

Paper:

Pose Estimation of Swimming Fish Using NACA Airfoil Model for Collective Behavior Analysis

Hitoshi Habe*, Yoshiki Takeuchi*, Kei Terayama**, and Masa-aki Sakagami***

*Kindai University

3-4-1 Kowakae, Higashi-osaka, Osaka 577-8502, Japan

E-mail: habe@kindai.ac.jp

**Yokohama City University

1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

E-mail: terayama@yokohama-cu.ac.jp

***Kyoto University

Yosida Nihonmatsu-cho, Sakyo-ku, Kyoto 606-8316, Japan

E-mail: sakagami.masaaki.6x@kyoto-u.ac.jp

[Received January 4, 2021; accepted March 17, 2021]

We propose a pose estimation method using a National Advisory Committee for Aeronautics (NACA) airfoil model for fish schools. This method allows one to understand the state in which fish are swimming based on their posture and dynamic variations. Moreover, their collective behavior can be understood based on their posture changes. Therefore, fish pose is a crucial indicator for collective behavior analysis. We use the NACA model to represent the fish posture; this enables more accurate tracking and movement prediction owing to the capability of the model in describing posture dynamics. To fit the model to video data, we first adopt the DeepLabCut toolbox to detect body parts (i.e., head, center, and tail fin) in an image sequence. Subsequently, we apply a particle filter to fit a set of parameters from the NACA model. The results from DeepLabCut, i.e., three points on a fish body, are used to adjust the components of the state vector. This enables more reliable estimation results to be obtained when the speed and direction of the fish change abruptly. Experimental results using both simulation data and real video data demonstrate that the proposed method provides good results, including when rapid changes occur in the swimming direction.

Keywords: swimming fish, tracking, pose estimation, collective behavior analysis

1. Introduction

The behavior of fish in schools is yet to be elucidated completely [1–5]. For example, the location in which individual fish swim in a torus-like school or their communication mechanism is yet to be clarified. Fish positions and movement trajectories provide essential data for understanding these mechanisms. From a practical perspective, movement information is useful for monitoring fish

growth in aquariums and aquaculture sites. However, observing a school of fish swimming in water by human eyes is unrealistic owing to the significant amount of time required for the task. Therefore, researchers have proposed methods that can automatically acquire fish movements using video processing [6, 7].

Herein, we propose a method for estimating the position and pose of swimming fish in a school. The pose of swimming fish can provide important information for understanding their ecology as well as for the state in which the fish are swimming. Moreover, their collective behavior can be understood based on their posture changes. Therefore, fish pose is a crucial indicator for collective behavior analysis.

We used the National Advisory Committee for Aeronautics (NACA) model [8] to represent the posture of the fish; this enables more accurate tracking and movement prediction because the NACA model can represent time-series changes of the pose. To fit the NACA model to the fish shown in video data, we used DeepLabCut (DLC) [9] and a particle filter (PF) [10]. In DLC, which is widely used in ecology, position information can be obtained accurately from a video by selecting several parts of the target object. A PF was used to predict the parameters of the NACA model. The location information obtained by the DLC was used in the likelihood calculation of the prediction. Combining these two methods enables us to stably obtain the location and trajectory of the fish, including when the swimming direction and speed change rapidly. The proposed method was applied to silhouette images of fish captured from the top or bottom of a water tank, as shown in **Fig. 1**. This paper is organized as follows. Related studies are introduced in Section 2, and our tracking and pose estimation method for multiple fish is described in Section 3. Our experimental results obtained using simulation data and videos recorded in an aquarium are provided in Section 4. Finally, we summarize this paper and present our plans for future studies in Section 5.



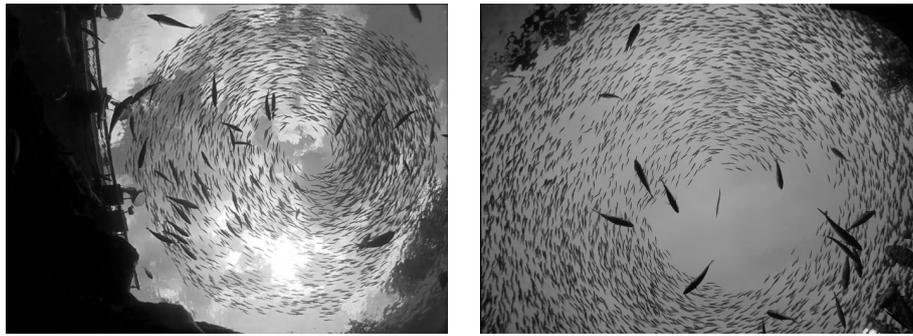


Fig. 1. Fish school images captured from bottom of the aquarium.

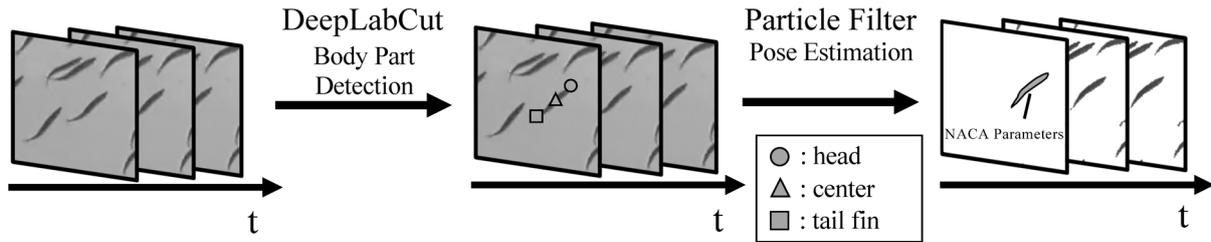


Fig. 2. Overview of proposed method.

