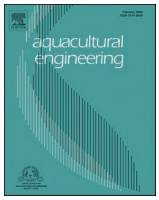
Aquacultural Engineering 97 (2022) 102227

Contents lists available at ScienceDirect

Aquacultural Engineering

journal homepage: www.elsevier.com/locate/aque



Automatic length estimation of free-swimming fish using an underwater 3D range-gated camera

Petter Risholm [a,*], Ahmed Mohammed [a], Trine Kirkhus [a], Sigmund Clausen [b], Leonid Vasilyev [b], Ole Folkedal [c], Øistein Johnsen [b], Karl Henrik Haugholt [a], Jens Thielemann [a]

[a] SINTEF Digital, Smart Sensor Systems, Oslo, Norway
[b] NORCE Technology, NORCE Norwegian Research Centre AS, Haugesund, Norway
[c] Animal Welfare Research Group, Institute of Marine Research, Matre, Norway

ARTICLE INFO

Keywords:
Farmed fish
Fish size estimation
Welfare indicator
Fisheries
3D imaging
Biomass

ABSTRACT

Camera estimation of fish size in aquaculture production and fisheries is crucial for enabling knowledge-based decision making. Automatic estimation of fish lengths on an individual and population basis are key steps to automate biomass estimation. We propose to use an underwater 3D range-gated camera for accurate fish-length estimation of free-swimming fish. The proposed algorithm requires no manual work or contact with the fish and is done in-situ. A robust algorithmic pipeline consisting of detection, tracking and fish length estimation stages is proposed. We show the accuracy of the proposed system on free-swimming fish, both in terms of individual fish lengths as well as the population distributions. The results show that the proposed system achieves length estimation errors in the order of 1% of manual-measured fish length.

1. Introduction

By using emerging technologies, the aquaculture industry and fisheries are moving from experience-based towards knowledge-based fish farming and capture (Føre et al., 2018; Li et al., 2020; Saberioon et al., 2017). Underwater (3D) vision systems are a key enabling technology to facilitate this transition, mainly because they can enable automatic behaviour analysis (Papadakis et al., 2012; Oppedal et al., 2011), species identification (Iqbal et al., 2021), fish count (Sharif et al., 2015), fish-length estimation (Costa et al., 2006; Shi et al., 2020), and biomass estimation (Li et al., 2020). In fisheries, this type of information can be used to decide whether to haul in a catch or not, or for automatic sorting of the catch before hauling it in. In aquaculture, biomass estimation is important in production planning, for feeding practice and pre-harvest sale of the fish (Li et al., 2020). As a result, these industries will increase profits by reducing the need for manual labour and optimize their operations and reduce bycatch.

While species identification and fish counts can be performed with a standard underwater 2D camera, accurate fish length estimation is very difficult to estimate without having high-resolution depth[1] information available (Shortis et al., 2016). One approach for fish-length estimation with a 2D camera is to install infrastructure, e.g. a tunnel in the water, so the fish always swim past the camera at the same calibrated distance (Miranda and Romero, 2017). However, if the goal is to have a more flexible vision system which does not depend on external infrastructure, a high-resolution 3D underwater camera which is able to estimate both an intensity and a depth value for each pixel, is required (Shortis et al., 2013).

Numerous optical 3D vision systems have been proposed for the purpose of fish length estimation, such as structured light (Lopes et al., 2017) and stereo-vision (Shortis et al., 2013). Stereo-vision has the advantage of being a relatively simple and cost-effective approach for gathering underwater 3D data of free-swimming fish. Many authors have proposed alternative stereo-vision systems for the purpose of fish length estimation, but they tend to struggle with limited viewing distance and water turbidity which smears out the details which are used in the feature matching part of the disparity estimation procedure.

In this study, we propose a range-gated 3D camera, which is described in detail in (Risholm et al., 2018), for the purpose of length estimation of free-swimming fish. The advantage of the 3D range-gated

* Corresponding author.
E-mail address: petter.risholm@sintef.no (P. Risholm).

[1] A 3D/depth sensor provides a metric distance (depth) measurement along each pixel ray of the camera. The depth should not be confused with the depth in the water column which is not relevant for the work presented here.

https://doi.org/10.1016/j.aquaeng.2022.102227
Received 8 April 2021; Received in revised form 18 January 2022; Accepted 18 January 2022
Available online 20 January 2022

camera is the combination of depth resolution, field of view and real-time 3D acquisition (10 Hz), and its resistance to turbidity makes it ideal for real-time monitoring of fish in e.g. fish-farms or on trawls.

Our algorithmic approach to length estimation constitutes a 3-stage pipeline of fish-detection, tracking, and length estimation. We propose to use DBSCAN (Ester et al., 1996), a versatile clustering algorithm, for detection and segmentation of the fish followed by a global tracking algorithm which aims at producing a set of reliable tracks without occlusions. In the final step, we rely on the depth map to compute a centreline and corresponding fish length estimates for each frame in a track. From each track/fish, we propose a robust filtering approach to obtain the final length estimate given the per-frame results.

We show, on data acquired of free-swimming fish in tanks and fish-farms, that the combination of the 3D range-gated camera and the robust algorithms provide accurate fish length estimates for both individual fish and a distribution within fish groups.

2. Materials and methods

2.1. Overview of pipeline for length-estimation

To provide reliable length estimates of fish swimming in an unconstrained environment, we propose a three-step process pipeline. Fig. 1 shows an overview of the algorithmic pipeline we propose for length estimation based on the range-gated 3D data.

The first step is to apply a detection algorithm to find and segment the fish in a single frame. Based on the depth of the fish, the orientation, and the location in the image, we apply a tracking algorithm to establish reliable tracks. For each frame in each established track, we estimate the metric length of the fish. Each track consists of a set of noisy length estimates which we filter to produce an accurate length estimate for each track. The distribution of track-based length estimates is summarized to generate accurate fish population statistics.

2.2. A range-gated underwater 3D camera

Capturing high quality underwater fish video can be challenging due to poor visibility and technological limitations in sensor design. Recently, the range gating 3D system UTOFIA (Risholm et al., 2018) as shown in Fig. 1, has provided an opportunity to capture high-resolution underwater intensity and depth images. We include a short description here for completeness.

