Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 54-16 (2021) 438–443

# Vision-based pose estimation for autonomous operations in aquacultural fish farms ⋆

**Christian Schellewald** * **Annette Stahl** ** **Eleni Kelasidi** *

* *SINTEF Ocean AS, Brattørkaia 17C, 7010 Trondheim, Norway (e-mail: Christian.Schellewald@sintef.no, Eleni.Kelasidi@sintef.no).*

** *Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), O. S, Bragstads plass 2D, 7491 Trondheim, Norway (e-mail: Annette.Stahl@ntnu.no).*

**Abstract:** There is a largely increasing demand for the usage of Unmanned Underwater Vehicles (UUVs) including Remotely Operated Vehicles (ROVs) for underwater aquaculture operations thereby minimizing the risks for diving accidents associated with such operations. ROVs are commonly used for short-distance inspection and intervention operations. Typically, these vehicles are human-operated and improving the sensing capabilities for visual scene interpretation will contribute significantly to achieve the desired higher degree of autonomy within ROV operations in such a challenging environment. In this paper we propose and investigate an approach enabling the underwater robot to measure its distance to the fishnet and to estimate its orientation with respect to the net. The computer vision based system exploits the 2D Fast Fourier Transform (FFT) for distance estimation from a camera to a regular net-structure in an aquaculture installation. The approach is evaluated in a simulation as well as demonstrated in real-world recordings.



*Keywords:* FFT, ROV, pose estimation, distance estimation, net detection, calibration

## 1. INTRODUCTION

In order to face future challenges that come with an increasing food demand of an increasing world population, fish-farming (Føre et al., 2018) and the development of algorithms that help to increase the autonomous capabilities of underwater robots is crucial ((Balchen, 1991)). Adapting advanced technological solutions such as intelligent sensors and using underwater robotic systems (Bogue, 2015; Kelasidi et al., 2016) will contribute to address the challenges of the aquaculture industry. These include the optimization of costs but also other aspects like minimizing escapees, reducing sea lice, reducing the environmental impact among others (Føre et al., 2018). Currently, mostly ROVs are used in salmon fish farms and basic tasks towards more autonomous behaviour like hovering or traversing at a certain distance from the cage are requested features. Cameras provide, compared with other sensors a cost effective solution for vehicle navigation, pose estimation, orientation, station keeping and drift correction. In addition, underwater positioning system's like Ultra-short Baseline (USBL) or Dopper Velocity Log (DVL) are not able to provide the relative position of the vehicle from the observed structure (Rundtop and Frank, 2016) or are disturbed by fish, respectively. Within this paper, we suggest a cost effective computer vision based method that only requires a monocular camera to estimate distance and pose of the robot to the net in a net cage. The obtained results can therefore be used as inputs to the control strategies for the autonomous navigation of UUVs (Gafurov and Klochkov, 2015) during inspection and intervention operations in fish farms.

### 1.1 Motivation

The automation of aquaculture operations is highly desired by the industry (Føre et al., 2018), but many additional challenges – compared to land based automation efforts – arise from the fact that farming operations today are mostly performed in the sea. Furthermore, farming of Atlantic salmon in exposed areas (Bjelland et al., 2015) poses unique challenges to operations. Many of the operational challenges seen at present sheltered sites are likely to be amplified when moving production to more exposed locations. There is, however, a strong Norwegian industrial interest in utilizing such areas. This includes for example that net cages are flexible structures that change with the ocean current, tide and different weather conditions (Lader et al., 2008) meaning that the environment where, for example, a UUV/ROV (Antonelli, 2014) is supposed to operate is constantly changing. In addition, ordinary Global Positioning Systems (GPS) fail to provide location and time information under water as radio signals from the satellites do not penetrate in water very far as they are heavily damped (Paull et al., 2014; Taraldsen et al., 2011), and acoustic systems in noisy and reflective environments tend to have a lower accuracy. Alternative solutions based on vision sensor systems are commonly more cost effective, accurate, and deliver environmental scene information in

⋆ This work was financed by the Research Council of Norway through the project: Development of technology for autonomous, bio-interactive and high-quality data acquisition from aquaculture net cages (CageReporter, project number 269087)


