

EXPRESS PAPER

Open Access



Multiple fish tracking with an NACA airfoil model for collective behavior analysis

Kei Terayama^{1*}, Hitoshi Habe² and Masa-aki Sakagami³

Abstract

We propose a visual tracking method with an NACA airfoil model for dense fish schools in which occlusions occur frequently. Although much progress has been made for tracking multiple objects, it remains a challenging task to track individuals due to factors such as occlusion and target appearance variation. In this paper, we first introduce a NACA airfoil model as a deformable appearance model of fish. For occluded fish, we estimate their positions, angles, and postures with template matching and simulated annealing algorithms to effectively optimize their parameters. To improve performance of tracking, we repeatedly track fish with the parameter estimation algorithm forwards and backwards. We prepared two real fish scenes in which the average number of fish is over 25 in each frame and multiple fish superimpose over 50 times. Experimental results for the scenes show that fish are practically tracked with our method compared to a tracking method based on a mixture particle filter. Over 75 % of fish in each scene have been tracked throughout the scene, and the average difference is less than 4 % of the mean body length of the school.

Keywords: Visual tracking, NACA airfoil model, Collective behavior

1 Introduction

The tracking of multiple fish in a tank to measure their behaviors has many important applications in various fields of natural science, such as animal behavior and neuroscience [1–3]. Automatic surveillance of fish in aquariums and fish farms is also important for observing the growth and health of fish in order to improve their survival rate.

A number of multiple target tracking methods have been investigated, most of which are intended for tracking humans, e.g., [4–6]. Methods for tracking multiple fish in a shallow tank have also been developed, e.g., [3, 7–10]. However, it is quite difficult to track targets when they are homogeneous and their density is high, such as the fish in the school shown in Fig. 1a. Videos of multiple fish pose many difficulties for visual tracking: fish frequently overlap with each other, their textures are weak, they deform their bodies by beating with their tails, and identification is difficult because they are homogeneous. The detection of fish, i.e., counting their number and estimating their positions and directions, in a cluster of fish such as that

shown in Fig. 1a is difficult, even manually. Therefore, it is not appropriate to simply apply the tracking-by-detection frameworks that have been widely used to track multiple targets, e.g., [5, 6], in this situation.

Terayama et al. tracked multiple fish in such a dense school using their appearance model based on the images of fish in a video [11]. They showed that if the number of fish in a cluster of fish is known, their positions and other parameters can be estimated by matching all of the combinations of the possible parameters. However, their algorithm is quite slow because of the number of their parameter combinations, and their model is not parameterized.

In this paper, we propose a novel multiple fish tracking method for a dense school of fish. First, we introduce a parameterized appearance model based on the NACA0012 airfoil model¹, which has been adopted in biomechanics and computational fluid dynamics research, e.g., in [12], to represent a fish body. The model is simple but can effectively represent the deformations of fish caused by tail beating using small parameters as compared to the models in [7–9]. The results of our experiments, in which two types of swimming event were easily detected, show the effectiveness of this model. Second, we propose a practical tracking method, which estimates

*Correspondence: terayama@cbms.k.u-tokyo.ac.jp

¹ Graduate School of Frontier Sciences, the University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa-shi, Chiba 277-8561, Japan

