TRACKING UNDULATORY BODY MOTION OF MULTIPLE FISH BASED ON MIDLINE DYNAMICS MODELING

Shuo Hong Wang[1], *Xi En Cheng*[1,2], *Yan Qiu Chen*[1*]

[1]School of Computer Science, Shanghai Key Laboratory of
Intelligent Information Processing, Fudan University, Shanghai, China
[2]Jingdezhen Ceramic Institute, Jindezhen, Jiangxi, China
{sh_wang, xcheng, chenyq}@fudan.edu.cn

ABSTRACT

Accurately and reliably tracking the undulatory motion of deformable fish body is of great significance for not only scientific researches but also practical applications such as robot design and computer graphics. However, it remains a challenging task due to severe body deformation, erratic motion and frequent occlusions. This paper proposes a tracking method which is capable of tracking the midlines of multiple fish based on midline evolution and head motion pattern modeling with Long Short-Term Memory (LSTM) networks. The midline and head motion state are predicted using two LSTM networks respectively and the predicted state is associated with detections to estimate the state of each target at each moment. Experiment results show that the system can accurately track midline dynamics of multiple zebrafish even when mutual occlusions occur frequently.

***Index Terms*—** Multi-object tracking, fish school, LSTM networks, midline dynamics modeling

1. INTRODUCTION

The development of high speed camera makes it possible to capture the complete undulatory motion of fish body, which is of great value for scientific researches such as fish behavior and hydrodynamics [1, 2], bio-inspired robot design [3] and computer graphics applications [4]. The challenges of accurately and reliably tracking the waving bodies of multiple fish mainly lie in: (I) The erratic, highly deformable motion of their bodies; (II) The difficulties of segmenting fish body when occlusions occur; (III) The association ambiguities caused by similar appearance and frequent occlusions.

Particle image velocimetry (PIV) technique is widely used by scientists to quantitatively measure the motion of fish body and wake hydrodynamics around swimming fish [5]. However, this is not a direct way to track the motion of fish body, and the measurement accuracy is limited by the space resolution of the view of the wake [6]. In addition, they are appropriate for just a few kinds of behavior researches.

Using high speed camera to track the locomotion of fish is the most effective way to directly obtain motion data of their waving bodies. Existing fish tracking systems that are able to track the body shape of each individual mainly employ two strategies to model the fish body. The first one is to describe fish body by an image blob, these systems track the contour of each fish [7]. To accurately describe the highly deformable fish body, complex contour model should be used, which makes such systems suffer from high time cost and rely on boundary accuracy. Most of the fish related experiments use a top view setup, *i.e.*, the camera is located right above the water tank, thus, in the captured images, the fish appears as an elongated belt. Inspired by this observation, several fish tracking systems describe shape of the fish body by its midline. Mirat *et al.* [8] simply separates fish body into two parts, *i.e.*, head and tail part, the tail part is represented by two endpoints, the joint angle of tail part and head part is used to measure the deformation of fish body, such tail angle model results in insufficient accuracy for describing the specific shape of fish body. Fontaine [9] models fish midline with B-spline basis functions and applies iterated Kalman Filter (IKF) to track the locomotion of fish by tracking the B-spline parameters. The defects of the it are: (I) It suffers from relatively high state space dimensionality and requires high frame rate (3000fps), resulting in difficulties of long term tracking; (II) It relies on strict illumination condition to guarantee fish boundary extraction accuracy; (III) It needs manual initialization; (IV) It is not capable of resolving occlusions because it doesn't include a well-designed data association strategy.

To reliably track the waving body of each individual in a fish school even when occlusions occur that makes fish body difficult to be neatly segmented, a fish model that can accurately describe the undulatory body of fish and a motion model that can predict the midline dynamics are essential. This paper proposes a novel fish model, the fish body is described by the joint angle chain of midline segments. Two Long Short-Term Memory (LSTM) networks are employed

*Corresponding author. Thanks to National Natural Science Foundation of China, Grant No. 61175036 for funding.

