Estimating Dimensions of Free-Swimming Fish Using 3D Point Distribution Models

Robin Tillett, Nigel McFarlane, and Jeff Lines

Silsoe Research Institute, Wrest Park, Silsoe, Bedford MK45 4HS, United Kingdom
E-mail: robin.tillett@bbsrc.ac.uk

Monitoring the growth of farmed fish is an important task which is currently difficult to carry out. An underwater stereo image analysis technique offers the potential for estimating key dimensions of free-swimming fish, from which the fish mass can be estimated. This paper describes the development of a three-dimensional point distribution model to capture the typical shape and variability of salmon viewed from the side. The model was fitted to stereo images of test fish by minimizing an energy function, which was based on probability distributions. The minimization was an iterated two-step method in which edges were selected for magnitude, direction, and proximity to the model, and the model was then fitted to the edges. A search strategy for locating the edges in 3D was devised. The model is tested on two image sets. In the first set 19 of the 26 fish are located in spite of their variable appearance and the presence of neighboring fish. In the second set the measurements made on 11 images of fish are compared with manual measurements of the fish dimensions and show an average error in length estimation of 5%. © 2000 Academic Press

Key Words: model-based image processing; stereo; measurement; deformable objects.

1. INTRODUCTION

This paper describes progress toward an automatic vision-based system for estimating the weight of free-swimming salmon. Estimation of fish size and weight is of significant interest to the fish farming industry since it provides key data with which feeding, grading, and harvesting operations are controlled.

The most widespread method for estimating the weight of farmed salmon in a sea cage relies on a small sample of fish being caught in a hand net and weighed. This is a stressful and potentially injurious process for the sample fish. It is also a process which provides relatively poor weight estimates for the population because of the small size of the sample (frequently only 10 fish from a population of 20,000) and because it samples only the fish

close to the surface and within easy reach of the sampling net. The alternative method in current use is the submersible biomass counter. This is a device somewhat like a thick picture frame through which the fish can swim. As they pass through this frame they break a finely spaced lattice of infrared beams, thus generating a shadow image from which features such as maximum height and length can be estimated. Problems with this device are both the reluctance of many fish to swim through such a constricted space and the limited number of features which can be extracted from the data. A vision-based system need not require a fish to swim in a constricted zone and could offer far more information on which to base an estimation of biomass. In support of this objective Hockaday *et al.* [1] have developed several models relating the separation of landmark points on Atlantic salmon and other species of fish to their mass, strain, and maturity status. This data was based on hand measurements made on anesthetized fish [2]. Beddow *et al.* [3] have shown that the masses of individual salmon can be measured to within ±9% from stereo images using manually identified points.

Images of salmon taken in production sea cages can vary significantly in quality due to the direction and strength of the surface illumination and the effect of water surface motion on the subsurface illumination. Light scatter due to particles in the water and the natural camouflage coloring of the salmon combine to give complex low-contrast images which may have multiple apparent edges close in position and similar in magnitude and direction to the real boundaries of the fish. These fish may also be against a nonuniform background which may be lighter or darker than the fish and which can include other, partially occluded fish.

The nature of these images requires that a model-based technique be used for image segmentation. Methods for fitting models to image data include those which measure the image only at the current positions of the template [4], such as snakes [5–8]. These attempt to slide into position by following local image gradients. This approach was inadequate in this application, because the initial landmark positions were, in general, too far away from their correct positions for the correct edges to have any influence, even if the model was initialized exactly in the center of the fish. Stochastic methods such as Markov chain Monte Carlo [9] can escape from local minima and featureless parts of the image by random jumps, but such methods tend to be slow [10].

A point distribution model (PDM) [11] appeared to be the most suitable technique for segmenting the images. However, a common approach with PDMs is to locate the edges by thresholding and then to attract the landmarks to the nearest edges [12]. This allows the PDM to be pulled by more distant edges, but the assumption that the correct edge was the nearest was not always valid in this application. In this work, a fitting method was developed in which the edges were located by thresholding, but the PDM was attracted not just to the nearest edge, but to edges selected by a combination of edge strength and proximity, allowing the PDM to be attracted to stronger edges beyond weaker but closer ones.

One of the problems of fitting models such as snakes is that the energy functions which are minimized contain constants which need to be "tuned" to the application, and which are difficult to estimate [13, 14]. This can be a somewhat arbitrary exercise and may compromise the general applicability of the model. In this work, arbitrary tuning was avoided by training the model on a set of images. The constants in the energy function were all related to probability distributions in this set of images.

This paper describes the techniques developed to apply a three-dimensional point distribution model to stereo images of salmon. The reliability with which a visual fit of the model to the fish in the images is achieved is discussed, with comments on the sensitivity

to the initial start position of the model. Key dimensions of the fish are extracted from the model and compared with dimensions obtained from visually identified points in the stereo images and also from caliper measurements taken from the fish.

The images to which this analysis was applied were taken in tanks rather than in the sea cages in order to maintain greater control over the system. However, many of the problems of images taken in sea cages are also found in these images. Some analysis of one set of images has been previously presented [15].

2. COLLECTING IMAGES AND GROUND TRUTH MEASUREMENTS

The stereo pair images of free-swimming fish were collected in a tank as shown in Fig. 1. The tank was 3.5 m in diameter and 1 m deep. The stereo camera pair was vertically oriented with a baseline of 0.5 m and with the optical axes converging 1.5 m in front of the cameras. For some images a black backboard was placed 1.8 m from the cameras. A black baseboard was also used because the fish often swam near the floor of the tank. The field of view of the cameras was approximately 2 m wide at the backboard.

Fish images were collected on two separate dates. The first set consisted of 36 stereo pairs of Atlantic salmon (*Salmo salar*) using the tank side as the background. The direction of the ambient lighting cast some shadows onto the tank side. These images were 320×256 pixels. The second set, also of Atlantic salmon, consisted of sequences of stereo pairs of the salmon against a black backboard. The ambient light was diffused by a covering of shade netting. These images were 384×288 pixels. The salmon in these images were tagged to allow individual identification. Typical stereo pairs from the two sets are shown in Fig. 2.

The stereo cameras were calibrated from images of a submerged grid of known dimensions using Tsai's method [16].

After the collection of the second set of images the tagged fish were anesthetized, removed from the water, and measured using calipers and a rule. The measurements taken are shown in Fig. 3. This set of lengths is known as a truss network. Methods for estimating mass from this network of lengths have been identified elsewhere [1–3]. Errors may be expected in these length measurements since the fish, although anesthetized, still have involuntary spasms, making precise measurement difficult. The measurements are likely to have errors of up to ±5 mm.

3. BUILDING A 3D POINT DISTRIBUTION MODEL

The shape of a fish, for this work, was defined as the shape, size, orientation, and position of its self-occluding boundary, or outline, as viewed from two directions by the stereo cameras. In defining the shape in this way, it was assumed that each fish was approximately flat in cross section, and hence that the points perceived to be on the self-occluding boundary were the same in both cameras.

