Ching Soon TAN, Phooi Yee LAU, Paulo L. CORREIA, Aida CAMPOS, 2018. Automatic analysis of deep-water remotely operated vehicle footage for estimation of Norway lobster abundance. *Frontiers of Information Technology & Electronic Engineering*, 19(8): 1042-1055. https://doi.org/10.1631/FITEE.1700720

Automatic analysis of deep-water remotely operated vehicle footage for estimation of Norway lobster abundance

Key words: Object detection; Object tracking; Feature extraction; Remotely operated vehicle (ROV)

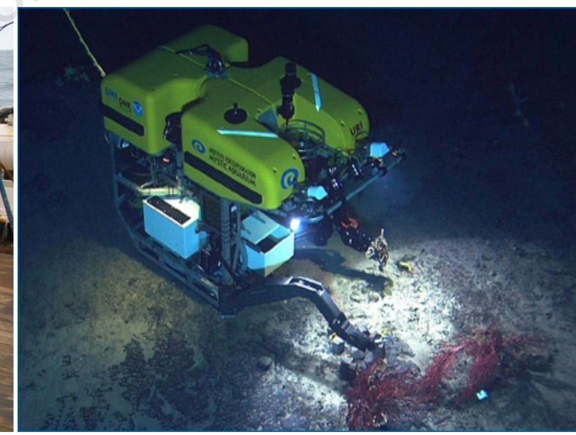
Corresponding author: Phooi Yee LAU
E-mail: laupy@utar.edu.my
ORCID: http://orcid.org/0000-0002-6329-4558

Motivation (1/2)

- ❑ Underwater imaging is a technique of measuring or recording dynamic event in seafloor or in the water column, which is usually done with specialized equipment or method as below.

Diving

Towed sledge



Remote operated vehicle (ROV)

Shallow water

Deep water

Motivation (2/2)

Manual approach

Weakness:

1. Require high trained human operator, usually marine biologist;

2. Become bottlenecked when processing a large pool of data;

3. Counting bias (low-visibility video, human fatigue (capability of concentration, repeatability))

Automatic approach

Strength:

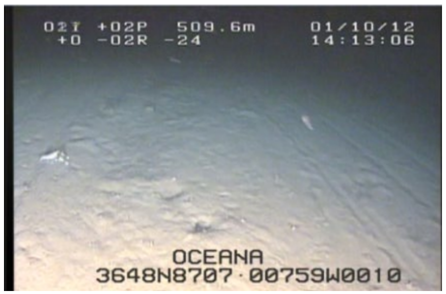
1. Minimize human intensive workload;

2. Time saving;

3. Ensure the certainty of the results.

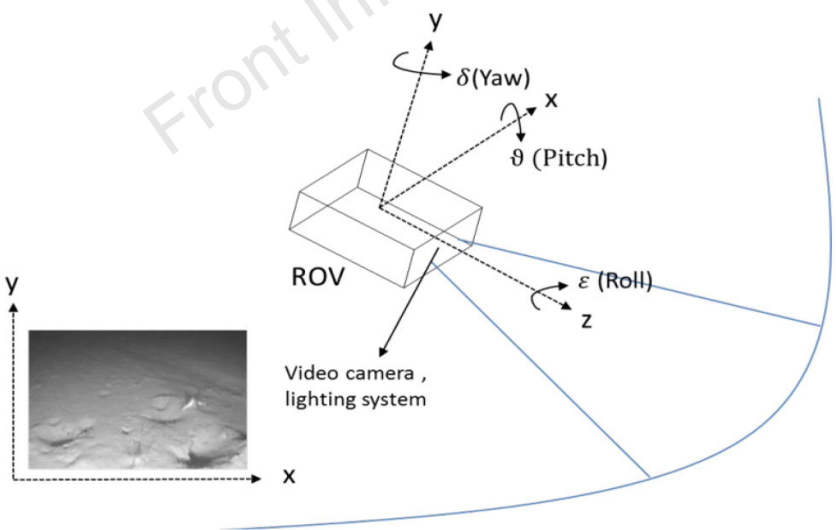
Challenges of the underwater video processing

- Lack of natural illumination in underwater scene.