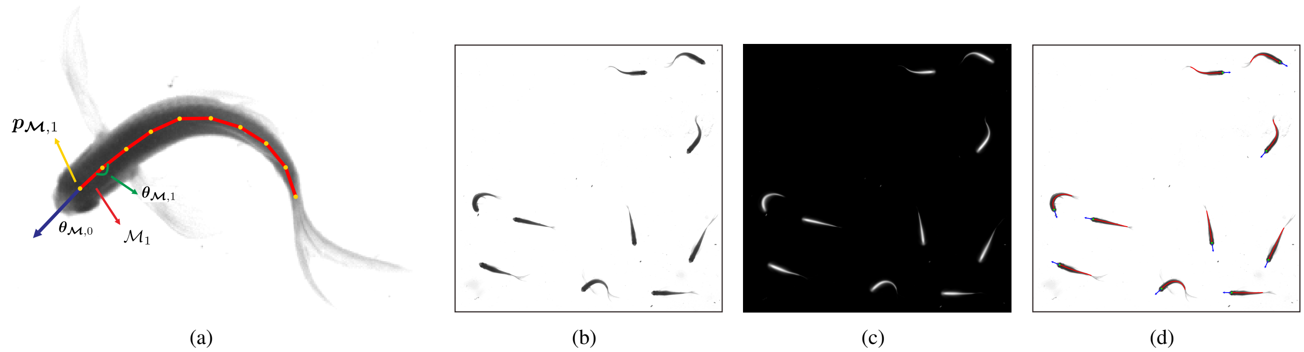

Fig. 1. (a). The proposed fish model, midline segments are plotted as red segments, yellow points are the joint points, blue arrow indicates the orientation of fish head, green arc indicates the joint angle of adjacent midline segments; (b). One sample image captured by high speed camera; (c). Visualization of resulting ρ of (b); (d). Detection results of (b), midlines are plotted as red segments, blue arrows and green points indicate the orientation and center point of each fish head.

to model the motion of fish head and evolution of midline respectively. In detection stage, midline of each fish is extracted by a scale-space method. In tracking stage, the state of fish head and midline is first predicted by LSTM networks, then data association is accomplished by associating the predicted state and detection results.

The main contributions of the proposed method lie in:

- Proposing a novel fish model composed of position, orientation of fish head and joint angle chain of discrete midline segments.
- Applying two LSTM networks to model the evolution of fish midline and motion of fish head respectively.
- Implementing the two LSTM networks in a fish tracking method to predict the state of fish in each frame in tracking process, the proposed tracking method can track the waving bodies of multiple fish even when occlusions frequent occur.

2. MOTION MODELING WITH LSTM NETWORKS

In a tracking system, a motion model is essential because the posterior density of the target's motion at each time step needs to be calculated which is used to determine the probability of a hypothetical motion sequence. From the Bayesian perspective, the motion of a target is modeled as a dynamic system and inference on the system subjects to first-order Markov chain rule. In the fish tracking case, the state parameters of fish head and midline depend on a motion process in several consecutive frames, which is the characteristic of motion patterns of living organisms. Hence, the first-order Markov assumption is not suitable for fish tracking systems.

Several existing tracking systems model motion of the target as random walk [10], however, if the difference of motion state between consecutive frames is large, the variance needs

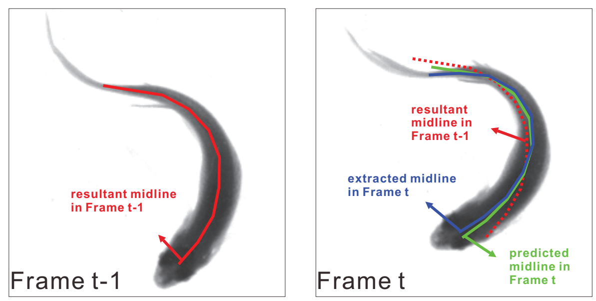

Fig. 2. The visualization of midline prediction. Red segments in the left image plot the resultant midline in frame $t-1$, identical to the red dotted segments in the right image. Green and blue segments in the right image plot the predicted midline by LSTM network and extracted midline by the scale-space midline extraction method respectively.

to be set to a large value, which makes it difficult to accurately predict the state in the next frame.