A 532 nm solid-state laser with an active Q-switch combined with a fast range-gated black and white CMOS (1280 ×1024) chip makes up the vital parts of the range-gated system. The camera includes an onboard sequencer in firmware which allows for a fine-tuned control of the shutter opening in steps of 1.67 ns (~18 cm) in relation to the triggering of the laser pulse. The camera can acquire images at 1000 Hz. A set of images acquired with increments of 1.67 ns between the triggering of the laser pulse and the opening of the shutter is used to estimate the distance to objects in the field of view. Another benefit of the range gating is that we can use the image gated right in front of objects in the scene to remove the effect of backscatter and thereby increase the image contrast. An example of gated signal traces for pixels in an image are shown in Fig. 2.

Typically, we use 100 frames to generate a 3D frame, and consequently generate 3D frames at 10 Hz. We then organize the range-gating such that we image 25 ranges which covers 4 m (25 *18 cm). At each range we average four images to increase the signal to noise ratio (SNR). We can further increase the SNR by binning pixels. In this paper we use intensity images which are binned 2 × 2 and the range gated images are binned 4 × 4 before we compute the depth map. A super-resolution depth-estimation algorithm is used to achieve a depth-resolution of about 1 cm depending on the SNR.

The intrinsic camera parameters were estimated using Zhang's calibration method (Zhang, 2000) applied to checkerboard patterns imaged underwater. We found the field of view of the camera to be 50 degrees horizontally and 29 degrees vertically.

2.3. Fish detection and segmentation

Detection and segmentation of fish in 2D video streams can be difficult when one fish is partially occluding another fish. Typically, object detection algorithms will merge the two fish together because their texture will be similar. However, with a 3D representation of the scene, two fish swimming in front of each other are easy to separate out because they will swim at different depths.

Generally, there are two types of object detection methods to be used for detecting individual fishes: supervised and unsupervised. The supervised approaches require high quality labelled data to train machine learning models that can detect the target objects. Since manually labelling individual fishes is cumbersome, we rely on unsupervised clustering techniques. As we do not know how many clusters/fish we have in a frame, partition-based clustering such as k-means will not work. Instead we apply the density based clustering algorithm DBSCAN (density-based spatial clustering of applications with noise) (Ester et al., 1996) to detect and separate out arbitrary number of fish in the scene based on the depth map.

Before applying DBSCAN we filter out points in the depth map that have a low confidence/weak signals and apply a median filter with filter size 3. The depth map is converted into a point cloud by using the intrinsic camera parameters (focal length, principal point and lens

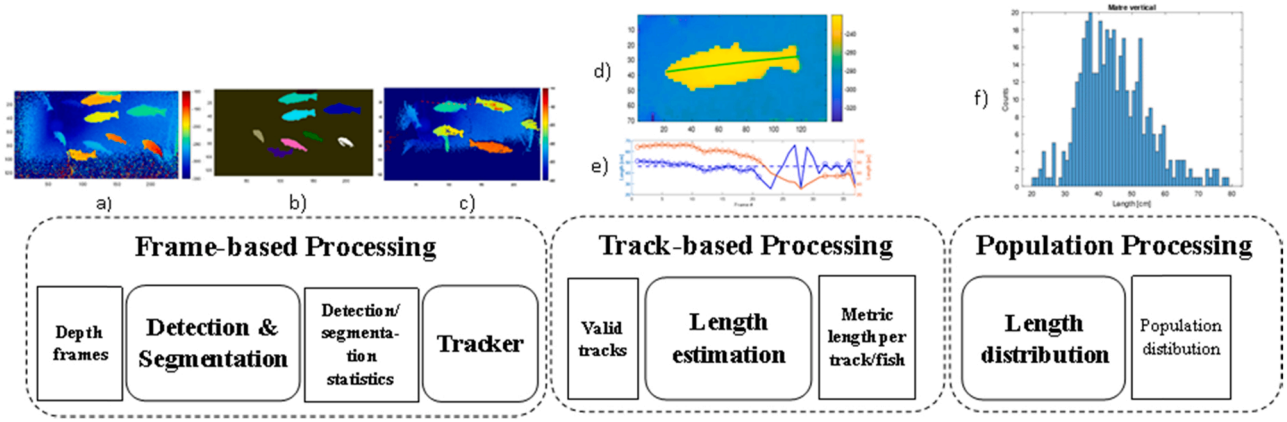
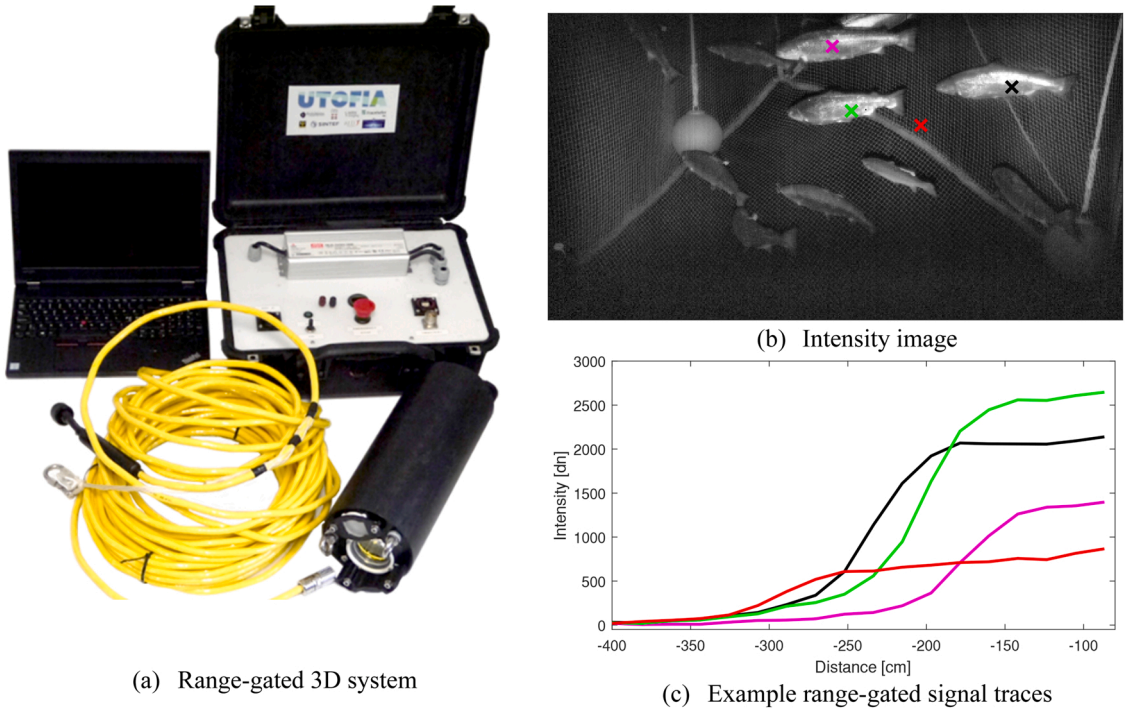


Fig. 1. Algorithmic pipeline. A DBSCAN clustering algorithm takes a depth frame (a) as input and produces a segmentation for each detected fish as output (b). A tracker takes a batch of detections as input and produces a set of tracks (c). Centrelines are estimated for each detection (d) and length estimates are computed for each frame in a track (e). A robust filtering method estimates a length estimate for each track to produce a population distribution (f).