2. Related Studies

Many multitarget tracking methods have been proposed, most of which are intended for tracking humans [11]. Additionally, methods for tracking multiple fish in a shallow tank have been developed [6, 12]. Recently, tools have been developed to easily acquire the location information of target animals more easily [13]. For underwater environments, Okuda et al. proposed a method for detecting and tracking marine life in images captured using movable sensor nodes [14]. The above-mentioned methods enable numerous animal movement trajectories to be obtained more easily, thereby contributing to advancements in the field of ecology.

After the location information is acquired, pose is to be estimated to reveal more detailed information regarding a target person or animal. Human pose estimation has been an active research area [15, 16], and the results of pose estimation can be applied to sports motion analysis and motion recognition.

Furthermore, pose estimation is used to understand the ecology of animals. DLC [9] is an efficient tool for the three-dimensional (3D) markerless pose estimation of animals based on transfer learning using deep neural networks. Although it has been used for various types of animals, few studies have focused on the change in posture of fish or introduced a parametric model to represent such changes, unlike the current study.

Terayama et al. tracked multiple fish in a dense school using their appearance model based on fish images in a video [7]. The appearance model was parameterized, which enabled future motions to be estimated and unusual behaviors to be detected. However, their algorithm is relatively slow because the optimization process was based

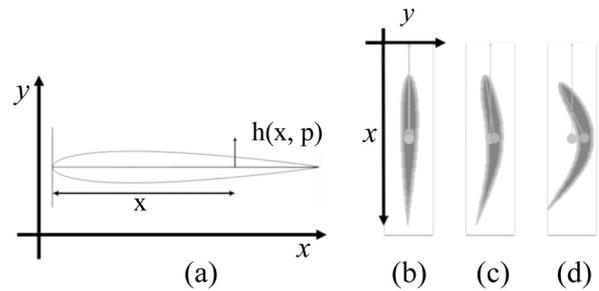


Fig. 3. (a) NACA0012 airfoil model. (b) Fish shadow model when $A = 0.01, p = 0$; (c) $A = 0.1, p = 0$; (d) $A = 0.33, p = 0$.

on simulated annealing [17], which is a severe problem for practical applications.

The method proposed herein uses a parameterized appearance model to represent the pose and dynamics of fish while a PF is applied to fit the parameters. If a PF estimates all parameters, it may not be able to manage rapid changes in the swimming direction. Hence, we first obtained the location information using DeepLabCut and applied a PF to estimate the time-series parameters representing the dynamics of the fish postures. This allows us to achieve a stable posture estimation.

3. Proposed Framework

Figure 2 shows an overview of the proposed method. In this study, we adopted the NACA0012 airfoil model to represent the appearance of a fish (Fig. 3). As shown in

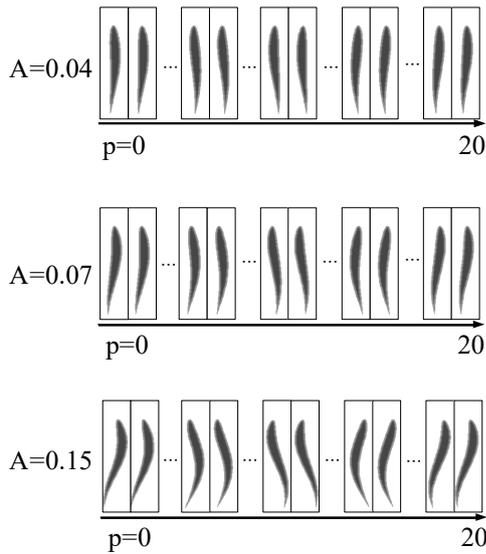


Fig. 4. Time-varying fish silhouette model.

Fig. 4, this model represents the silhouette of a swimming fish, and changes in the silhouette during swimming can be represented by changes in the parameters.

3.1. Appearance Model of Fish

To fit the NACA0012 model to a swimming fish video, we used DLC and a PF. DLC is a tool that is widely used in the field of ecology to acquire the motion information of a target animal from video images. First, we provided the positional information of multiple parts of the target animal for certain images. Subsequently, the locations of the parts in the remaining images were accurately estimated using machine learning.

Because DLC only estimates the location of the target objects, additional processing is required to obtain the parameters representing the time-series changes in the NACA0012 model. Therefore, we used a PF, which is widely used to predict the time series of multiple variables. Additionally, the location information obtained using DLC was used to calculate the likelihood of the estimated variables.

We employed the NACA0012 airfoil model as a basis for a deformable appearance model of fish. Fig. 3(a) shows the NACA0012 model. As in [7], we define the deformation equation $h(x, p)$ around the initial center line in Fig. 3(a) as

$$h(x, p) = A \left(-(x-1)^2 + 1 \right) \cos \frac{2\pi}{\lambda} (x - v \cdot p), \quad (1)$$

where the parameters A , λ , v , x , and p represent the maximum amplitude, wave length, phase velocity, and position from the head as shown in Fig. 3, and the phase of a single beat cycle, respectively. For each scene, we first calculated the average brightness of the fish and constructed 92 standard appearance models based on the NACA0012 model and Eq. (1), where we changed A and p by filling

in the form of the brightness. We refer to these models as the NACA model. Figs. 3(b) and (c) show the examples of the NACA model. We set both λ and v to 2, and changed A from 0.01 to 0.3 and p from 0 to 1. In Fig. 4, A indicates the deformation extent of a fish, and the beating of a fish while it is swimming is represented by the changing p . In our pose estimation method, we estimated A and p , whereas λ and v were fixed to predetermined values. To manage a large deformation (bending), we added a few significantly deformed models based on $h(x, p)$ with a large amplitude. Fig. 3(d) shows an example of a significantly deformed model.