Long Short-Term Memory (LSTM) network is a special kind of recurrent neural network (RNN) architecture proposed by Hochreiter and Schmidhuber [11] that replaces the conventional neurons in hidden layer with memory blocks. It has shown to be more powerful than standard RNN for temporal processing tasks [12] thanks to its ability of learning long-term dependencies. According to these considerations, we employ two LSTM networks denoted as $\mathcal{L}_1$ and $\mathcal{L}_2$ to model the evolution of midline and motion of fish head respectively. Modeling the motion of fish head enables the tracking method to predict position of fish head and modeling the midline of fish makes it possible to predict the body shape in the next frame. It's worth mentioning that we use velocity sequence in the LSTM network instead of position sequence for the consideration that the spatio-temporal data is more robust than positional data. The architecture of the LSTM block in $\mathcal{L}_1$ and

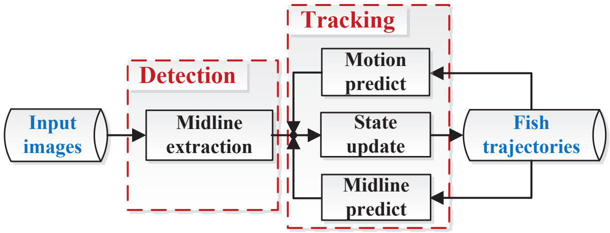

Fig. 3. Workflow of the proposed tracking method.

$\mathcal{L}_2$ is based on [13]. For an input sequence $L_{1:T}$, the LSTM layer will generate a representation sequence $\boldsymbol{h}_{1:T}$, then only $\boldsymbol{h}_T$ is fed to a sigmoid layer to obtain the prediction of L at time $T+1$. The visualization of midline prediction and detection result of one fish in frame t accompanied with the resultant midline in frame $t-1$ is shown in Figure 2. This section introduces details of the proposed fish model and motion dynamics modeling including midline evolution modeling and head motion modeling.

2.1. Fish model

In the proposed method, a fish is described by its head position, orientation and midline, as shown in Figure 1(a). The midline is discretized into n_m+1 midline points $\boldsymbol{p}_{\mathcal{M},k}, (k = 1, ..., n_m + 1)$ which constitute n_m segments denoted as $\mathcal{M}_k, (k = 1, ..., n_m)$ with equal length starting from the center point of fish head. The joint angle between each pair of adjacent segments along midline is denoted as $\boldsymbol{\theta}_{\mathcal{M},k}, (k = 1, ..., n_m - 1)$ and the length of each midline segment is $l_{\mathcal{M}}$. The rigid fish head part occupies about 20% of the total body length [9], meaning that segment $\mathcal{M}_1$ is totally in the rigid part, hence the orientation of $\mathcal{M}_1$ can be used to represent the orientation of fish head, specially denoted as $\boldsymbol{\theta}_{\mathcal{M},0}$. Then the $n_m \times 1$ vector $\boldsymbol{\theta}_{\mathcal{M},k}, (k = 0, ..., n_m - 1)$ plus $l_{\mathcal{M}}$ can be applied to accurately describe the waving fish body.

2.2. Head motion modeling

Assume the state vector of fish i's head at frame t is defined as $X^i_{h,t} = (x^i_{h,t}, y^i_{h,t}, \theta^i_{h,t})^T$, in which $\boldsymbol{p}^i_{\mathcal{M},1} = (x^i_{h,t}, y^i_{h,t})^T$ is the coordinate of fish head center, $\theta^i_{h,t}$ is the orientation of fish head, identical to $\boldsymbol{\theta}^i_{\mathcal{M},0,t}$. The input of LSTM network $\mathcal{L}_1$ is the velocity sequence of fish head in consecutive T frames, denoted as $L^i_{\mathcal{L}_1,1:T}$, defined as:

$$L^i_{\mathcal{L}_1,t} = \begin{bmatrix} \|\boldsymbol{v}^i_{h,t}\| \\ \delta^i_{\theta_{v_h},t} \end{bmatrix}, \ \ (t = 1, ..., T) \tag{1}$$

in which $\|\boldsymbol{v}^i_{h,t}\|$ is the norm of fish head velocity at frame t, $\delta^i_{\theta_{v_h},t} = \theta^i_{v_h,t} - \theta^i_{v_h,t-1}$ is the change rate of velocity direction. Output of $\mathcal{L}_1$ written as $\boldsymbol{h}^i_{\mathcal{L}_1}$ is the hypothetical velocity and change rate of velocity direction at time $T+1$.

2.3. Midline evolution modeling

As introduced in Sec. 2.1, fish midline can be characterized by the joint angle chain of discrete midline segments $\boldsymbol{\theta}_{\mathcal{M},k}, (k = 1, ..., n_m - 1)$, length of each midline segment $l_{\mathcal{M}}$ and head orientation $\boldsymbol{\theta}_{\mathcal{M},0}$, the state of fish i's midline at time t is defined as $X^i_{\mathcal{M},t} = \{\boldsymbol{\theta}^i_{\mathcal{M},k,t} | k = 0, ..., n_m - 1\}$. The input of $\mathcal{L}_2$ is the variation of $\boldsymbol{\theta}_{\mathcal{M}}$ between adjacent frames, formulated as Equ. (2), in which $\delta^i_{\boldsymbol{\theta}_{\mathcal{M}},k,t} = \boldsymbol{\theta}^i_{\mathcal{M},k,t} - \boldsymbol{\theta}^i_{\mathcal{M},k,t-1}$. The output of $\mathcal{L}_2$ denoted as $\boldsymbol{h}^i_{\mathcal{L}_2}$ corresponds to the hypothetical variation of $\boldsymbol{\theta}_{\mathcal{M}}$ between time $T+1$ and T.

$$L^i_{\mathcal{L}_2,t} = [\delta^i_{\boldsymbol{\theta}_{\mathcal{M}},k,t}], \ \ (k = 0, ..., n_m - 1, t = 1, ..., T) \tag{2}$$

3. THE TRACKING METHOD

The ultimate goal of a tracking system is to estimate the state sequence of the target based on observation sequence. In the proposed method, for each coming frame, detection and tracking are performed sequentially. In detection step, midline of each fish is extracted using a scale-space midline extraction method. In tracking step, the state of fish written as $X = (X_h, X_{\mathcal{M}})$ (includes both head motion and midline) is first predicted using two LSTM networks respectively, then the predicted state $\tilde{X} = (\tilde{X}_h, \tilde{X}_{\mathcal{M}})$ and detection results written as $Z = (Z_h, Z_{\mathcal{M}})$ (detected fish head and midline) are associated to update the state of each fish. The workflow of the proposed tracking method is shown in Figure 3.

3.1. Midline extraction

In detection step, midline of each fish needs to be accurately localized. In many biological experiments however, illumination condition is not strictly controlled, leading to shadow of fish body in the captured image and image blur near fish boundary. And fish may have varying sizes. Hence a robust midline detection method should not be dependent on fixed fish size or boundary extraction accuracy. In the proposed method, a midline extraction method based on scale-space ridge detection is employed to solve the above problems.

