems, must be overcome to automate the process. This section describes a new deformable 2D model for identifying Bluefin tuna that adapts to the movements and variable sizes of fish.

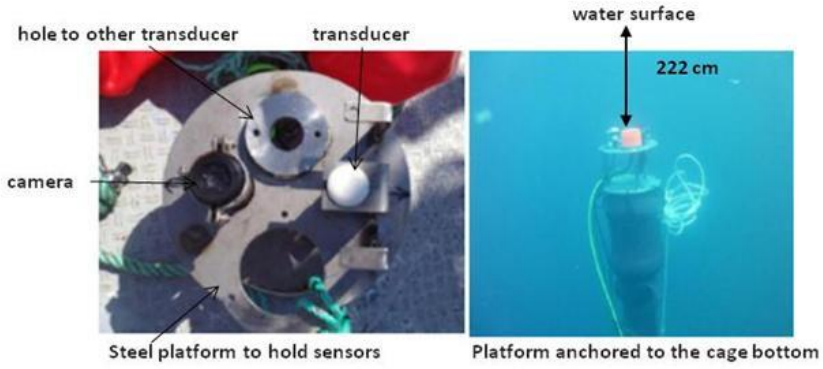

Figure 2: Details of the arrangement of the underwater equipment. On the right, steel platform held by buoys that provide neutral buoyancy

2.1 Video system and image acquisition

The video films used in this work were taken in grow-out cages installed in Spanish waters in the Mediterranean when the fish were swimming freely. The sequences were acquired with a camera anchored at the bottom of the grow-out cage, and pointing towards the surface as shown in Figure 2. The cages are cone shaped with a circle with a diameter of 50 meters on the water surface and 30 meters tall. The videos were acquired at 222 cm from the water surface (Figure 2) and the cage contained about 400 adult tuna which were between 120 and 210 centimetres long. One of the films (VideoA) was acquired on a sunny day in summer (June) and the other one (VideoB) on a cloudy day in autumn (November).

The acquisition system comprised a Sony SNC-CH210 (3 Megapixel) single IP video camera, encapsulated to immerse, connected by Ethernet 100Base-TX and powered via PoE, (see Figure 2 left). Lens focal length was 3.3mm for a horizontal field of view

of twice the working distance. Recording was coded in Mpeg4 with a resolution of 1280x1024 pixels, at 20 frames per second (fps) in VideoA and 30 fps in VideoB.

2.2 Segmentation process

Figure 1 shows the effect of sunlight that acts like a backlighting emitter and brightness varies widely across the image due to the refraction of sunlight through the surface (Lines et al. 2001). Consequently background luminosity is non-uniform, fish tone can vary when it crosses the sunlight spot and can even vary from head to tail, although the fish always are darker than background in our images. The situation can deteriorate even further because the camera may move slightly due to underwater currents.

Our application needs compact regions or blobs (large binary objects) which are the candidates for adjusting the model and then to decide whether or not the blob is a whole fish. So that we used two different region based segmentation approaches: (i) a global technique based on background subtraction and (ii) a local technique based on local thresholding. The background subtraction technique uses a background model that captures the spatial variability of the light. Local thresholding, however, is a fast computation method that behaves well in non-uniform background scenarios.

2.2.1 Background subtraction

Background subtraction compares a video frame F_t against a background model B and identifies candidate pixels to be foreground pixels from the input frame (Piccardi, 2004). Relative difference rather than absolute difference is used to emphasize the contrast in dark areas, and foreground pixels are estimated as:

$$\frac{|F_t(x,y) - B(x,y)|}{B(x,y)} > T_r$$

where $F_t(x,y)$ and B(x,y) denote the luminance pixel and its background estimate at spatial location (x,y) and time t, while T_r represents a threshold value.

Stationary techniques compute model B_s starting with a set of frames and maintain the same model throughout the process. However we use the median rather than the average intensity level because it is a nonlinear process useful for preserving edges in an image while reducing random noise. The median intensity value of the pixels in window frames becomes the output intensity of the pixel being processed. Thus the background model can be defined as:

$$B_S(x,y) = Median\,(F_1(x,y), \ldots, F_n(x,y))\,;$$

Where n is the number of buffer frames, which usually correspond to n first frames or n randomly chosen frames. Besides close and open morphological operations have been introduced to obtain the best results when some objects are near each other and when small holes appear.

2.2.2 Local thresholding

Local thresholding examines statistically the intensity values of the local neighbourhood of each pixel assuming that illumination is uniform in the neighbourhood. Fast approaches include the median value, the mean of the local intensity distribution, or the mean of the minimum and maximum values (Petrou and Bosdogianni, 1999). The statistic is then used as a local threshold to determine if the current pixel is selected as foreground or background. The most appropriate statistic depends largely on the input image. We carried out some heuristic supervised tests with our data videos and concluded that the best choice was to use the mean with a neighbourhood size large enough to cover sufficient foreground and background pixels. In our case a pixel is selected as foreground if its value is below the local statistic and the local threshold can be expressed as:

$$M_{ij}(x,y) = \frac{1}{w*w}\sum_{l=i-\frac{w}{2}}^{i+\frac{w}{2}}\sum_{k=j+\frac{w}{2}}^{j+\frac{w}{2}} p_{lk}\ ; \quad \forall\, p_{ij} \in F_t; \begin{cases} p_{ij} \geq M_{ij}\ ;\ p_{ij}\ is\ background \\ p_{ij} < M_{ij};\ p_{ij}\ is\ foreground \end{cases}$$

where *w* is the size neighbourhood and p_{ij} and $p_{lk} \in \mathrm{F_t}$,. To overcome border problems, pixel values outside the bounds of the image are computed by mirror-reflecting the pixels across the image border.