Deep water images

- Dynamic and unstructured observation scene.



Main idea

Deep water marine abundance study

- Implement an object detection algorithm for the Norway lobster and burrow detection.
- Design a visual tracking scheme to address the dynamic observation scene.
- Evaluate and test the performance of the proposed framework using deep water video sequences.

Method (1/11)

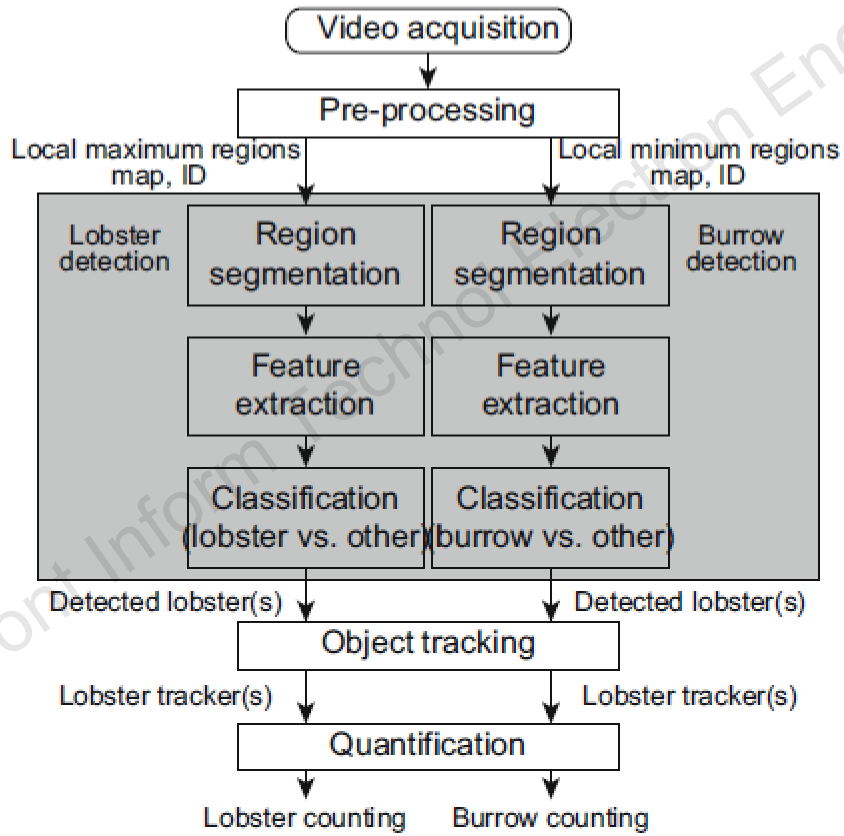


Fig. 3 Architecture of the proposed system

Method (2/11)
Pre-processing & region segmentation

Pre-processing

Step 1.1 underwater video acquisition → Step 1.2 crop the image → Step 1.3 convert RGB to gray scale image → Step 1.4 Background subtraction

$$\mathbf{IB} = G_1 \cdot I - G_2 \cdot I,$$

$$\mathbf{ID} = G_2 \cdot I - G_1 \cdot I,$$

where ***G*** is Gaussian function

Region segmentation

Step 2.1 Region segmentation

Lobster

$$\mathbf{CL}(x,y) = \begin{cases} 255, & \mathbf{IB}(x,y) \geq \frac{m+I_{\max}}{2}, \\ 0, & \text{otherwise}, \end{cases}$$

Burrow

$$T_2(x,y) = m(x,y)\left[1 + k\left(\frac{\delta(x,y)}{R} - 1\right)\right],$$

$$\mathbf{CB}(x,y) = \begin{cases} 255, & \mathbf{ID}(x,y) \geq \max(T_2, T_3), \\ 0, & \text{otherwise}. \end{cases}$$

where $m(x,y)$ is the mean of the local window W, which is calculated using integral image.