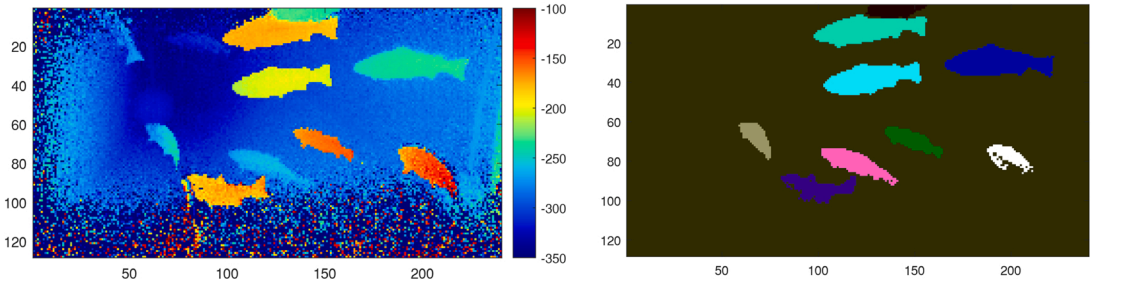


(a) Range-gated 3D system

(b) Intensity image

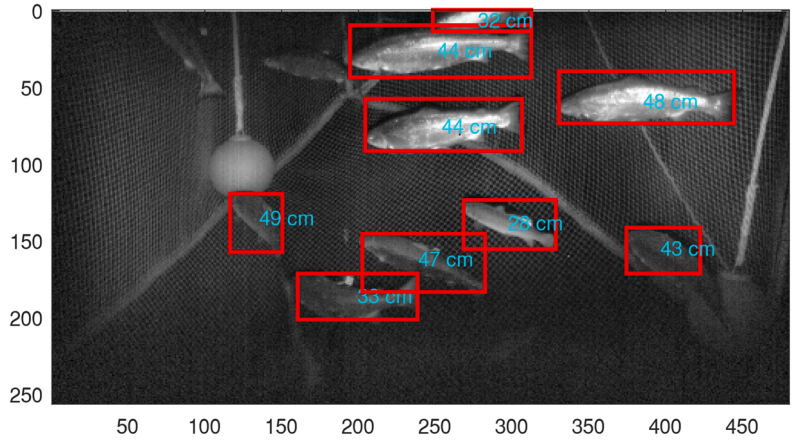
(c) Example range-gated signal traces

Fig. 2. Range-gated 3D system. (a) shows an image of the range-gated 3D system. The full system includes a top-side box, Gigabit Ethernet cable, PC, and a cylindrical underwater housing. (b) A gated intensity image generated by the camera. (c) Example range-gated intensity traces of the points marked with an 'x' in (b). The point at which the signal increases fastest represent the distance to the object.



(a) Depth map

(b) Label map output from DBSCAN clustering



(c) Bounding boxes with the maximum metric length of cluster

Fig. 3. DBSCAN clustering for fish segmentation. (a) Input depth map to DBSCAN. (b) Label map output from DBSCAN. (c) Intensity image overlaid with the DBSCAN bounding boxes and the maximum metric span of the associated cluster. (a)Depth map(b)Label map output from DBSCAN clustering(c)Bounding boxes with the maximum metric length of cluster.

distortion parameters) which are found through a standard checkerboard image calibration procedure (Zhang, 2000).

The output of the DBSCAN algorithm is a separation of all the points into different fish clusters. An example output from DBSCAN is shown in Fig. 3.

2.4. Fish tracking

A single fish swimming in front of the camera will be present in several consecutive frames. Tracking of individual fish is important to avoid counting the length of a single fish more than once. Since each track consists of several frames with the same fish, we can extract more robust individual fish estimates by clever filtering of the noisy frame-based length-estimates.

There are generally three approaches to multi-object tracking (Luo, 2021); sequential trackers, global trackers and deep learning based tracking. As global trackers have shown to be robust towards occlusions and re-identification, and are relatively flexible, fast and easy to implement, we based our implementation on a state-of-the-art global tracker (Wang et al., 2020). However, we do not anticipate that the length estimates would differ much by basing our implementation on one of the other main multi-object tracking approaches.

A global tracker uses all the detections in a batch of frames and tries to link the detections into trajectories. They tend to organize the detections as nodes in a directed acyclic graph (DAG) and solve for a maximum a-posteriori (MAP) estimate using the minimum-cost flow or circulation formulation. This makes the MAP problem computationally tractable (the time required for a computer to solve the problem is a simple polynomial function of the size of the input). We base our implementation on the minimum cost circulation (CINDA) framework (Wang et al., 2020).

Assume $X = \{x_i\}$ is a set of detections in a batch of frames. A DAG is constructed where the nodes are represented by the negative log likelihood $C_i = \log\frac{\beta_i}{1-\beta_i}$ of a detection x_i. We set $\beta_i = 0.0$ if we have found that a detection was occluded to exclude it from being included in any tracks. Otherwise we set $\beta_i = 0.9$ to make it likely that the detection is included in a track. An extension of this scheme would adjust β_i according to the confidence we have in the detection. We connect each node to a source/enter node $C_i^{en} = -\log P_{enter}(x_i)$ and sink/exit node $C_i^{ex} = -\log P_{exit}(x_i)$, where the edge values define the negative log likelihood of that node being the start or end of a track. We define each of these likelihoods the same way, where the likelihood increases with a diminishing distance to the boundary of the image and the further away from the camera the detection is. The edges connecting the nodes are modelled by the negative log likelihood of two detections belonging to the same track $C_{i,j} = -\log P(x_i|y_j)$. We model $C_{i,j}$ with three independent Boltzmann distributions which accounts for the distance in depth between the detection, the distance in image-space between the detections and the frame horizon distance. We have used a frame horizon of 10 frames, where a detection is connected to all detections in the next 10 frames, but with a diminishing probability the further out in time the detection is. This is to allow for some missed detections, e.g. caused by occlusions or noise.