Assuming that we want to detect the contour of the foreground objects, neighbourhood size has to be large enough to include some foreground pixels and some background pixels when the contour pixels of objects are being processed. Choosing neighbourhoods which are too large, however, can violate the assumption of approximately uniform illumination introducing noise and artefacts that do not correspond to real objects. A right segmentation was observed using neighbourhood sizes between 15x15 and 19x19 as indicated in Section 3.3.

Open and close morphological operations complete the segmentation process.

The result of both these segmentation methods is a binary image showing blobs that represent the objects detected. Then we use the contour of these blobs to fit our tuna model introduced in the next section.

2.3 Tuna fish model

The real shape (Figure 3) of Bluefin tuna and some data on kinematics of other related species available in Dewar and Graham 1994 and Hawkins et al. 2003 were studied to design our landmark-based model.

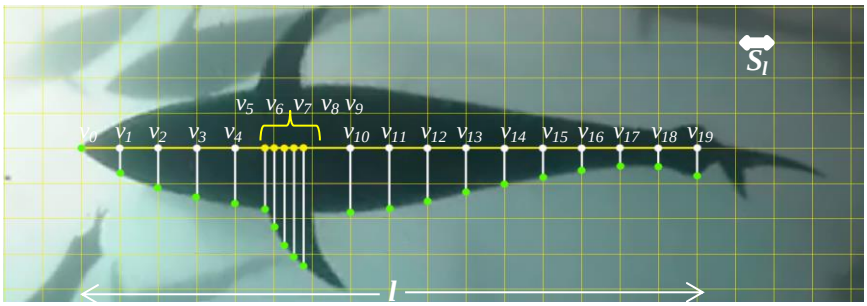

Figure 3: Tuna contour model. White and yellow points are reference points along the vertebral column. Green points correspond to left (lower) side landmarks. Grid size represents the unit measure S_l.

2.3.1 Obtaining the landmark points for the model

To build the landmark set we chose a middle-distant standard adult tuna shape in a straight pose (see Figure 3). Our tuna model comprises a set K of 39 landmark points for the tuna contour, taking 19 landmarks to represent each side of the tuna body and one for the tip of the snout. The caudal fin contour was not modelled because its shape varies widely over video frames due to swimming movement. The first step to design the K landmarks was to consider the longitudinal axis of the fish, which ranges from the tip of the snout to the end of the caudal peduncle, with length l. This axis was divided into sixteen equally spaced sections (with length $S_l = l/16$) and this section length S_l was taken as a unit measure to define reference vertebral column positions v_l and 15 vertebral segments. In Figure 3, the size of the grid represents the S_l unit measure while the v_l positions are marked as white points (green for the tip of the snout) and the landmarks as green points. These v_l positions were defined to locate the K set of landmarks corresponding to the body side. Then, in order to define the pectoral fin profile, five v_j additional vertebral column positions (marked in yellow in the figure) were considered. The distance between these additional points was reduced

to $0.25 * S_l$ to achieve adequate detail. The first of these points (v_5) is located at $4.7 * S_l$ from the tip of the snout, which was found to be the characteristic position that invariably corresponds to the pectoral fin profile starting point. In summary, we obtained an ordered set of 20 vertebral column reference positions $\boldsymbol{v_i}$, from head to tail, that can be written as: $V = (\boldsymbol{v_0}, \boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \boldsymbol{v_4}, v_5, v_6, \boldsymbol{v_7}, \boldsymbol{v_8}, \boldsymbol{v_9}, \boldsymbol{v_{10}}, \boldsymbol{v_{11}}, \boldsymbol{v_{12}}, \boldsymbol{v_{13}}, \boldsymbol{v_{14}}, \boldsymbol{v_{15}}, \boldsymbol{v_{16}}, \boldsymbol{v_{17}}, \boldsymbol{v_{18}}, \boldsymbol{v_{19}})$. And their respective unitary x-component coefficients are $\Delta \boldsymbol{x_i^v} = [0, 1, 1, 1, 1, 0.7, 0.25, 0.25, 0.25, 025, 1.3, 1, 1, 1, 1, 1, 1, 1, 1, 1]$ where $\sum_{i=0}^{19} \Delta x_i^v = 16$ whereby the x_i^v , y_i^v coordinates for each vertebral $\boldsymbol{v_i} = (x_i^v, y_i^v)$ are obtained as:

$$x_i^v = S_l \sum_{j=0}^{i} \Delta x_i^v; \quad y_i^v = 0; \qquad i = 0,..,19$$

Then, two aspects were deduced from the fish represented in Figure 3 and assumed to obtain the K landmark points: i) the thickness or width of the tuna body is proportional to its length and consequently to S_l; ii) the tuna body is symmetrical in relation to its vertebral column $\boldsymbol{v}$. Thus for each $\boldsymbol{v_i}$ reference vertebral points we look for its corresponding contour points in the $\widehat{\boldsymbol{u_i}}$ normal direction to the vertebral column, to define the location of the 19 landmark points corresponding to one side of the silhouette (green points in Figure 3). Next, for the selected standard tuna shape, we computed the normal distance from reference vertebral positions and their landmark points and we obtained a vector $\boldsymbol{c^u}$ of distance coefficients. Each coefficient c_i of $\boldsymbol{c^u}$ represents the proportionality between S_l and the $d(v_i, k_i)$ distances from v_i vertebral point to its corresponding landmark k_i. The values for the vector of distance coefficients are: $\boldsymbol{c^u} = [0, 0.7, 1.15, 1.35, 1.55, 1.65, 2.0, 2.5, 2.8, 3.15, 1.7, 1.55, 1.35, 1.1, 0.9, 0.7, 0.55, 0.45, 0.5, 0.75]$. The tip of the snout is vertebral point (v_0) and landmark point (k_0) at the same time, so $c_0 = 0$ and also $d(v_0, k_0) = 0$. Having taken the symmetry of the fish body, we consider the same vector $\boldsymbol{c^u}$ to locate the 19 landmark points on the other side of the fish silhouette. So the x_i^k , y_i^k coordinates for landmark $\boldsymbol{k_i} = (x_i^k, y_i^k)$ can be obtained as:

$$x_i^k = x_i^v; \quad y_i^k = S_l * c_i; \qquad i = 0,..,19$$

$$x_i^k = x_{i-19}^v; \quad y_i^k = -S_l * c_{i-19}; \qquad i = 20,..,39 \;\; \forall v_i \in v, \; c_i \in \boldsymbol{c^u}$$

In short, once the vertebral positions v_i have been determined for a fish image their k_i landmarks points can be located at distance $d(v_i, k_i) = c_i * S_l$ in the $\hat{u}_l$ normal directions on both sides of the vertebral column in the fish body edges.

2.3.2 Bending the model

During the swimming motion, tunas make a global flexion that is not uniformly distributed along the length of their bodies. We have used the fundamentals presented in Dewar and Graham 1994 and Hawkins et al. 2003 to model the distribution of flexion along Bluefin vertebral segments correctly.

The global bending of a tuna body can be defined as the angle θ between the first v_0 and the nineteenth v_{19} of the vertebral column segments. Consequently:

$$\sum_{i=0}^{i=19} d\theta_i = \theta;$$

where $d\theta_i$ represents the angle between two consecutive segments that correspond to vertebral positions v_i and v_{i-1}. The set of nineteen $d\theta_i$ values can be represented as a vector $\boldsymbol{d\theta}$. Considering an equitable distribution of bending where each point of the vertebral column makes a contribution to global flexion in relation to its reference positions, $d\theta_i$ can be defined as:

$$d\theta_i = \frac{\theta \ \Delta v_i}{\sum_{i=0}^{19} \Delta v_i}; \qquad \theta_i = \sum_{j=0}^{i} d\theta_j; \qquad where \quad \theta_{19} = \theta;$$

where Δv_i is the unitary coefficient mentioned above with $\sum_{i=0}^{19} \Delta v_i = 16$, and θ_i is the bending angle of each v_i in relation to the fish head.

In real fish, however, the degree of flexion is concentrated in the central part of body to tail. So, a model flexion distribution with nineteen $\Delta\theta_i$ unitary flexion coefficients, one for each vertebral v_i where $\sum_{i=0}^{19} \Delta\theta_i = 16$, has been defined. In our case, the numerical values of each $\Delta\theta_i$ coefficient are based on the graphics of maximum flexion presented in Hawkins et al. 2003. The unitary vector $\Delta\boldsymbol{\theta}$ of flexion coefficient contributions (head to tail order) we use is $\Delta\boldsymbol{\theta} = [\,0,\ 0,\ 0.64,\ 0.64,\ 0.48,\ 0.48,\ 0.0,\ 0.0,\ 0.0,\ 0.0,\ 0.48,\ 0.64,\ 0.8,\ 0.96,\ 1.12,\ 1.28,\ 1.44,\ 1.92,\ 2.4,\ 2.72\,]$ where $\sum_{i=0}^{19} \Delta\theta_i = 16$. It can be observed that the highest flexibility corresponds to the segments close to the tail $(\Delta\theta_{19} = 2.72)$ while the head segments have no possibility of flexion $(\Delta\theta_0 = 0)$. Figure

4 shows an example of $d\theta_i$ values for only sixteen vertebral segments in a non-equitable bending distribution. For this kind of non-equitable distribution, the partial flexion angles $d\theta_i$ of each vertebral v_i can be obtained as:

$$d\theta_i = \frac{\theta * \Delta\theta_i}{\sum_{i=0}^{19} \Delta\theta_j} \; ; \qquad \sum_{i=0}^{19} \Delta\theta_i = 16 \; ; \; \forall \, v_i \in v;$$

To maintain rigidity of lateral fin shapes, similar flexion coefficients for vertebral positions $v_5,\ v_6,\ v_7,\ v_8,\ v_9$ were used.

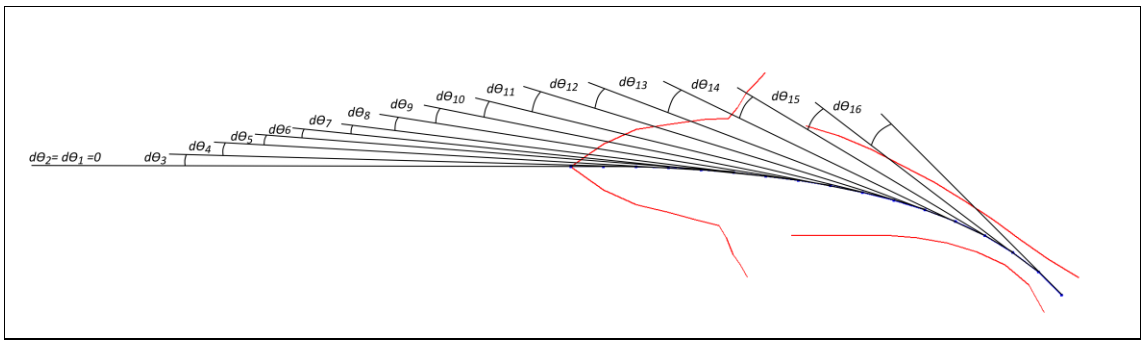

Figure 4: Modelling non equitable distribution of flexion along vertebral column segments.