Step 2.2 Morphology close operation in CL and CB → Step 2.3 connected component analysis for contour extraction

Method (3/11)

Object detection—feature extraction

F1 curvature (burrow)

I.
Rotate the region and make it to be horizontal

II
Fit parabola model using least square method: $y = ax^2 + bx + c$

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \\ \vdots & \vdots & \vdots \\ x_K^2 & x_K & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_K \end{bmatrix}$$

III.
$$\mathrm{F1} = \mathrm{a}$$

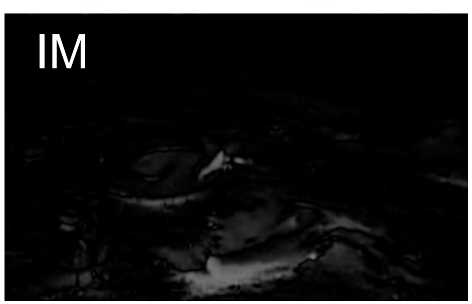
$$f(a) = \begin{cases} upward, a > 0 \\ downward, a < 0 \\ straight, a = 0 \end{cases}$$

F2 Local intensity contrast (Lobster and burrow)

$$\mathrm{IM} = |\mathrm{I} - (\mathrm{H} * \mathrm{I})|$$

$$\mathrm{F2} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{IM}(\mathrm{x_i}, \mathrm{y_i}),\ \mathrm{IM}(\mathrm{x_i}, \mathrm{y_i}) \in \mathrm{C}(\mathrm{x_i}, \mathrm{y_i})$$

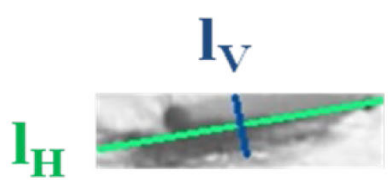
Where H is the median blur with kernel size WH, C is the contour pixel of the region.



F3 Elongated structure (Lobster and burrow)

$$\text{Burrow: } \mathrm{F3} = \frac{\mathrm{l_H}}{\mathrm{l_V}} \leq 1.0$$

$$\text{Lobster: } \mathrm{F3} = \frac{\mathrm{l_{major}}}{\mathrm{l_{minor}}} > 1.5$$



Method (4/11)

Object Detection—feature extraction—cont.

F4 Diameter (Burrow)

$$\mathrm{F4} = \mathrm{l_H}/W_I$$

To reject false alarms F4 should take values in the range $[0.05, 0.5]$ of the image width (W_I).

F5 Contour area (Lobster)

$$\mathrm{F5} = A_{lobster}/A_I$$

To exclude such regions from further analysis, each candidate region's area is analysed, with lobster regions' area, $A_{lobster}$, being expected to be between 0.02 and 0.3 of image area, A_I .

F6 Orientation (Lobster)