3.2. DeepLabCut (DLC)

To represent the posture change of a swimming fish in a video using the NACA model, we must first identify the position of the fish, followed by the appropriate parameters. It is extremely challenging to accurately determine the location of fish swimming in a dense school (see Fig. 1); therefore, we used DLC.

DLC is a widely used tool in the field of ecology to analyze the behavior of target animals. First, the user manually specifies the positions of the animal's legs, head, and other parts of the body in several images from a video of the target animal. Subsequently, the system builds a model to detect them through machine learning. By applying the model to the remaining images, the system can detect the body part positions of the animal.

Our preliminary experiments show that DLC can yield sufficient detection accuracy; hence, we used it as basic data to estimate the parameters of the NACA model. To obtain the parameters of the NACA model, not only the position of the individual but also its swimming direction must be determined. Hence, we tracked the head, body center, and tail fin of a single fish, as shown in Fig. 5.

For training DLC, these three points were specified manually by the annotators. The center of the body was set to be equidistant from the head and the tail fin. Although the center point did not exhibit a clear image feature compared with the other two points, it can still be tracked with sufficient accuracy using DLC (as will be presented in Section 4.2.1).

In Section 4, we present the tracking performance results of DLC obtained from detailed evaluation experiments.

3.3. Particle Filter (PF)

Because the NACA model represents the time-series variations in the appearance of fish, we used a PF to estimate its parameters. PFs are widely used in time-series filtering and can estimate the state vector accurately, including noisy or missing observations. In this framework, we define the state vector $\mathbf{x} = (A, p, \mathbf{c}, \theta)$, where A and p are the maximum amplitude and phase of one beat cycle in the NACA model described in Section 3.1, respectively. The swimming direction θ and center position \mathbf{c} were obtained from three points detected using DLC, as shown in Fig. 6. We calculated θ from the estimated positions

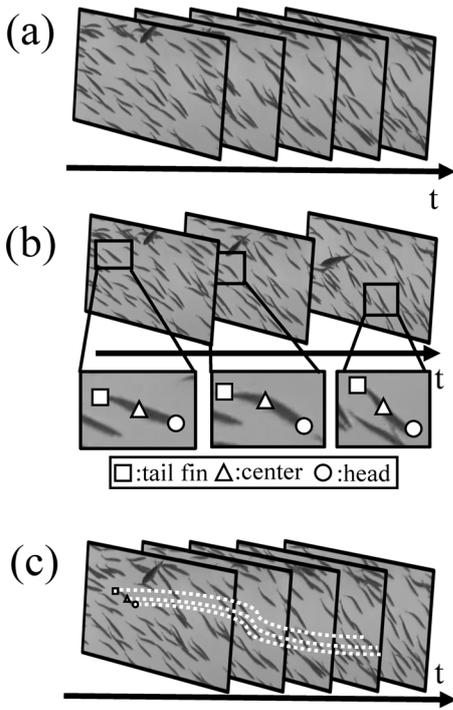


Fig. 5. Body part detection using DeepLabCut. (a) Original video. (b) Labeling of the head, center, and tail fin. (c) Learning and detection.

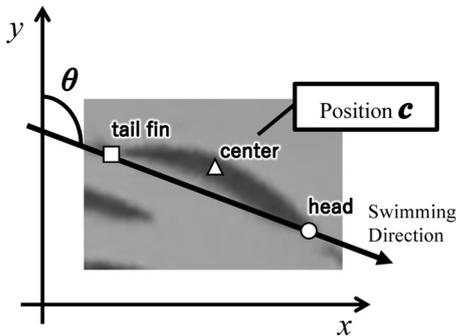


Fig. 6. Fish position and swimming direction obtained from DLC detection results.

of the head and tail fin of the swimming fish, which corresponded to the direction of the line passing through the two points. The center of the fish body is denoted as \mathbf{c} . Finally, we set the observation vector \mathbf{y} be I in the binarized observed image, which is used to calculate the likelihood of the state vector.

The flow of the PF process was as follows: (a)–(e) correspond to the symbols shown in **Fig. 7**.

- (a) Determine the initial state vector at the beginning of tracking. For (\mathbf{c}, θ) , we used detection results from DLC.
- (b) Particles were generated. As in the standard definition of a PF, each particle represents one possible value of the state vector $\mathbf{x}_t^{(i)} = (A_t^{(i)}, p_t^{(i)}, \mathbf{c}_t^{(i)}, \theta_t^{(i)})$.

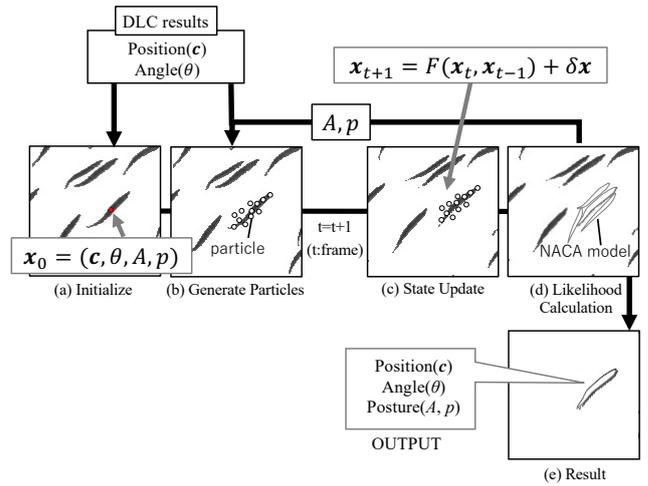


Fig. 7. Pose estimation using particle filter.