Full list of author information is available at the end of the article

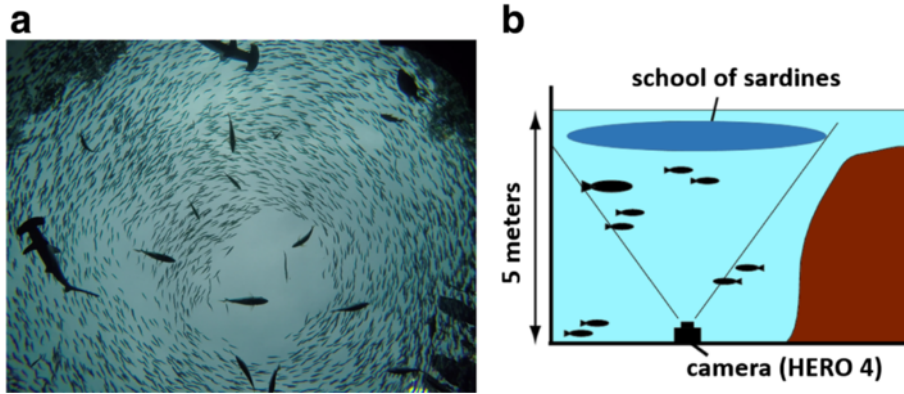


Fig. 1 **a** Snapshot of a school of sardines. **b** Camera setup

the parameters of fish with in a realistic time by using simulated annealing (SA) [13]. The approach for parameter estimation is based on that in [11]. However, their algorithm is unrealistic because of the combination of parameters. Since it is difficult to estimate the number and positions of fish in a cluster, in the proposed method, we begin to track only isolated fish that do not overlap with others. Therefore, we cannot track a fish that is initially occluded and in the middle of the video is isolated at the beginning of its trajectory. Finally, to deal with this problem, we propose a forward-backward tracking algorithm. This algorithm corresponds to manual tracking, in which we track fish in a cluster where the fish overlap by playing the video forward and backward repeatedly.

In the rest of the paper, we describe our tracking method for multiple fish in Section 2. We show the results of experiments using movies recorded in an aquarium and our event detection results in Section 3. Finally, we summarize this paper and state the plan for future work in Section 4.

2 Our method

In this section, we explain the details of our method: our appearance model and tracking algorithm using SA, and forward-backward tracking. Figure 2a shows overviews of our tracking algorithm.

2.1 Appearance model of fish

We employ the NACA0012 airfoil model as the basis of a deformable appearance model of fish. Figure 3a shows the NACA0012 model. As in [12], we define the deformation equation $h(x, \phi)$ around the initial center line in Fig. 3a as

$$h(x, \phi) = A(-(x-1)^2 + 1) \cos \frac{2\pi}{\lambda}(x - c\phi), \quad (1)$$

where the parameters A , λ , c , x , and ϕ represent the maximum amplitude, wave length, phase velocity, position from the head as shown in Fig. 3a, and phase of one

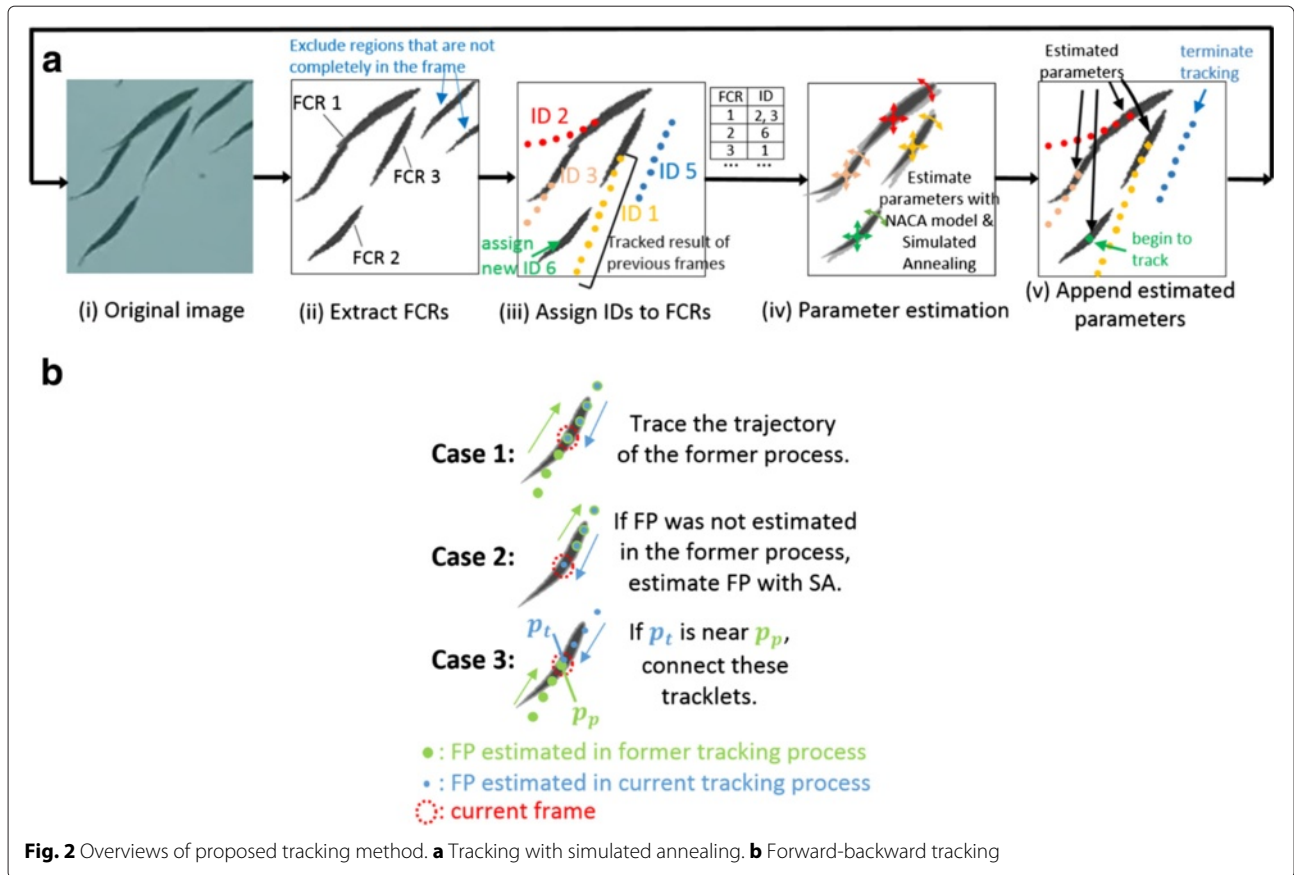
beat cycle, respectively. For each scene, we first calculate the averaged brightness of the fish and construct 92 normal appearance models based on the NACA0012 model and Eq. (1), changing A and ϕ by filling in the form with the brightness. We call these models the NACA model. Figure 3b, c shows examples of the NACA model. We set λ to 2 and c to 2 and change A from 0.01 to 0.3 and ϕ from 0 to 1. In order to deal with large deformation (bending), we add some largely deformed models based on $h(x, \phi)$ with a large amplitude. Figure 3d shows an example of a largely deformed model.