10.1016/j.ifacol.2021.10.128

high resolution (Massot-Campos M, 2015). In addition, the application of simultaneous localization and mapping (SLAM) technology to the underwater realm (Leonardi and Stahl, 2018) have yielded new possibilities in the field of navigation and localization (Paull et al., 2014). A fundamental task for navigation, to allow a higher level of autonomy within aquaculture operations that can support advanced control algorithms (Fossen, 2011) to steer ROVs/UUVs, is to estimate the distance and the relative orientation of the camera to an object with high precision/accuracy. State-of-the-art vision technology relies mostly on stereo vision or RGB-D systems, in order to provide comprehensive 3D information (Leonardi et al., 2017). Monocular systems, which are even lower in cost and which provide a solution for low payload and small size form factor systems rely on concepts like structure from motion (SFM) (Saputra et al., 2018) to calculate distances (Davison et al., 2007). Three-dimensional measurements from the surrounding scene are retrieved by moving the camera from one viewpoint to the next. The camera pose and 3D structure of the scene can be estimated through a set of feature correspondences, detected from multiple images. Absolute scale of objects rely thereby on the assumption that the baseline of the motion or the geometry of the observed object is known. SFM implementations are rather complex and computationally expensive (Fraundorfer and Scaramuzza, 2012). An alternative vision-based approach, not relying on the SFM concept is presented by (Duda et al., 2015). The main limitation of their proposed method is that the success of the approach is restricted to situations where the fishnet knots are clearly visible which is for example not the case if the fishnet is partly covered by seaweed or occluded by fish. In addition in shallow waters, most SFM methods have difficulties to select/track feature points, because of the caustics, visible as fast moving illuminated patterns created by the sun and the surface's wavelets. In order to overcome this issue and to provide a non-complex and computational inexpensive solution, we propose to exploit the Fourier Transform to detect the presence of a net (i.e. a regular grid structure) in images/videos of aquaculture net cages recorded by a monocular camera. In addition, the knowledge of the grid-structure is exploited to estimate the relative distance and orientation (pose) of the camera to the net.

### *1.2 Main idea*

In the following, we explain in detail our approach to automatically detect the presence of net-structure and to determine its distance and orientation based on a monocular camera mounted on an underwater ROV in a salmon net cage. A squared region of interest (ROI) of the camera video stream is analysed in order to detect regular peaks in the Fourier Transform (FT) indicating the presents of a fishnet in the considered ROI. Once a fishnet is detected a single mesh is reconstructed from the regular peaks in the FT. Knowing the camera parameters and the real mesh-size one can compute which distance and which orientation the net has with respect to the camera. The main steps of this approach are illustrated in Figure 1.

### *1.3 Contributions*

A major issue in realizing autonomous underwater vehicles for fish net inspections or intervention tasks is to estimate in real-time and with high precision the relative distance

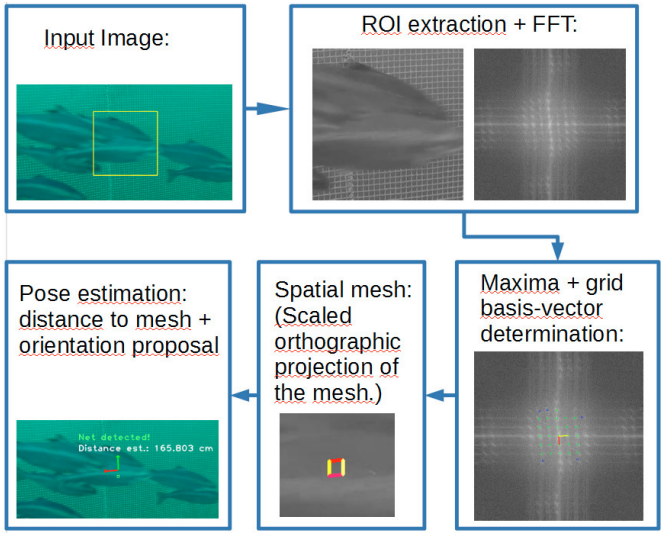


Fig. 1. Illustration of the main steps of the proposed net-pose estimation approach. The Fourier Transform of an image is analysed and searched for a regular grid pattern of detected local maximum peaks. Found base grid vectors correspond to a single spatial mesh for which the pose can be determined.