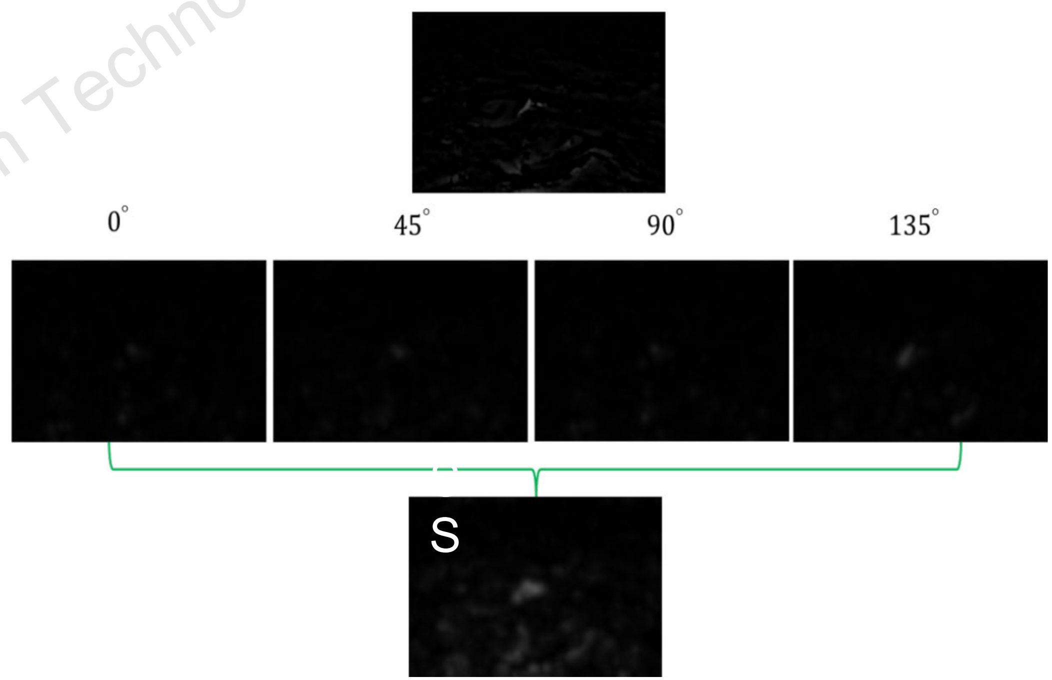
$$G_{u,v}(y, \mathrm{x}) = \frac{f_u^2}{\pi\gamma\eta} e^{-\left(\frac{f^2}{\gamma}x'^2+\frac{f^2}{\eta}y'^2\right)} e^{j2\pi f_u x'})$$

$$S = \sum_u^3 \sum_v^4 G_{u,v}$$

$$\mathrm{F6} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{S}(\mathrm{x_i}, \mathrm{y_i}),$$

$$\mathrm{S}(\mathrm{x_i}, \mathrm{y_i}) \in \mathrm{C}(\mathrm{x_i}, \mathrm{y_i})$$

where $G_{u,v}$ is the Gabor filter response at the scale u and orientation v, C is the contour pixel of the region; u= {0, 1, 2}; v={0, $\frac{1}{4}\pi$, $\frac{1}{2}\pi$, $\frac{3}{4}\pi$}



Method (5/11)