2.2 Multiple fish tracking with simulated annealing

Initially, we track only isolated fish, because in a cluster of overlapping fish it is difficult to estimate the number of fish. When two or more fish tracked by our method begin to overlap, we estimate their parameters by matching the overlapped image and the image drawn from their parameters by applying the NACA model using SA. The details of our tracking algorithm are as follows.

We estimate not only the positions of targets but also their direction angles, parameters A and ϕ of the NACA model, length scale and thickness scale. We call these fish parameters FPs. Table 1 summarizes the parameters used in our method.

For each frame t in a scene, we first binarize the frame and extract fish candidate regions (FCRs) using the binarized image, as shown in image (ii) in Fig. 2a. To each FCR, we assign fish IDs from tracking results of the previous frames by calculating the minimum of the similarity² between the FCR and all tracked fish. The image (iii) in Fig. 2a shows examples of the assignment of IDs to each FCR. If no IDs are assigned to an FCR and its area size is in the range $[a_l, a_m]$, we assign a new ID and begin to track a new fish as the FCR. We do not assign IDs to an FCR and terminate tracking if there is little or no overlap between the FCR and any images drawn from the FPs.



For an FCR that consists of multiple fish, it is difficult to estimate the number of fish in the FCR and their FPs simultaneously. However, if we know the number of fish in the FCR, we can accurately estimate their FPs by minimizing the sum of absolute differences (SAD) between the FCR and the image drawn from the FPs and the NACA model, as shown in [11]. We minimize the SAD using SA [13] to accelerate tracking process, although all the combinations of parameters were matched in [11].

In this paper, we define the neighbor of an FP as the FP in which one of the original parameters is changed according to their units. We define the cooling rate γ

and the acceptance probability function $p(s, s', T)$ for the current similarity s , a candidate new similarity s' , and the temperature T as

$$\gamma = \alpha^{1/m} \tag{2}$$

$$p(s, s', T) = \begin{cases} 1 & \text{if } s' < s \\ \exp(-(s' - s)/T) & \text{otherwise,} \end{cases} \tag{3}$$

where m is the number of IDs assigned to the FCR and α is constant close to, but smaller than, 1. Note that we employed the SAD as the similarity measure. We define the threshold $th_s(lp)$ of the termination of the