The shape was modeled by a point distribution model [11] consisting of 26 landmark points. Figure 4 shows a stereo image pair with the landmarks indicated. Some of the landmarks corresponded to physical features of the fish, such as the tip of the tail or the junction of a fin with the body. Others were added to fill in the gaps along featureless parts of the shape. All of the landmarks were points on the self-occluding boundary. Further work might include adding internal features such as the eye or the gill slit to the model.
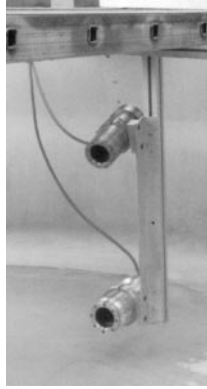
FIG. 1.   Aerial view of the image collection equipment and close-up of the stereo cameras.

In this work, the fins, which were not always visible, were not included as part of the shape.

The training images were taken from image set 1. This set was divided into 18 images for training and 18 images for an independent test set. The training set contained 25 fish which were suitable for training the PDM, in that they were unoccluded in both stereo images and oriented sidelong to the cameras. For each fish in the training set, the landmarks were placed by hand in both images of the stereo pair. Then the two sets of image coordinates were used to calculate the 3D positions of the landmarks in world coordinates.
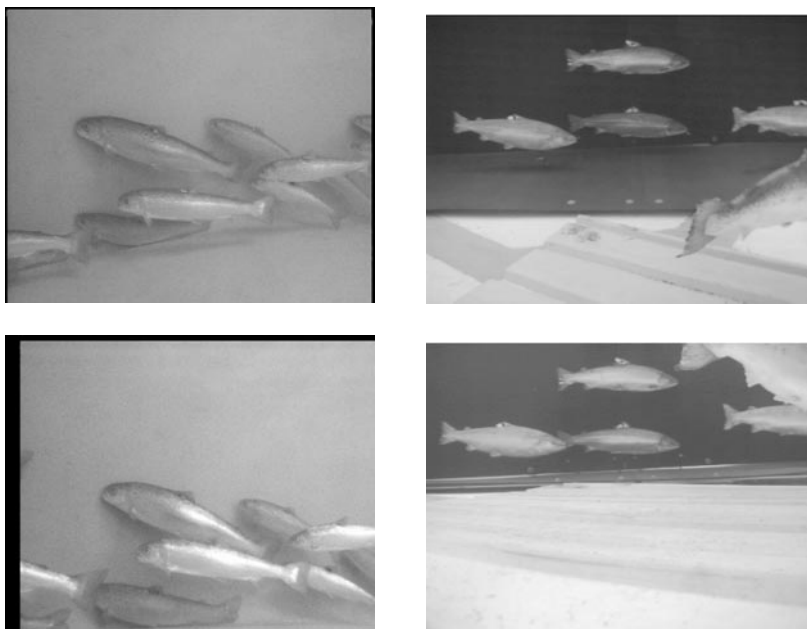
FIG. 2. Stereo pairs from image set 1 (left) and image set 2 (right).

The 25 example fish shapes were then normalized to the same scale, rotation, and translation. The translation parameters, x_0, y_0, and z_0, were respectively the mean x, y, and z real-world positions of the points; the rotation parameters, α, β, and γ, were respectively the Euler angles [17] describing the roll, pitch, and yaw of the principal axes; and the scale parameter, s, was the r.m.s. distance of the shape points from the center (x_0, y_0, z_0). Reflection was not included in the model, because the fish in this training set were all swimming in the same direction.

The values of the parameters for each shape were recorded, and each shape was then normalized so that x_0, y_0, and z_0 were all zero, α, β, and γ were all zero, and s was unity. This omitted the minimization step of Cootes *et al.* [11], in which all the shapes in the training set were translated, rotated, and scaled to optimize their fit to the collective mean shape. In this application, the principal axes of the shapes were well defined enough that there was little difference between the parameters derived by direct calculation and those
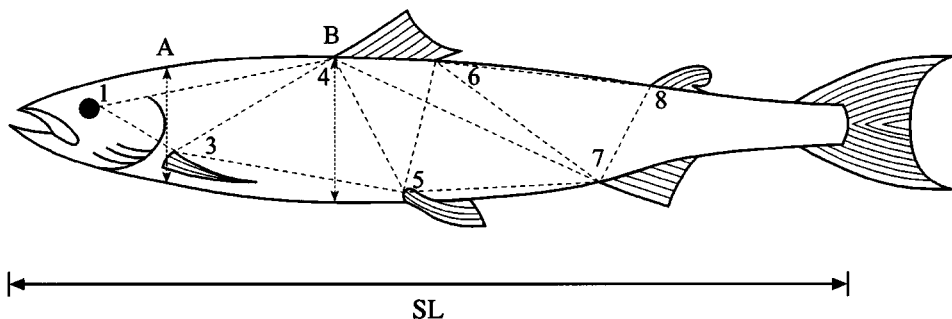

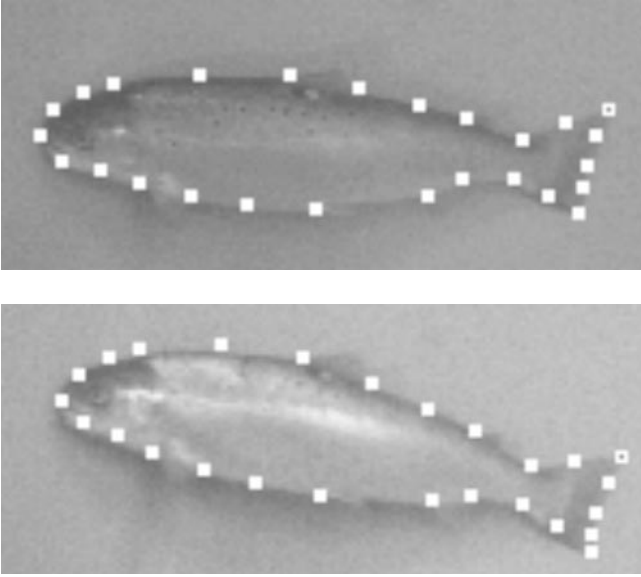**FIG. 3.** Diagram of the truss measurements taken manually on anesthetized salmon.

FIG. 4. Stereo pair with landmark points placed by hand.

derived by optimized fit, but the shapes produced by the optimal fit were not completely normalized, having some residual variation in scale and rotation. In this application, the variances in scale and rotation were potentially of interest, so the direct calculation, which separated the variations in scale, rotation, and translation entirely from the variations in pure shape, was preferred.