Object detection—region classification

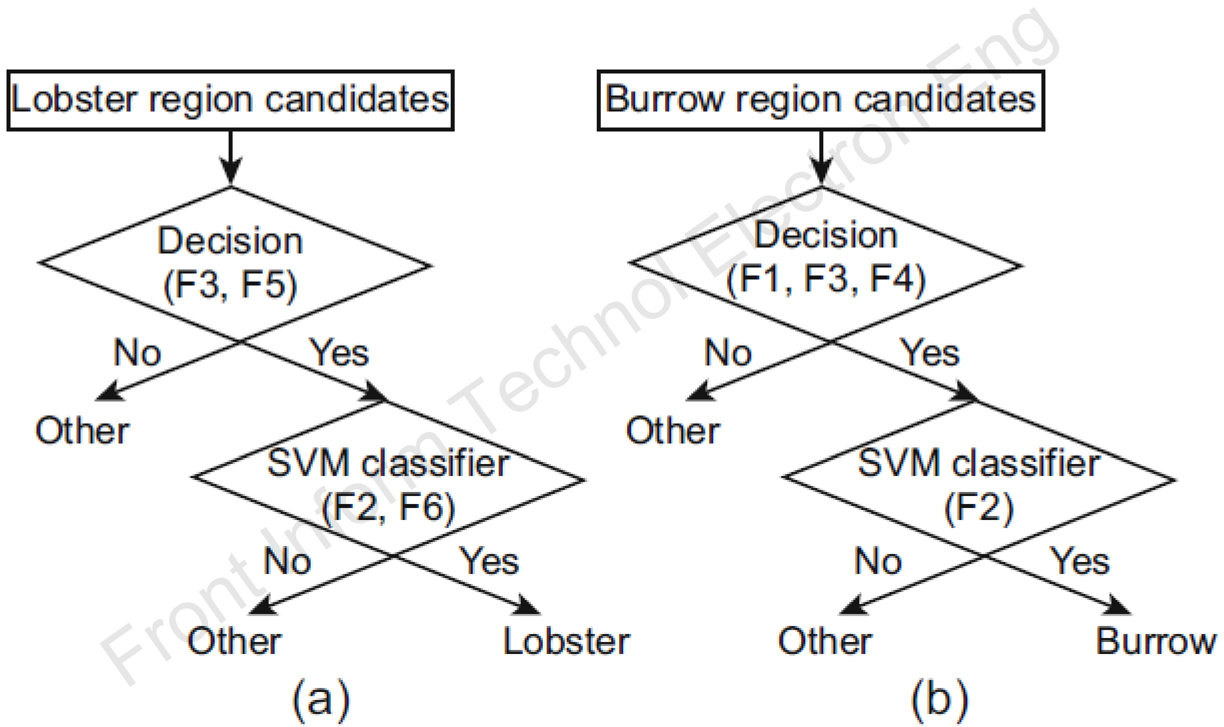
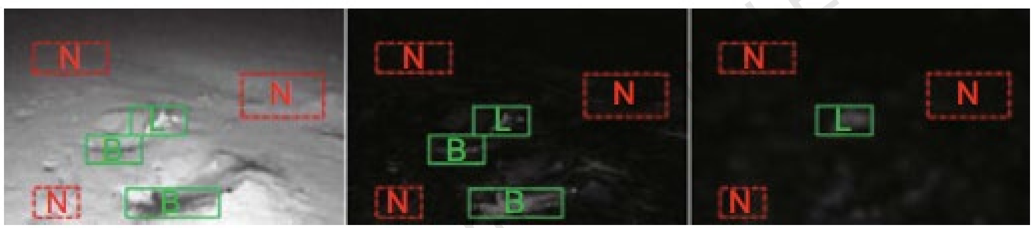


Fig. 11 Proposed classification schemes: (a) lobster; (b) burrow

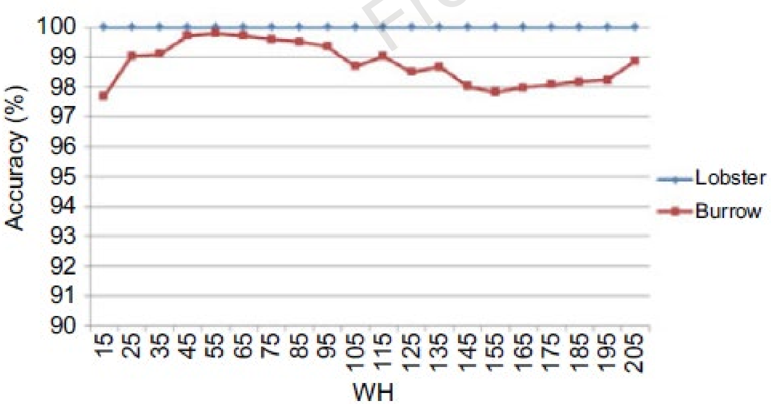
Method (6/11)

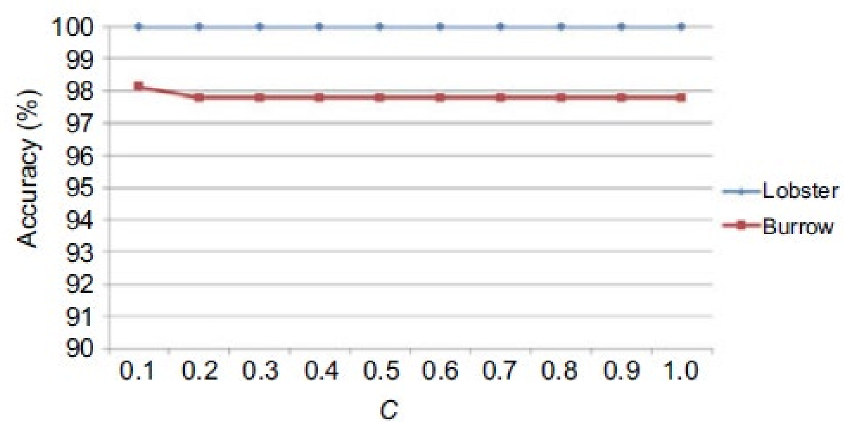
Object detection—region classification—cont.

