

Group behavior tracking of *Daphnia magna* based on motion estimation and appearance models

Zhitao Wang^a, Chunlei Xia^{a,b,*}, JangMyung Lee^{a,*}

^a Department of Electronics Engineering, Pusan National University, Busan 46241, Republic of Korea

^b Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, China

ARTICLE INFO

Keywords:

Multiple individual tracking
Occlusion segmentation
SORT
Behavioral monitoring
Motion estimation

ABSTRACT

Daphnia magna is a widely adopted biological indicator for studying the environmental toxicity in aquatic ecosystems. Due to the tiny size and appearance similarity of *Daphnia* most behavioral tracking systems can only record the behavioral movement of a single individual. In this work, a novel behavioral tracking scheme is proposed to track the group movement of *Daphnia* and record the accurate trajectory of each individual. Individual position, motion estimation and appearance similarity are incorporated to solve the multiple objects tracking problem. Optical flow is utilized to estimate individual motion and identify each individual in occlusions. Normalized cross correlation is employed to measure appearance similarity of *Daphnia* individuals. According to these feature movement trajectories are assigned to each individual by linear optimization. The main advantage of the proposed method is to segment individual *Daphnia* from occluded images according to the appearance and motion characters. Consequently, the success rate of tracking under occlusions could be significantly improved. The proposed method was tested on 8 video clips which captured in two different observation conditions. The correct identification ratio reached 94.74% and the average success ratio was 87.68%. The comparison tests proved that our method was superior to a well-known multiple object tracking method in dealing with occlusions. The proposed *Daphnia* tracking scheme was reliable and could provide accurate movement information for behavioral and environmental studies.

1. Introduction

There has been a growing research interest in group animal behavior due to its high values in science and a wide range of applications, for example in environment pollution (Häder and Erzinger, 2017), toxicity testing (Moldaenke, 2000). *Daphnia magna* is one of the model organisms due to their sensitivity to toxics or pollutants. And *Daphnia* can be easily cultured in laboratory conditions. Toxicological studies usually require acute toxicity data for evaluation of lethal concentration for mortality or effective concentration for immobilization (Bownik, 2017). To investigate the mechanisms or deep understanding of the behavioral responses affected by toxics or environmental changes precise individual response should be obtained, such as swimming time (Ferrão-Filho et al., 2014), speed (Cano et al., 2017; Noss et al., 2013; Stanley et al., 2016), hopping frequency (Lovren et al., 2007), cumulative distance (Hansen and Roslev, 2016; Zein et al., 2015) are required. Benefited by recent advances in computer vision techniques, vision-based behavioral monitoring system can simultaneously measure several types of

movement parameters by visual tracking. Vision-based behavioral monitoring system is an efficient and fundamental tool and more ecologically relevant and economical than traditional methods (Bae and Park, 2014).

Most current *Daphnia* behavioral monitoring systems could only accurately record the movement of a single individual. Multiple individual tracking system still faces many challenges because the *Daphnia magna* body is not rigid, and the body shape changes during swimming. In addition, *Daphnia magna* often mutually occlude which has brought great difficulties for tracking multiple individuals. A number of visual tracking methods have been proposed to solve the multiple individual tracking problem. These methods can be divided into two categories: tracking by prediction and tracking by detection.

In the first category, the tracking algorithm predicts the position of each individual in current frame based on the motion state calculated by analyzing previous frames (Bewley et al., 2016; Qian et al., 2014; Qian et al., 2016). These methods can track a large number of individuals; however, the tracking accuracy and stability will decrease when

* Corresponding authors at: Department of Electronics Engineering, Pusan National University, Busan 46241, Republic of Korea.

E-mail addresses: clxia@yic.ac.cn (C. Xia), jmlee@pusan.ac.kr (J. Lee).

<https://doi.org/10.1016/j.ecolinf.2021.101238>

Received 29 June 2019; Received in revised form 6 November 2020; Accepted 7 November 2020

Available online 20 January 2021

1574-9541/© 2021 Elsevier B.V. All rights reserved.

occlusion occurs. In addition, these motion prediction methods commonly by applying linear predictor such as Kalman filter, cannot be effectively adapted to track animals with uncertain mobility. For example, Qian et al. (2014) uses a constant velocity model to predict motion between consecutive frames, but a compensation window has to be used to make up for the shortage. Similarly, Xia et al. (2016) presented a behavioral monitoring of multiple fish based on the constraint of fish posture.

In the second category, the main idea is associating detection based on the appearance information. For example, Rodriguez et al. (2017) proposed ToxId to solve the tracklets linking problem caused by occlusions, the tracklets are linked based on their intensity histogram and Hu-moments. Bai et al. (2018) proposed an algorithm to tracking multiple Zebrafish by training classifier to classify the improved HOG features, extracted from fish back, then, establishes the data association. These methods increase the tracking accuracy after occlusion because the appearance information is stable, however, they may not be suitable for tracking of a large number of *Daphnia magna* due to the similarity of appearance and, especially, the huge appearance variance during the

interaction.

Occlusion among biological individuals is the most challenge of vision-based behavioral monitoring system. In this work, motion estimation and appearance feature of *Daphnia* are utilized to improve the tracking accuracy, especially, under occlusions. Motion analysis of each individual is implemented by optical flow. According to the optical flow field occlusion events could be detected and individuals could be identified in the occluded images. In addition, appearance similarity of each individual is described by normalized cross-correlation. By combining individual positions, motion estimation and appearance similarity movement trajectory of each individual is generated.

The contributions of this work are:

1. Individuals could be identified and accurately segmented from the occluded images. Most previous works could not identify individuals when occlusion occurs and the tracker will lose the targets. With the proposed occlusion segmentation, the tracker has high probability to follow the correct target even in the occlusion process. It could significantly enhance the stability and robustness of tracking.