- (c) Update the state vector for each particle, $\mathbf{x}_{t+1}^{(i)} = F(\mathbf{x}_t^{(i)}, \mathbf{x}_{t-1}^{(i)}) + \delta \mathbf{x}$, where the function F represents the dynamics of the NACA model, and $\delta \mathbf{x}$ denotes Gaussian noise. In our implementation, F is simply defined as follows. The center point \mathbf{c} is assumed to move in a linear motion of constant speed and is updated as $\mathbf{c}_{t+1}^{(i)} = 2\mathbf{c}_t^{(i)} - \mathbf{c}_{t-1}^{(i)} + \delta \mathbf{c}$, where $\delta \mathbf{c}$ represents Gaussian noise. For the other components, $(A_t^{(i)}, p_t^{(i)}, \theta_t^{(i)})$, Gaussian noise is added to $(A_{t+1}^{(i)}, p_{t+1}^{(i)}, \theta_{t+1}^{(i)}) = (A_t^{(i)}, p_t^{(i)}, \theta_t^{(i)}) + \delta$ in the update process.

For updating (\mathbf{c}, θ) , we used the DLC-derived $\mathbf{c}_t^{(i)}$, $\mathbf{c}_{t-1}^{(i)}$, and $\theta_t^{(i)}$ instead of the results of the PF in the previous frames.

- (d) Calculate the likelihood of each particle. We computed the difference between the binarized observed image and the silhouette generated using Eq. (1).

The implementation above differs slightly from the typically used PF. The value of (\mathbf{c}, θ) is updated based on the observation from the DLC; therefore, reliable results are yielded.

4. Experiments

To demonstrate the effectiveness of our method, we conducted experiments using the following two types of datasets:

- A set of images generated by manually setting the parameters of the NACA model (Experiment 1).
- Real fish school images captured from an aquarium (Experiment 2).

The former evaluates the performance of the proposed method when the ground truth of all the parameters of

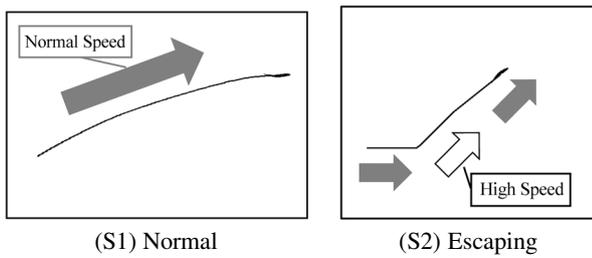


Fig. 8. Simulation images for Experiment 1.

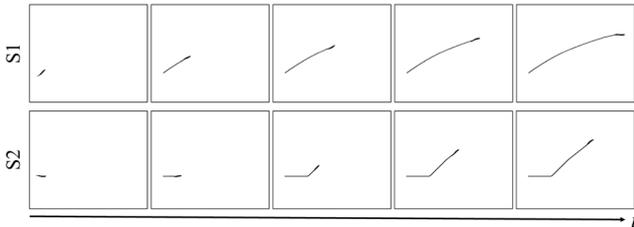


Fig. 9. Sample snapshots of dataset for Experiment 1.

the NACA model is known. The latter is used to confirm the effectiveness of the proposed method for actual swimming scenes. Because the ground truths of all the NACA model parameters were not provided, we conducted a quantitative evaluation for some of the parameters in addition to a qualitative review.

4.1. Experiment 1: Simulation Data

To conduct a quantitative evaluation, we prepared simulation images for two situations, as shown in Fig. 8. S1 simulates regular swimming, where a fish swims in a constant direction and speed. S2 simulates a situation in which a fish changes its swimming direction abruptly to escape from an external stimulus. Fig. 9 shows snapshots of S1 and S2. Estimating the position and posture of S2 is more challenging than that of S1.

In the experiment, we compared the following four methods.

- **PF:** A PF was used for all components of $\mathbf{x} = (A, p, \mathbf{c}, \theta)$.
- **DLC:** (\mathbf{c}, θ) was calculated from the output of DLC, as described in Section 3.2.
- **DLC (pos.) w/ PF:** The DLC results were used for the PF process described in Section 3.3, but only the location data \mathbf{c} were used.
- **Proposed:** The proposed method which is described in the previous section was used.

We did not compare this method with Terayama's method [7], which also uses the NACA model. Terayama's method uses the SA to optimize the parameters; therefore, the optimization results are much better, although more time is required. This paper focuses on our evaluation of the methods above, which provides results within a reasonable amount of time. The proposed

Table 1. Average error for data S1.

| Average errors | \mathbf{c} [pixel] | θ [rad] | A [pixel] | p [rad] |
|------------------|----------------------|----------------|--------------|--------------|
| PF | 0.762 | 0.014 | 0.010 | 0.165 |
| DLC | 1.722 | 0.056 | N/A | N/A |
| DLC (pos.) w/ PF | 0.783 | 0.028 | 0.012 | 0.195 |
| Proposed | 0.710 | 0.049 | 0.019 | 0.315 |

Table 2. Average error for data S2.

| Average errors | \mathbf{c} [pixel] | θ [rad] | A [pixel] | p [rad] |
|------------------|----------------------|----------------|--------------|--------------|
| PF | 2.730 | 0.109 | 0.032 | 0.568 |
| DLC | 1.799 | 0.049 | N/A | N/A |
| DLC (pos.) w/ PF | 1.611 | 0.109 | 0.032 | 0.744 |
| Proposed | 0.883 | 0.046 | 0.017 | 0.846 |

method requires approximately 3–4 h to complete computations using a Core i7-5960X CPU and GTX1070Ti GPU.