Once the vertebral profile ($\boldsymbol{d\theta}$) corresponding to a given flexion (θ) is determined, we compute the curved vertebral column profile and then determine normal directions at reference points ($\boldsymbol{v_i}$) in the curved vertebral column to obtain landmark point positions taking the bending into account. The x_i^v, y_i^v coordinates for each vertebra $\boldsymbol{v_i} = (x_i^v, y_i^v)$ in the flexed vertebral column can be obtained as:

$$x_i^v = x_{i-1}^v + S_l \, c_i \, \cos(\theta_i) \; ; \quad y_i^v = y_{i-1}^v + S_l \, c_i \, \sin(\theta_i) \; ; \quad i = 1, \ldots, 19$$

with $x_0^v = 0$ and $y_0^v = 0$. And the x_i^k , y_i^k coordinates for 39 landmarks $\boldsymbol{k_i} = (x_i^k, y_i^k)$ taking bending into account can be obtained as:

$$x_i^k = x_i^v - S_l \, c_i \, \sin(\theta_i) \; ; \quad y_i^k = y_i^v + S_l \, c_i \, \cos(\theta_i) \; ; \quad i = 0, \ldots, 19$$

$$x_i^k = x_{i-19}^v + S_l \, c_{i-19} \, \sin(\theta_{i-19}) \; ; \quad y_i^k = y_{i-19}^v - S_l \, c_{i-19} \, \cos(\theta_{i-19}) \; ; \quad i = 20, \ldots, 39$$

As an example, Figure 5 shows the v_i vertebral column segments, normal segments, K landmark points and resulting contours generated from our model for global flexions with $\theta = 15°,\ \theta = 30°$ and $\theta = 45°$.

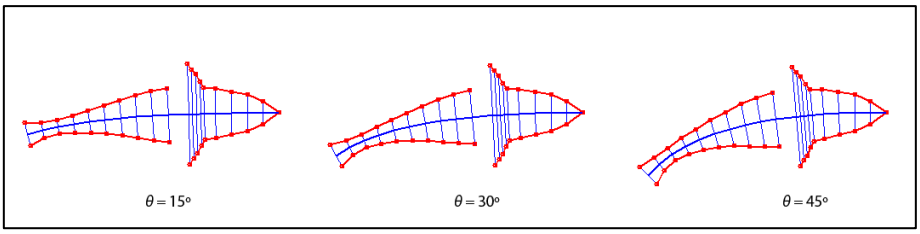

Figure 5: Contours generated by our model for flexions $\theta = 15°, \theta = 30°$ and $\theta = 45°$.

To achieve insensitivity to scale, translation and rotation our model $\boldsymbol{M}$ of tuna fish is finally defined by a vector of five parameters $\boldsymbol{M} = [s_x, s_y, l, \alpha, \theta]$ (see Figure 6) where: translation parameters s_x and s_y give the image location of the snout tip; l is the length of the vertebral column ($l = 16\, S_l$), which gives the *scale* factor; α denotes the rigid *rotation* of the model, defined as the angle of the fish head in relation to the horizontal axis, and θ is the angle of global flexion of the vertebral column as defined above. If $\alpha = 0$ then the fish head is directed to the right of the image and its head is completely horizontal. And the rigid transformation matrix can be written as:

$$\begin{pmatrix} x_i'^v \\ y_i'^v \end{pmatrix} = \begin{pmatrix} S_x \\ S_y \end{pmatrix} + \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x_i^v \\ y_i^v \end{pmatrix}$$

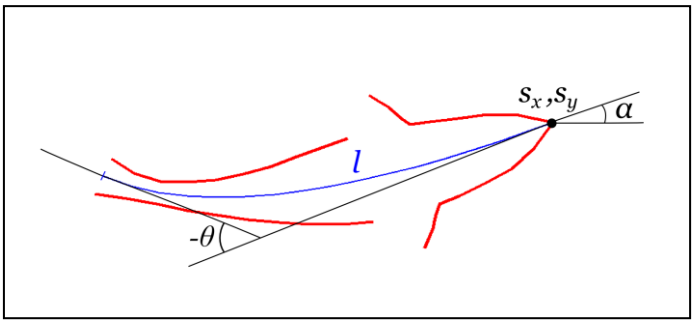

Figure 6: The five parameters $\boldsymbol{M} = (s_x, s_y, l, \alpha, \theta)$ that define our model.

2.3.3 Model fitting and Fitting Error Index (FEI)

The objective of the fitting process is to obtain the optimum model parameters $\boldsymbol{M^{op}}$ for a candidate blob I_{blob}. Model-to-image discrepancy is defined as a fitting error index (FEI) based on the quadratic distances $d(k_i, k_i^{blob})$ that occur between the modeled positions of landmark points k_i for a given set of parameter values and the corresponding blob boundary points k_i^{blob}, and it can be written as:

$$FEI = \frac{100}{l}\sqrt{\frac{1}{m}\sum_{i=1}^{m} d(k_i, k_i^{blob})^2} \qquad with\ 35 < m \leq 39$$

Where l is the model parameter for the estimation of the length of the vertebral column, used to obtain scale invariance, $d(k_i, k_i^{blob})$ is the distance from model landmark point k_i to the nearest blob border element k_i^{blob} and m is the number of silhouette landmark points that were successfully matched to an image border element. These border elements are searched along line segments normal to the modeled fish silhouette centered at landmark points k_i. The length L of these exploration segments is proportional to the modeled fish length l, $L = l/5$. Figure 7.b shows the line exploration segments to find the k_i^{blob} landmarks. Figure 7.c depicts (in red) the m error distances found for each shape and model $\boldsymbol{M}$ used in (b).