Assume one pixel in scale space is denoted as (x, y, s), the Hessian matrix at this point denoted as $\mathcal{H}(x, y, s)$ is calculated by convolution with second order derivative of Gaussian at scale s. And λ_k is the kth largest eigenvalue (absolute value) of Hessian matrix $\mathcal{H}$. In the captured gray scale image, fish body appears as a darker blob with tubular structure compared to background pixels, resulting in high positive λ_1 and low λ_2. $\mathcal{R} = \lambda_1/|\lambda_2|$ reflects eccentricity of the second order ellipse at that point. $\mathcal{S} = \sqrt{\lambda_1^2 + \lambda_2^2}$ reflects the second order structureness. Pixels with large $\mathcal{S}$ correspond to fish body pixels and small $\mathcal{S}$ correspond to background pixels. Inspired by [14], the measurement that reflects distance from the pixel to the nearest midline of fish body can be measured by Equ. (3), in which Ψ is all the possible scales in scale space. Point with larger ρ is nearer to a midline.

$$\rho(x,y) = \max[\rho(x,y,s)], s \in \Psi$$
$$\rho(x,y,s) = \begin{cases} 0, & \lambda_1 < 0 \\ 1 - exp[-(\alpha\mathcal{R}^2 + \beta\mathcal{S}^2)], & else \end{cases} \tag{3}$$

ρ is first binarized using a thresholding method and midlines are extracted by a skeleton algorithm based on the binarized image. Then equal distance sampling is performed on the skeleton points to discretize the midline into n_m segments with equal length $l_{\mathcal{M}}$. The visualization of resulting ρ of one input image Figure 1(b) is shown in Figure 1(c), the extracted midlines are plotted as red segments in Figure 1(d).

3.2. Data association and state update

Data association is an essential component in multiple object tracking systems that aims to associate the detections and targets at each moment. In the proposed method, both motion continuity and midline similarity are taken into consideration in data association step to guarantee a higher association accuracy. The distance between the predicted and detected head position is measured by Gaussian density function, so that the larger the distance is, the less possible the prediction and detection will be associated, which is computed as:

$$\mathcal{D}(\tilde{X}_h^i, Z_h^j) \propto \mathcal{N}(Z_h^j; \tilde{X}_h^i, \Sigma_h^i) \tag{4}$$

Assume the predicted and detected joint angle chain of midline segments are $\tilde{X}_{\mathcal{M}}$ and $Z_{\mathcal{M}}$ respectively, cosine similarity of the two vectors can be employed to measure the similarity of the predicted and detected midline, computed as:

$$\psi(\tilde{X}_{\mathcal{M}}^i, Z_{\mathcal{M}}^j) = \frac{\tilde{X}_{\mathcal{M}}^i \cdot Z_{\mathcal{M}}^j}{\|\tilde{X}_{\mathcal{M}}^i\| \|Z_{\mathcal{M}}^j\|} \tag{5}$$

And the data association problem can be formulated as a global optimization problem as Equ. (6) and solved by Kuhn-Munkres algorithm [15].

$$\begin{aligned} & C(\tilde{X}^i, Z^j) = \begin{cases} -Inf, & |l_{\mathcal{M}}^i - l_{\mathcal{M}}^j| >= \varepsilon_{l\Delta} \\ \mathcal{D}(\tilde{X}_h^i, Z_h^j) * \psi(\tilde{X}_{\mathcal{M}}^i, Z_{\mathcal{M}}^j), & else \end{cases} \\ & \tau = \arg\max_{\Lambda} \sum_{i=1}^{n} \sum_{j=1}^{m} C(\tilde{X}^i, Z^j) * \Lambda(\tilde{X}^i, Z^j) \\ & s.t. \\ & \sum_{i=1}^{n} \Lambda(\tilde{X}^i, Z^j) = 1 \;\; and \;\; \sum_{j=1}^{m} \Lambda(\tilde{X}^i, Z^j) = 1 \end{aligned} \tag{6}$$

For each target i, if it is associated with detection j, then its state is updated as the detection j, written as $X^i = Z^j$, otherwise if target i is not associated with any detection, then a midline verification strategy (see Sec. 4.2) is employed to check if the predicted midline is correct. If so, the state of fish i is updated as the predicted value, written as $X^i = \tilde{X}^i$.