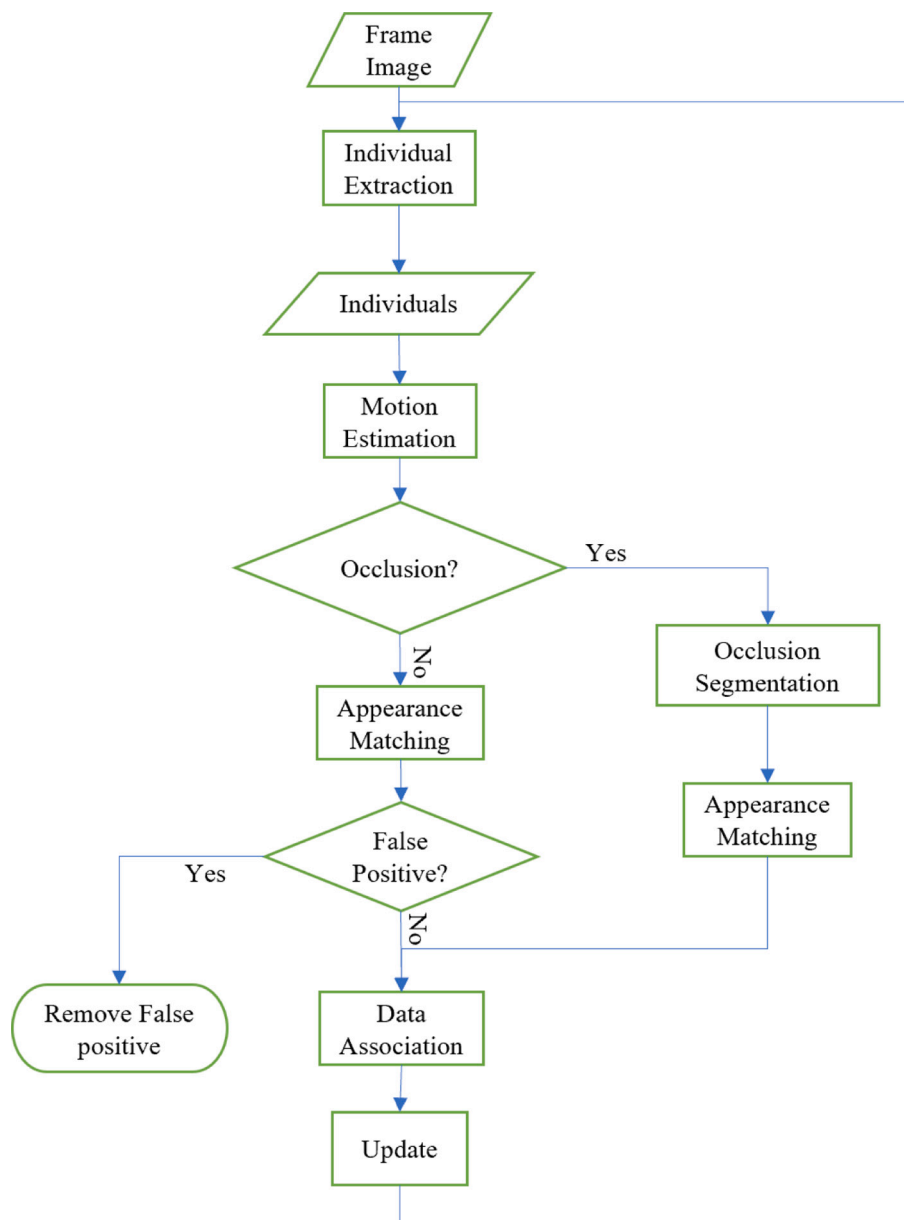


Fig. 1. Overall structure of the proposed tracking method.

2. Motion estimation and appearance feature are incorporated to track multiple individuals. An appearance update strategy is implemented according to the tracking results. Appearance similarity of each individual is dynamically calculated in every frame to ensure the accuracy of matching individual. Therefore, the appearance matching could enhance the tracking accuracy.

2. Multiple individual tracking of *Daphnia magna*

2.1. Overall structure of proposed tracking scheme

A multiple individual tracking scheme is proposed by integrating positions, motion estimation and appearance features. The proposed method is inspired by the principle of tracking-learning-detection framework. The original Tracking-Learning-Detection (TLD) framework was proposed by Kalal et al. (2011) and was designed for long-term tracking of a single unknown object in a video stream. It can be divided into three components as follows: tracking, detecting, and learning. The creative idea of the TLD framework is that the learning component, trained by the positive and negative appearance model, estimates the detector's error then avoids the errors in the future to increase the tracking performance. In our method, tracking scheme was constructed based on both motion (optical flow) and appearance feature to adapt for the group behavior tracking applications. The flowchart of the proposed multiple individual tracking is presented in Fig. 1. First of all, *Daphnia* individuals are extracted from the background. Motion analysis is performed to obtain the optical flow field and occlusion could be determined according to the optical flow field. After occlusion segmentation, appearance similarity of each individual is examined. False positives are removed based on motion and appearance matching result. To associate the detected individuals in current frame to their tracker, a linear assignment method is performed according to the combination result of estimated motion and appearance similarity. In the meantime, the appearance model of each individual is updated as well. The trackers are updated frame by frame according to the optimal association result.

2.2. Individual detection

For the laboratory study of group animal behavior, the captured video contains only moving animals and a nearly static background. In our case, most of the *Daphnia magna* stay only for a short time in an area. Therefore, it is possible to detect the moving region by modeling the background with the time domain-based median filtering method. Firstly, the median images of the first n frames in the observed video can be selected as a background image, and then the moving regions can be segmented by setting a threshold for the differential image of the background image and the input image. One frame of the captured aquarium image with *Daphnia magna* (raw image) is shown in Fig. 3a. The background image Fig. 3b is calculated by computing the median image of the first 500 successive frames. After background subtraction, the pixels within the *Daphnia magna* body area are significantly different from the background pixels. In addition, we use the size constraint to remove most false positives, then the blobs extracted as shown in Fig. 3c. All the extracted blobs can be divided into three categories: one blob including one individual, one blob including multiple individuals, and false positive.

All the candidates detected in the blob extraction stage are classified into three categories according to the estimated motion (Farnebäck, 2003; Kalal et al., 2010) and appearance similarity. Let B_t^i and T_j denote the i^{th} blob in frame t and the j^{th} tracker, respectively, and $OF(i,j)$ and $NCC(i,j)$ indicate the probability of i^{th} blob moved from j^{th} tracker and the appearance similarity between i^{th} blob and j^{th} tracker, respectively.

$$r_{ij} = \begin{cases} 1, & OF(i,j) > Thr_{of} \text{ and } NCC(i,j) > Thr_{ncc} \\ 0, & \text{otherwise} \end{cases} \quad (i = 1, \dots, n; j = 1, \dots, m) \quad (1)$$