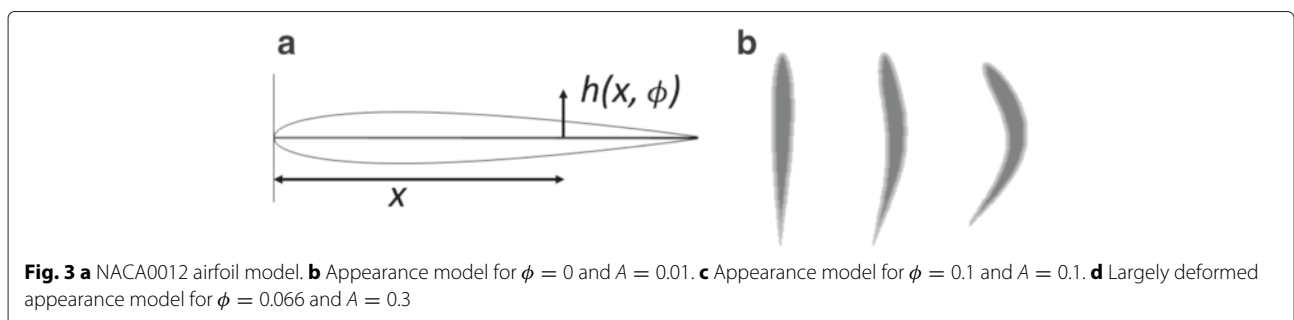


Table 1 Parameters of our method

Parameter	Units	Value explored
Position (x, y)	Pixels	The entire image
Direction angle	Degrees	Omnidirectional
A (amplitude)		0.01, 0.04, 0.07, 0.10, 0.15, 0.22, 0.30
ϕ (phase of beat cycle)	0.05	0–1
Length scale	1 %	75–150 %
Thickness scale	1 %	70–130 %

optimization process for the number of loops lp and threshold parameters th_{\min} , th_{\max} , th_{Δ} , lp_0 and lp_{\max} as

$$th_s(lp) = \begin{cases} th_{\min} + th_{\Delta} \times lp & \text{if } lp \leq lp_0 \\ th_{\max} & \text{if } lp < lp_{\max}. \end{cases} \quad (4)$$

We also terminate the optimization process if $lp \geq lp_{\max}$.

Note that our optimization process is more practical than that in [11], because the order of our algorithm for an FCR that has m assigned IDs is $\mathcal{O}(m)$ from Eq. (2). We refer to the proposed tracking method with SA as SAT (SA Tracking).

2.3 Forward-backward tracking

We repeatedly apply the SAT process to the same scene in reverse, i.e., we track the fish forward-backward. We call the former tracking process the former process. Figure 2b shows an overview of a reverse tracking.

During reverse tracking, for an FCR, if the FPs were appropriately estimated in the former process, we simply trace the FPs (case 1 in Fig. 2b). If the FPs were not estimated in the former process for the FCR, we calculate a novel FPs according to the SAT (case 2 in Fig. 2b).

When the estimated tracklet tracked in the current process reaches the tracklet estimated in the former process, we connect these tracklets when they are close (case 3 in Fig. 2b). We calculate the distance d between the FPs p_p estimated by temporal tracking and p_t estimated by the former process in the same frame defined by

$$(p_p, p_t) = \sqrt{(x_p - x_t)^2 + (y_p - y_t)^2 + (\theta_p - \theta_t)^2}, \quad (5)$$

where x_p , x_t , y_p , and y_t are positions and θ_p and θ_t are direction angles. We integrate the trajectories when the distance is smaller than d_{cn} .

We repeat the forward-backward tracking process in order to improve the tracking performance until no new estimated FPs appear. We call the process FBT.

3 Experimental results

We conducted experiments to show the effectiveness of the proposed method. We first describe the dataset used in the experiments. In order to compare the tracking performance of the proposed method, we also performed tracking using an implementation based on [4]. We call the implementation MPF (Mixture Particle Filter). We

prepared 30,000 and 40,000 particles for scene A and B in order to assign hundreds of particles to each fish. The parameters of a particle such as positions, direction angles, and NACA model parameters and the likelihood function of MPF are also based on the NACA model, as is the proposed SAT.

3.1 Dataset

We recorded videos of schools of sardines at Kujukushima Aquarium Umikirara, Nagasaki, Japan in March 2015. The videos were recorded at 30 fps using a HERO4 video camera. Figure 1a, b shows a snapshot of the video and the camera setup, respectively.