FEI obtains values in the $[0..10]$ range, $FEI = 0$ denotes a perfect fit between the segmented blob I_{blob} and model $\boldsymbol{M}$.

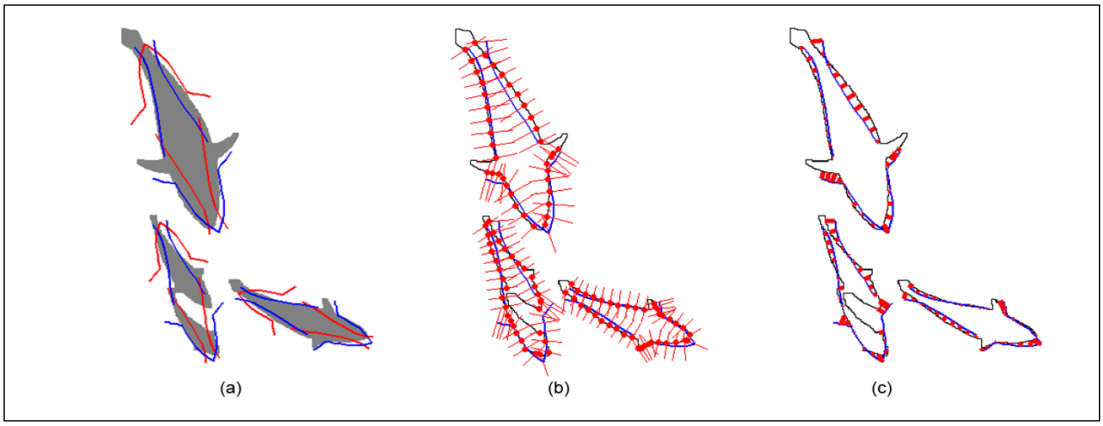

Figure 7: Summary of the fitting procedure. (a) Estimating initial fits and head position. This is done by using the centroid and major axis of the blob. As fish orientation is yet unknown, both hypotheses are considered (blue: tail-head; red: head-tail) (b) Obtaining the nearest image borders to the model along scanning segments normal to modeled contour at landmark points for the "blue" hypothesis. (c) Measuring fitting error index FEI. Red segments show the distances between model and actual contour points.

Furthermore, a small amount of not-found border points are allowed (10%) to achieve some tolerance at small silhouette discontinuities that appear in the segmented blobI_{blob}. The fitting procedure is only successful if at least 36 landmarks (m>35) are found in the blob. FEI is used in the experiments section to decide whether the segmented object is a well-defined fish or not.

2.3.4 Initial fit estimation and fitting procedure

The fitting procedure uses an iterative method that successively refines an initial model estimation $\boldsymbol{M_0}$ to converge at optimum model $\boldsymbol{M^{op}}$ that minimizes the FEI. The I_{blob} features used to obtain an initial $\boldsymbol{M_0}$ are: centroid (c_x, c_y), major axis length l_{blob} and major axis orientation φ. So the corresponding model $\boldsymbol{M_0}$ parameters are:

$$(s_x, s_y)_0 = (c_x, c_y) \pm \frac{l_{blob}}{2}\,\widehat{\boldsymbol{u_\varphi}}; \qquad l_0 = l_{blob}; \quad \alpha_0 = \varphi; \quad \theta_0 = 0°;$$

where $\widehat{\boldsymbol{u_\varphi}}$ is the unit bidimensional vector oriented in direction φ.

An important question to resolve is the location of the head. A priori, the head may be at either blob axis end but the successful or unsuccessful matching of the model depends strongly on this decision. Given that the sign of the unit vector that corresponds to the true fish orientation cannot be known a priori, both signs ($\pm$) are tried, leading to a twofold estimation of the initial hypothesis. Figure 7.a depicts the initial $\boldsymbol{M_0}$ estimations obtained with this method for a set of given shapes with the initial hypothesis ($\pm$) marked in red and blue.

To achieve $\boldsymbol{M^{op}}$, our iterative fitting procedure uses a sequential quadratic programming (SQP) method (Fletcher et al. 1963) (Fletcher 1980). The unconstrained approach we use leads to the computation of a quasi-Newton approximation to the Hessian of the Lagrangian. We define a set of lower and upper bounds on the objective function variable $\boldsymbol{M}$ (model parameters), so that the solution must always be in the range $\boldsymbol{lb} \leq \boldsymbol{M^{op}} \leq \boldsymbol{ub}$, as follows:

$$(s_x, s_y)_0 - (20,20) \leq (s_x, s_y)_t \leq (s_x, s_y)_0 + (20,20)$$

$$l_0 - 10 \leq l_i \leq l_0 + 10$$

$$\alpha_0 - 10° \leq \alpha_i \leq \alpha_0 + 10°; \quad \theta_0 - 45° \leq \theta_i \leq \theta_0 + 45°$$

where angles are expressed in degrees and positions and lengths in pixel counts. The index t is used here to denote the iteratively refined solution valuesThe algorithm usually converges to the optimum solution $\boldsymbol{M^{op}}$ in a variable number of iterations (20-50) for true fish shapes and correct hypothesis about the direction of fish orientation. In other cases, the search is often aborted in a low number of iterations (typically six) as the result of early divergence.

3. Experiments and Results

The aim of these experiments is to evaluate the accuracy of our model to fit Bluefin tuna in images acquired in real conditions. A ground truth was generated using two different underwater videos (Video-A and Video-B) with complex scenes (live fish in continuous movement, low contrast, murky water, overlapped fish, variable lighting conditions and crowded situations). The FEI index, shown in the previous section, was used to discriminate whether or not the object detected I_{blob} (candidate blob) is an individual fish.