4. IMPLEMENTATION DETAILS

4.1. Training LSTM networks

Training samples should be prepared to train the two LSTM networks before applying them in a fish tracking system. To reduce the manual work of preparing training samples, we employ a semi-automatic strategy. Firstly, fish velocity sequence and midline evolution data are obtained using a simple fish tracking method without any LSTM networks. Then the correctness of the tracking results (including the correctness of head position, orientation and midline) are checked and each of the incorrect ones is corrected manually. The two networks are all trained with Backpropagation Through Time (BPTT) under a matrix-based batch learning paradigm [18], which achieves better performance of convergence compared to stochastic gradient descent (SGD) method [19].

4.2. Midline verification

When severe occlusion occurs, the scale-space extraction method may fail to correctly extract the midline, if this situation happens, the state of fish will be updated as the predicted value by LSTM networks. Hence, it is necessary to judge: (1) If the extracted midline in detection step is correct; (2) If the predicted midline and head state by LSTM are correct.

The correctness of a midline $\mathcal{M}$ is verified by its segment length and the distance between it and the corresponding fish blob $\mathcal{B}$ in the image. The distance between $\mathcal{M}$ and $\mathcal{B}$ is defined as sum of Euclidean distance of each pixel in the blob $\mathcal{B}$ to the nearest midline segment, formulated as:

$$\mathcal{D}(\mathcal{B}, \mathcal{M}) = \sum_i d(\mathcal{B}_i, \mathcal{M}), \;\; \mathcal{B}_i \in \mathcal{B} \tag{7}$$

If the extracted midline $\mathcal{M}$ satisfies Equ. (8), then the midline will be judged as a correct one.

$$\begin{cases} \varepsilon_{l_{min}} < l_{\mathcal{M}} < \varepsilon_{l_{max}} \\ \mathcal{D}(\mathcal{B}, \mathcal{M}) < \varepsilon_d \end{cases} \tag{8}$$

5. EXPERIMENTS

5.1. Experiment setup

To evaluate the performance of the proposed tracking method, we tested it on a zebrafish (*Danio rerio*) school. Two video clips were captured with different fish density. The experiment setup includes: (1) One high speed monochrome camera (100fps) mounted right above the water tank; (2) One 20cm×20cm transparent acrylic water tank, the depth of water is about 8cm; (3) One light source placed above the camera. With such experiment setup the fish body appears as a dark blob in the captured image. The proposed tracking method is implemented with MATLAB$^{\text{TM}}$, the implementation of LSTM network is adapted from [18].

Table 1. Evaluation of tracking performance compared with other methods.

Data Set	GT	OP(%)	Method	Precision	Recall	F1	Frag	IDS	MA(%)
V1	10	4.4	Ours	**0.991**	**0.999**	**0.995**	0.8	0.6	**98.8**
			Ours without LSTM	0.930	0.990	0.959	4.8	0.6	92.3
			Qian *et al.* [16]	0.963	0.983	0.973	3.3	0.7	—
			idTracker [17]	0.925	0.979	0.951	**0.5**	**0.4**	—
V2	20	19.3	Ours	**0.987**	**0.990**	**0.988**	**1.3**	**0.8**	**96.2**
			Ours without LSTM	0.826	0.982	0.897	9.5	1.2	82.5
			Qian *et al.* [16]	0.926	0.973	0.949	6.1	1.3	—
			idTracker [17]	0.852	0.956	0.901	4.2	2.6	—

5.2. Experiment results and discussions

The proposed method was tested on two video clips V1 and V2, each contains 2000 frames with 10 and 20 zebrafish respectively. The metrics adopted to evaluate the performance include: Groundtruth Trajectories (GT), Precision, Recall, F1-score (F1), Fragments (Frag) and ID switches (IDS) as defined in [20] and other two self-defined metrics, Occlusion Probability (OP): the probability that one fish is in an occlusion event, Midline Accuracy (MA): the accuracy of resultant midlines. Supplementary material shows more evaluation details. The tracking performance of the proposed method was compared with other two state-of-the-art fish tracking systems and the proposed method without LSTM prediction. Qian *et al.*'s [16] method is based on motion continuity and appearance similarity for data association. idTracker [17] obtains the identity of each fish by recognizing each individual based on an intensity distribution feature. However, both systems cannot track fish body, so when calculating the metrics except MA, the correctness of midline is not considered.