of the vehicle from the fishnet implying also the detection of the net. This is needed in order to maintain distance to the net during automated operations. Especially, regions of fishnets with little or no marine growth are extremely difficult to detect and further to track with visual sensors. The reason for this is on one side the fine structure of the fishnet and on the other side the repetitive pattern resulting in a high similarity between different net regions. Thus, feature based matching methods are prone to generate large consistent sets of outliers resulting in wrong distance and pose estimations (Duda et al., 2015). This paper presents a novel computationally non-complex and inexpensive computer vision based approach for regular pattern detection as well as orientation, and distance estimation based on an analysis of the images/videos in the spectral domain. This avoids the tracking of image features and therefore does not suffer in situations with repeated scene structures. The method is proven to be robust against occlusions. In addition, the scale is directly estimated from the fishnet using the FFT eliminating the main disadvantage of monocular camera systems which can generally reconstruct scenes only up to scale. The outcome is a low cost vision based detection system to support autonomous operations of underwater vehicles.

## 2. THEORETICAL BACKGROUND

In this section, we present the theoretical building blocks that we exploit to efficiently determine the distance and the orientation to a regular grid like a fish-net seen in an image with respect to the camera.

### *2.1 Fourier transformation of periodic patterns*

We denote the Fourier Transform of an image $I$ as $\mathcal{F}\{I\}$. The Discrete Fourier Transform (DFT)

$$F(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} I(x,y)e^{-i2\pi(\frac{ux}{N}+\frac{vy}{N})} \quad (1)$$

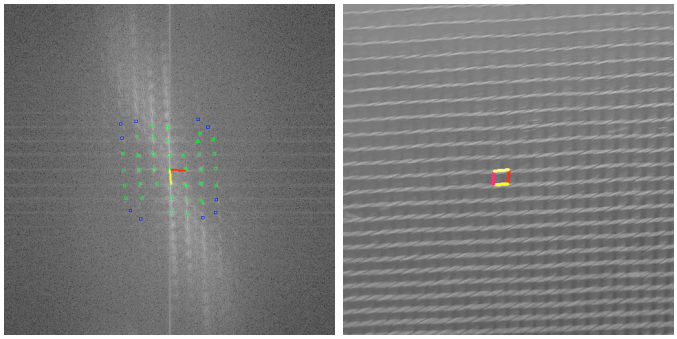

Fig. 2. **Left:** Magnitude image of the Fourier Transform of the right image. The inner local maximums are used to determine the grid structure and are marked by green rectangles. The found base grid vectors are shown in red and yellow. **Right:** A 512x512 pixel large area showing the net structure. The approximated mesh size of the net is reconstructed based on the found base grid vectors in the magnitude image.

of an $N \times N$ image ($u, v \in [0, ..., N-1]$) can be computed efficiently using the Fast Fourier Transform (Cooley and Tukey, 1965) assuming periodic border conditions and $N = 2^m$ with $m$ being an integer (e.g. $N$= ..., 128, 256, 512, ...). Note, that due to its "separability", the FFT of a 2D-image has complexity $O(N^2 log N)$. As $\mathcal{F}$ is in general complex valued, we visualize just the magnitude $|\mathcal{F}\{I\}|$ of the output image (with size $N \times N$) to illustrate the results of the FT (see e.g. (Gonzalez et al., 2004)).

A periodic structure leads to peaks in the FT at spatial frequencies of the repeated texture. We are particularly interested to extract these repeating patterns from the Fourier transformed image of the scene. We do that by searching for local maximums in the magnitude image $|\mathcal{F}\{I\}|$. In a following step candidate basis-vectors of the grid structure are determined and subsequently checked for consistency and regularity of the local maximums. If a large fraction (for example 0.5) of the observed local maximums lie on the grid we assume that regular structure (e.g. a net) is present in the ROI. This results in two grid basis vectors $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$. Let $d_{k_i} = |\boldsymbol{k}_i|$ be the magnitude of one basis vector $\boldsymbol{k}_i = (u, v)$ measured in pixels from the origin (in the center) to an observed frequency intensity maximum within the Fourier transformed image. It can be interpreted as *wave number* indicating the number of waves or cycles per *unit distance*, which is here the length of the image side $N$. Then the reciprocal space length $d_{s_i}$ of the associated periodic structure can be computed as