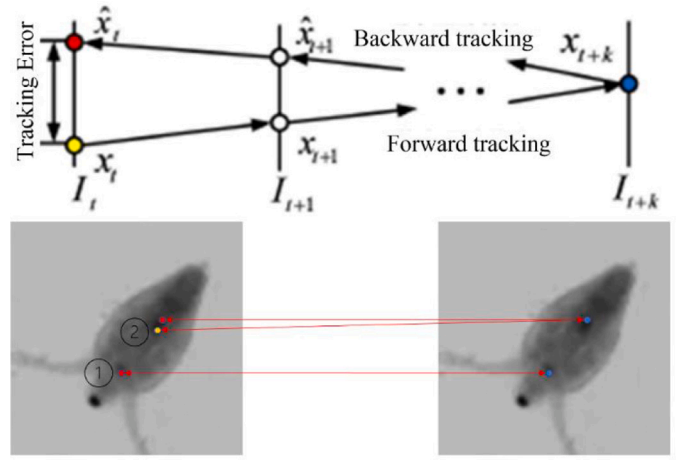


Fig. 3. *Daphnia magna* Observations and Segmentation: (a) raw image, (b) background image, (c) background subtraction, (d) occlusion segmentation.

where $r_{ij} = 1$ means the blob i is related to j^{th} tracker; $r_{ij} = 0$ means the blob i is not related to j^{th} tracker, and Thr_{of} and Thr_{ncc} denote the thresholds of motion measured and appearance similarity, respectively.

$$count_i = \sum_{j=1}^m r_{ij} \quad (i = 1, \dots, n) \quad (2)$$

where $count_i$ to indicate the number of trackers related to i^{th} blob. Due to occlusion among individuals, three cases may occur for $count_i$ during the detection stage. Details of these cases and corresponding processing are given below:

- (1) $count_i = 0$: the i^{th} blob cannot relate to an existed tracker. It is considered as false positive.
- (2) $count_i = 1$: the i^{th} blob only relates to one existed tracker. It is considered as true positive, which will be remained to be processed in the tracking component.
- (3) $count_i > 1$: the i^{th} blob relates to multiple existed trackers due to occlusion. In this case, the blob i includes more than one individual, therefore, it will be re-segmented in occlusion segmentation stage.

2.3. Appearance similarity and motion estimation

To measure the appearance similarity, the Pearson's normalized correlation coefficient (NCC) is utilized between the individual i and tracker j , which is also known as mean shifted cross correlation. Therefore, it is robust under illumination change and slight partial occlusion. Firstly, the trackers T is defined as follows:

$$T = \{T_j | j = 1, \dots, m\} \quad (3)$$

where m is the total number of individuals in the initial frame, and T_j is j^{th} tracker represents a single individual, defined as:

$$T_j = \{(x', y'), w', h', I\} \quad (4)$$

where (x', y') denotes the coordinate of the left top corner of the individual bounding box, w' and h' denote the width and height of the bounding box, respectively, in the previous frame. I , the appearance model, is a set of pixels of individual. Similarly, let $D_i = \{(x, y), w, h, P\}$ be the i^{th} individual of detected individuals, $D = \{D_i | i = 1, \dots, n\}$, in current frame, where n is the number of detected individuals. The $NCC(i,j)$ is determined by the following equation:

$$NCC(i,j) = \max\{S(dx, dy)\} \quad (5)$$

where $S(dx, dy)$ is:

$$S(dx, dy) = \frac{\sum_{dx', dy'} (I'(dx', dy') \cdot P'(dx + dx', dy + dy'))}{\sqrt{\sum_{dx', dy'} I'(dx', dy')^2 \cdot \sum_{dx', dy'} P'(dx + dx', dy + dy')^2}} \quad (6)$$

where

$$\begin{aligned} I'(dx', dy') &= I(dx', dy') - \frac{1}{w' \cdot h'} \cdot \sum_{dx'', dy''} I(dx'', dy'') \\ P'(dx + dx', dy + dy') &= P(dx + dx', dy + dy') \\ &\quad - \frac{1}{w \cdot h} \cdot \sum_{dx'', dy''} P(dx + dx'', dy + dy'') \end{aligned} \quad (7)$$

The above equations indicate that, firstly, appearance model I of tracker j and image patch P of detected individual i are subtracted by their mean value of all pixels, then, sliding through the image patch P , compares the overlapped patches of size $w' \times h'$ against appearance model I according to eq. 4. We choose the maximum to represent the similarity between appearance model I and image patch P .

On the other hand, the motion of each individual in consecutive frames is estimated by Farneback's method (Farneback, 2003; Kalal et al., 2010). Let the $OF(i, j)$ indicate the ratio of pixels come from tracker j to all pixels in detected individual i . It is formulated as:

$$OF(i, j) = \frac{Num(P')}{Num(P)} \quad (8)$$

where $Num(P)$ denotes the total number of pixels in i^{th} detected individual, $Num(P')$ denotes the number of pixels come from tracker j . P' subjects to:

$$\begin{aligned} dy' &\in [y', y' + h'] \\ dx' &\in [x', x' + w'] \end{aligned} \quad (9)$$

where x' and y' are the left top coordinate of bounding box of tracker j , respectively, w' and h' are the width and height of the bounding box, respectively. And, dy' and dx' are defined as:

$$dy' \sim dy + flow(dy, dx)[1] dx' \sim dx + flow(dy, dx)[0] \quad (10)$$

where dx and dy are the coordinate of pixel in i^{th} detected individual, dx' and dy' are the corresponding coordinate in j^{th} tracker, respectively. Flow (dy, dx) represents the motion estimated by Farneback's method.

In addition, the forward-backward (FB) error prediction method (Kalal et al., 2010) is applied to reduce the motion estimation error. A tracking trajectory is obtained by the forward tracking, and a verification trajectory is gained by backward tracking. If the error between the two trajectories is less than a threshold, the estimated trajectory of the pixel is considered successful, otherwise the trajectory fails. Fig. 3 shows the prediction process for the FB error predictor.