3.1 Ground truth

Video-A and Video-B (Figure 1) were acquired at 20 fps and 30 fps, respectively, so only one frame in ten was considered because consecutive frames do not provide significant differences. Finally, a set of 703 frames contributed to the ground truth out of a total of 7036 (4788+2248) video frames.

The sequence for achieving the ground truth was: i) to segment the image and obtain foreground blobs, ii) to automatically discard blobs that touch the image border (border-blob) and blobs smaller than a considered minimum area (small-blob), iii) the remaining segmented blobs were labelled in a supervised way by three different human operators as good-fish (whole and well-defined fish) or bad-fish. The blobs which do not contain a whole tuna fish or include overlapping fish were considered bad-fish. Table 1 summarizes the number of blobs obtained for VideoA and VideoB using two different segmentation processes and two different minimum blob area sizes.

Table 1: Ground truth obtained with VideoA and VideoB.

Ground truth	**Minimum size 2000 pixels**		**Minimum size 3000 pixels**	
	Local Thresholding	**Background model**	**Local Thresholding**	**Background model**
Blobs	**VideoA + VideoB**	**VideoA + VideoB**	**VideoA + VideoB**	**VideoA + VideoB**
Good-fish	0707 + 0288 = 00995	0435 + 0050 = 0485	0302 + 0287 = 00589	0189 + 0050 = 0239
Bad-fish	0656 + 0520 = 01176	1736 + 0521 = 2257	0455 + 0301 = 00756	0614 + 0418 = 1032

Figure 8 left shows an example of ground truth frame labelled by operators. White objects correspond to border-blobs while black objects correspond to small-blobs that usually correspond to fish far away from the camera. Green and red objects correspond to good and bad fish blobs, respectively.

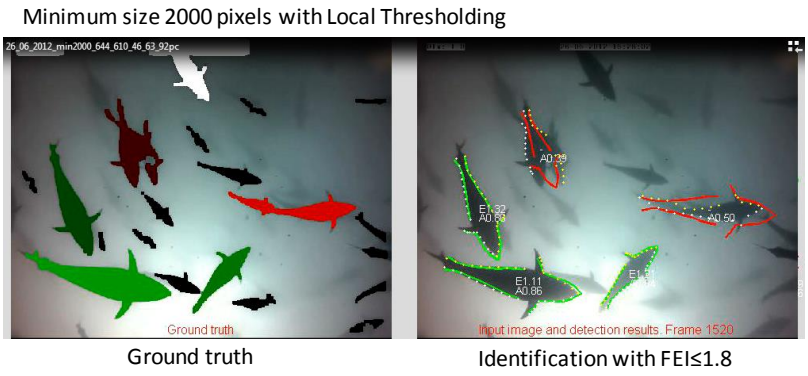

Figure 8: One labelled frame of the ground truth in the left (green: good-fish, red: bad-fish, black: small-blob, white: border-blob). The right image shows the good performance of our model.

3.2 Experiments

The experiments were designed to find the values of FEI which permit us to discriminate a blob as good-fish with good accuracy. Figure 9 shows the steps of this process where the border-blobs and small-blobs are discarded before applying our fitting algorithm.

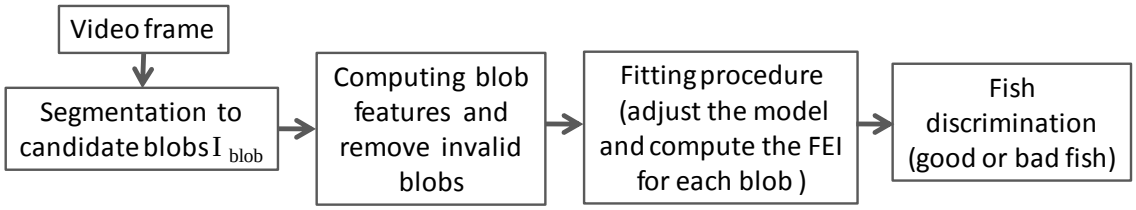

Figure 9: Sequence of steps followed to discriminate tuna individuals.

All the ground truth blobs were classified in each experiment and a confusion matrix was used to compute the following accuracy measures:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN};$$

$$AP = \frac{TP}{TP + FP}; \qquad AR = \frac{TP}{TP + FN};$$

where TP (True Positives) are the good blobs classified as good, TN (True Negatives) the bad blobs classified as bad, FP (False Positives) the bad blobs classified as good and FN (False Negatives) the good blobs classified as bad. So Accuracy = 1 indicates a success rate of 100% in classification. AR (accuracy of recall) estimates how effective the test is when used on positive individuals so AR = 1 indicates that there are

no FN. AP (accuracy of precision) represents the proportion of truly positive cases and AP = 1 indicates that there are no FP.

3.3 Results

As already mentioned in Section 2.2, the segmentation methods used are local thresholding and the background model. A set of tests were conducted to heuristically decide the most appropriate neighborhood size to apply the local thresholding. It was observed that the best segmentation results were achieved with neighborhood sizes between 15x15 and 19x19. The differences obtained in results using these sizes were not significant so finally we decided to use a size of 15x15 because it supposed to assume more uniform illumination between neighbours. Tests were also conducted to decide the best background model. In these tests median intensity provided better results than average intensity. Two sizes of minimum blob area with 2000 pixels and 3000 pixels (Table 1) were tested to compare results.