$$d_{s_i} = N/d_{k_i}. \tag{2}$$

The orientation of the vector $\boldsymbol{k}_i$ is perpendicular to the direction of the associated spacial grid/lattice. An example of a determined grid in a FFT pair of images is shown in Fig. 2. The image on the left side is the magnitude image of the FFT of the image on the right side showing a net cage. The local maximum peaks, in an inner circular area (diameter is $N/3$) are marked by green rectangles. Two wave number vectors, providing the base grid vectors of the grid structure, are shown in red and yellow. In the right image a single mesh reconstructed from the base grid vectors is overlaid. The length (in pixels) of the edges of the mesh is computed by (2) and their orientation is perpendicular to the corresponding wave number vector. This reconstructed idealized single mesh – in form of a parallelogram/quadrilateral – approximates the mesh seen in the image and is used in the following step to estimate the distance of the net to the camera. As the geometry and the size of the mesh (in our case a flat square with a side length of $1.5cm$) is known, the internal camera parameters can be used to determine the distance and orientation of the camera to the idealized mesh. We note that the parallelogram/quadrilateral only approximates the perspective projection of a single mesh as a scaled orthographic projection. In addition, whenever a grid structure can be verified to be present in the FFT one knows that a regular structure (in the considered application a fish net) is visible in the image. So, we can also exploit this as fish-net detection algorithm. Fig. 3 illustrates that the regular grid pattern is still present in the FFT image even if the net is partly occluded, as the FFT constitutes a global operation. This property contributes to the robustness of the approach against occlusion.

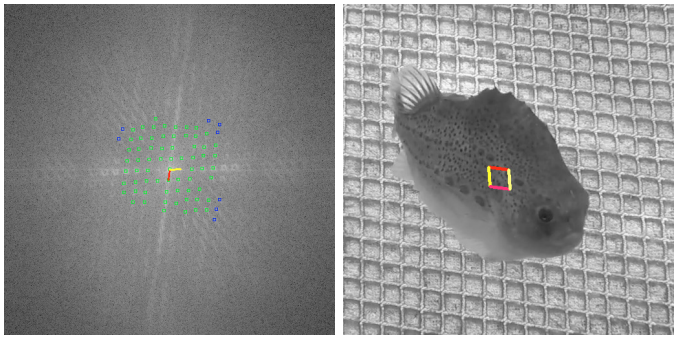

Fig. 3. **Left:** Magnitude image of the Fourier Transform of the right image. **Right:** An image showing the regular net structure occluded in the center by a cleaner fish. Still the mesh can be reconstructed based on the found base grid vectors in the magnitude image.

### *2.2 Camera calibration*

The internal or intrinsic camera calibration refers to the determination of camera specific parameters that define the configuration of the pin hole camera model (perspective projection) along with distortion parameters (compare (Hartley and Zisserman, 2000)). The perspective projection can be described by the intrinsic camera calibration matrix

$$K = \begin{bmatrix} f & s & o_x \\ 0 & fa & o_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where $f$ denotes the focal length and the point $(o_x, o_y)$ refers to the optical center (i.e., the intersection of the optical axis with the image plane) of the camera. Note that for today's cameras we most often can assume that the skew parameter is zero ($s = 0$) and that the pixels represent a square grid with an aspect ratio of one ($a = l_y/l_x = 1$). Here $l_x$ and $l_y$ are indicating the horizontal and vertical size of the pixels (i.e. measured in pixels per unit length [meter, cm, mm, etc.]). The intrinsic camera parameters can be obtained by using a flat chessboard pattern with known geometry for calibration. In Fig. 4, an underwater image containing the used $7 \times 4$ calibration checkerboard is shown. We used the OpenCV library (Itseez, 2020)

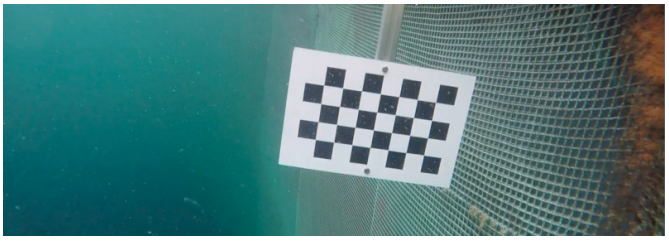

Fig. 4. An example image from a video used for determining the intrinsic camera parameters with a checkerboard calibration board ($7 \times 4$ inner crossings).