- Training samples are selected from random frames of a train video clip, with a set of regions being manually operated. A total of 600 samples were considered for the training stage, with 300 positive and 300 negative examples, respectively.



- Find the optimal value for *C* of SVM and WH used for feature *F*2 extraction by using a 10-fold cross-validation technique.





Method (7/11)

Visual tracking scheme

Algorithm: Our overall framework

Input: Underwater video sequence

FOR t=0:N frame

(2.4.2) Object Detection as described in chapter 2.

(3) Tracking by our proposed tracking scheme in this chapter

IF t>=1 THEN

(3.3.2) State Prediction ($LTr_{1:t}^{(i)}$, $BTr_{1:t}^{(j)}$)

(3.3.2.1) Lobster tracking state prediction and update($LTr_{1:t}^{(i)}$)

(3.3.2.2) Burrow tracking state prediction and update ($BTr_{1:t}^{(j)}$)

END IF

(3.3.3) Data association ($LTr_{1:t}^{(i)}$, $BTr_{1:t}^{(j)}$)

IF t ==N THEN

(3.3.4) Quantification strategy ($LTr_{1:t}^{(i)}$, $BTr_{1:t}^{(j)}$)

END IF

END FOR

Output: Video-based count number of lobster and burrow

**$LTr_{1:t}^{(i)}$ is the lobster tracker at time t; $BTr_{1:t}^{(i)}$ is the burrow tracker at time t

Method (8/11)

State prediction - Lobster state prediction

Step 1

Propagate N=50 particles according to transition model:

$$x_t = v + x_0$$
$$v = Ax_{t-1} + Bx_{t-2} + CG(0,1)$$

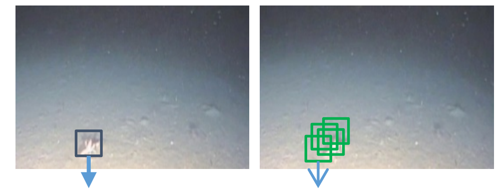
where {A,B,C} is the autoregressive coefficient ; G is Gaussian noise with zero mean and standard deviation 1; The coefficients are commonly defined as $A = 2, B = -1, C = 1$.

Step 2

Update the weight, $w^{(i)}$ of each particles according to the observation model, where the observation model is created based on intensity feature.

$$D(\alpha, \beta) = \sqrt{1 - \sum_{i}^{N} \frac{\sqrt{\alpha(i)\beta(i)}}{\sqrt{\sum_{i=1}^{N} \alpha(i) \sum_{i=1}^{N} \beta(i)}}}$$

$$w^{(i)} \propto D(\alpha, \beta)$$



Reference histogram obtained at t-1th frame

histogram of particles at tth frame

Where α is the histogram of the object, β is the histogram of the referred object, N is the total of bins

Step 3

Resample the old particles

Step 4

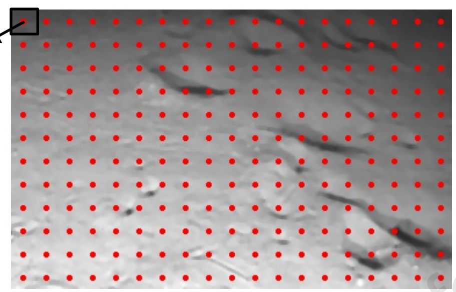
The particle with the lowest weight is picked to be the observation state candidate, x. If $w^{(x)} > 0.3$, we will consider that the lobster is out of scene so that the observation state will not be assigned.