According to the results shown in Table 1, the proposed method achieves better performance on most metrics compared with other methods. As an identity preserving method, idTracker outperforms others and achieves lower Frag and IDS under low fish density, however, when fish density doubles, its performance greatly deteriorates. Compared with the proposed method without LSTM, the method with LSTM prediction achieves significant improvement on most metrics and the performance is almost not affected by high fish density with high OP. Figure 4 shows two occlusion cases. For each case, the first row shows the tracking result of the proposed method without implementing LSTM network and the second row shows the result with LSTM prediction. In both cases, the method without LSTM network fails to correctly segment the occluded fish individuals but the implemented one can successfully do it. It can be concluded that the state prediction step accomplished by midline evolution and motion state modeling with LSTM networks enables the tracking system to cope with frequent occlusions and obtain more accurate midlines. The supplementary video shows the resultant midlines of 250 consecutive frames in V1.

6. CONCLUSIONS

In this paper we have proposed a tracking framework capable of tracking the undulatory body motion of multiple fish. LSTM networks are employed to model the evolution of midline and head motion pattern. For each coming frame, midline of each fish is first extracted using a scale-space method. Then the midline parameters and head motion state at this frame are predicted using two LSTM networks respectively. At last, the predicted state and detection results are associated to update the state of each individual.

The proposed method has been implemented and performance evaluation results on data sets with different zebrafish density show that the proposed method can accurately and reliably track the midlines of multiple fish with frequent occlusions. The method can also be applied to track other species of fish and belt-like objects with undulatory motion.

7. REFERENCES

[1] M. J. McHenry, C. A. Pell, and J. H. Long, "Mechanical control of swimming speed: stiffness and axial wave form in undulating fish models," *J. Exp. Biol.*, vol. 198, no. 11, pp. 2293–2305, 1995.

[2] J. Y. Cheng, T. J. Pedley, and J. D. Altringham, "A continuous dynamic beam model for swimming fish," *Philos. Trans. R. Soc. B-Biol. Sci.*, vol. 353, no. 1371, pp. 981–997, 1998.

[3] J. Liu and H. Hu, "Biological inspiration: From carangiform fish to multi-joint robotic fish," *J. Bionic Eng.*, vol. 7, no. 1, pp. 35–48, 2010.

[4] X. Tu and D. Terzopoulos, "Artificial fishes: Physics, locomotion, perception, behavior," in *SIGGRAPH*. ACM, 1994, pp. 43–50.

[5] J. Sakakibara, M. Nakagawa, and M. Yoshida, "Stereo-piv study of flow around a maneuvering fish," *Exp. Fluids*, vol. 36, no. 2, pp. 282–293, 2004.