for a C++ standard implementation of Zhang's (Zhang, 2000) calibration method. It first finds the coordinates of all the checkerboard corners in the camera image for all the captured checkerboard orientations. Then the intrinsic camera parameters and distortion parameters are computed determining the linear mappings (homographies) from the checkerboard model points to the observed 2D image points using a closed-form (linear) solution. The coefficients for two distortion models are estimated by a linear least-square minimization which is followed by a final nonlinear optimization that refines the results. The distortion coefficients $k_1, k_2$ and $k_3$ are used to describe the model for a radial lens distortion (Visible as "barrel" or "pin cushion" distortion):

$$x_{distorted} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (4)$$

$$y_{distorted} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (5)$$

Here $x$ and $y$ refer to the coordinates of the undistorted pixel and $r^n$ with $n = 2, 4, 6$ is computed as $r^n = x^n + y^n$. The tangential distortion model is described by the parameters $p_1$ and $p_2$.

$$x_{distorted} = x + [2p_1 xy + p_2(r^2 + 2x^2)] \quad (6)$$

$$y_{distorted} = y + [p_1(r^2 + 2y^2) + 2p_2 xy] \quad (7)$$

Correcting the video-streams for the measured distortion allows to employ the perspective camera model (pinhole camera model) to estimate the pose (distance and orientation) of the approximated single mesh.

### 2.3 Pose Estimation

The above obtained quadrilateral is the projection of a small flat square onto the camera image which approximates a single mesh of the net that has a known size, in our case, $s_m \times s_m = 1.5cm \times 1.5cm$ (Note, that this size needs to be measured at the actual net used in the net cage). As the intrinsic camera parameters are known, one can compute the pose (i.e distance and orientation) of this mesh relative to the camera by first describing the mesh as square in the real world by its four coplanar and non collinear corner points (i.e. $X_1 = [-b, -b, 0]^T, X_2 = [-b, b, 0]^T, X_3 = [b, b, 0]^T, X_4 = [b, -b, 0]^T$). The parameter $b = s_m/2.0$ is half of the mesh-size $s_m$, putting its origin in the center. Then as the projection matrix $P = K[R|t]$ encodes the transformation from real world coordinates to pixel-coordinates in the image (Hartley and Zisserman, 2000) and by knowing the correspondences of the corners in the real world and its projected corners on the image one is able to reconstruct the pose of the mesh in terms of the translation vector $t \in \mathbb{R}^3$ and the rotation matrix $R \in SO(3)$. The full problem can be written as

$$\begin{bmatrix} hx_i \\ hy_i \\ h \end{bmatrix} = \begin{bmatrix} f & 0 & o_x \\ 0 & f & o_y \\ 0 & 0 & 1 \end{bmatrix} [R|t] \begin{bmatrix} X_i \\ Y_i \\ 0 \\ 1 \end{bmatrix}, \quad (8)$$

or shorter as

$$\mathbf{x}_i = P\mathbf{X}_i = K[R|t]\mathbf{X}_i, \quad (9)$$

where $\mathbf{x}_i$ is an image point represented by a homogeneous 3-vector and $\mathbf{X}_i$ is the corresponding world point represented by a homogeneous 4-vector. In order to decompose or solve (9) for the unknown pose of the object ($R$ and $t$ of the mesh relative to the camera) Perspective-n-Point (PnP) algorithms like suggested in (Lepetit et al., 2009; Oberkampf et al., 1996; Schweighofer and Pinz, 2006; Xiao-Shan Gao et al., 2003) can be employed to determine the distance and orientation of the net mesh.

We exploited the OpenCV PnP implementation, which is based on (Lepetit et al., 2009; Oberkampf et al., 1996), in our system for solving (9).

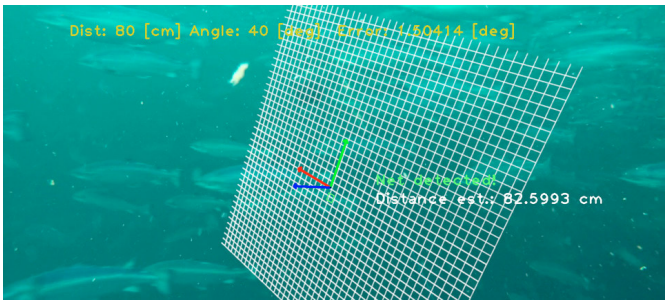


Fig. 5. An artificial net is projected into the video scene with varying distance and angle (here $80cm$ and $40°$ respectively) and overlaid to a video sequence. The snapshot shows a reconstructed distance of $82.6cm$ and an orientation angle error of $1.5°$.