The FB error has been shown in the top half of Fig. 2. A tracking trajectory $T_f^k = (x_t, x_{t+1}, \dots, x_{t+k})$ is obtained by the forward tracking of the local tracker in the image series $S = (I_t, I_{t+1}, \dots, I_{t+k})$, where x is the position of the tracker. The verification trajectory $T_b^k = (\hat{x}_t, \hat{x}_{t+1}, \dots, \hat{x}_{t+k})$ can be gained by backward tracking, where $\hat{x}_t = x_{t+k}$. The FB error is the distance of the two trajectories: $FB(T_f^k | S) = \text{distance}(T_f^k, T_b^k)$, where

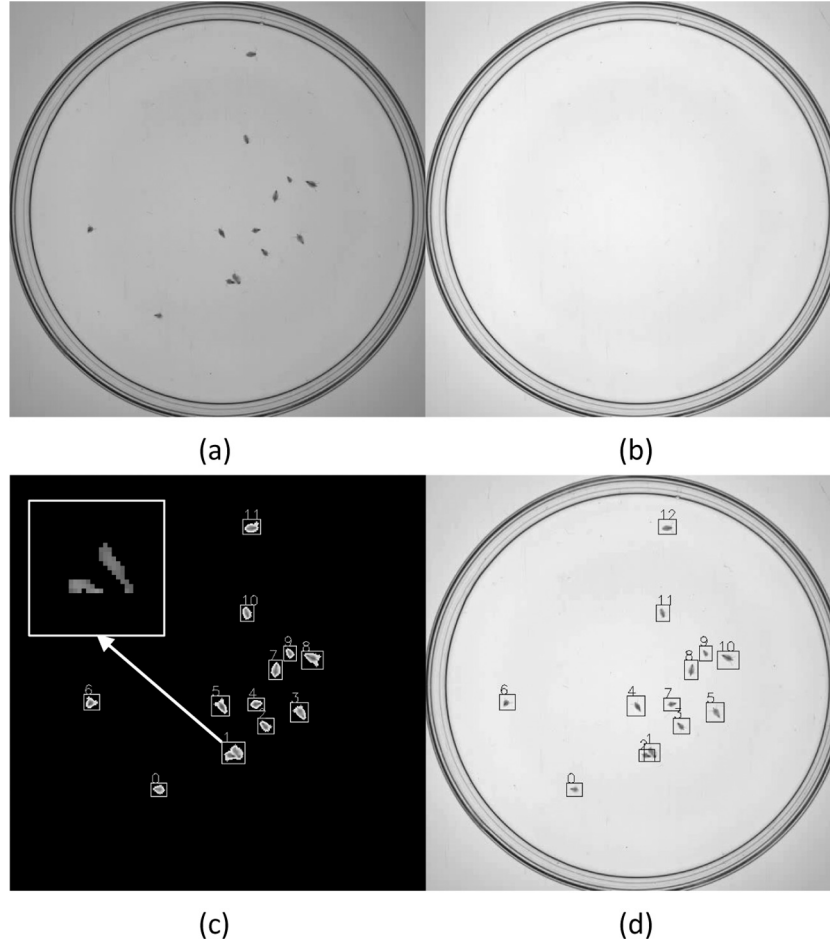


Fig. 2. Process of FB Error Predictor.

distance(T_f^k, T_b^k) = $\|x_t - \hat{x}_t\|^2$. In the bottom half of Fig. 2, trajectory of Point 1 works consistently forward and backward. Trajectory of Point 2 are in consistent, that indicates estimated motion failing if the distance was greater than a threshold.

2.4. Occlusion segmentation

According to the classification result for every blob in frame t , blobs, which include multiple individuals, are re-segmented using the adaptive thresholding algorithm (Sauvola and Pietikäinen, 2000). Due to adaptive thresholding algorithm estimates the threshold for each pixel from its neighboring area, which significantly reduces the segmentation failure in the blob extraction stage caused by partial occlusion. In Fig. 3c, the rectangle 1 shows the occlusion of two individuals, resulting in segmentation failure. The rectangle on the left top of Fig. 3c shows the occlusion segmentation result and Fig. 3d shows the result after appearance similarity checking. On the other hand, because the classification result and the segmented blobs may be incorrect, a similarity of appearance checking process is performed, in which the similarity of the pixels in new blob and pixels in each tracker is measured by a cross correlation method. If $NCC(i, j)$ is greater than the threshold indicates that the occlusion segmentation successes. Otherwise, the blob is remained and shared by multiple trackers.

2.5. Trajectory assignment and Learning appearance feature

After the detection of individuals in current frame with the above method, we will associate each detected individual in current frame with each tracker. Commonly, data association follows the one to one criterion, which means that one tracker should be associated with at most one detection and each detection should be associated with at most one tracker. However, due to the occlusion, resulting in fewer detected individuals than trackers. We let the tracker shares the same detected individual.

In order to associate detected individual D_i in current frame to tracker T_j , we formulate the data association task as a global optimization problem and employ Hungarian algorithm (Kuhn, 1955) to calculate a global optimum solution. The cost matrix is defined as:

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1j} \\ c_{21} & c_{22} & \dots & c_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ c_{i1} & c_{i2} & \dots & c_{ij} \end{bmatrix} \quad (11)$$

where c_{ij} represents the cost of detected individual i being associated with tracker j , defined as:

$$c_{ij} = 1 - \max\{\lambda_o \cdot OF(i, j), \lambda_n \cdot NCC(i, j)\} \quad (12)$$

$$\lambda_o = \begin{cases} 1, & \text{if } OF(i, j) > Thr_{of} \\ 0.5, & \text{otherwise} \end{cases} \quad (13)$$

$$\lambda_n = \begin{cases} 1, & \text{if } NCC(i, j) > Thr_{ncc} \\ 0.5, & \text{otherwise} \end{cases} \quad (14)$$

the objective function is:

$$\delta = \underset{A}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^m c_{in} \cdot A_{ij} \quad (15)$$

subject to:

$$\sum_{i=1}^n A_{ij} = 1 \text{ and } \sum_{j=1}^m A_{ij} = 1 \quad (16)$$

where $OF(i, j)$ represents the possibility of correct association between

detected individual i and tracker j according to the motion estimated by employing Farneback's Optical Flow algorithm with Kalal's FB error detection algorithm, $NCC(i, j)$ is Pearson's normalized correlation coefficient, which measures the appearance correlation, similarity, between the individual i and tracker j , which is also known as mean shifted cross correlation. Therefore, it is robust under illustration change and slight partial occlusion.

Following the steps above, there are three cases may occur during the tracking processing due to occlusion among individuals. Details of these cases and corresponding processing are given below:

- (1) $m = n$: The current frame has the same number of individuals as trackers and are fully matched.
- (2) $m < n$: The current frame has more individuals than the trackers. This implies that false negative occurred. In this case, among the n detected individuals, false positive can be detected and removed as described in detection component.
- (3) $m > n$: The current frame has fewer individuals than the trackers. This means that occlusion occurs among the individuals. In this case, a sub-process, occlusion segmentation, applies based on the appearance model.