For our experiments, we extracted 5-s scenes A and B (400×300 pixels) from the movies. Figure 4a, b shows the first frames of the scenes. We set parameter a_l to 100, a_m to 450, α to 0.995, th_{\min} to 40, th_{\max} to 70, th_{Δ} to 4, lp_0 to 5, lp_{\max} to 20, and d_{cn} to 10.

We manually prepared the ground-truth (GT) trajectories of all the fish in the scenes. In this study, we tracked only the fish having a connected (occluded) component that is completely contained in the frame of the scenes. For example, the fish in the white dotted oval in Fig. 4b is not tracked in this frame.

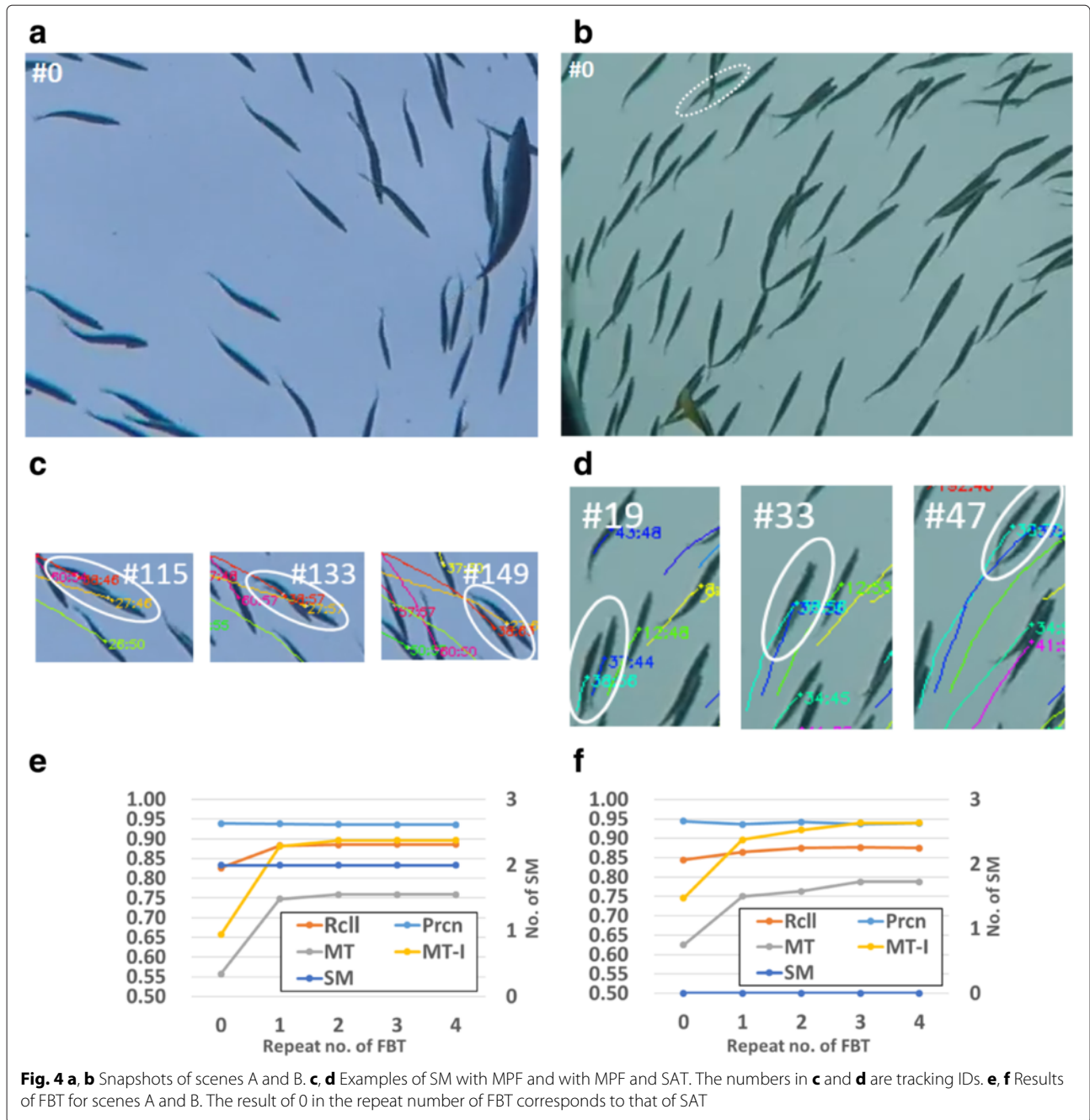
Table 2 summarizes the basic data of the scenes: the average number of fish in each frame (AN), the total number of trajectories (NT) in each scene and the total number of overlappings with others (NO) seen from each trajectory.