The modes of variation of the normalized shapes were then calculated, using principal component analysis, and arranged in order of descending eigenvalues. The eigenvectors of the covariance matrix represented the principal modes of variation of the shape, and the eigenvalues represented the variances along these modes. The shape model used in this work was thus given by

$$\mathbf{u}' = \mathbf{u}_m + \mathbf{Pb}, \tag{1}$$

where $\mathbf{u}' = (x_1 \ldots x_n, y_1 \ldots y_n, z_1 \ldots z_n)^{\mathrm{T}}$, $\mathbf{u}_m$ was the mean normalized shape, $\mathbf{P}$ was the $3n \times n_t$ matrix of the n_t eigenvectors with the largest eigenvalues, and $\mathbf{b}$ was the vector consisting of the n_t components of shape resolved along these eigenvectors; and by

$$\mathbf{x} = s\mathbf{R}(\alpha, \beta, \gamma)\mathbf{x}' + \mathbf{x_0}, \tag{2}$$

where $\mathbf{x}$ was the calculated shape, in the form of a $3 \times n$ matrix, $\mathbf{x}'$ was the $3 \times n$ matrix of coordinates corresponding to $\mathbf{u}'$, s was the scale parameter, $\mathbf{R}(\alpha, \beta, \gamma)$ was the Euler rotation matrix, and $\mathbf{x_0}$ was the translation vector $(x_0 \ y_0 \ z_0)^{\mathrm{T}}$. In this work, the value of n_t was 6, which was a fairly arbitrary cutoff which assumed that all but the first 6 modes of variation were negligible. Hence the shape model was described by 13 parameters: 6 modes of shape variation and 7 parameters of scale, rotation, and translation.

Figures 5a to 5c show the first mode of variation of the fish shape. Figures 5a and 5c show the shape modified by plus and minus 2 standard deviations about the mean shape $\mathbf{u}_m$, shown in Fig. 5b.
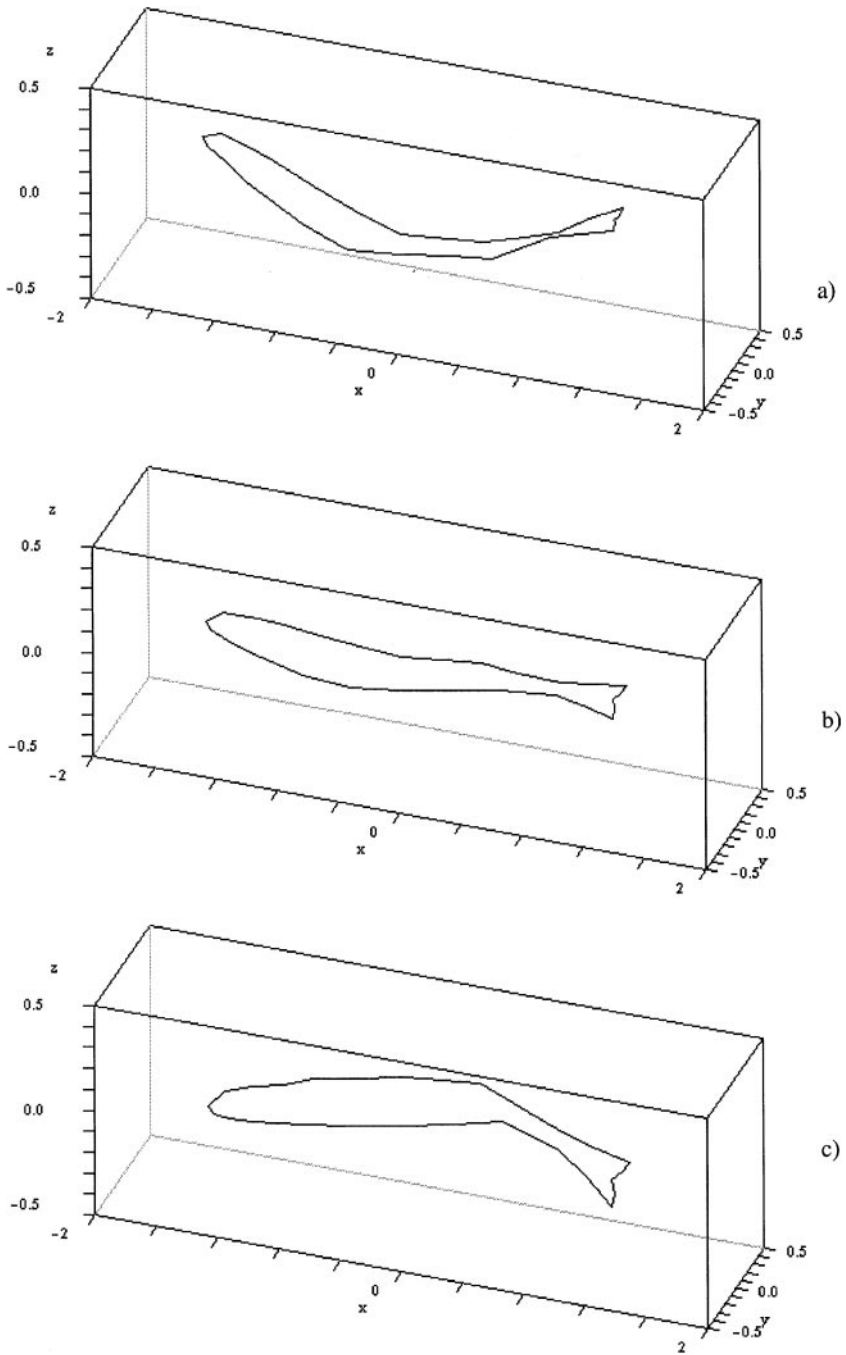
FIG. 5.   First mode of variation of the point distribution model.

4. FITTING THE MODEL

4.1. Calculating the Energy Function

The PDM was fitted to the test fish by minimizing an energy function. The energy function consisted of three components which represented the energy required to deform and to move

the model to a particular shape and position, the image energy pulling the model to specific edge points, and a residual energy due to the distance between the model and the edge points. Each component was configured in terms of a probability distribution or in a compatible form, so that no empirical weightings between the energy terms were required.

The components of the energy function were $E_{pdm}(\mathbf{p}, \mathbf{b})$, the energy of the PDM position and shape, where $\mathbf{p}$ was the vector comprising the seven components of scale, rotation, and translation, and $\mathbf{b}$ was the vector comprising the six components of shape variation; $E_{im}(\mathbf{e})$, the energy of the two stereo images at the world position $\mathbf{e}$ of a candidate edge; and $E_{resid}(d[\mathbf{x}, \mathbf{e}])$, the energy due to the residual distance $d[\mathbf{x}, \mathbf{e}]$ between the landmark at $\mathbf{x}$ and the corresponding candidate edge point at $\mathbf{e}$.