The results are shown in Tables 1 and 2, where the average errors of the parameters are shown as $\mathbf{x} = (A, p, \mathbf{c}, \theta)$. A and p are not available in the second line because they (the posture information) cannot be obtained from DLC alone.

Dataset S1 was relatively simple and easy to track and estimate. The results presented in Table 1 show that the four methods do not differ significantly. Although the proposed method does not necessarily provide the best results, the difference is negligibly small and does not pose a significant problem.

By contrast, dataset S2 was more challenging to track because of the rapid changes in swimming direction and speed. The table shows a significant difference in position \mathbf{c} in this case because the PF (first row) failed to monitor the rapid changes. By contrast, tracking via DLC yielded relatively stable location information (second row), unlike tracking using the PF. The combination of position information from DLC and time-series filtering by the PF, which is a feature of the proposed method, improved the estimation accuracy. However, among the DLC results, DLC (pos.) w/ PF, which is shown in the third row, did not improve the accuracy sufficiently. When using the three points of the head, body center, and tail fin (proposed method, fourth row), better estimation results were obtained.

Figure 10 shows some snapshots of the results of the PF and proposed method, in addition to the ground truth of the silhouettes. As shown by these results, although the difference between the two results for S1 did not differ significantly, the proposed method generated better silhouettes than the PF for S2, particularly during burst swimming.

These results demonstrate that the proposed method is effective. It is noteworthy that the body length of the target fish was approximately 50–60 pixels. Compared with the body length, the average error in the position was relatively low and reliable for further analysis.

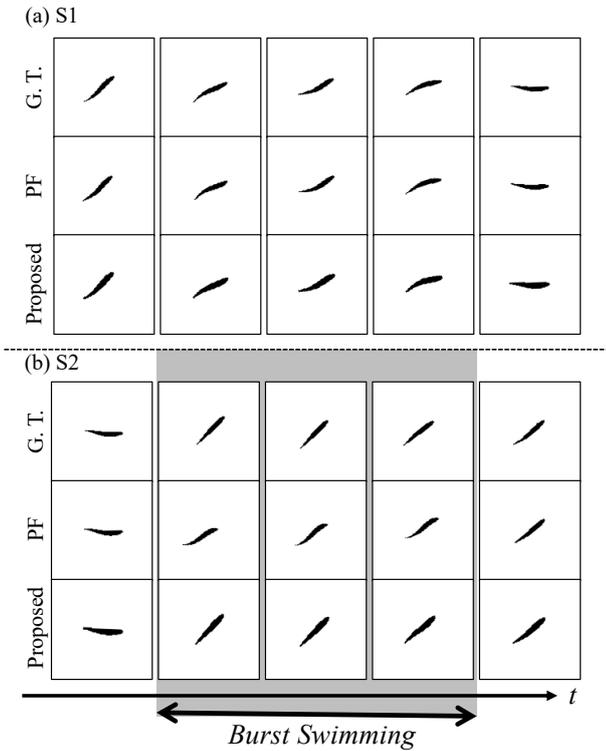


Fig. 10. Experimental results for simulation data. Ground truth, results of proposed method (combination of DLC and PF), and PF are shown. For each of simulation dataset S1 and S2, snapshots of ground truth (upper row), PF results (middle row), and results of proposed method (lower row) are shown. For S2, columns 2–4 correspond to burst swimming, in which fish changes swimming direction and speed.

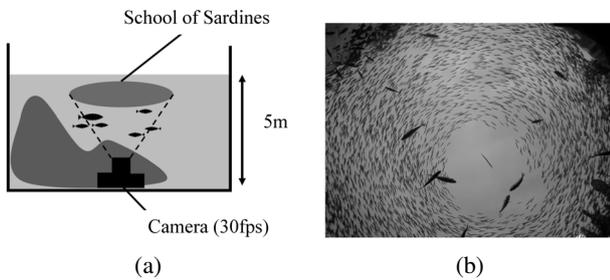


Fig. 11. Dataset for experiments of real scenario: (a) camera setup and (b) snapshot of fish school video.

4.2. Experiment 2: Real Scenario

The effectiveness of the proposed method was analyzed using actual images of fish schools. We recorded videos of schools of sardines (*Sardinops melanostictus*) at Kujukushima Aquarium Umikirara, Nagasaki, Japan. The videos were recorded at 30 fps using a GoPro HERO4 video camera. **Figs. 11(a)** and **(b)** show the camera setup and a snapshot of the video, respectively.

We extracted three video clips from the video as the target of the experiments. We focused on a single fish in the video clip and qualitatively evaluated the tracking results. **Fig. 12** shows snapshots of the three video clips, where the trajectories of the target fish are depicted by

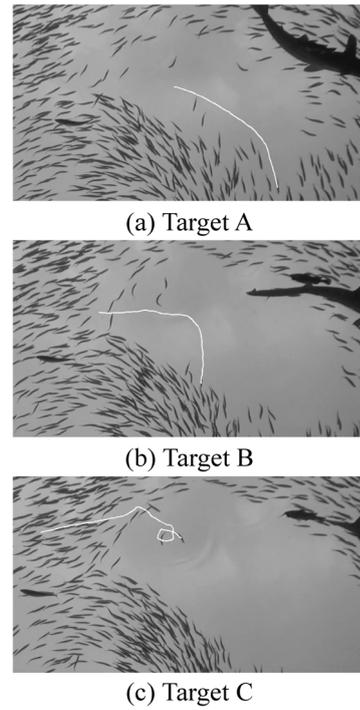


Fig. 12. Snapshots of extracted target videos for Experiment 2. White lines show trajectories of target fish.