A block of 400 experiments was conducted to find the FEI values that provide both good Accuracy (as good as possible) and high balanced AP and AR values. The 100 values of FEI tested were from 0.1 to 10 in increments of 0.1 $(0.1, 0.2, ..., 9.9, 10)$ and the most significant values are shown in Tables 2 and 3. In our case the best result was achieved with values between FEI=1.8 and FEI=2.2 in all cases. There is a 90.6% success ratio with local thresholding and minimum area of 2000 pixels for FEI=2.2, see Table 2. This ratio was 89.7% with 3000 pixels for FEI=2.0. In the case of the background model method, see Table 3, 90.6% Accuracy is achieved using a minimum size of 2000 pixels and 91.4% when using 3000 pixels. Thus, at first glance, the background model performs slightly better than local thresholding but if we compare the associated AP and AR results it is clear that local thresholding is better than the background model. An increase in the minimum size of blobs does not improve results with local thresholding and only slightly with the background model.

Table 2: Identification results with Local Thresholding: Accuracy measures varying FEI.

Local Thresholding, (VideoA + VideoB)										
	Minimum size 2000					**Minimum size 3000**				
FEI	**TNR**	**FPR**	**Accuracy**	**AP**	**AR**	**TNR**	**FPR**	**Accuracy**	**AP**	**AR**
1.8	0.921	0.079	0.889	0.901	0.850	0.901	0.099	0.891	0.873	0.878
2.0	0.906	0.094	0.898	0.888	0.888	**0.882**	**0.118**	**0.897**	**0.859**	**0.917**
2.2	**0.891**	**0.109**	**0.906**	**0.878**	**0.925**	0.864	0.136	0.897	0.843	0.941
2.4	0.875	0.125	0.906	0.865	0.943	0.847	0.153	0.894	0,829	0.956

Table 3: Identification results with Background model: Accuracy measures varying FEI.

Background model, (VideoA + VideoB)										
	Minimum size 2000					**Minimum size 3000**				
FEI	**TNR**	**FPR**	**Accuracy**	**AP**	**AR**	**TNR**	**FPR**	**Accuracy**	**AP**	**AR**
1.6	0.982	0.018	0.911	0.873	0.579	0.974	0.026	0.914	0.853	0.657
1.8	0.962	0.038	0.910	0.792	0.668	**0.954**	**0.046**	**0.914**	**0.790**	**0.741**
2.0	**0.939**	**0.061**	**0.908**	**0.729**	**0.767**	0.934	0.066	0.913	0.742	0.820
2.2	0.918	0.082	0.902	0.684	0.831	0.916	0.084	0.909	0.707	0.879

Figure 10 illustrates the performance of our model for identifying good-fish. In this case plotting ROC curves shows the TP rate against the FP rate. All our experiments achieve results above the no discrimination line so the FEI index can be considered a good parameter for good/bad-fish classification. In an exhaustive analysis of the ROC curves we found that the best accuracy values are located close to the perpendicular to the no discrimination line, and the FEI of these points ranges from 1.8 to 2.2.

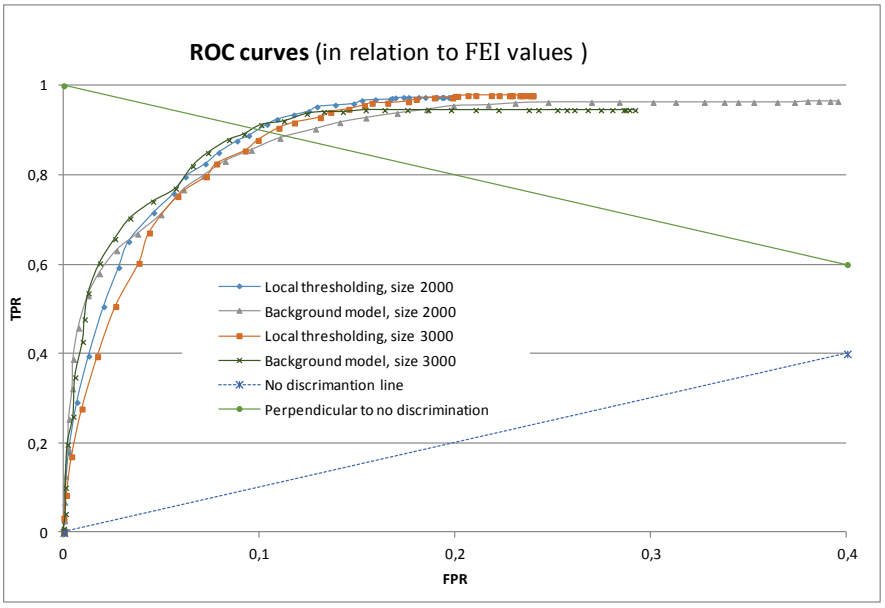

Figure 10: ROC curves for the test results.

4. Discussion

We want to emphasize the complexity of the videos used in our experiments to identify individual fish and also that our proposed model obtains the landmarks automatically, so that a 90.6% success rate is a very promising result. In previous studies such as

Zion et al. 2007 and Tillett et al. 2000 the results were achieved with images acquired in semi-controlled environments, although the fish were swimming, to demonstrate the performance of their system. Zion et al. 2007 classify 94% of grey mullet, St Peter´s fish and carp on images acquired while swimming through a narrow channel and using background illumination. Tillett et al. 2000 perform a model-based approach to estimate biomass. The images used in these trials are collected using the tank side as background and the algorithm needs some manual initiation to fish location. Also, the authors report that the model converges on the fish in 19 out of 26 cases (73%) and one priority for their future work is to link the classifier with the initial fish location.