In this work, the energy of an event occurring, for example, a particular set of values of $\mathbf{p}$ and $\mathbf{b}$, was minus the log likelihood of the event. The more probable the event, the lower the energy. E_{pdm} was simple to calculate, since the model described the components as varying according to independent Gaussian distributions. Hence, E_{pdm} was given by

$$E_{pdm}(\mathbf{p}, \mathbf{b}) = \frac{1}{2}\left(\sum_{k=1,7}\frac{\Delta p_k^2}{\sigma_k^2} + \sum_{k=1,6}\frac{b_k^2}{\lambda_k}\right), \tag{3}$$

where Δp_k was the deviation from the mean of position parameter p_k, σ_k was the standard deviation of p_k, and λ_k was the kth eigenvalue of the PDM, equal to the variance of b_k. The additive constants have been ignored throughout this work. E_{resid} was similarly easy to calculate, but in this case, a Gaussian distribution of likelihood vs distance was not suitable, because of the presence of outlier points. The edge candidates were sometimes incorrect and occurred at greater distances from their corresponding landmarks than would have been allowed for by a Gaussian distribution. To avoid a small number of outliers exerting too much influence on the overall fit of the PDM to the points, a robust energy function, corresponding to the Cauchy distribution [18], was used. This was given by

$$E_{resid}(d) = \ln\left(1 + d^2/2d_0^2\right), \tag{4}$$

where d_0 was a length constant analogous to the standard deviations in Eq. (3).

Finally, the image energy $E_{im}(\mathbf{e})$ was a function of the edge magnitude and direction at the image coordinates corresponding to the world point $\mathbf{e}$. All of the landmark points in this work corresponded to edge features, and therefore needed to be fitted to edges in the images. An edge was assumed to be better the larger the magnitude of the image gradient and the more closely its direction matched that of the normal to the PDM at the corresponding landmark point. The absolute gray levels were not used, because of the danger of learning features which were only applicable to the current lighting conditions and camera angles. Knowledge of whether the fish were lighter or darker than the background was also not used, because this tended to vary from point to point around each fish, and hence required a model of the gray level pattern on the body of the fish. Such models have been used to model the gray level patterns on pigs [19, 20] or the colors of leaves [21], but the gray level patterns in the images used here included highlights and shadows which were specific to the imaging configuration. The image energy at world point $\mathbf{e}$ was given by

$$E_{im}(x_T(\mathbf{e}), y_T(\mathbf{e}), x_B(\mathbf{e}), y_B(\mathbf{e})) = \frac{255 - g_T}{g_0} + \frac{\Delta\theta_T^2}{2\theta_0^2} + \frac{255 - g_B}{g_0} + \frac{\Delta\theta_B^2}{2\theta_0^2}, \tag{5}$$

FIG. 6. Typical edge magnitudes from a Sobel operator.

where x_T, y_T, x_B, and y_B were the coordinates in the top and bottom stereo images, g_T and g_B were the magnitudes of the image gradients in the images, and $\Delta\theta_T$ and $\Delta\theta_B$ were the differences between the observed gradient directions and the normals to the PDM. The gradient magnitudes and directions were calculated for the images using a Sobel operator. A typical edge image is shown in Fig. 6. The direction normal to the PDM projected into an image, at a particular landmark, was the normal to the vector drawn through the two neighboring landmark points. An angle θ was regarded as the same as $\theta + \pi$ for calculating the angular differences, because it was possible for parts of the fish to be brighter or darker than the background, and therefore no distinction was made between gradient vectors pointing inward or outward from the fish. The constant θ_0 was the standard deviation of the observed directions about the PDM normals, assuming a Gaussian distribution. The distribution of the gradient magnitude had no single preferred value, so a linear energy function was used for simplicity. The constant g_0 was a measure of the expected magnitude of the correct edges compared to that of the image noise.

4.2. Finding the Minimum Energy Fit to an Image

The energies described in Section 4.1 were combined into a total energy function E given by

$$E = E_{\text{pdm}}(\mathbf{p}, \mathbf{b}) + \sum_{i=1,n} \min_{j} (E_{\text{im}}(\mathbf{e}_{ij}) + E_{\text{resid}}(d[\mathbf{x}_i(\mathbf{p}, \mathbf{b}), \mathbf{e}_{ij}])), \qquad (6)$$

where $\mathbf{x}_i(\mathbf{p}, \mathbf{b})$ was the position of the ith landmark point and $\mathbf{e}_{ij}$ was the position of the jth edge candidate corresponding to the ith landmark. The method used to fit the PDM to the fish was an iterated two-step method.

In step 1, the PDM was held fixed, while a set of candidate edges $\mathbf{e}_{ij}$ was identified for each landmark point $\mathbf{x}_i$. In each image, a search was performed along the normals to the landmarks, recording edges for which the magnitudes and directions were within thresholded limits. The thresholds were not essential for the algorithm to work, but they reduced the processing time required by removing obvious nonedges at an early stage. At this point, the edge candidates for each landmark consisted of two lists of 2D image coordinates, which were combined by converting every pair of points (x_T, y_T) and (x_B, y_B) into a world point $\mathbf{e}_{ij}$. In general, the normal search directions in the images did not coincide with epipolar lines, so the reprojection of $\mathbf{e}_{ij}$ into the images was a point close to, but no longer the same as, (x_T, y_T) and (x_B, y_B). This was a potential problem, because the image energy would be measured at image points different from those at which the original edges had been recorded. However, if it was assumed that the edges were approximately straight over the distances involved, then if (x_T, y_T) and (x_B, y_B) were points on the same edge, the reprojected points should also fall on the edge. The candidate edges $\mathbf{e}_{ij}$, having been obtained, the function

$$E_1 = E_{\text{im}}(\mathbf{e}_{ij}) + E_{\text{resid}}(d[\mathbf{x}_i(\mathbf{p}, \mathbf{b}), \mathbf{e}_{ij}]) \tag{7}$$

was then minimized over j for each landmark, so that each landmark $\mathbf{x}_i$ was associated with one edge point $\mathbf{e}_i$.

In step 2, the best edge candidates were held fixed while the PDM was fitted to them by minimizing the function

$$E_2 = E_{\text{pdm}}(\mathbf{p}, \mathbf{b}) + \sum_{i=1,n} E_{\text{resid}}(d[\mathbf{x}_i(\mathbf{p}, \mathbf{b}), \mathbf{e}_i]) \tag{8}$$

over all components of $\mathbf{p}$ and $\mathbf{b}$. Powell's method [18] was used for the minimization of E_2. Steps 1 and 2 were iterated several times.

At the start of the iterations, it was not known how close the PDM was to its correct position, but at the end, assuming that the PDM had converged properly, the landmarks were expected to be close to their corresponding image edges. Hence it was reasonable to "anneal" Eq. (4) by gradually reducing the distance d_0 on each iteration, and hence changing the relative weight given to the strength and the distance of the edge candidates. The value of d_0 on iteration k was given by

$$d_0(k) = d_{00}/\mu^k, \tag{9}$$

where d_{00} was the initial value and μ was an annealing rate constant.