3.2 Evaluation metrics

We used five metrics to evaluate the tracking results based on the metrics in [6]. For each scene, we calculated the average ratio of the fish detected correctly using our tracking methods to the GT (RcII) and to all fish detections that may contain failures (Prcn). We considered a pair of an estimated parameter and GT correctly matched if the metric distance between the estimated head position and the GT position is less than five pixels. To measure the tracking performance, we calculated the ratio of GT trajectories that are covered by the estimated tracklets for more than 95 % of their length to all the GT trajectories (MT). Our method cannot track a fish which is overlapped for the entire frames in the scene. We also measured the MT except for such fish (MT-I). To measure tracking failures, we counted the total number of ID switchings during fish crossing and particle migrations (switchings and migrations (SMs)).

3.3 Experiment 1

To evaluate the effectiveness of our method as compared to the MPF, we show the results of the SAT and the MPF in Table 3. Figure 5a shows a snapshot of the tracking results in scene B using SAT. All isolated fish, except for one using the MPF in scene A, are correctly tracked with the SAT



and the MPF. SM numbers using the SAT are smaller than those using the MPF. Figure 4c, d shows examples of SM with the MPF and with both methods, respectively.

We tracked fish, repeating the FBT process five times. The results are shown in the SAT+FBT row in Table 3. Figure 5c, d shows space-time trajectory plot of the entire

Table 2 Basic data of scenes A and B

	AN	NT	NO
A	28.61	80	104
B	40.05	114	117

Table 3 Quantitative comparison of tracking results

Scene	Method	Rcll	Prcn	MT	MT-I	SM
A	MPF	0.779	0.753	0.575	0.687	6
	SAT	0.844	<i>0.944</i>	0.625	0.746	0
	SAT+FBT	0.875	0.939	0.788	<i>0.940</i>	0
B	MPF	0.810	0.796	0.519	0.612	9
	SAT	0.826	<i>0.936</i>	0.557	0.657	2
	SAT+FBT	<i>0.866</i>	<i>0.936</i>	0.760	<i>0.896</i>	2