The MAP solution is found with the CINDA framework which results in a set of trajectories with a corresponding cost. The cost can be used to filter out unreliable (low-cost) trajectories. As the global tracker acts on a batch of frames, an online version would e.g. work on a batch of the latest N number of frames. In the current study we have used all available frames.

2.5. Fish length estimation

This section describes the approach we apply to estimate a metric length for each fish track. The overall pipeline is shown in Fig. 1. For each frame in a track, we first extract the centreline in the DBSCAN segmentation going from head to tail. Next, we estimate the metric length by converting the centreline to world coordinates using the depth values. Based on the per-frame length estimates, we compute a robust length estimate for each track/fish. These steps are explained in more detail in the following sections.

2.5.1. Centreline extraction

The centreline estimation approach is shown in Fig. 4. It can be difficult to extract the centreline robustly and efficiently for arbitrarily oriented fish in an image. We assume that the fish are represented by elongated structures and can be approximated by an ellipsoid. The fish segmentation is rotated according to the orientation of the major axis of the fitted ellipsoid such that the major axis is oriented across the columns of the image grid.

A distance map (distance from the boundary of the fish segmentation) is computed along the columns. For each column, the fish centreline is located at where the distance from the boundary is largest (highest value of the distance map along a column). The DBSCAN segmentation has problems with very thin or narrow structures – which may occur on the tail and head of the fish. Applying DBSCAN segmentation directly may therefore lead to under-segmentation of fishes resulting in a shortened centreline. Consequently, we have added an extra search forward and backward of the fish to see if there may be extra pixels that should be classified as part of the fish. Extra pixels are added to the centreline if the difference between the depth values right outside the segmentation are within a threshold T_{ext}. Empirically we have found that $T_{ext} = 10cm$ works well to include the thin/narrow structures.

In some instances, we may have other fish swimming in front of the fish we are trying to estimate the length of. We label these cases as occluded if we observe depth values at the extrapolated centreline edges being closer to the camera than the fish we are currently processing. Nearer depth values will most likely represent another fish and occluded fish will most likely be estimated with a too short length estimate.

A 2nd order polynomial is fitted to the centreline coordinates to regularize the centreline. We explored different approximations (1st and higher order polynomials and splines), where a 2nd order polynomial provided the best compromise between insensitivity to outliers and flexibility of fitting to a bending fish.

2.5.2. Length-estimation per detection

We sample the depth map along the fitted centreline. The depth values are noisy and may have large outliers – especially at the head and tail side of the fish. We apply a robust 1st order line fit to the noisy data by using the mid-80 percentile depth values. A 1st order polynomial (line) fit is used because we are not interested in the geodesic distance along the surface of the fish which may be better approximated by a higher order polynomial fit.

Given the pixel coordinates and associated depth value, we use the intrinsic camera parameters to convert the centreline into Euclidean coordinates. We integrate along the centreline Euclidean coordinates to determine the length of the fish.

We have found that with large slopes of the line fit (fish is swimming away or towards the camera), the length estimate is underestimated. The perspective foreshortening is the main cause of this. We tend to miss pixels on the head and tail of the fish the more it is angled along the optical axis, and around the area of the fish furthest from the camera (Fig. 5).

2.5.3. Length-estimation per track

Fig. 6 shows a set of intensity frames extracted from tracking an individual salmon over a total of 36 frames. Fig. 7 shows different estimates, including the fish length, slope of the line fit to the depth profile and the curvature of the centreline, extracted from the fish segmentation at each frame in the track. Based on the per-frame estimates, the aim is to extract a per-track length estimate. It is obvious that at certain frames,

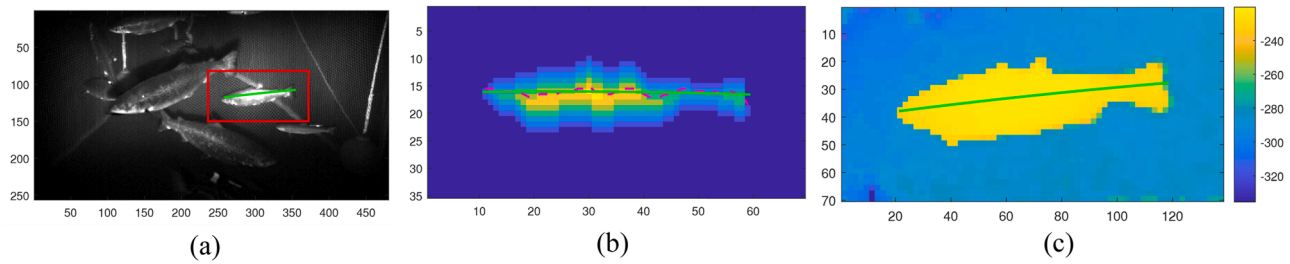


Fig. 4. Centreline extraction. (a) Intensity image overlaid with the detection bounding box and the final estimated centreline. (b) The distance transform of the segmentation mask rotated such that the major axis of the ellipsoid resides along the x-axis. The raw centreline is represented by the stapled magenta line, while the fitted 2nd order polynomial is represented by the green line. (c) The fitted centreline is rotated back into image space and overlaid on the depth map.

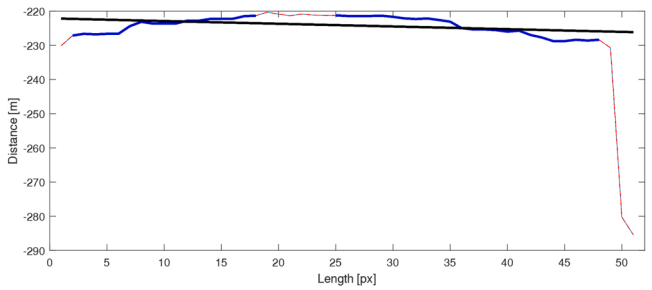


Fig. 5. Length estimation. The depth values along the centreline shown in Fig. 4(c) are shown in red. Robust values used in line-fit are shown in solid blue, while the line fit is shown in black. The slope of the curve is − 4 cm and the median distance of the fitted line is 224 cm. The estimated fish length using the depth values from the fitted line is 41.8 cm.

the length estimates become unreliable because the fish is quickly turning (e.g. at frames 25–37 where it is turning away from the camera). Consequently, the slope of the depth profile fit, and curvature of the 2nd order polynomial fit, are good discriminators for whether the length estimate is reliable. We propose a simple filtering technique to exclude these unreliable estimates from the calculation of the per-track fish length: (1) the slope should be stable over a neighbourhood of 5 frames (should not differ more than 10 cm) and the absolute slope should be less than 40 cm. (2) the curvature of the fitted centreline should be less than 0.1.