5. RESULTS

5.1. Fitting the Model to Image Set 1

The constants which were not learned during the training of the PDM were set to the following values:

$$g_0 = 20 \text{ gray levels per pixel}$$
$$\theta_0 = 0.175 \text{ rad}$$
$$d_{00} = 50.0 \text{ mm}$$
$$\mu = 1.2.$$

The fitting algorithm was tested on the 26 unoccluded fish in the 18 test images from image set 1. These images had not been used in training the model, but were of the same group of fish. For each of the 26 fish the PDM parameters were set to their mean training set values, and the PDM was then moved by hand so that its translated position was close to the center of the test fish. The scale, rotation, and shape were not changed. The algorithm described in Section 4 was then executed for 34 iterations. Figure 7 shows the initial position of the PDM in a stereo pair of test images, Fig. 8 shows the best candidate edges detected from the initial position, and Fig. 9 shows the final position of the PDM.

The success of the algorithm in converging on the fish was judged by eye. The PDM was judged to have converged on the fish if most of the landmarks were on the occluding boundary of the fish in both images. In 19 of the 26 tests, the PDM converged successfully, and in 5 of these cases, all of the landmarks were positioned correctly. Complete failure to converge (as shown in Fig. 10) could usually be attributed to one of three reasons:

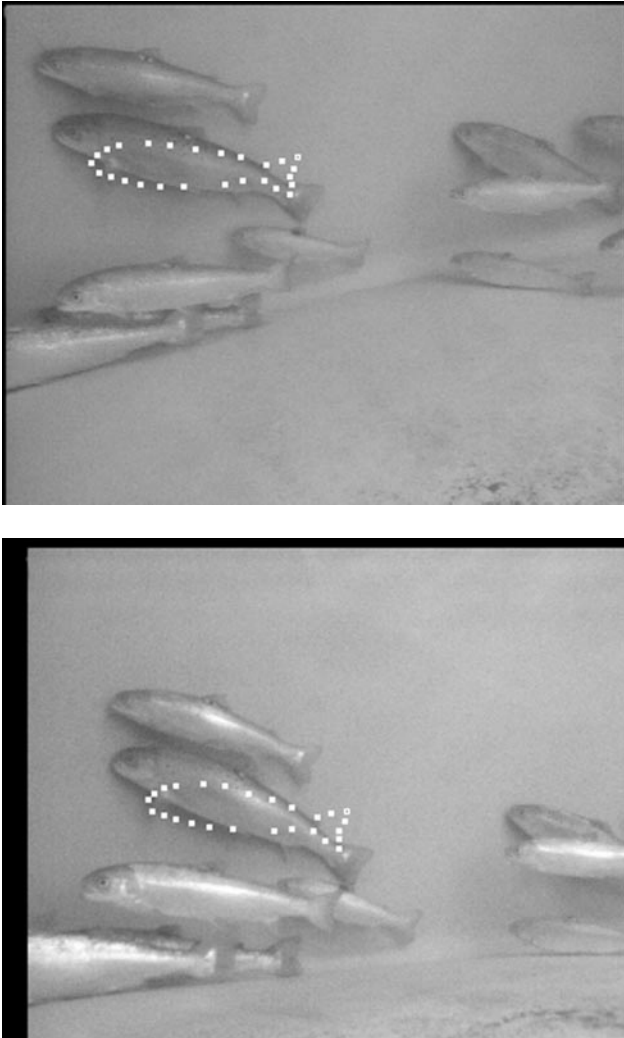- The PDM was pulled out of position by attempting to fit to the edges of a neighboring fish.

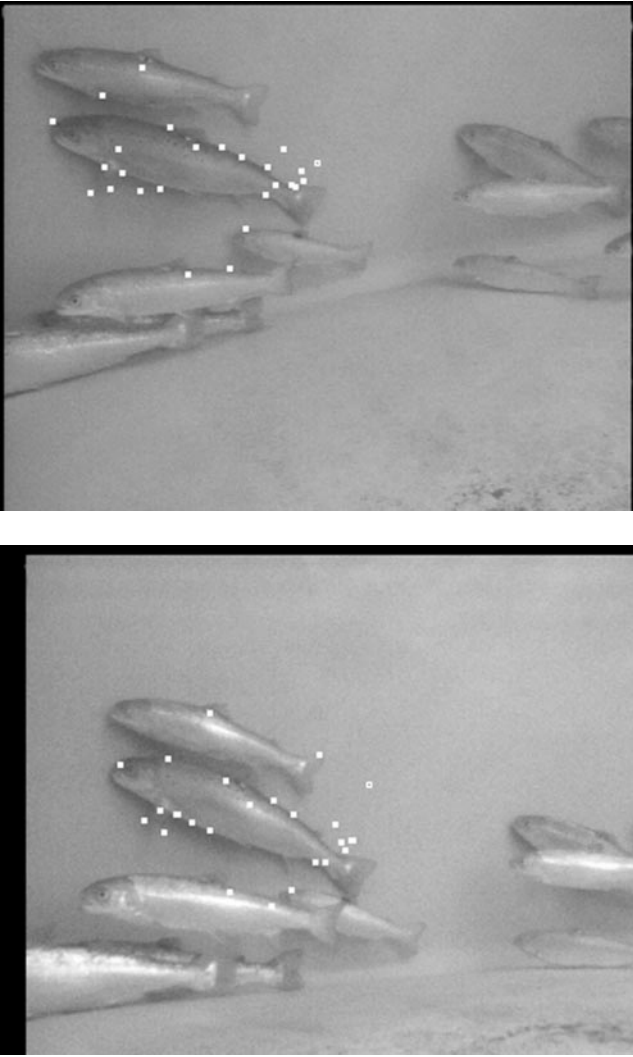**FIG. 7.** Initial position of the PDM.

FIG. 8. Candidate edge points found from the initial position.

• The orientation of the test fish was so different from that of the initial PDM that the edges of the fish were ignored because they were not in an appropriate direction.

• The fish were much smaller than the initial PDM. This may have been due to the limits placed on the edge searching, which prevented edges from being detected too far away in the inward direction. This limit was supposed to prevent the opposite sides of the PDM from pursuing the same edges, but it may also have made it difficult for the PDM to shrink.

5.2. Fitting the Model to Image Set 2

Image set 2 contained images of tagged fish for which manual measurements had also been taken. Eleven fish in five images were selected where the fish were well presented, isolated from neighbors, and against a black background.