### 2.4 Experiments

In order to evaluate the described method we projected – using the camera parameters from the real world experiment– an artificial net into an underwater video scene. An example for this is shown in Fig. 5. We defined a flat 3D grid with mesh size of $1.5cm$ and placed it virtually at different distances between $40cm$ and $200cm$ and with different tilt angles in front of the camera and investigated the performance of the approach proposed in this paper over 200 video images for each distance-angle pair. Table 1 shows the mean measured distance and mean deviation of the estimated orientation vector of the net using the proposed approach in this paper for a typical working range of distances and orientations. Based on the results in Table 1, one sees that the precision of the distance is well-suited for autonomous inspection of a net cage by underwater vehicles since accurate measurements from 40cm are possible compared to the conventional positioning systems such as DVL where the minimum measurement distance is 1.5m (see e.g. (Rundtop and Frank, 2016)). In the artificial experiments we designed, the determined distances are in good agreement with the ground truth and have an observed mean maximum deviation of 4% for larger tilt angles of the net. The orientation estimation shows that the estimated normal vector and the ground truth normal vector of the net deviates usually between $4° - 6°$ but that it reached $16°$ indicating that the orientation ambiguity

(Zhou et al., 2018) led to a larger number of non-desired pose estimations in our initial implementation. A possible strategy to select the correct pose solution is to estimate the pose at more than one ROI-location and testing for consistency assuming that the visible net-patch is relatively flat. Note also that due to the reciprocal relationship of the FFT space and the real space one can obtain a higher error for net-structures that are closer to the camera and appear spatial larger as the related structure appears smaller in the FT. The simplicity to employ our developed method along with the observed accuracy that is high enough to enable a closed online control loop during autonomous navigation makes our approach a highly suitable candidate for autonomous net inspection tasks that require a systematic traversing of the net-structure even when fish or seaweed partly disturb the view.

Table 1. An artificial net is projected into an underwater video scene with varying known distances and angles (ground truth). 200 images of this video sequence were used to determine the mean measured distance and mean deviation of the estimated orientation vector.

| **Z** | **Net Tilt Angle** | | |
|---|---|---|---|
| **[cm]** | Tilt: 0° | 20° | 40° |
| **40** | $39.97 \pm 0.25$ (9.0°) | $39.73 \pm 1.2$ (15°) | - |
| **80** | $78.56 \pm 0.18$ (4.9°) | $80.06 \pm 0.50$ (16°) | $79.13 \pm 9.81$ (5.3°) |
| **120** | $118.22 \pm 0.61$ (4.7°) | $119.21 \pm 1.15$ (11.6°) | $119.05 \pm 0.85$ (2.8°) |
| **160** | $156.15 \pm 1.94$ (6.6°) | $158.75 \pm 12.10$ (1.6°) | $153.38 \pm 13.72$ (5.9°) |
| **200** | $198.50 \pm 0.88$ (5.2°) | $200.76 \pm 0.61$ (4.2°) | $197.51 \pm 0.88$ (6.2°) |

### *2.5 Real World Experiments*

The described algorithm was developed for a ROV performing inspection tasks within a commercial net cage and the test recordings were obtained by the ROV camera and also by an underwater consumer camera mounted to the ROV. A calibration board (7×4 inner crossings) with square size of $3.11cm \times 3.11cm$ (compare Fig. 4) was used for the calibration of the cameras. In Fig. 6 a snapshot of a video taken by a ROV in a commercial aquaculture net cage is shown. The proposed approach is capable of detecting when the net is present and is also determining the distance and orientation in real world scenarios. In Fig 7 an example is shown where we can detect the net even though it is largely occluded by salmon.

In particular, the Argus Mini ROV (Argus, 2021) has been used to obtain the videos in this paper to demonstrate experimentally the usability of the proposed method. The vehicle has been manually controlled by an experienced ROV operator to approach the net structure and also moving along the net cage to capture videos. The videos were processed after recording and based on the processing rate of 2.5 FPS (frames per second) on a desktop PC with non-optimised code, real-time control algorithms will be developed and implemented in a simulation environment (i.e FhSim (2021) ) and then tested and evaluated in real world net cage experiments.