Table 3. Average errors of c [pixel] for Targets A–C.

| Estimation method | Target A | Target B | Target C |
|-------------------|--------------|--------------|--------------|
| PF | 2.395 | 8.530 | N/A |
| DLC | 5.540 | 4.031 | 3.526 |
| DLC (pos.) w/ PF | 1.703 | 4.323 | 3.665 |
| Proposed | 1.604 | 2.010 | 3.033 |

white lines. Target A overlapped slightly with other fish and exhibited a simple movement. Similarly, Target B overlapped slightly with the other fish, but it changed its swimming direction in the middle of the trajectory. We expect this motion to be challenging to track. Finally, Target C is more difficult to track because it overlaps with other fish, and it swims in a circle at the end of the trajectory. We evaluated the performance of the proposed method using samples of different difficulty levels.

4.2.1. Quantitative Evaluation

We first conducted a quantitative evaluation on the three datasets. Among the parameters estimated by the proposed method, only the center point c can be manually obtained from the image. Hence, we calculated the error in c between the results estimated using the proposed method and the correct values obtained manually.

Table 3 shows the average errors of c , i.e., the center of the fish body. We compared four methods, as in Experiment 1. The calculated average errors for the four methods are presented in **Table 3**. For Target C, we could not compute the errors of the PF because it failed to track a fish that changed its swimming direction rapidly – the details are shown in **Fig. 13**. These results revealed the same

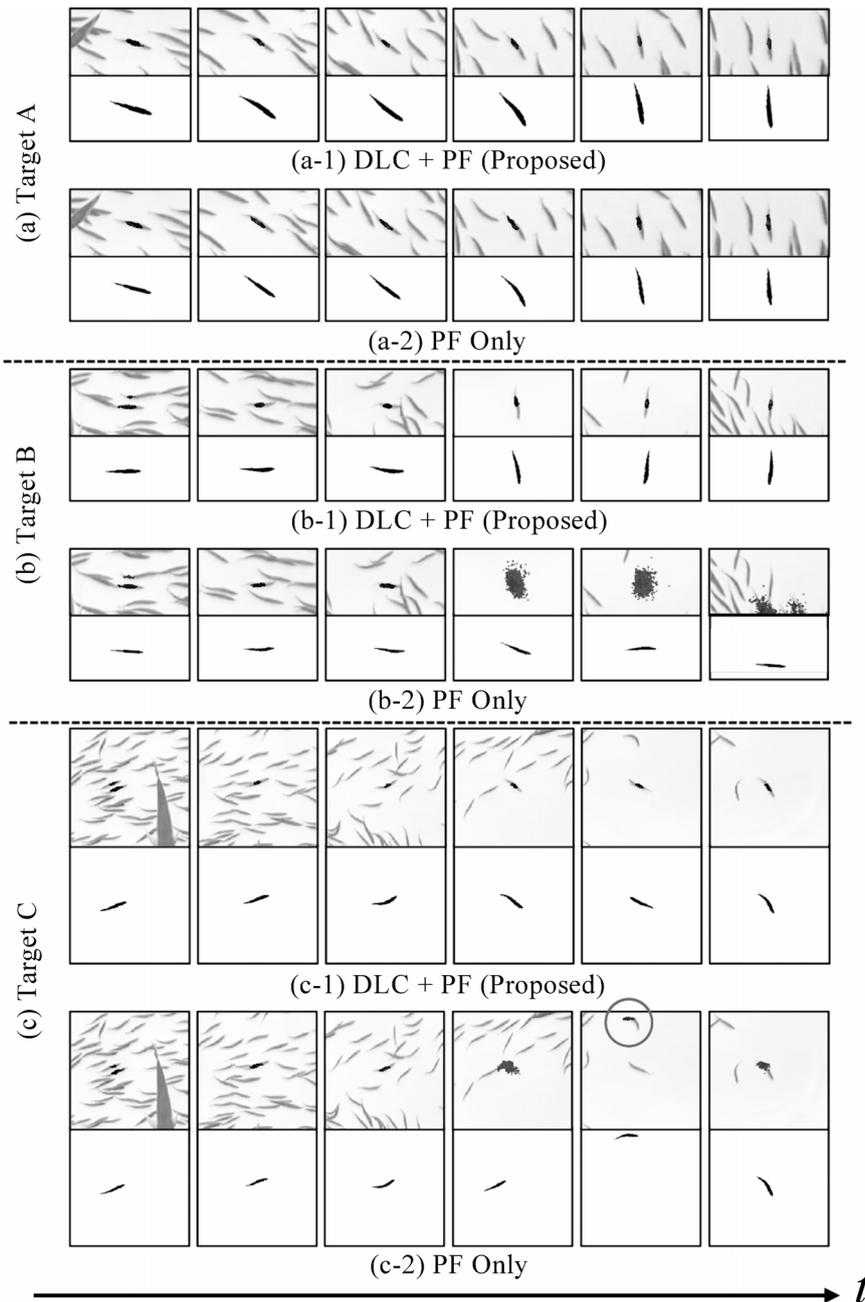


Fig. 13. Experimental results for real videos. Proposed method (a combination of DLC and PF) and PF are compared. For each of three Targets (a)–(c), results of proposed method are shown in upper row, and results of PF are shown in lower row. For results of each frame, distribution of \mathbf{c} among state vectors of particles is shown above, and silhouette generated from particle with highest likelihood is shown below.

characteristics as those shown in Experiment 1. Compared with the head and tail fins, the center of the body is more ambiguous for tracking; however, it was tracked with good accuracy even when using DLC alone. In addition, the accuracy improved further when the DLC results were combined with particle filtering, and the proposed method demonstrated the best accuracy.