Method (9/11)

State prediction—burrow state prediction

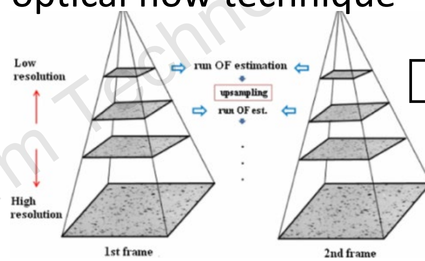
t-1th frame $\quad$ t^{th} frame

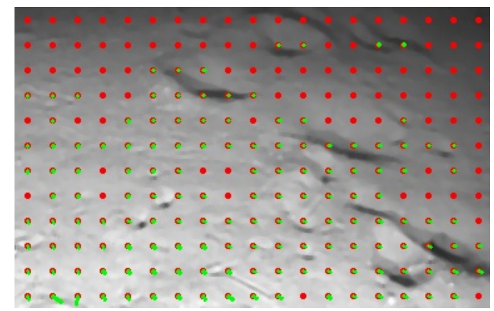
Flow point



Divided into N grids with a size of 15x15. Center pixel of each grid is used as flow point.

Iterative Lucas-Kanade optical flow technique





Find 6 parameters of Affine transformation with the estimated optical flows (x_{t-1}, y_{t-1}) and (x_t, y_t).

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} b_{00} \\ b_{10} \end{bmatrix}$$

Let the centroid of the tracked region be (x_c, y_c).

$$\begin{bmatrix} x_o \\ y_o \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} b_{00} \\ b_{10} \end{bmatrix}$$

Predicted observation state : (x_o, y_o).

Method (10/11)

Data association

- Let $\mathrm{L_t}$ and $\mathrm{B_t}$ be the detected lobster and burrow in frame t, $\mathrm{LT_{1:t-1}}$ and $\mathrm{BT_{1:t-1}}$ are the lobster and burrow tracker, $\mathrm{L_t}$ links to $\mathrm{LT_{1:t-1}}$; $\mathrm{B_t}$ links to $\mathrm{BT_{1:t-1}}$; Each of the linkage should be one-to-one correspondence. The data assignment between detected items and trackers can be solved by Hungarian algorithm (Kuhn, 1955).
- If $Cost_{\mathrm{i}\to j} \geq 20$, it will be consider as “New Object” ; Otherwise be considered as the item I is same to the item j.

	Detect Item A	Detect Item B	....	Detect Item N
Tracker 1	$Cost_{1\to A}$	$Cost_{1\to B}$	....	$Cost_{1\to N}$
Tracker 2	$Cost_{2\to A}$	$Cost_{2\to B}$	....	$Cost_{2\to N}$
Tracker 3	$Cost_{3\to A}$	$Cost_{3\to B}$	...	$Cost_{3\to N}$
⋮	⋮	. . .	. . .	. . .
Tracker M	$Cost_{M\to A}$	$Cost_{M\to B}$	....	$Cost_{M\to N}$

Method (11/11)

Quantification strategy

Strategy 1 (lobster and burrow)

$$N_{\text{Detection}} \geq \delta, \tag{21}$$

where $N_{\text{Detection}}$ is the number of frames in which the lobster candidate is successfully tracked, and δ is a threshold corresponding to the minimum value to consider that the object does not correspond to a false detection. Herewith, δ is set to 10 after a set of preliminary tests, which means that the object will be detected in at least 10 frames.

Strategy 2 (burrow)

Two variables, N_B, N_O for each burrow trajectory.

Situation 1: At frame t, when the tracked burrow is detected again in the visible image area then its N_B value is increased by one.

Situation 2: If the tracked burrow centroid is supposed to be found in the image area but it is not detected then N_O is increased by one.

$$f = \frac{N_B}{N_B + N_O}$$

If $f \geq 0.5$, consider to be counted as a detected burrow.