The fitting algorithm was used exactly as described in Section 5.1 with no adjustment of the parameters. A visually good fit could be achieved for all the fish from a manually selected
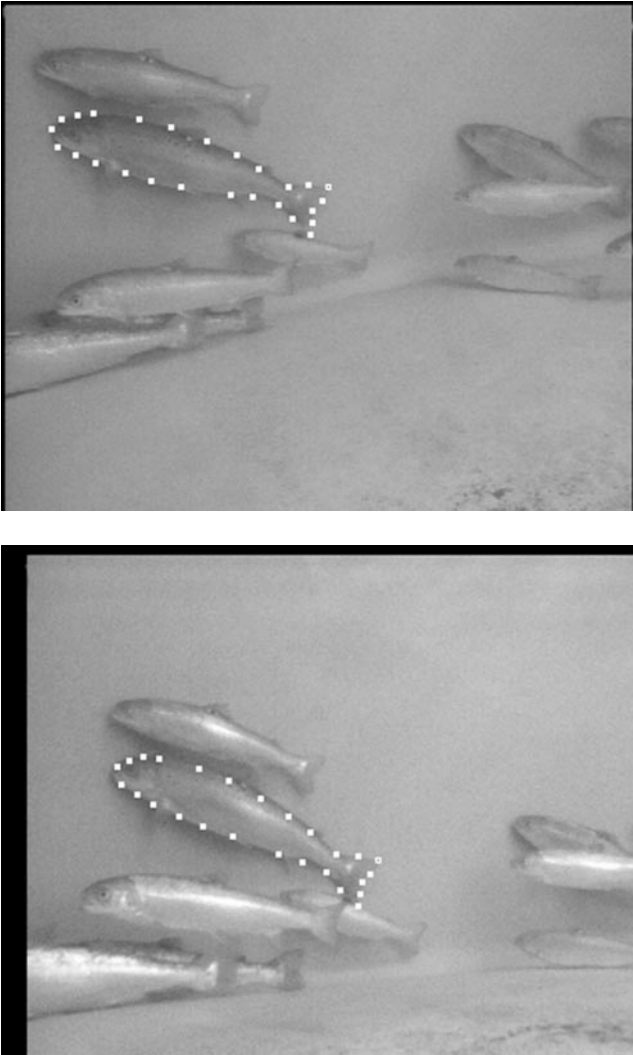
FIG. 9. Final position of the PDM.

starting point. Figure 11 shows the results for four of the fish. The bottom right example shows a fish with significant flexing of its tail. The fitted model is flexed but not as much as the actual fish. This is because the training set did not contain such large movements, and because certain sections of the edge start to be occluded by the fish itself.

The results shown in Fig. 11 appear to show a consistent difference in shape between the fish in image set 1 and the fish in image set 2. The fish in image set 1 appear to be longer and thinner. The model was built on image set 1 and when applied to image set 2 tends to converge with the top and bottom edges inside the apparent edge of the fish.

5.3. Sensitivity to Starting Position

The design of the model-based approach and iterative fitting method was aimed at developing a reasonably reliable technique for locating and measuring fish without requiring exhaustive search on each image. In the fitting trials described above it was clear that the
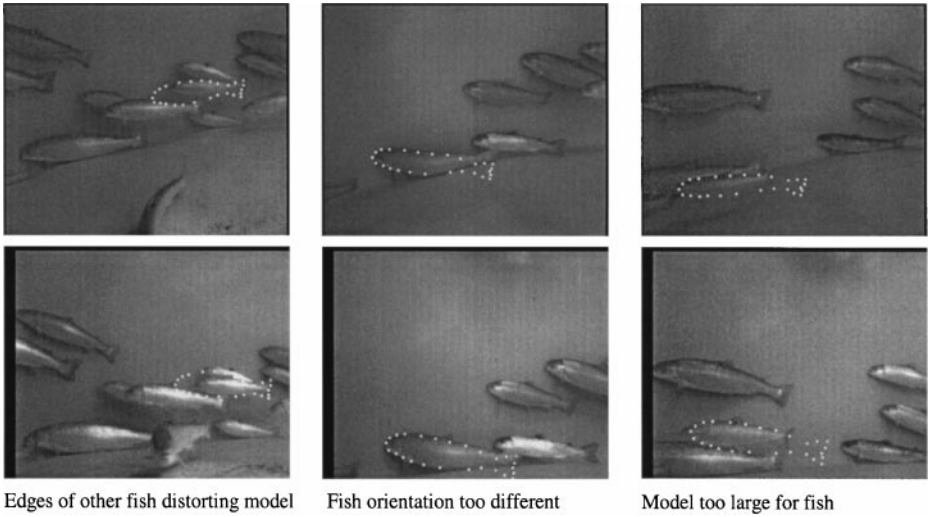
Edges of other fish distorting model   Fish orientation too different   Model too large for fish

FIG. 10. Failure of the PDM to converge correctly.

outcome of the model fitting was sensitive to the initial placement of the model and that the sensitivity was highly dependent on the image content. To demonstrate the nature of this sensitivity we have run the fitting algorithm for a range of starting positions for fish 1 and fish 8 (as shown in Fig. 11).

The initial placement of the model was varied from a manually selected point close to the fish center by steps of 20 mm in the real-world coordinate frame in the x, y, and z directions separately. In each direction the position was moved six steps, giving a distance of up to ± 120 mm from the initial point. This gave 37 separate starting positions. The roll, pitch, yaw, and scale of the fish were not altered.

For fish 1, 27 of the 37 starting positions resulted in a well-positioned fitted model, as judged by eye. The model converged correctly for $-40 \leq x$, $-100 \leq y \leq 40$, and $z \leq 100$.

For fish 8, only 3 of the 37 starting positions resulted in a well-positioned fitted model. For this fish there is a distinctive change in illumination along the middle of its side, particularly in the lower image. The sequence of starting positions was repeated with the scale of the model increased by 15%. In this case, 12 of the 37 starting positions resulted in well-positioned fitted models. The model converged correctly for $-80 \leq x \leq 0$, $0 \leq y \leq 60$, and $0 \leq z \leq 80$.

The range of acceptable starting positions for fish 8 is much smaller than that for fish 1. This is due to the much higher level of distracting edges both within and close to the outline of fish 8.

5.4. Estimating the Dimensions of Free-Swimming Fish

The primary motivation for developing model-based image analysis for fish was to allow the estimation of the fish mass. Therefore it is important to assess whether this technique has potential for estimating the key dimensions of fish.