4.2.2. Qualitative Evaluation

Figure 13 shows the tracking results. We compared the proposed method (a combination of DLC and PF) with a

method in which a PF was used for all components of the state vector, which is the same as the method shown in the first row of Tables 1–3.

We observed the following from the figure:

- For Target A, both methods yielded good results.
- In the case of Target B, when the swimming direction changed, the tracking results of the PF were affected. Eventually, the PF failed to track and output an inaccurate silhouette at the frame shown on the rightmost side.

- The same pattern was observed for Target C; the fish could not be tracked by the PF in the area where it was swimming in a circular direction.

This observation is consistent with the simulation results presented in Section 4.1. When using a PF alone, tracking fails when the swimming direction changes rapidly, and tracking cannot be restarted easily. By contrast, DLC employs a detection process based on machine learning, which reduces tracking failures. Therefore, by combining DLC with PF-based pose estimation, we can perform a more accurate position and pose estimation. In summary, we confirmed that the proposed method can effectively detect and track fish.

5. Conclusion

Herein, we proposed a method to track fish swimming in a school and estimate their posture. The proposed method utilizes the NACA model to represent the pose and dynamic variations of fish. The model is a parametric representation that can estimate future fish movements and detect abnormal behaviors, such as escape. To fit the model to video data, we first adopted the DLC toolbox to detect body parts (i.e., head, center, and tail fin) in an image sequence. Subsequently, a PF was applied to fit a set of parameters of the NACA model. In this filtering process, the results from DLC, i.e., three points on a fish body, were used to adjust the components of the state vector. This method enables more reliable estimation results to be yielded, including when abrupt changes in speed and direction occur.

To evaluate the effectiveness of the proposed method, we conducted experiments using simulation data and real fish school videos. The results indicated that the proposed method stably tracked a fish and estimated its posture information.

A future study will be conducted to investigate a method to achieve a fully automated detection and tracking of large-scale fish schools, since the current implementation requires manual annotation for each scene, and our focus was on tracking a single fish. To track a large number of fish in a fish school, the algorithm used must be improved such that it can accurately track the fish even when they overlap each other. We believe that the posture estimation results obtained using the proposed method can be used in data association for multiple-object tracking [18, 19]. Additionally, 3D information acquired using two or more cameras [20] may facilitate in tracking overlapping fish in a large-scale school.

Additionally, we will focus on applying the proposed method to various types of collective behaviors and analyze the mechanism using the pose estimation results. We plan to analyze the mechanism of collective behavior by processing the captured images offline because the proposed method requires several hours to yield the results. Another future direction includes implementing this technique in autonomous underwater vehicles [21].

Acknowledgements

We wish to express our gratitude to Director Akihiro Kawakubo of Kujukushima Aquarium Umi Kirara, Nagasaki, Japan, and the staff of the aquarium for their assistance in conducting this study. Mr. Kouki Hongou at Kindai University assisted with some of the experiments. This study was partially supported by JSPS KAKENHI JP19H04939.

References:

- [1] J. Toner and Y. Tu, "Flocks, herds, and schools: A Quantitative Theory of Flocking," *Physical Review E*, Vol.58, No.4, pp. 4828-4858, 1998.
- [2] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks, "Collective Memory and Spatial Sorting in Animal Groups," *J. of Theoretical Biology*, Vol.218, No.1, pp. 1-11, 2002.
- [3] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin, "Effective Leadership and Decision-making in Animal Groups on the Move," *Nature*, Vol.433, No.7025, pp. 513-516, 2005.
- [4] T. Vicsek and A. Zafeiris, "Collective Motion," *Physics Reports*, Vol.517, No.3, pp. 71-140, 2012.
- [5] A. D. Liliya and V. G. Artyom, "Research and Study of the Hybrid Algorithms Based on the Collective Behavior of Fish Schools and Classical Optimization Methods," *Algorithms*, Vol.13, Issue 4, 85, 2020.
- [6] J. Delcourt, M. Denoel, M. Ylief, and P. Poncin, "Video Multi-tracking of Fish Behavior: a synthesis and future perspectives," *Fish and Fishers*, Vol.14, No.2, pp. 186-204, 2013.
- [7] K. Terayama, H. Habe, and M. Sakagami, "Multiple Fish Tracking with an NACA Airfoil Model for Collective Behavior Analysis," *IPSI Trans. on Computer Vision and Applications*, Vol.8, No.4, pp. 1-7, 2016.
- [8] H. Akimoto and H. Miyata, "Finite-volume simulation of a Flow about a Moving Body with Deformation," *Proc. 5th Int. Symp. Comp. Fluid Dynamics*, Vol.1, pp. 13-18, 1993.
- [9] M. Alexander, M. Pranav, M. C. Kevin, A. Taiga, N. M. Venkatesh, W. M. Mackenzie, and B. Matthias, "DeepLabCut: marker-less pose estimation of user-defined body parts with deep learning," *Nature Neuroscience*, Vol.21, Issue 9, pp. 1281-1289, 2018.
- [10] I. Michael and B. Andrew, "CONDENSATION – Conditional Density Propagation for Visual Tracking," *Int. J. of Computer Vision*, Vol.29, No.1, pp. 5-28, 1998.
- [11] T. N. Duc, L. Wanqing, and O. O. Philip, "Human detection from images and videos: A survey," *Pattern Recognition*, Vol.51, pp. 148-175, 2016.
- [12] T. Fukunaga, S. Kubota, S. Oda, and W. Iwasaki, "GroupTracker: Video tracking system for multiple animals under severe occlusion," *Computational Biology and Chemistry*, Vol.57, pp. 39-45, 2015.
- [13] O. Yamanaka and R. Takeuchi, "UMATracker: an intuitive image-based tracking platform," *J. of Experimental Biology*, Vol.221, jeb182469, 2018.
- [14] Y. Okuda, H. Kamada, S. Takahashi, S. Kaneko, K. Kawabata, and F. Takemura, "Method of Dynamic Image Processing for Ecology Observation of Marine Life," *J. Robot. Mechatron.*, Vol.25, No.5, pp. 820-829, 2013.
- [15] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.43, No.1, pp. 172-186, 2021.
- [16] S. Kreiss, L. Bertoni, and A. Alahi, "PifPaf: Composite Fields for Human Pose Estimation," *2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 11969-11978, 2019.
- [17] K. Scott, C. D. Gelatt, and P. V. Mario, "Optimization by simulated annealing," *Science*, Vol.220, No.4598, pp. 671-680, 1983.
- [18] C. Ma, Y. Li, F. Yang, Z. Zhang, Y. Zhuang, H. Jia, and X. Xie, "Deep Association: End-to-end Graph-Based Learning for Multiple Object Tracking with Conv-Graph Neural Network," *Proc. of the 2019 on Int. Conf. on Multimedia Retrieval (ICMR '19)*, pp. 253-261, 2019.
- [19] T. Kikuchi, K. Nonaka, and K. Sekiguchi, "Moving Horizon Estimation with Probabilistic Data Association for Object Tracking Considering System Noise Constraint," *J. Robot. Mechatron.*, Vol.32, No.3, pp. 537-547, 2020.
- [20] K. W. Xiaolin and C. Jinxiang, "Modeling 3D human poses from uncalibrated monocular images," *Int. Conf. on Computer Vision*, pp. 1873-1880, 2009.
- [21] T. Ura, "Development Timeline of the Autonomous Underwater Vehicle in Japan," *J. Robot. Mechatron.*, Vol.32, No.4, pp. 713-721, 2020.