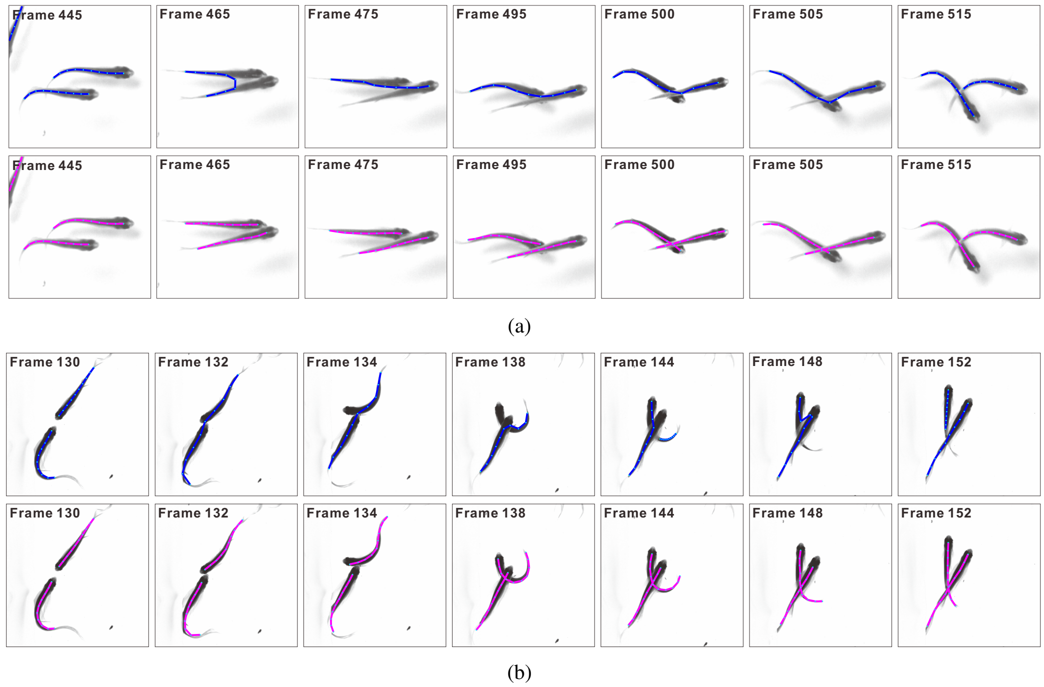

Fig. 4. Tracking results of two occlusion cases. The first and second row of (a) and (b) show the resultant midlines of the proposed method without and with LSTM prediction respectively. In both cases, the method without LSTM prediction fails to correctly resolve the occlusion case but the implemented one can successfully segment the occluded fish individuals.

[6] G. Wu, Y. Yang, and L. Zeng, "Novel method based on video tracking system for simultaneous measurement of kinematics and flow in the wake of a freely swimming fish," *Rev. Sci. Instrum.*, vol. 77, no. 11, pp. 114302, 2006.

[7] K. Branson and S. Belongie, "Tracking multiple mouse contours (without too many samples)," in *CVPR*, 2005.

[8] O. Mirat, J. R. Sternberg, K. E. Severi, and C. Wyart, "Zebrazoom: an automated program for high-throughput behavioral analysis and categorization," *Front. Neural Circuits*, vol. 7, no. 107, pp. 1–12, 2013.

[9] E. I. Fontaine, *Automated visual tracking for behavioral analysis of biological model organisms*, Ph.D. thesis, California Institute of Technology, 2008.

[10] D. A. Ross, J. Lim, R. S. Lin, and M. H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, no. 1-3, pp. 125–141, 2008.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[12] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.

[13] A. Graves, *Supervised sequence labelling with recurrent neural networks*, vol. 385, Springer, 2012.

[14] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever, "Multiscale vessel enhancement filtering," in *MICCAI98*, pp. 130–137. Springer, 1998.

[15] H.W. Kuhn, "The hungarian method for the assignment problem," *Nav. Res. Logist.*, vol. 2, no. 1-2, pp. 83–97, 1955.

[16] Z.M. Qian, X.E. Cheng, and Y.Q. Chen, "Automatically detect and track multiple fish swimming in shallow water with frequent occlusion," *PLoS ONE*, vol. 9, no. 9, pp. e106506, 09 2014.

[17] A. Pérez-Escudero, J. Vicente-Page, R.C. Hinz, S. Arganda, et al., "idTracker: tracking individuals in a group by automatic identification of unmarked animals," *Nat. Methods*, vol. 11, no. 7, pp. 743–751, JUL 2014.

[18] Q. Lyu and J. Zhu, "Revisit long short-term memory: An optimization perspective," 2015.

[19] A. Graves, N. Jaitly, and A. Mohamed, "Hybrid speech recognition with deep bidirectional lstm," in *ASRU*. IEEE, 2013, pp. 273–278.

[20] Y. Li, C. Huang, and R. Nevatia, "Learning to associate: Hybridboosted multi-target tracker for crowded scene," in *CVPR*. IEEE, 2009, pp. 2953–2960.