Three sets of measurements can be considered for each fish image in image set 2. The first is the truss measurement made using calipers on the anesthetized fish. These truss measurements can be regarded as ground truth provided that due consideration is given to
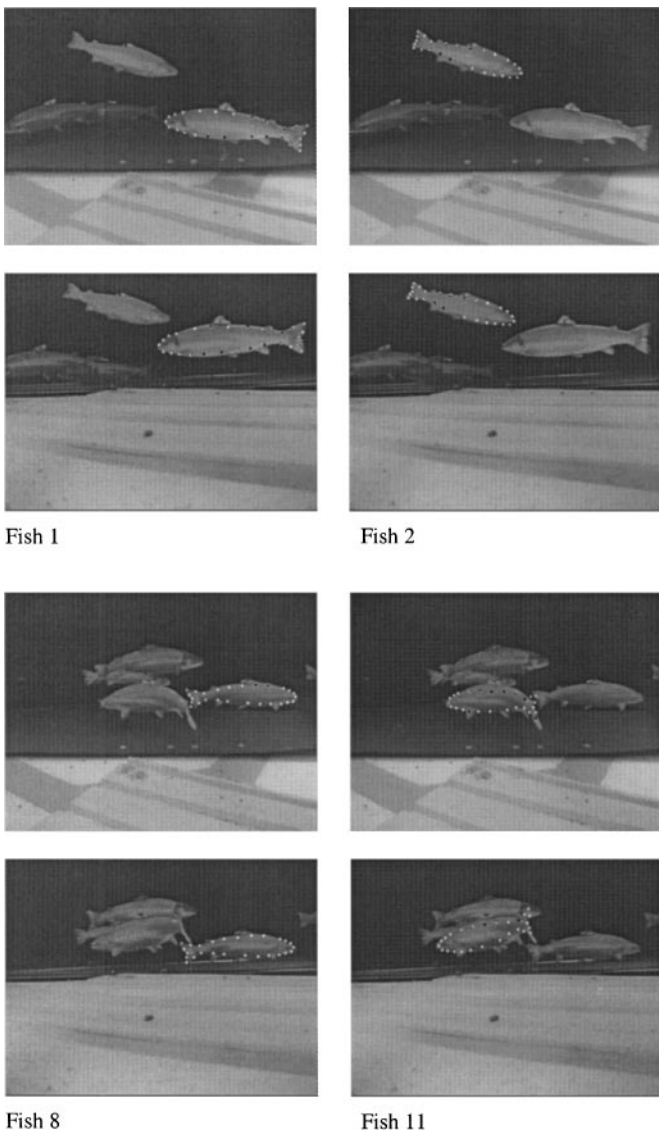
Fish 1　　　　　　　　Fish 2

Fish 8　　　　　　　　Fish 11

FIG. 11. Examples of the fitted model on four fish from image set 2.

the previous comments on their likely accuracy. The second set of data was collected by manually positioning a cursor on the landmark points in each image, calculating the world coordinates of each landmark using the stereo calibration, and calculating the Euclidean distances between the landmark points. It was not always easy to identify all the landmarks; for example, the pectoral fin which marks point 3 (Fig. 3) was often difficult to see. The third set of data is from the fitted point distribution model. For this data it was necessary to estimate the position of some landmarks from combinations of the model points, in particular the eye, which is not on the boundary of the fish.

Figure 12 shows an example of the world points recovered from the stereo image pair of fish 10. The projections onto the $x–y$ plane and the $x–z$ plane are shown. The solid markers show the landmark points (as in Fig. 3) and the open markers show the final fit of the model.
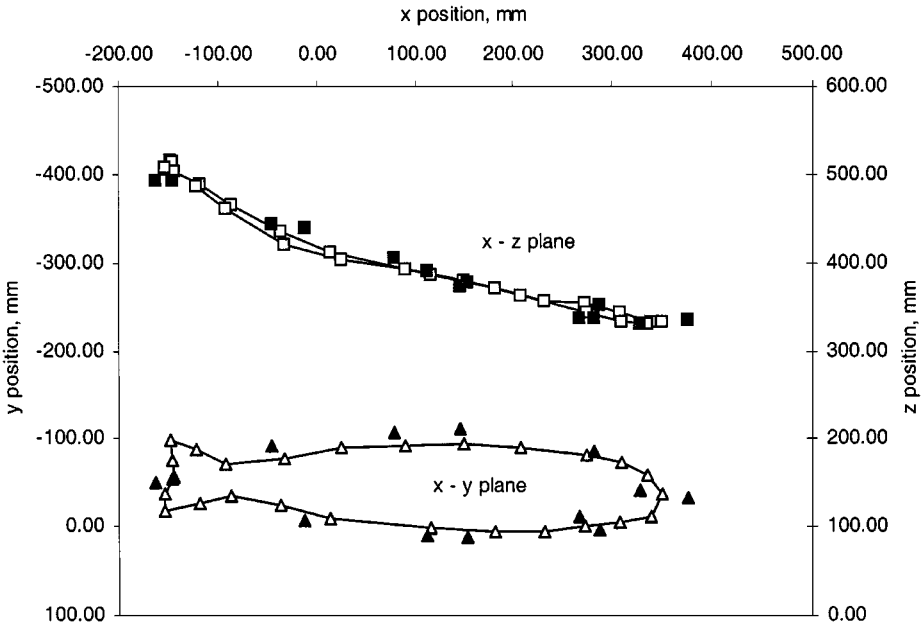
FIG. 12.   Positions of hand-picked landmark points (solid markers) and fitted model (open markers) for fish 10.

The smoothness in the depth estimation (i.e., in the z direction) and the high correlation between manual and model-based estimates of the world position were similar for all the fish.

The world points were used to recover estimated lengths and these are shown for the fish length (SL) and the fish depth (B) in Fig. 13. The hand-picked points gave a reasonable estimate of the lengths with mean errors of 1% for SL and 3% for B. The model-based estimates showed a consistent offset, giving high mean errors (5% for SL and 25% for B) but low standard deviations for those errors (2.8% for SL and 2.3% for B). The equivalent standard deviations in the errors for the hand-picked points were 3.2% for SL and 4.4% for B. The general trend was the same for the other truss length estimates.
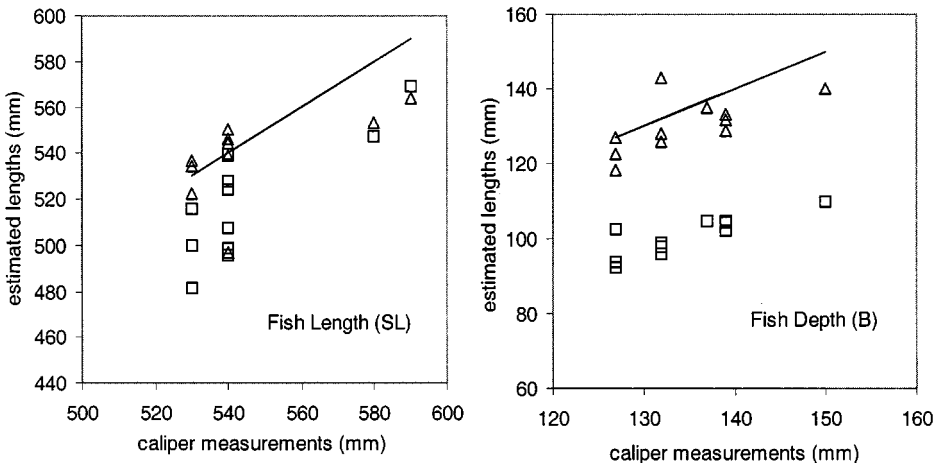

FIG. 13.   Comparison of caliper measurements of fish length (SL) and depth (B) with estimated values from hand-picked points (△) and points identified by model (□).