We estimate a reliable fish/track estimate by calculating the mean and standard deviation over the valid estimates. For the case shown in Fig. 7, valid estimates are shown with a circle.

3. Results

We have performed two experiments to validate the performance of the proposed fish-length estimation. The first experiment validates the accuracy of the individual fish length estimation, while the second experiment validates the generated length-distribution over a population of fish.

3.1. Three free-swimming fish with known lengths

With this experiment we were interested in establishing the accuracy of the camera calibration and the length estimation. This was achieved by being able to estimate a precise fish length for tracks of fish that we know the exact length of. Consequently, we establish the accuracy of the system by estimating the length of three manually measured fish that can be uniquely identified in the video. Fig. 8 shows example frames from the video where three Atlantic cod *(Gadus morhua)* are swimming freely in a tank. The centreline of the three fish have been manually annotated in each frame in the two different video segments. We used the manually annotated fish to validate that the data produced by the system can be used to accurately estimate the fish length. Two tracks can be extracted for each fish. We estimate the fish length for each frame in a track and the aggregate track-based length estimate. In Fig. 9 we show examples of the fish length estimates for the three fish, while Table 1 summarizes the aggregate results for the two track sequences for each fish.

Our results show that the camera system is well calibrated, both in terms of the checkerboard calibration to determine the field of view of the camera and that it reports correct depth estimates. A 10% error in the depth estimate would correspond to a 10% error in the length estimate. We find that the relative error in the length estimate is less than 2% based on the manually annotated centrelines.

3.2. Free-swimming Atlantic salmon with known population statistics

With this experiment we establish the validity of the population length-estimation of free-swimming salmon *(Salmo salar)*. We base the experiment on a dataset acquired from a fish farm on the west-coast of Norway. Fig. 10 shows example frames of the dataset (a total of ∼3000 frames/5 min recording) with overlaid centrelines for a selection of fish in the frame. A small net cage (5 m × 5 m and 4 m deep) contained 45 salmon. All the 45 salmon were manually measured to the closest 0.5 cm. We found that the average length was 49.3 cm with a standard deviation of 7.0 cm (see Table 2 for more robust statistics as well).

We established population statistics by both manually labelling 44 fish and by applying the algorithm. With the proposed algorithm we

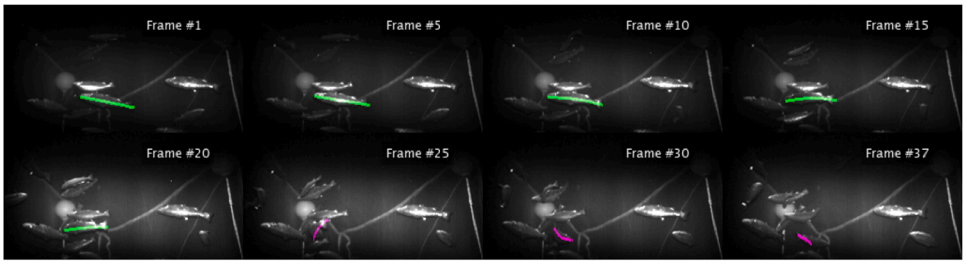
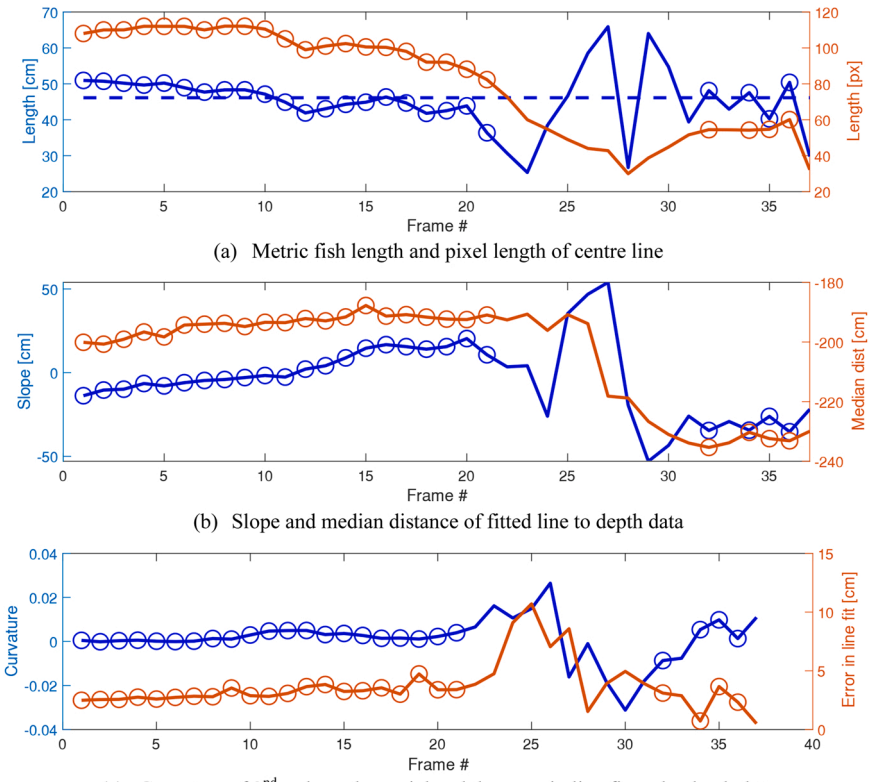


Fig. 6. Individual frames extracted from a track with overlaid centreline. The green centreline are valid samples used to estimate the composite length of the tracked fish.



(a) Metric fish length and pixel length of centre line

(b) Slope and median distance of fitted line to depth data

(c) Curvature of 2nd order polynomial and the error in line fit to the depth data.