Therefore, considering that the number of tracker m stays constant, if the number of individuals in the initial frame is equal to m , the number of associated targets for all subsequent frames remains the same. It solves the problem of trajectory breaks that occur due to track merging and splitting when individuals occlude. Based on the association result and appearance similarity, the most recent appearance is updated as I of tracker in the learning component, where I is a 11×11 pixel image patch at the center of detected individual image. We use two constraints to decide if update the appearance model of each tracker. One is the threshold Thr_{ncc} of appearance similarity between detected individual i and tracker j , the other is threshold Thr_{dis} of distance error err_{dis} defined as:

$$err_{dis} = \sqrt{\left(dx - \left(x + w/2\right)\right)^2 + \left(dy - \left(y + h/2\right)\right)^2} \quad (17)$$

where dx and dy are the coordinate of the center of the image patch, which is the most similar to the appearance model of tracker j , x and y are the coordinate of the left top corner of the bounding box of individual I , w and h are the width and height of the bounding box, respectively.

3. Experiments

3.1. Experimental conditions

In this work *Daphnia magna* was obtained from local market. A video monitoring facility was constructed to observe the two-dimensional behavior of *Daphnia*. As shown in Fig. 4, the observation facility was composed by a Petri dish, a high-resolution digital camera, light source and a personal computer. *Daphnia* was released to the Petri dish. A digital camera was installed over the Petri dish and captured the images of *Daphnia* from top view. The camera was placed 20 cm above the petri dish that the field of view of the camera could cover the arena of *Daphnia*. A while LED panel was installed under the petri dish and contrast between the *Daphnia* and background was optimized by adjusting the light controller. Uniform illumination was obtained in the observation area and reflection and shadow were eliminated. The proposed tracking scheme was implemented by OpenCV and Python on a personal computer. The details of the development environment are listed in Table 1. The proposed tracking system was developed under ubuntu 20.04 and it was cross-platform which can work properly under other operating systems, such as Microsoft® Windows. The digital camera was from Hikvision (MV-CH120-10UM) with 1200 M Pixel. The

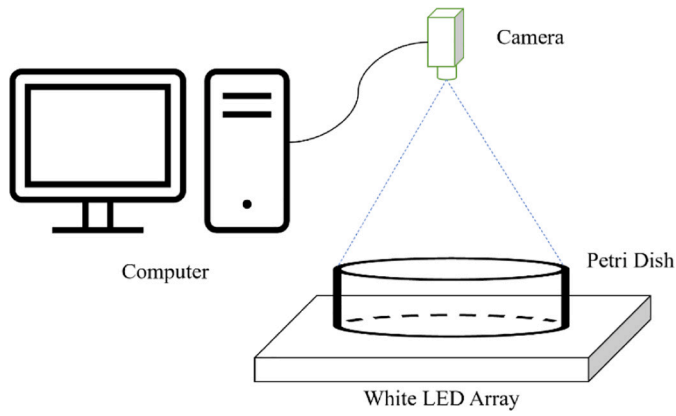


Fig. 4. Structure of Observation System.

Table 1
Details of Experimental Utilization.

Observation facilities		Computer configuration		Software configuration
Camera	HikVision (MV-CH120-10GU)	CPU	AMD 3800× / 4.1GHz	Python 3.8
Light Source	20 cm*20 cm LED panel with white light	Memory	16G / 3200 MHz	OPENCV 4.2
Observation Arena	9 cm glass Petri dish	Hard Disk	Seagate 1 T Hard Drive	Ubuntu 20.04

image resolution of the camera was 3000*4000 pixels. The observation area was limited to the petri dish region which was cropped by setting ROI of the camera. The image resolution of the observation area was 2232*2232 pixels. To reduce the computational cost the video clip was resized to 558*558 pixels.

In this work, three video clips were chosen to evaluate the propose tracking scheme. These video clips were named as D1, D2 and D3. The individual number appeared in D1 was 13 and 14 *Daphnia* was recorded in D2 and D3. The water depth was set to 5 mm, 10 mm and 15 mm. The length of the videos was approximate 30 s. An example of the *Daphnia* image is given in Fig. 5a. To demonstrate the robustness of our proposed tracking scheme additional 5 video clips from the internet were included in our tests. These 5 video clips were captured in transparent flow-through cell with backlighting. The image resolution was 640*480 but the image clarity was much poor than our data (Fig. 5b). These video clips were named as D1 to D5 and 5 individuals

were presented in each video clip. The details and analysis of the datasets are given in Table 2. In addition to the basic information (e.g., frame rate, length and resolution), the occlusion level was examined to investigate the influence of occlusion on tracking performance. Occlusion proportion and occlusion events were calculated in Table 2. Occlusion proportion is defined as the proportion of the frames presenting occlusions among all frames in a video clip. Occlusion proportion indicates the length of occlusions while occlusion events is the number of occlusions.

3.2. Evaluation metrics

To evaluate the performance of our proposed tracking scheme precision-recall analysis and correction identification ratio (CIR) (Wang et al., 2016) was utilized to quantify the tracking accuracy and stability. Precision-recall analysis is widely used in evaluating object recognition and object detection tasks (Powers, 2011). In this work, precision and recall are applied to evaluate the detection accuracy of individual *Daphnia*. Precision is the proportion of correct detections among all the detected results (Eq. (18)). Precision represents the accuracy of detected results and reflects the fluence of false alarms. While recall calculates the proportion of correct detections among all the *Daphnia* individuals. Recall presents the detection ability on target objects (Eq. (19)).

$$Precision = \frac{TP}{TP + FP} \quad (18)$$

$$Recall = \frac{TP}{TP + FN} \quad (19)$$

True positive (TP) is defined as the correct detection of *Daphnia* individuals. False positive (FP) is the Non-*Daphnia* objects are incorrectly *Daphnia* individuals. Usually, false positives are from the background image noises or mirrored image on the container. False Negative (FN) is the number of missed *Daphnia* individual in the detection process.