The italicized values are the best scores among the methods in each scene

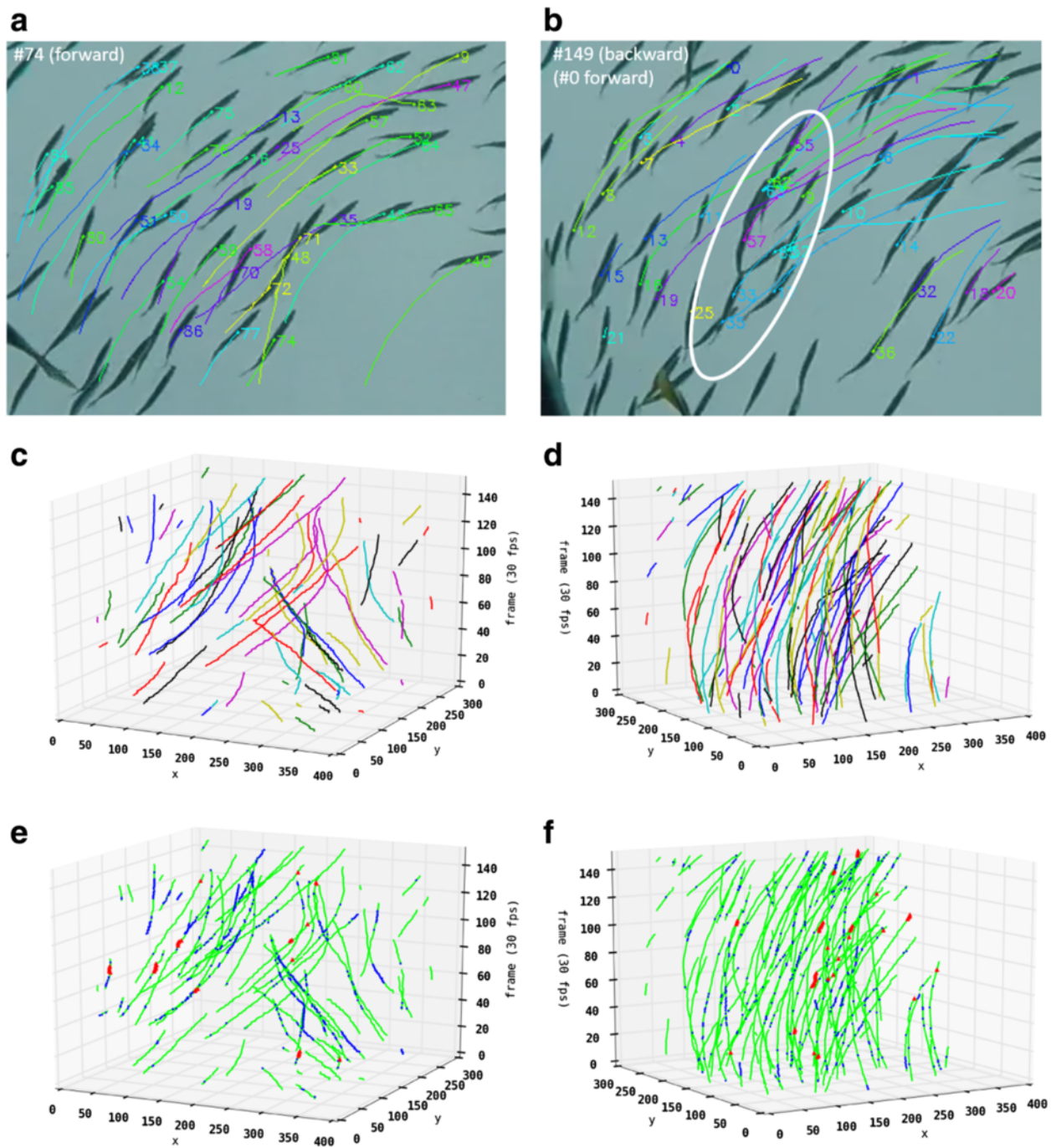


Fig. 5 **a** Snapshots of tracking results in scene B obtained using our method. The fish parameters of nine occluded fish in the white oval in **b** can be estimated. The numbers in **a** and **b** are tracking IDs. **c, d** Space-time trajectory plot of the entire sequence of scenes A and B. **e, f** Gliding (blue point) and bending (red point) events of scenes A and B. In **c, d, e,** and **f,** the X and Y axes represent 2D space

sequence of scenes A and B. By virtue of backward tracking, we can estimate the FPs of a cluster of nine fish in the white oval in Fig. 5b in the first frame. Figure 4e, f shows the improvement in the tracking performance using FBT. Over 75 % (approximately 90 % in MT-1) of fish

in each scene are correctly tracked by FBT, and the average differences between GT and estimated positions are less than 4 % of the mean body length in each scene. The experiments are performed using our non-optimized implementation in Python and OpenCV. The

average SAT computation times for processing one frame of scenes A and B were about 4 and 6 min.

3.4 Experiment 2

Since we employed the parametrized appearance model, we can easily find events that are useful for collective behavior analysis, such as bending (Fig. 3d) and gliding. Gliding is a swimming phase in which there is no beating and the fish sometimes strongly bend their bodies to change the direction of their movement. We extracted such events using the amplitude parameter A . The blue points in Fig. 5e, f show gliding events. The fish are bending at the red points in Fig. 5e, f.

From the viewpoints of biomechanics and animal behavior, such measurements are essential, because gliding and bending are related to the energy consumption of swimming [14] and constitute a type of information transmission in a school [2, 15].

4 Conclusion

In this paper, we proposed an appearance-based tracking method for multiple fish tracking. Over 75 % (approximately 90 % in MT-1) of the fish in two scenes were successfully tracked. The experimental results indicate that our method is practical for multiple fish tracking and collective motion analysis.

Our method is suitable for fish filmed from the bottom because the NACA model represent a fish viewed from the bottom and top. We would like to extend the applicable range of our method for movies taken from other directions. Our future work also includes improving the tracking performance by introducing data association frameworks and interaction models between fishes to estimate the states in the next frame. Furthermore, it is worth accelerating our algorithm in order to track thousands of fishes in schools.