Fig. 7. Example of estimates extracted from a single track/fish. Images for a selection of the frames are shown in Fig. 6. Valid length estimate points are shown with a circle.

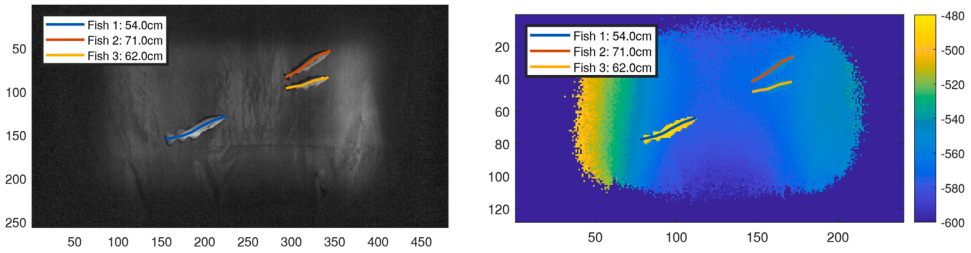


Fig. 8. Example frames from the cod dataset. Left: Intensity image of the three fish overlaid with the manually labelled centrelines. Right: Depth map with the overlaid centrelines. We have filtered away all the non-confident depth estimates. The depth is reported in centimetres. Notice that the longest fish (fish 2) has a centre line which is shorter in pixel length than the shortest fish (fish 1), hence the need for accurate distance estimates to convert the pixel coordinates to metric distances.

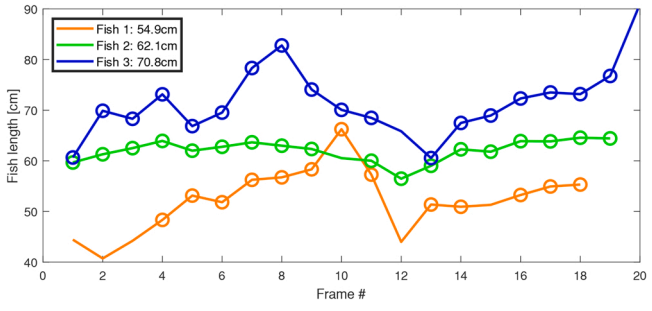


Fig. 9. Length estimates along tracks for the three cods (sequence 2). Circles represent valid points which are used to estimate the per-track length estimates.

established 106 fish tracks. The population statistics are reported in Table 2 and show that the three estimated populations match well. With the experiment in Section 3.1 we established that the camera was

Table 1
Comparison of the estimated versus ground truth length estimates for the three cods.

	Ground truth length	Estimated mean	Estimated std	% error
Fish 1 – sequence 1	54 cm	53.6 cm	5.8 cm	0.7%
Fish 1 – sequence 2	54 cm	54.9 cm	5.5 cm	1.7%
Fish 2 – sequence 1	62 cm	62.1 cm	3.9 cm	0.2%
Fish 2 – sequence 2	62 cm	62.1 cm	2.1 cm	0.2%
Fish 3 – sequence 1	71 cm	70.2 cm	1.5 cm	1.3%
Fish 3 – sequence 2	71 cm	70.8 cm	4.5 cm	0.2%

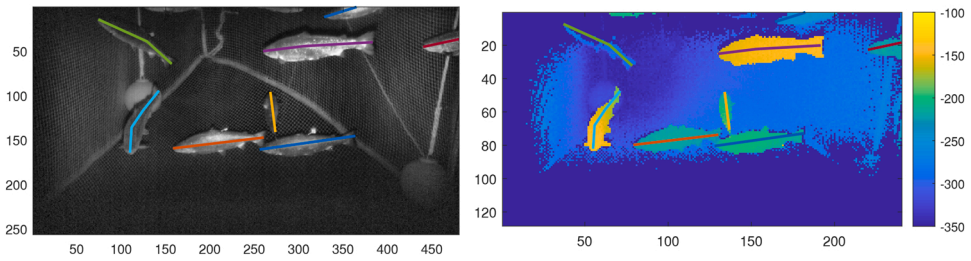


Fig. 10. Example frame from the free-swimming salmon dataset. The manually labelled centrelines are overlaid. The depth map is shown in centimetres and low confidence depth values have been removed.

Table 2
Population estimates for the free-swimming salmon.

Measurement method	Mean	Mean Error	Standard deviation	Minimum	Median	Maximum
Manually measured (n = 45)	49.3 cm	–	7.0 cm	33.2 cm	50.7 cm	58.2 cm
Manually labelled tracks (n = 46)	49.8 cm	1.0%	8.0 cm	29.6 cm	50.2 cm	67.7 cm
Proposed algorithm (n = 106)	49.2 cm	0.2%	10.2 cm	24.4 cm	50.9 cm	69.0 cm

calibrated, and that the length estimation procedure produced accurate and precise estimates. Because the three population distributions (ground truth, manually labelled and the full algorithmic pipeline) reported here are similar, we have established the accuracy and validity of the automatic length estimation pipeline including the detection, centreline extraction and tracking.

4. Discussion

Systems for automatic monitoring of free-swimming fish are important to enable precision feeding, behaviour analysis and biomass estimation in aquaculture. Accurate fish length estimates, both on an individual and population basis, is an essential technological feature of such a system. Fish lengths can be used to monitor growth and weight of the fish and be used as input to automatic feeding systems. We have presented a system for automatic length estimation of free-swimming fish which produces population mean length estimates with a mean error in the order of 1% of the ground truth population.

The proposed algorithm produces a population distribution which has somewhat fatter tails than the ground truth population distribution. The outliers at the top of the length scale (see Fig. 11) are caused by two fish swimming close together where DBSCAN has segmented them as one fish. Verifying that a detection represents a single fish can be achieved by e.g. matching it with a fish template such as the Point Distribution Model(Tillett et al., 2000). At the lower length scale, the outliers are caused by tracks of fish swimming far away from the camera where it is difficult to get reliable depth of the tail/head because it covers few pixels. Another difficult case is when the fish are swimming away or towards the camera where the head/tail is occluding for the other end of the fish. Both cases cause a foreshortening of the estimated length. In future work we plan to regress a scaling factor to adjust for the foreshortening caused by these different effects. It may also be that some fish are more likely to swim by the camera more often than others which will induce a bias in the distribution (e.g. Folkedal et al., 2012; Nilsson et al., 2013).