The tracking accuracy was evaluated by Frequency of Incorrect Tracking without Occlusions (FITO) and CIR. Occlusion is one of the most concerns in multiple object tracking which is the significant disturbance in decreasing the tracking accuracy. In this work, our tracking scheme is evaluated from two aspects: tracking without occlusions and tracking under occlusions. Here, frequency of incorrect tracking without occlusions (FITO) is proposed to evaluate the performance of tracker without considering the influence by occlusion. FITO calculates the total number of tracking error (e.g., trajectory switch) in all frames without occlusions. On the other side, CIR presents the proportion of the correct identification of all individuals after occlusion. The calculation of CIR is given in Eq. (20):

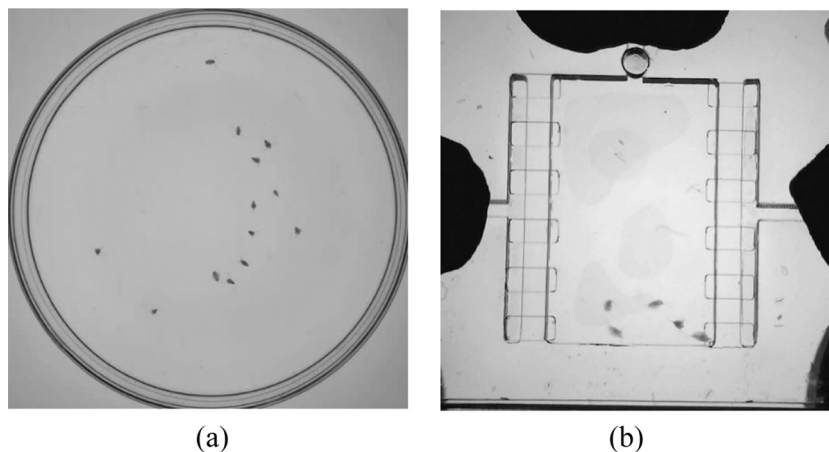
Fig. 5. Example of *Daphnia* Data Sets: (a) our observed video, (b) internet downloaded video.

Table 2
Detailed Information of *Daphnia* Data Sets.

	Individuals	Total Frame	Frame Rate	Resolution	Occlusion Proportion (%)	Occlusion Events
D1	13	567	20	558*558	10.93	19
D2	14	426	20	558*558	23.94	32
D3	14	390	20	558*558	26.67	32
D4	5	958	14	640*480	25.99	58
D5	5	961	14	640*480	8.32	16
D6	5	961	14	640*480	9.57	19
D7	5	962	14	640*480	9.98	23
D8	5	962	14	640*480	7.48	18

$$CIR = \frac{T^{Occ}}{N} \quad (20)$$

where T^{Occ} represents the number of individuals correctly identified after occlusions and N is the total number of occlusion events. CIR represents the stability of the tracking scheme against occlusions. High CIR value indicates the tracking scheme is robust to occlusion.

4. Results and discussions

4.1. Individual detection accuracy

Detection accuracy of *Daphnia* individuals were initially evaluated by precision-recall analysis. The analysis results are presented in Table 3. In the tests, false positives were mainly caused by mirrored images on the inner side of the container. An ROI was applied to reduce the image noises and improve the tracking accuracy. Therefore, precision with / without ROI were separately calculated to demonstrate the effectiveness of setting ROI. Table 3 shows that precision values without ROI ranged from 76.74% to 98.35% in all datasets while precisions were exceeded 99.96% with ROI setting. All false positives were eliminated in our recorded datasets (D1, D2 and D3). These results prove that the ROI can effectively reduce the frequency of false positives. Recall values are in the range of 94.43% - 99.31%. Table 3 shows that recall is highly related to the occlusion event. Since *Daphnia* individuals were attached together during occlusion and only one blob could be extracted from the background. The detected number of individuals was less than the expected number of targets that false negatives (miss-detection) were occurred in this situation. Therefore, the recall values were decreased by occlusions.

4.2. Performance evaluation of multiple individual tracking

Tracking performance was evaluated from two aspects: tracking without occlusions and tracking with occlusions. The tracking performance was quantified by calculating FITO and CIR as described in Section 2.3. FITO indicates the number incorrect tracking without occlusions, i.e., the switch of trajectories of two individuals. CIR is the tracking accuracy after occlusion which represents the stability of tracker under occlusions. Tracking performance is given in Table 4. By our proposed tracking scheme, 3 videos had no tracking error without occlusions and 4 videos had only one trajectory switch. In video D8, 3 tracking errors were observed. On the other side, CIR values ranged in 77.78% to 94.74% in all datasets. The tracking results were most stable in D1, D5 and D6 which CIR were 94.74%, 93.75% and 94.74%. The

proposed tracker presented lowest tracking accuracy in D8 that the CIR was only 77.78%.

To demonstrate the advantages of our proposed tracking scheme a comparison test was conducted by one of the well-known tracking methods, Simple Online and Realtime Tracking (SORT). The original SORT was proposed to solve the data association problem by Kalman filter and Hungarian algorithm (Kalman, 1960; Kuhn, 1955). The cost matrix is constructed by calculating the intersection-over-union (IOU) between each detected candidate and all predicted bounding boxes of the existing targets. In the original SORT, a pre-trained object detector was employed to obtain the individual positions and bounding boxes (Bewley et al., 2016). In this work, SORT was combined with our individual detector for conducting the *Daphnia* tracking. The *Daphnia* detection results were given as the input of SORT and the parameters were tuned to optimize the tracking performance in the group tracking of *Daphnia*. The tracking performance of SORT is given in Table 4. Two videos had no tracking errors, 3 videos showed only 1 tracking errors and 3 videos had 2 tracking errors. While the CIR of SORT were presented in 51.72% - 78.95% which was significantly reduced by occlusions.

The comparison results show that the proposed tracking scheme and SORT are comparable in tracking *Daphnia* individuals without occlusions. It is notable that our proposed method was outperformed SORT under occlusions since appearance feature was considered in our method. The average CIR of the proposed method was 87.68% while the CIR was only 64.80% for SORT (Table 4). On average, our method had 22.88% of success rate more than SORT in dealing with tracking multiple individuals under occlusions. Especially in D4, the CIR value of our method was 34.49% higher than SORT. The experimental results demonstrated that our proposed method was stable and robustness in tracking group behavior of *Daphnia*.

The main advantage of the proposed tracking scheme is that appearance of *Daphnia* was utilized to identify of individuals from occlusions and integrated to data association. Optical flow was employed to estimate the individual movement according to their appearance. Optical flow provided motion estimation of each individual to ensure high accuracy of individual assignment in the tracking process. It is worth to note that optical flow could effectively match the appearance feature from the images of attached or occluded *Daphnia* individuals. Consequently, the proposed method could separate each individual from occlusions. Identifying individuals from occlusion and estimating their movement direction could significantly improve the tracking accuracy and stability. The process of tracking individuals under occlusion is illustrated in Fig. 6.