The data available only covered a narrow spread of fish sizes and so these results only give an initial indication of the potential of this technique. However, it appears that the model-based approach will position the landmarks reliably so that key dimensions of the fish can be estimated. A specific calibration may be required to allow for bias in how the model fits to the fish images.

6. DISCUSSION

The work presented here has indicated the potential for using a model-based approach to estimate fish biomass. These trials used images from tanks and manual initiation of the fitting algorithm, so it is relevant to indicate whether the approach can be extended to a fully automatic approach for use in the sea cage.

Images collected in the sea cage will show greater variability in terms of fish posture and position, occlusion, water quality, and illumination. The technique chosen for this work was a model-based approach specifically to allow for poorly defined or missing information around the fish outline. The fitting technique uses an iterative approach allowing weaker edges to be selected if they fit well with the converging model position. These techniques should be appropriate to sea cage images, although it will also be important to reject images where fish are not well presented and to check for convergence of the model.

Associated work is investigating the use of an n-tuple pattern classifier to give an initial estimate of fish position in the image [22, 23]. The classifier is applied to thresholded difference images from a sequence. Initial results indicate some promise in locating possible fish head positions, and further work will link the two techniques. If the performance of this combination of techniques is sufficiently reliable, then a fully automatic biomass estimation system can be proposed.

The priorities for future work are linking with the classifier for initial fish location, developing methods for rejecting common failures in the model fitting, and assessing the overall performance of the proposed approach. Other interesting areas such as improving the estimates by using a whole image sequence may be investigated , but may not be necessary for this application.

7. CONCLUSIONS

A model-based approach which allows the estimation of dimensions of free-swimming fish has been developed.

The point distribution model successfully modeled the shapes of the fish in 3D with six modes of variation.

The two-step fitting algorithm allowed the PDM to locate and move toward edges which were some distance away. In selecting the edge points to move toward, a combination of edge strength and distance was used, which enabled the PDM to locate the best rather than just the nearest edges. The constants in the fitting algorithm were all related to probability distributions of shape, gray level, direction, and distance and so their magnitudes were easily measured or estimated.

The model converged on the fish in image set 1 in 19 of 26 cases (73%). This was in the presence of image noise in the form of shadows, neighboring fish, and shading patterns on the fish themselves.

On a second set of images of clearly presented fish the model always converged well when the initialization point was close to the fish position.

The length estimates for well-presented fish made by the model gave an average error of 5%. The standard deviation of the errors was 2.8%, indicating a reasonable correlation of the estimate with the true length, but with a consistent offset.

ACKNOWLEDGMENTS

REFERENCES

1. S. Hockaday, L. G. Ross, and R. D. Tillett, Using stereo image pairs to measure mass in strains of Atlantic salmon (*Salmo salar* L), in *Sensors and Their Applications VIII, Section A Environmental and Biomedical Sensors, 7–10 Sept. 1997*, pp. 21–26.

2. T. A. Beddow and L. G. Ross, Predicting biomass of Atlantic salmon from morphometric lateral measurements, *J. Fish Biol.* **49**(3), 1996, 469–482.

3. T. A. Beddow, L. G. Ross, and J. A. Marchant, Predicting salmon biomass remotely using a digital technique, *Aquaculture* **146**, 1996, 189–203.

4. L. H. Staib and J. S. Duncan, Boundary finding with parametrically deformable models, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(11), 1992.

5. M. Kass *et al.*, Snakes: Active contour models, *Internat. J. Comput. Vision* **1**, 1988, 321–331.

6. J. B. Waite and W. J. Welsh, An application of active contour models to head boundary location, in *Proc. 1st British Machine Vision Conf., Oxford, UK, 24–27 Sept. 1990*, pp. 407–412.

7. P. Radeva *et al.*, A snake for model-based segmentation, in *Proc. 5th Int. Conf. on Computer Vision, Cambridge, USA, 20–23 June 1995*, Chap. 162, pp. 816–821.

8. S. R. Gunn and S. M. Nixon, A robust snake implementation; a dual active contour, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(1), 1997.

9. K. V. Mardia, The art and science of Bayesian object recognition, in *Int. Conf. on Image Fusion and Shape Variability Techniques, Leeds, UK, 3–5 July 1996*.

10. K. M. A. de Souza, J. T. Kent, K. V. Mardia, and R. D. Tillett, Estimation of objects in highly variable images using Markov chain Monte Carlo, in *Proceedings of the British Machine Vision Conference, BMVC 1997, University of Essex, UK, 8–11 September 1997*, Vol. 2, pp. 460–469.

11. T. F. Cootes *et al.*, Training models of shape from sets of examples, in *Proc. 3rd British Machine Vision Conf., Leeds, UK, 22–24 Sept. 1992*, pp. 9–18.

12. T. F. Cootes and C. J. Taylor, Data driven refinement of active shape model search, in *Proc. 7th British Machine Vision Conf., Edinburgh, UK, September 9–12, 1996*, pp. 383–392.

13. X. Gang *et al.*, Robust active contours with insensitive parameters, *Pattern Recog.* **7**, 1994, 879–884.

14. R. Samadani, Adaptive snakes: Control of damping and material parameters, in *Proc. Conf. on Geometric Methods in Computer Vision, San Diego, 25–26 July 1991*.

15. N. J. B. McFarlane and R. D. Tillett, Fitting 3D point distribution models of fish to stereo images, in *Proceedings of the British Machine Vision Conference, BMVC 1997, University of Essex, UK, 8–11 September 1997*, Vol. 1, pp. 330–339.

16. R. Y. Tsai, An efficient and accurate camera calibration technique for 3D vision, in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Miami Beach, June 22–26, 1986*.

17. H. G. Goldstein, *Classical Mechanics*, 2nd ed., Addison–Wesley. Reading, MA, 1980.

18. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed., Cambridge Univ. Press, Cambridge, UK, 1995.

19. C. M. Onyango, J. A. Marchant, and B. P. Ruff, Model based location of pigs in scenes, *Comput. Electronics Agriculture* **12**, 1995, 261–273.

20. C. M. Onyango and J. A. Marchant, Modelling grey level surfaces using three-dimensional point distribution models, *Image Vision Comput.* **14**, 1996, 733–739.

21. C. M. Onyango and J. A. Marchant, Flexible colour point distribution models, *Image Vision Comput.* **14**, 1996, 703–708.

22. D. Chan, R. D. Tillett, L. G. Ross, and S. Hockaday, A trainable *n*-tuple pattern classifier and its application for monitoring fish underwater, in *Seventh International Conference on Image Processing and Its Applications, University of Manchester, Manchester, UK, 12–15 July 1999*, pp. 255–259.

23. D. Chan, R. D. Tillett, L. G. Ross, and S. Hockaday, Factors affecting the training of a WISARD classifier for monitoring fish underwater, in *Proc. British Machine Vision Conference 1999, University of Nottingham, Nottingham, UK, 13–16 Sept. 1999*, pp. 340–351.