### *2.6 Conclusion and Discussion*

In this paper, we presented an approach which is capable to detect and estimate the pose of a cage net recorded by a monocular camera. It is based on the spectral analysis of the image in Fourier space from which we obtain the geometry of a single mesh if the regular net structure is


Fig. 6. Video image showing the distance and orientation estimation (visualised by the back-projected coordinate systen) to a net within a commercial aquaculture cage recorded by a monocular ROV camera. The observed pose estimation agrees with the dynamics seen in the video. Here the distance to the net is approximately $96cm$ and the single mesh used for the pose estimation is shown as well.


Fig. 7. The net of the cage is detected even though a large fraction of the net is occluded by salmon.

visible and detected in the image. This in turn allows – in combination with the known camera parameters – to estimate the pose of the mesh in real world coordinates providing a distance and orientation to the net. In our implementation, the search for a regular net-structure is organized hierarchically meaning that we search at different scales for a regular pattern in the Fourier space. The algorithm works best if motion blur can be avoided by maintaining a short enough shutter speed that results in sharp views of the net. The accurate and efficient monocular camera based pose estimation of the ROV relative to the net (or vice versa) allows the integration of our approach into control approaches for maintaining automatically a certain distance to the net cage thereby avoiding a collision between ROV and the cage. It is also easy to integrate the suggested positioning measurement approach into existing ROV solutions as no additional hardware at the ROV is required. The proposed method is computationally less complex than the commonly used SFM/SLAM implementations in the localization context. This low cost solution compared with multiple sensor settings for navigation and localization for example DVL/USBL systems, is capable to produce more accurate results and is more robust with respect to occlusion compared to existing commercial solutions such as DVL. Based on the suggested method our next steps will include the development and implementation of an online control mechanism for autonomous net-inspection tasks and we will evaluate its capability to traverse and scan a net cage at a predefined distance in different weather conditions.

## ACKNOWLEDGEMENTS

We would like to thank Water Linked AS, Sealab AS, Norsk Havservice AS and Norwegian University of Science and Technology (NTNU) for their contributions in this project. We would also like to thank SINTEF ACE for enabling us to perform the trials at the full-scale laboratory

facility in Rataran, Norway. In addition we wish to thank Biao Su, Magnus Oshaug Pedersen, Walter Caharija and Terje Bremvåg for their help during the ROV experiments.

## REFERENCES


Antonelli, G. (2014). *Underwater Robots*. 3rd ed., ser. Springer Tracts in Advanced Robotics. Springer International Publishing.

Argus (2021). Argus AS argus mini rov. `https://www.argus-rs.no/argus-rovs/11/argus-mini`. Accessed: 2021-09-29.

Balchen, J.G. (1991). Possible roles of remotely operated underwater vehicles (ROV) and robotics in mariculture of the future. *Modeling, Identification and Control*, 12(4), 207–217. doi:10.4173/mic.1991.4.3.

Bjelland, H.V., Føre, M., Lader, P., Kristiansen, D., Holmen, I.M., Fredheim, A., Grøtli, E.I., Fathi, D.E., Oppedal, F., Utne, I.B., and Schjølberg, I. (2015). Exposed aquaculture in norway. In *OCEANS 2015 - MTS/IEEE Washington*, 1–10. doi: 10.23919/OCEANS.2015.7404486.

Bogue, R. (2015). Underwater robots: a review of technologies and applications. *Industrial Robot: An International Journal*, 42.3, 186–191.

Cooley, J.W. and Tukey, J.W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90), 297–301.

Davison, A.J., Reid, I.D., Molton, N.D., and Stasse, O. (2007). Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6), 1052–1067. doi: 10.1109/TPAMI.2007.1049.

Duda, A., Schwendner, J., Stahl, A., and Rundtop, P. (2015). Visual pose estimation for autonomous inspection of fish pens. In *OCEANS 2015 - Genova*, 1–6. doi: 10.1109/OCEANS-Genova.2015.7271392.

FhSim (2021). FhSim homepage simulation of marine operations and systems. `https://fhsim.no/`. Accessed: 2021-09-29.

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., and Berckmans, D. (2018). Precision fish farming: A new framework to improve production in aquaculture. *biosystems engineering*, 173, 176–193.