To produce true biometric measurements in the near future, we will need to process the synchronized video acquired for both cameras in a configured stereoscopic system and acoustic data obtained with transducers may also be taken into account. In this kind of application an important factor to consider is the False Positives (FP). A high FP rate may lead to inaccurate estimations of fish size and thus a biometric mass will be computed that is very far from the real catch to the detriment of fishermen or government control. Our model can achieve a very low FP when considering blobs which have a very demanding FEI (close to 1.0) so in this case it can ensure good biometric measurements. And because the application can run for hours, a sufficient number of blobs can be obtained to ensure representative measurements.

As we saw in the experiments section, the FEI index shows a remarkable capacity to obtain good fits. This index performs well even when the segmented blob includes a fish body and small portions of other fish or if the blob presents holes due to inaccurate segmentation. The model is therefore able to overcome some segmentation problems. Although this is a positive point for any automatic fish finding application, human operators tend to classify these poorly segmented shapes as bad-fish in the ground-truth, leading to some questionable misclassifications in the form of false positives when the classifier is being tested.

5. Conclusions and future work

This research is expected to contribute to an automatic method for identifying individual fish in underwater real conditions. We propose a novel deformable tuna fish model that fits the fish body. Our model is adaptive and deformable because it takes fish length and flexion of the tuna during swimming into account. The initial tuna model is based on five parameters obtained automatically from the segmented blob of the image. The proposed procedure adjusts automatically to fish shape and size, bending to fit their

flexion motion. The proposed FEI (Fitting Error Index) has proved robust enough to overcome possible segmentation inaccuracies. When the fish has been modelled it will be possible to extract good measurements of fish length and other features. In the near future we could incorporate processing of the synchronized stereoscopic video in order to transform the length and thickness obtained by our model to true biometric measurements.

Although our model has proven able to correctly identify individuals whose segmented blobs included two or more tunas or one tuna with part of another individual, we still need to resolve the problem of overlapping tunas. For example, we hope to resolve correctly in the near future the identification of individuals whose heads are oriented in the same direction and with about 50% body area overlapping.

Another improvement on our model could be the definition of a new thickening parameter that allows us to carry out studies on growth control and tuna fattening.

In our experiments, the videos are highly complex because they were acquired in natural conditions, so we have worked with crowed scenarios where fish overlap, with wide variability of light intensity from one part to another of the same image, with poorly contrasted images due to murky water, fish in different planes and away from the camera, and of course, with continuously moving live fish. Furthermore, our proposal used landmarks obtained automatically without the need for human intervention. Considering all the above factors, the 90.6% success rate is a very promising result.

Acknowledgements

This work was partially supported by the EU Commission [2013/410/EU] and it has been possible thanks to the collaboration of the IEO (Spanish Oceanographic Institute).

We acknowledge the support of the Spanish company Balfego Group S.L. in supplying boats and divers to acquire underwater video in the Mediterranean Sea.

References

AKVA Group, 2014. http://www.akvagroup.com/products/land-based-aquaculture/camera-systems/biomass-estimator (accessed September 2014).

AQ1 Systems, 2013. http://www.aq1systems.com/products (accessed Sept. 2014).

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

Dewar H. and Graham, 1994, Studies Of Tropical Tuna Swimming Performance in a Large Water Tunnel – Kinematics, The Journal of Experimental Biology, 192(1):45-59.

Espinosa V., Soliveres E., Cebrecos A., Puig V., SainzPardo S. and de la Gándara F., 2011. Growing monitoring in sea cages: TS measurements issues. Proceedings of the 34th Scandinavian Symposium on Physical Acoustics, 2011.

Fletcher R. and Powell M.J.D., 1963. A Rapidly Convergent Descent Method for Minimization. Computer Journal, 6:163-168.

Fletcher R., 1980. Practical Methods of Optimization. Vol. 1, Unconstrained Optimization, John Wiley and Sons.

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

Hawkins J.D., Sepulveda, C.A., Graham, J.B. and Dickson, K.A., 2003. Swimming performance studies on the eastern Pacific bonito Sarda chiliensis, a close relative of the tunas (family Scombridae). II. Kinematic. The Journal of Experimental Biology, 206: 2749-2758.

Lee D., Schoenberger R.B., Shiozawa D., Xu X. and Zhan P., 2004. Contour Matching for Fish Species Recognition and Migration Monitoring. Proceedings of the SPIE, Volume 5606, pp. 37-48.

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, Elserveir, Computers and Electronics in Agriculture, 31:151–168.

Martínez-de-Dios R., Serna C. and Ollero A., 2003. Computer vision and robotics techniques in fish farms. Robotica. Vo. 21. No. 3. Editor Cambridge University Press. pp. 233-243.

Petrou M. and. Bosdogianni P., 1999. Image Processing: The Fundamentals. Wiley.

Piccardi. M., 2004. Background subtraction techniques: a review. IEEE International Conference on Systems, Man and Cybernetics. pp. 3099–3104.

Shortis, M., Harveyb E. and Seagerb J., 2007. A Review of the Status and Trends in Underwater Videometric Measurement. IS&T/SPIE Electronic Imaging, Videometrics IX.

Shortis M. R., Mehdi Ravanbakskh, Faisal Shaifat, Euan S. Harvey, Ajmal Mian, et al, 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.

Spampinato C., Giordano D., Di Salvo R., Chen-Burger Y., Fisher R. B. and Nadarajan, G., 2010. Automatic Fish Classification for Underwater Species Behavior Understanding. ARTEMIS'10, 2010, Firenze, Italy.

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.

Zion B., Alchanatis V., Ostrovsky V., Assaf Barki and Karplus I., 2007. Real-time underwater sorting of edible fish species. Computers and Electronics in Agriculture 56:34–45.

Zion B., 2012. The use of computer vision technologies in aquaculture – A review. Computers and Electronics in Agriculture 88:125–132.