Endnotes

¹The NACA airfoils are models of shapes for aircraft wing sections originally developed by the National Advisory Committee for Aeronautics (NACA). The digits represent the parameters of the shapes.

²We employed the sum of absolute differences (SAD) as the similarity measure.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KT initiated and led the study and was the main author of the manuscript. He designed the study, developed the algorithm, acquired the data, conducted experiments, and analyzed the results. HH participated in the design of the study and drafted the manuscript. MS participated in the design of the study and acquired the data. All authors read and approved the final manuscript.

Acknowledgements

This work were supported by JSPS KAKENHI Grant Number 26240023 and 26610114. The authors sincerely appreciate the cooperation of Kujukushima Aquarium Umikirara to take movies of the school of sardines.

Author details

¹Graduate School of Frontier Sciences, the University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa-shi, Chiba 277-8561, Japan. ²Faculty of Science and Engineering, Kindai University, Higashi-osaka, Japan. ³Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan.

Received: 28 April 2016 Accepted: 17 June 2016

Published online: 02 August 2016

References

1. Vicsek T, Zafeiris A (2012) Collective motion. *Phys Rep* 517(3):71–140
2. Strandburg-Peshkin A, Twomey CR, Bode NW, Kao AB, Katz Y, Ioannou CC, Rosenthal SB, Torney CJ, Wu HS, Levin SA, et al. (2013) Visual sensory networks and effective information transfer in animal groups. *Curr Biol* 23(17):709–711
3. Delcourt J, Denoël M, Ylieff M, Poncin P (2013) Video multitasking of fish behaviour: a synthesis and future perspectives. *Fish Fish* 14(2):186–204
4. Vermaak J, Doucet A, Pérez P (2003) Maintaining multimodality through mixture tracking. In: *Proc 9th IEEE Int Conf Comput Vis Vol. 2*. pp 1110–1116. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1238473&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D1238473
5. Okuma K, Taleghani A, De Freitas N, Little JJ, Lowe DG (2004) A boosted particle filter: Multitarget detection and tracking. In: *Proc 8th European Conf Comput Vis Vol. 1*. pp 28–39. http://link.springer.com/chapter/10.1007/978-3-540-24670-1_3
6. Wu B, Nevatia R (2006) Tracking of multiple, partially occluded humans based on static body part detection. In: *Proc IEEE Comput Soc Conf Comput Vis Pattern Recog Vol. 1*. pp 951–958. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1640854&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D1640854
7. Ukita N, Kitajima T, Kidode M (2004) Estimating the positions and postures of non-rigid objects lacking sufficient features based on the stick and ellipse model. In: *Proc Conf Comput Vis Pattern Recog Workshop*. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1384794&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D1384794
8. Mitsugami I, Kakusho K, Minoh M (2009) Efficient particle filtering for a non-rigid object based on PCA about changes of its shape and motion. *IEICE Trans Inf Syst* 192-D(8):1270–1278
9. Butail S, Paley DA (2012) Three-dimensional reconstruction of the fast-start swimming kinematics of densely schooling fish. *J R Soc Interface* 9(66):77–88
10. Fukunaga T, Kubota S, Oda S, Iwasaki W (2015) Grouptracker: Video tracking system for multiple animals under severe occlusion. *Comput Biol Chem* 57:39–45
11. Terayama K, Hongo K, Habe H, Sakagami M-a (2015) Appearance-based multiple fish tracking for collective motion analysis. In: *Proc 3rd IAPR Asian Conf Pattern Recog*. pp 361–365. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7486526>
12. Akimoto H, Miyata H (1993) Finite-volume simulation of a flow about a moving body with deformation. In: *Proc 5th Int Symp Comput Fluid Dynamics Vol. 1*. pp 13–18. https://www.researchgate.net/profile/Hiromichi_AKIMOTO/publication/260146162_Finitevolume_simulation_of_a_flow_about_a_moving_body_with_deformation/links/0deec52fc29c0a556000000.pdf
13. Kirkpatrick S, Vecchi MP, et al. (1983) Optimization by simulated annealing. *Science* 220(4598):671–680
14. Hemelrijk CK, Reid DAP, Hildenbrandt H, Padding JT (2014) The increased efficiency of fish swimming in a school. *Fish Fish* 16(3):511–521
15. Radakov DV (1973) *Schooling in the Ecology of Fish*. Wiley, New York