The novelty of the system lies in the use of a range-gated 3D imaging sensor which can produce high-resolution intensity and depth images over a large depth range (typically 1–8 m), even under turbid conditions (Risholm et al., 2018). The availability of high-resolution depth information makes the detection and segmentation of the individual fish relatively easy. Furthermore, the real-time imaging (10 Hz) enable image acquisition of fast swimming fish. Because the fish deforms as they swim which may cause self-occlusions and perspective foreshortening, it is important to have a tracks consisting of many detections of the same fish to robustly estimate its length.

Stereo-based approaches to length estimation (Shafait et al., 2017) have shown potential of providing length errors within 1% of the true length by using semi-automatic stereo-correspondence labelling. Robust automatic estimation of dense stereo-correspondences in the underwater environment is very difficult because of water turbidity, specular

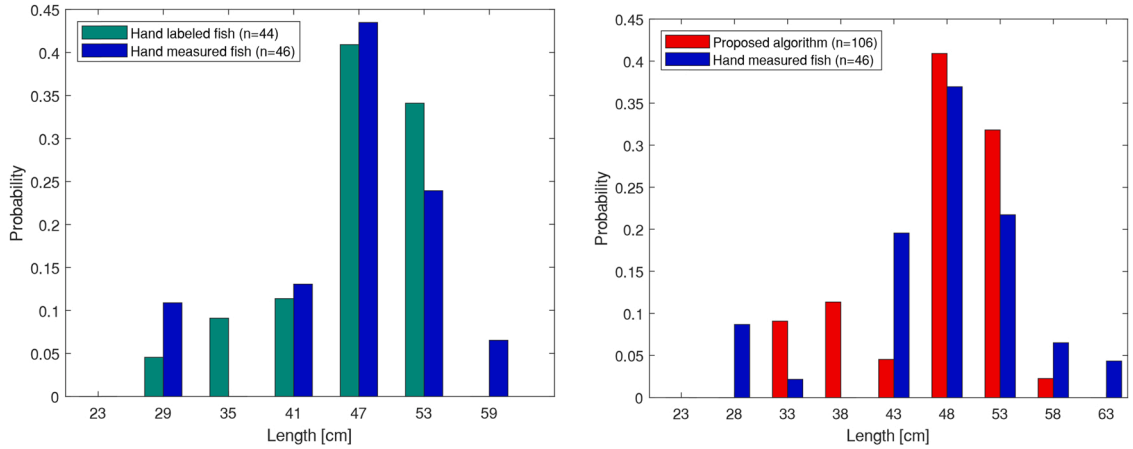


Fig. 11. Comparison of histograms of length estimates between the proposed algorithm and the manually measured fish.

repeating fish texture and varying ambient lighting. Artificial lighting is especially important for night-time observations or for deep-water observations. One approach to robustify the stereo-correspondence estimation is to use active stereo (Risholm et al., 2019), i.e. to project a non-repeating and possibly coded light signal. Such light would necessarily have to be in the visible light spectrum, preferably green/blue which is the least attenuated part of the spectrum as we use in this work, to be useful. The advantage of the range-gated system is that we are not looking for stereo-correspondences which quickly becomes difficult without artificial lighting and with deteriorating water conditions, but rather the returned light pulse for a specific pixel which is more robust to turbidity and difficult object/fish textures. However, increasing turbidity does deteriorate the quality of the depth estimates and causes forward scattering which causes blurring of the image. We defer evaluating the effect of turbidity on the length estimates to future work, but our initial approach will be to apply unsharp-filtering to reduce the blurring of the depth maps as described in (Risholm et al., 2018).

With robust and accurate length estimation established, the next goal will be to correlate it with weight to determine biomass. In future work we will develop algorithms to predict the weight by also taking the width of the cross section of the fish into account. We will also investigate if it is possible to fit a surface to the depth map and estimate the weight directly from the volume contained within the surface.

Behavioural information of free-swimming fish is valuable information for fish farmers. This type of information can be used to determine if the fish are healthy, hungry or stressed (Martins et al., 2012). One way to detect behavioural shifts is to study the swimming pattern of individual fish and the whole population as one (Oppedal et al., 2011). Tracking of fish in 3D facilitates recording of individual swimming speed. Swimming speed is a good indicator for buoyancy, it is relevant for submerged farming of salmonids coping with strong water currents as salmon then maintain a stationary position (Johansson et al., 2014), and sickness as indicated by low activity levels (Vindas et al., 2016). The high-resolution depth information facilitates easy detection and multi-object tracking of fish. We hypothesise that it should be possible to train a classifier based on the fish tracks to predict the cause of the fish behaviour.

CRediT authorship contribution statement

Petter Risholm: Conceptualization, Methodology, Software, Validation, Data curation, Writing, Visualization. **Ahmed Mohammed**: Conceptualization, Methodology, Software, Validation, Writing. **Trine Kirkhus**: Conceptualization, Methodology, Writing. **Sigmund Clausen**: Conceptualization, Methodology, Investigation, Writing. **Leonid Vasilyev**: Conceptualization, Methodology, Writing. **Ole Folkedal**: Resources, Writing, Project administration, Investigation. **Øistein Johnsen**: Funding, Project administration, Conceptualization, Investigation. **Karl Henrik Haugholt**: Conceptualization, Methodology, Resources, Writing, Investigation, Supervision. **Jens Thielemann**: Conceptualization, Methodology, Funding, Project administration, Writing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 773521; SFI Exposed (NCR 237790); and MarTERA Biosys under contract 283896.

References

Føre, M., Frank, K., Norton, T., Svendsen, E., Alfredsen, J.A., Dempster, T., Eguiraun, H., Watson, W., Stahl, A., Sunde, L.M., Schellewald, C., Skøien, K.R., Alver, M.O., Berckmans, D., 2018. Precision fish farming: A new framework to improve production in aquaculture. Biosyst. Eng. vol. 173, 176–193. https://doi.org/10.1016/j.biosystemseng.2017.10.014.

Li, D., Hao, Y., Duan, Y., 2020. Nonintrusive methods for biomass estimation in aquaculture with emphasis on fish: a review. Rev. Aquacult vol. 12 (3), 1390–1411. https://doi.org/10.1111/raq.12388.

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 vol. 9 (4), 369–387. https://doi.org/10.1111/raq.12143.