Name:
Hitoshi Habe

Affiliation:
Associate Professor, Department of Informatics,
Faculty of Science and Engineering, Kindai University

Address:
3-4-1 Kowakae, Higashi-osaka, Osaka 577-8502, Japan

Brief Biographical History:
1999-2002 Mitsubishi Electric Corp.
2002-2006 Assistant Professor, Kyoto University
2006-2011 Assistant Professor, Nara Institute of Science and Technology
2011-2012 Specially Appointed Lecturer, Osaka University
2012-2018 Lecturer, Kindai University
2018- Associate Professor, Kindai University

Main Works:
• “Behavior understanding based on intention-gait model.”
Human-Harmonized Information Technology, Vol.2, pp. 139-172, 2017.

Membership in Academic Societies:
• The Institute of Electrical and Electronics Engineers (IEEE)
• Association for Computing Machinery (ACM)
• The Institute of Electronics, Information and Communication Engineers (IEICE)
• Information Processing Society of Japan (IPSI)
• The Japanese Society for Artificial Intelligence (JSAI)
• The Institute of Electrical Engineers of Japan (IEEJ)
• The Japanese Society of Fisheries Science (JSFS)



Name:
Yoshiki Takeuchi

Affiliation:
Graduate School of Science and Engineering,
Kindai University

Address:
3-4-1 Kowakae, Higashi-osaka, Osaka 577-8502, Japan

Brief Biographical History:
2019- Graduate School of Science and Engineering, Kindai University

Main Works:
• “A Method for Pose Estimation in a Fish School Video Adaptive for Sudden Movements,” 2020 Vision Engineering Workshop (ViEW 2020), 2020.

Membership in Academic Societies:
• Information Processing Society of Japan (IPSI)



Name:
Kei Terayama

Affiliation:
Associate Professor, Graduate School of Medical Life Science, Yokohama City University

Address:
1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

Brief Biographical History:
2016-2018 Post-doctoral Researcher, The University of Tokyo
2018-2020 Post-doctoral Researcher, RIKEN
2018-2020 Specially Appointed Assistant Professor, Kyoto University
2020- Associate Professor, Yokohama City University

Main Works:
• “Pushing property limits in materials discovery via boundless objective-free exploration,” Chemical Science, Vol.11, pp. 5959-5968, 2020.

Membership in Academic Societies:
• The Institute of Electronics, Information and Communication Engineers (IEICE)
• The Japanese Society of Fisheries Science (JSFS)
• American Chemical Society (ACS)



Name:
Masa-aki Sakagami

Affiliation:
Professor, Department of Interdisciplinary Environment, Graduate School of Human and Environmental Studies, Kyoto University

Address:
Yosida Nihonmatsu-cho, Sakyo-ku, Kyoto 606-8316, Japan

Brief Biographical History:
1989-1996 Associated Professor, Fukui University
1996-2003 Associated Professor, Kyoto University
2003- Professor, Kyoto University

Main Works:
• “Multiple Fish Tracking with an NACA Airfoil Model for Collective Behavior Analysis,” IPSJ Trans. on Computer Vision and Applications, Vol.8, No.4, pp. 1-7, 2016.
• “Long-term Evolution of Stellar Self-Gravitating System away from the Thermal Equilibrium: connection with non-extensive statistics,” Phys. Rev. Lett., Vol.90, Issue 18, 2003.

Membership in Academic Societies:
• The Physical Society of Japan (JPS)