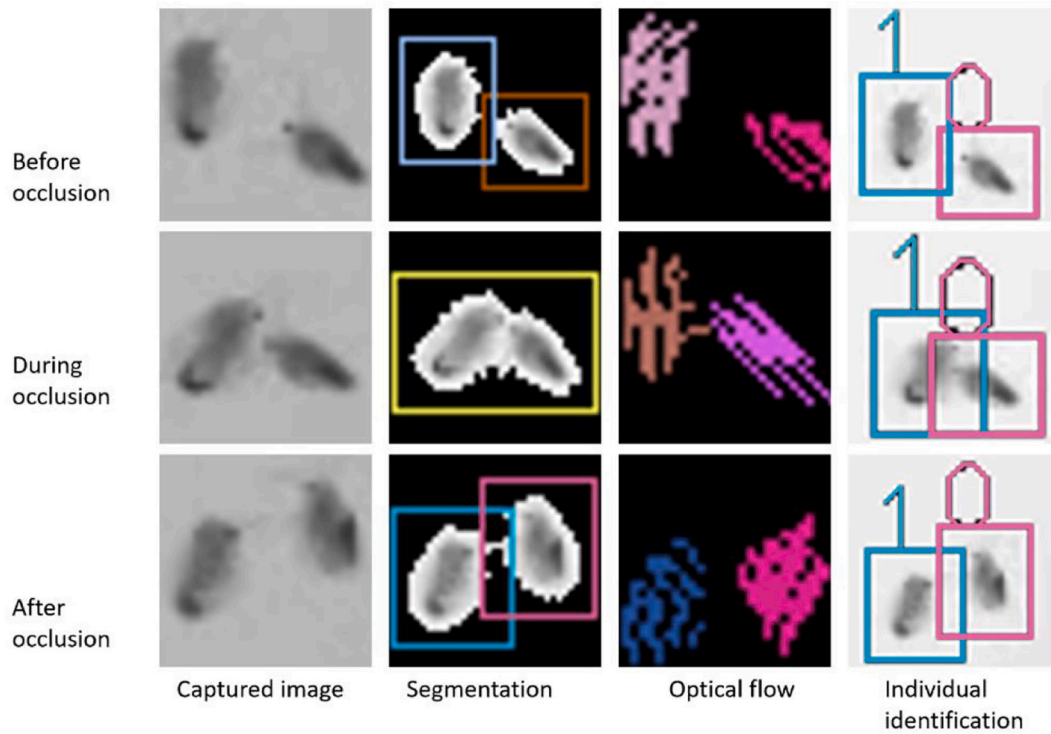
Table 3
Performance Evaluation of *Daphnia* Individual Detection.

	D1	D2	D3	D4	D5	D6	D7	D8
Occlusion proportion (%)	10.93	23.94	26.67	25.99	8.32	9.57	9.98	7.48
Occlusion event	19	32	32	58	16	19	23	18
Precision without ROI (%)	98.35	96.48	95.99	76.74	97.46	96.84	96.75	97.83
Precision with ROI (%)	100	100	100	99.98	99.96	99.98	99.81	99.92
Recall (%)	99.31	98.29	97.77	94.43	98.42	98.19	97.84	98.54

Table 4

Tracking Performance by Proposed Method and SORT.

		D1	D2	D3	D4	D5	D6	D7	D8	Average
FITO	SORT	1	0	1	2	0	1	2	2	1.13
	Proposed	0	0	1	1	1	0	1	3	0.88
CIR (%)	SORT	78.95	59.38	57.14	51.72	68.75	73.68	56.52	72.22	64.80
	Proposed	94.74	84.38	82.86	86.21	93.75	94.74	86.96	77.78	87.68

**Fig. 6.** Individual Tracking under Occlusion.

The three stages of individual tracking: before occlusion, during occlusion and after occlusion, are presented in each row in Fig. 6. And each column shows the visualized results of individual tracking, segmentation, optical flow and individual identification. The first column presets the tracking results. Each individual is numbered and marked by a bounding box in different colors. The second column is the segmented individuals. The results of optical flow and individual identification are followed in the third and fourth column. Although individuals are extracted as one blob during occlusions the optical flows show different direction on each individual in the attached image. Consequently, the two individuals could be effectively separated according to their optical flow field.

In our tests, optical flow was accurate and stable in matching the individuals across frames. However, there are still some cases that optical flow could lost the targets. For example, if a *Daphnia* has a sudden long jump the optical flow might fail to match the appearance. To solve this problem, NCC was applied to calculate the matching score of the same target across frames and ensuring the movement tracks could be correctly assigned to each individual when optical flow missed targets. On the contrary, SORT relied on the predicted position and IOU of targets in the consecutive frames. It is difficult to distinguish individuals by prediction and IOU. Therefore, SORT is vulnerable in tracking occluded targets.

The proposed tracking scheme has some limitations that could reduce the tracking accuracy. Long-time occlusion could lead individual separation failure by optical flow. Since in the long-time occlusion individuals probably attached together and moving towards to the same direction. In this case, the optical flow of the attached individual has

similar direction that individuals could not be extracted according to the estimated directions. It is still a challenge to segment individual from long-time occlusions. Without individual identification the movement tracks have high probability to be assigned to the incorrect individual. Blurred image is another issue of decreasing the tracking accuracy. Appearance feature could not be accurately presented in blurred image. Subsequently, appearance feature descriptor, e.g., optical flow or NCC, would be weak in dealing with the blurred images. The appearance descriptor would be failed either when an individual is completely occluded. Moreover, uniformly illumination is required to ensure the effectiveness of background subtraction for individual detection. In the further work, more robustness appearance descriptors such as convolutional feature should be considered to improve the ability of individual identification (Xu and Cheng, 2017). And learning unique biometric feature could greatly ensure the robustness of multiple individual tracking (Bruslund Haurum et al., 2020; Pérez-Escudero et al., 2014). To deal with the varying of aquatic biological organisms modeling and measuring individual shape or pose could provide more precise information for matching individual from occlusions (Xia et al., 2014). Furthermore, trajectory re-link could be effective to verify and match the correct individual trajectories after occlusions (Bai et al., 2018; Qian et al., 2014).

5. Conclusions

This work proposed an effective method for detecting and tracking a group of *Daphnia* individuals. A motion estimation and appearance

model were to solve the multiple individual tracking problem. Individual identification from occlusions was implemented by analyzing optical flow field. It could significantly reduce the tracking error caused by partial occlusion and prevents tracking of false positives. Our method was effective to track multiple individuals of *Daphnia* by integrating optical flow and appearance similarity matching. By taking full advantage of the motion and appearance information of *Daphnia* the tracker was able to accurately follow each individual and could cope with their sudden jump or fast motion. The tracking accuracy and stability of the proposed method was proved by experiments on maximal 14 individuals. Experimental results showed that the proposed method had 22.88% of success rate more than one of the well-known multiple objects tracking method. The proposed method could be an efficient tool for studying group of *Daphnia magna* and this method could also be applied to other types of biological individuals with clear appearance features, such as fish.