Fossen, T. (2011). *Handbook of Marine Craft Hydrodynamics and Motion Control*. John Wiley and Sons, Ltd.

Fraundorfer, F. and Scaramuzza, D. (2012). Visual odometry : Part ii: Matching, robustness, optimization, and applications. *IEEE Robotics Automation Magazine*, 19(2), 78–90. doi:10.1109/MRA.2012.2182810.

Gafurov, S.A. and Klochkov, E.V. (2015). Autonomous unmanned underwater vehicles development tendencies. *Procedia Engineering*, 106, 141 – 148. doi: https://doi.org/10.1016/j.proeng.2015.06.017. Proceedings of the 2nd International Conference on Dynamics and Vibroacoustics of Machines (DVM2014) September 15 –17, 2014 Samara, Russia.

Gonzalez, R.C., Woods, R.E., and Eddins, S.L. (2004). *Digital image processing using MATLAB*. Pearson Education India.

Hartley, R. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press.

Itseez (2020). Open source computer vision library. `https://github.com/itseez/opencv`.

Kelasidi, E., Liljeback, P., Pettersen, K.Y., and Gravdahl, J.T. (2016). Innovation in underwater robots: Biologically inspired swimming snake robots. *IEEE Robotics Automation Magazine*, 23(1), 44–62. doi: 10.1109/MRA.2015.2506121.

Lader, P., Dempster, T., Fredheim, A., and Østen Jensen (2008). Current induced net deformations in full-scale sea-cages for atlantic salmon (salmo salar). *Aquacultural Engineering*, 38(1), 52 – 65. doi: https://doi.org/10.1016/j.aquaeng.2007.11.001.

Leonardi, M. and Stahl, A. (2018). Convolutional autoencoder aided loop closure detection for monocular slam. *IFAC-PapersOnLine*, 51(29), 159–164. doi: https://doi.org/10.1016/j.ifacol.2018.09.486. 11th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles CAMS 2018.

Leonardi, M., Stahl, A., Gazzea, M., Ludvigsen, M., Rist-Christensen, I., and Nornes, S.M. (2017). Vision based obstacle avoidance and motion tracking for autonomous behaviors in underwater vehicles. In *OCEANS 2017 - Aberdeen*, 1–10. doi:10.1109/OCEANSE.2017.8084619.

Lepetit, V., Moreno-Noguer, F., and Fua, P. (2009). Epnp: An accurate o (n) solution to the pnp problem. *International journal of computer vision*, 81(2), 155.

Massot-Campos M, O.C.G. (2015). Optical sensors and methods for underwater 3d reconstruction. *Sensors (Basel)*, 15(12), 31525–31557.

Oberkampf, D., DeMenthon, D.F., and Davis, L.S. (1996). Iterative pose estimation using coplanar feature points. *Computer Vision and Image Understanding*, 63(3), 495–511.

Paull, L., Saeedi, S., Seto, M., and Li, H. (2014). Auv navigation and localization: A review. *IEEE Journal of Oceanic Engineering*, 39(1), 131–149. doi: 10.1109/JOE.2013.2278891.

Rundtop, P. and Frank, K. (2016). Experimental evaluation of hydroacoustic instruments for rov navigation along aquaculture net pens. *Aquacultural Engineering*, 74, 143–156.

Saputra, M.R.U., Markham, A., and Trigoni, N. (2018). Visual slam and structure from motion in dynamic environments: A survey. *ACM Comput. Surv.*, 51(2). doi:10.1145/3177853.

Schweighofer, G. and Pinz, A. (2006). Robust pose estimation from a planar target. *IEEE transactions on pattern analysis and machine intelligence*, 28(12), 2024–2030.

Taraldsen, G., Reinen, T.A., and Berg, T. (2011). The underwater gps problem. In *OCEANS 2011 IEEE - Spain*, 1–8. doi:10.1109/Oceans-Spain.2011.6003649.

Xiao-Shan Gao, Xiao-Rong Hou, Jianliang Tang, and Hang-Fei Cheng (2003). Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8), 930–943. doi:10.1109/TPAMI.2003.1217599.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11), 1330–1334.

Zhou, K., Wang, X., Wang, Z., Wei, H., and Yin, L. (2018). Complete initial solutions for iterative pose estimation from planar objects. *Ieee Access*, 6, 22257–22266.