Papadakis, V.M., Papadakis, I.E., Lamprianidou, F., Glaropoulos, A., Kentouri, M., 2012. A computer-vision system and methodology for the analysis of fish behavior. Aquac. Eng. vol. 46, 53–59. https://doi.org/10.1016/j.aquaeng.2011.11.002.

Oppedal, F., Dempster, T., Stien, L.H., 2011. Environmental drivers of Atlantic salmon behaviour in sea-cages: A review. Aquaculture vol. 311 (1–4), 1–18. https://doi.org/10.1016/j.aquaculture.2010.11.020.

Iqbal, M.A., Wang, Z., Ali, Z.A., Riaz, S., 2021. Automatic fish species classification using deep convolutional neural networks. Wirel. Pers. Commun. vol. 116 (2), 1043–1053. https://doi.org/10.1007/s11277-019-06634-1.

Md.H. Sharif, F. Galip, A. Guler, and S. Uyaver, "A simple approach to count and track underwater fishes from videos," in *2015 18th International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, Dec. 2015, pp. 347–352. doi: ⟨10.1109/ICCITechn.2015.7488094⟩.

Costa, C., Loy, A., Cataudella, S., Davis, D., Scardi, M., 2006. Extracting fish size using dual underwater cameras. Aquac. Eng. vol. 35 (3), 218–227. https://doi.org/10.1016/j.aquaeng.2006.02.003.

Shi, C., Wang, Q., He, X., Zhang, X., Li, D., 2020. An automatic method of fish length estimation using underwater stereo system based on LabVIEW. Comput. Electron. Agric. vol. 173, 105419 https://doi.org/10.1016/j.compag.2020.105419.

Shortis, M.R., Ravanbakhsh, M., Shafait, F., Mian, A., 2016. Progress in the Automated Identification, Measurement, and Counting of Fish in Underwater Image Sequences. Mar. Technol. Soc. J. vol. 50 (1), 4–16. https://doi.org/10.4031/MTSJ.50.1.1.

Miranda, J.M., Romero, M., 2017. A prototype to measure rainbow trout's length using image processing. Aquac. Eng. vol. 76, 41–49. https://doi.org/10.1016/j.aquaeng.2017.01.003.

M.R. Shortis et al., "A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences," in *SPIE Optical Metrology,* Munich, Germany, May 2013, p. 87910G. doi: ⟨10.1117/12.2020941⟩.

F. Lopes, H. Silva, J.M. Almeida, C. Pinho, and E. Silva, Fish farming autonomous calibration system," in *OCEANS 2017 - Aberdeen*, Aberdeen, United Kingdom, Jun. 2017, pp. 1–6. doi: ⟨10.1109/OCEANSE.2017.8084565⟩.

Risholm, P., Thorstensen, J., Thielemann, J.T., Kaspersen, K., Tschudi, J., Yates, C., Softley, C., Abrosimov, I., Alexander, J., Haugholt, K.H., 2018. Real-time super-resolved 3D in turbid water using a fast range-gated CMOS camera. Appl. Opt. vol. 57 (14), 3927–3937. https://doi.org/10.1364/AO.57.003927.

M. Ester, H.-P. Kriegel, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, 1996, pp. 226–231.

Zhang, Z., 2000. A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. vol. 22 (11), 5–1334.

Luo, W., 2021. Multiple object tracking: A literature review. Artif. Intell. 293, 23.

C. Wang, Y. Wang, and G. Yu, "Efficient Global Multi-object Tracking Under Minimum-cost Circulation Framework," *(in press by IEEE Trans. on PAMI)*, Feb. 2020, (Online). Available: ⟨http://arxiv.org/abs/1911.00796⟩.

Tillett, R., McFarlane, N., Lines, J., 2000. Estimating dimensions of free-swimming fish using 3D point distribution models. Comput. Vis. Image Underst. vol. 79 (1), 123–141. https://doi.org/10.1006/cviu.2000.0847.

Folkedal, O., Stien, L.H., Nilsson, J., Torgersen, T., Fosseidengen, J.E., Oppedal, F., 2012. Sea caged Atlantic salmon display size-dependent swimming depth. Aquat. Living Resour. vol. 25 (2), 143–149. https://doi.org/10.1051/alr/2012007.

Nilsson, J., Folkedal, O., Fosseidengen, J.E., Stien, L.H., Oppedal, F., 2013. PIT tagged individual Atlantic salmon registrered at static depth positions in a sea cage: Vertical size stratification and implications for fish sampling. Aquac. Eng. vol. 55, 32–36. https://doi.org/10.1016/j.aquaeng.2013.02.001.

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. vol. 74 (6), 1690–1701. https://doi.org/10.1093/icesjms/fsx007.

Risholm, P., Kirkhus, T., Thielemann, J., Thorstensen, J., 2019. Adaptive Structured Light with Scatter Correction for High-Precision Underwater 3D Measurements. Sensors vol. 19 (5), 1043. https://doi.org/10.3390/s19051043.

P. Risholm, J.T. Thielemann, R. Moore, and K.H. Haugholt, A scatter removal technique to enhance underwater range-gated 3D and intensity images," in *OCEANS 2018 MTS/IEEE Charleston*, Charleston, SC, Oct. 2018, pp. 1–6. doi: ⟨10.1109/OCEANS.2018.8604613⟩.

Martins, C.I., Galhardo, L., Noble, C., Damsgård, B., Spedicato, M.T., Zupa, W., Beauchaud, M., Kulczykowska, E., Massabuau, J.C., Carter, T., Planellas, S.R.,

Kristiansen, T., 2012. Behavioural indicators of welfare in farmed fish. Fish. Physiol. Biochem vol. 38 (1), 17–41. https://doi.org/10.1007/s10695-011-9518-8.
Johansson, D., Laursen, F., Fernö, A., Fosseidengen, J.E., Klebert, P., Stien, L.H., Vågseth, T., Oppedal, F., 2014. The Interaction between Water Currents and Salmon Swimming Behaviour in Sea Cages. PLoS ONE vol. 9 (5), e97635. https://doi.org/10.1371/journal.pone.0097635.
Vindas, M.A., Johansen, I.B., Folkedal, O., Höglund, E., Gorissen, M., Flik, G., Kristiansen, T.S., Øverli, Ø., 2016. Brain serotonergic activation in growth-stunted farmed salmon: adaption versus pathology. R. Soc. Open Sci. vol. 3 (5), 160030 https://doi.org/10.1098/rsos.160030.