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.

Acknowledgments

This study was supported by Key Research and Development Program of Yantai (2018YT06000808), CAS Key Technology Talent Program and the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. 2019R1A2C2088859).

References

- Bae, M.J., Park, Y.S., 2014. Biological early warning system based on the responses of aquatic organisms to disturbances: a review. *Sci. Total Environ.* 466, 635–649.
- Bai, Y.X., Zhang, S.H., Fan, Z., Liu, X.Y., Zhao, X., Feng, X.Z., Sun, M.Z., 2018. Automatic multiple zebrafish tracking based on improved HOG features. *Sci. Rep.* 8 (1), 1–14.
- Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016, September. Simple online and realtime tracking. In: 2016 IEEE International Conference on Image Processing (ICIP). IEEE, pp. 3464–3468.
- Bownik, A., 2017. *Daphnia* swimming behaviour as a biomarker in toxicity assessment: a review. *Sci. Total Environ.* 601, 194–205.
- Bruslund Haurum, J., Karpova, A., Pedersen, M., Hein Bengtson, S., Moeslund, T.B., 2020. Re-identification of zebrafish using metric learning. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision Workshops, pp. 1–11.
- Cano, A.M., Maul, J.D., Saed, M., Shah, S.A., Green, M.J., Cañas-Carrell, J.E., 2017. Bioaccumulation, stress, and swimming impairment in *Daphnia magna* exposed to multiwalled carbon nanotubes, graphene, and graphene oxide. *Environ. Toxicol. Chem.* 36 (8), 2199–2204.
- Farneback, G., 2003, June. Two-frame motion estimation based on polynomial expansion. In: Scandinavian Conference on Image Analysis. Springer, Berlin, Heidelberg, pp. 363–370.
- Ferrão-Filho, A.S., Soares, M.C.S., Lima, R.S., Magalhães, V.F., 2014. Effects of *Cylindrospermopsis raciborskii* (cyanobacteria) on the swimming behavior of *Daphnia* (cladocera). *Environ. Toxicol. Chem.* 33 (1), 223–229.
- Häder, D.P., Erzinger, G.S., 2017. Daphniatox—online monitoring of aquatic pollution and toxic substances. *Chemosphere* 167, 228–235.
- Hansen, L.R., Roslev, P., 2016. Behavioral responses of juvenile *Daphnia magna* after exposure to glyphosate and glyphosate-copper complexes. *Aquat. Toxicol.* 179, 36–43.
- Kalal, Z., Mikolajczyk, K., Matas, J., 2010, August. Forward-backward error: Automatic detection of tracking failures. In: 2010 20th International Conference on Pattern Recognition. IEEE, pp. 2756–2759.
- Kalal, Z., Mikolajczyk, K., Matas, J., 2011. Tracking-learning-detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (7), 1409–1422.
- Kalman, R.E., 1960. A New Approach to Linear Filtering and Prediction Problems.
- Kuhn, H.W., 1955. The Hungarian method for the assignment problem. *Naval Res. Logist. Q.* 2 (1–2), 83–97.
- Lovern, S.B., Strickler, J.R., Klaper, R., 2007. Behavioral and physiological changes in *Daphnia magna* when exposed to nanoparticle suspensions (titanium dioxide, nano-C60, and C60HxC70Hx). *Environ. Sci. Technol.* 41 (12), 4465–4470.
- Moldaenke, C., 2000. Monitoring of surface water by ultrasensitive *Daphnia* toxiometer. *Environ. Toxicol.* 15 (5), 390–400.
- Noss, C., Dabrunz, A., Rosenfeldt, R.R., Lorke, A., Schulz, R., 2013. Three-dimensional analysis of the swimming behavior of *Daphnia magna* exposed to nanosized titanium dioxide. *PLoS One* 8 (11), e80960.
- Pérez-Escudero, A., Vicente-Page, J., Hinz, R.C., Arganda, S., De Polavieja, G.G., 2014. idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nat. Methods* 11 (7), 743–748.
- Powers, D.M., 2011. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation.
- Qian, Z.M., Cheng, X.E., Chen, Y.Q., 2014. Automatically detect and track multiple fish swimming in shallow water with frequent occlusion. *PLoS One* 9 (9), e106506.
- Qian, Z.M., Wang, S.H., Cheng, X.E., Chen, Y.Q., 2016. An effective and robust method for tracking multiple fish in video image based on fish head detection. *BMC Bioinf.* 17 (1), 251.
- Rodríguez, A., Zhang, H., Klaminder, J., Brodin, T., Andersson, M., 2017. ToxId: an efficient algorithm to solve occlusions when tracking multiple animals. *Sci. Rep.* 7 (1), 1–8.
- Sauvola, J., Pietikäinen, M., 2000. Adaptive document image binarization. *Pattern Recogn.* 33 (2), 225–236.
- Stanley, J.K., Laird, J.G., Kennedy, A.J., Steevens, J.A., 2016. Sublethal effects of multiwalled carbon nanotube exposure in the invertebrate *Daphnia magna*. *Environ. Toxicol. Chem.* 35 (1), 200–204.
- Wang, S.H., Cheng, X.E., Qian, Z.M., Liu, Y., Chen, Y.Q., 2016. Automated planar tracking the waving bodies of multiple zebrafish swimming in shallow water. *PLoS One* 11 (4), e0154714.
- Xia, C., Li, Y., Lee, J.M., 2014, December. A visual measurement of fish locomotion based on deformable models. In: International Conference on Intelligent Robotics and Applications. Springer, Cham, pp. 110–116.
- Xia, C., Chon, T.S., Liu, Y., Chi, J., Lee, J., 2016. Posture tracking of multiple individual fish for behavioral monitoring with visual sensors. *Ecol. Inf.* 36, 190–198.
- Xu, Z., Cheng, X.E., 2017. Zebrafish tracking using convolutional neural networks. *Sci. Rep.* 7, 42815.
- Zein, M.A., McElmurry, S.P., Kashian, D.R., Savolainen, P.T., Pitts, D.K., 2015. Toxic effects of combined stressors on *Daphnia pulex*: interactions between diazinon, 4-nonylphenol, and wastewater effluent. *Environ. Toxicol. Chem.* 34 (5), 1145–1153.