Major results (1/3)

Object tracking performance evaluation
—evaluation metric and results

Evaluation metric:

$$\mathrm{MOTA} = 1 - \frac{\sum_{i=1}^{N}\left(c_m(m_i) + c_f\left(f_{P_i}\right) + \log_e(id_{switches})\right)}{\sum_{i=1}^{N} G_i}$$

where (m_i) is number of miss track at ith frame, $c_f\left(f_{P_i}\right)$ is the number of false positive tracks at ith frame, $id_{switches}$ is the number of false identify switch and G_i is number of mapped object over entire trajectory at ith frame. In that case, the evaluation will begin at the time once the items are detected and tracked.

Table 4 Performance evaluation of the proposed system based on the multiple object tracking accuracy (MOTA) metric

Target	$\sum_{i=1}^{N} G_i$	MOTA	ML	FM	ID
Lobster	681	1.0000	0	0	0
Burrow	7036	0.9902	0	0.0098	0

ML: ratio of the missed tracks in total frames; FM: ratio of tracks with false positive in total frames; ID: ratio of mismatched id in total frames; G_i: number of mapped objects over the entire trajectory at the i^{th} frame

Major results (2/3)

Precision evaluation for the video-based abundance estimation—evaluation metric and results

Evaluation metric

$$\text{Recall} = \frac{N_{\text{true\_positive}}}{N_{\text{positive}}} \times 100\%, \tag{18}$$

$$\text{Precision} = \frac{N_{\text{true\_positive}}}{N_{\text{true\_positive}} + N_{\text{false\_positive}}} \times 100\%, \tag{19}$$

Table 5 Comparison between an automatic approach and a manual approach

Approach	Object	Manual count	Automatic count	True positive	False positive	False negative	Recall (%)	Precision (%)	F-score (%)
R1	Lobster	5	6	5	1	0	100.00	83.33	90.90
R1	Burrow	105	110	76	34	29	72.38	69.09	70.69
PS	Lobster	5	5	5	0	0	100.00	100.00	100.00
PS	Burrow	105	104	87	17	18	82.86	83.65	83.25

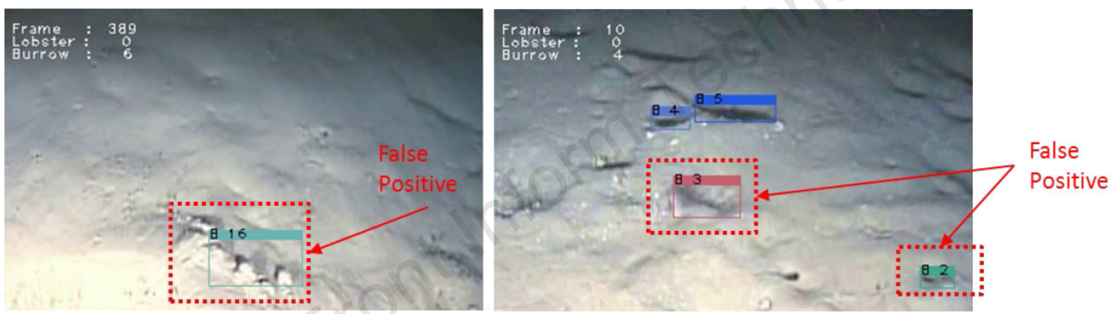
RI: methodological approach used in Tan et al. (2014) with the same tracking solution; PS: methodological approach used in this paper

Major results (3/3)

Precision evaluation for the video-based abundance estimation—result discussion

Some of the seabed sediments may confuse the decision making of the burrow detector as they seem like having a same characteristic of being a burrow but it does not exactly belong to the burrow in sense.

A lack a more robust feature or imprecision segmentation



Detection failure due to the violation of feature *F*1 curvature

Conclusions

An automated video analysis framework proposed here with facilitates the Norway lobster stock assessment by automatically detecting and tracking, and finally quantifying Norway lobster and its biogenic feature (main burrow's entrance) from the underwater video sequences obtained in deep-water crustacean grounds.