Paper Identifier: BISP-P4.9
Paper Number: 1008
Poster Session BISP-P4: Bioimaging and Microscopy
ICASSP 2017, New Orleans, USA

3D TRACKING SWIMMING FISH SCHOOL WITH LEARNED KINEMATIC MODEL USING LSTM NETWORK

Shuo Hong Wang[1], Jingwen Zhao[1], Xiang Liu[12], Zhi-Ming Qian[1], Ye Liu[3], Yan Qiu Chen[1*]
[1]School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai, China [2]Shanghai University of Engineering Science, China
[3]College of Automation, Nanjing University of Posts and Telecommunications, China

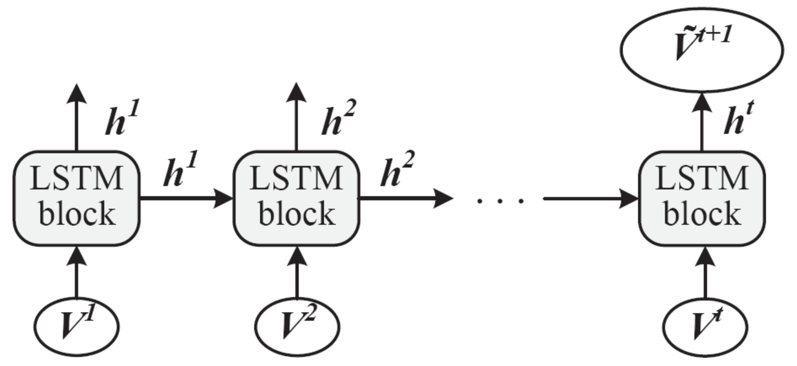
Introduction

This paper proposes a reliable 3D fish tracking method using a novel master-slave camera setup. The kinematic model is learned with a Long Short-Term Memory (LSTM) network, which predicts the 3D state of fish at each moment. We propose to use an innovative master-view-tracking-first strategy. The fish are first tracked in the master view. Cross-view association is then established utilizing motion continuity and epipolar constraint cues. Experiments on data sets of different fish densities show that the proposed method is effective and outperforms the state-of-the-art methods.

Learning kinematic model

- The state of fish depends on a motion process in several consecutive frames, first-order Markov assumption does not hold.
- LSTM network has shown superior power in processing sequential data with varying lengths and learning long-term dependencies

Hence, we model the fish's motion process by learning an LSTM network.

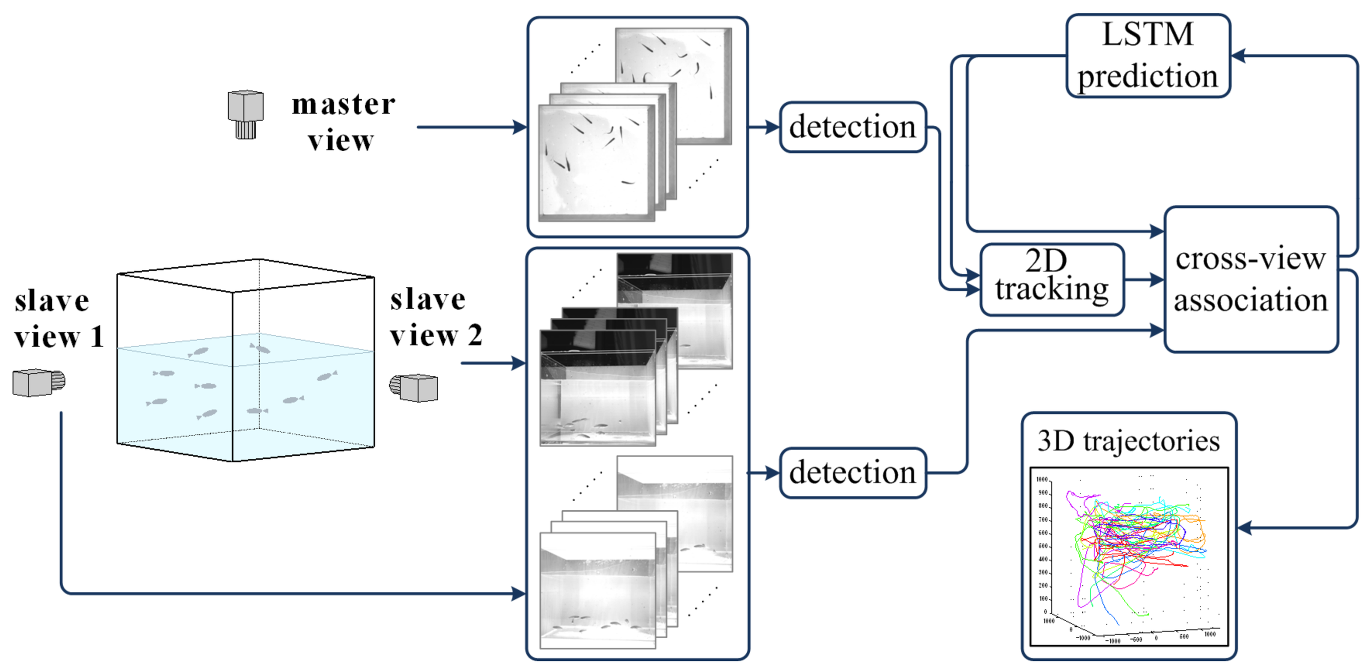


Network input: velocity sequence
Output: the hypothetical velocity vector at time t + 1

The predicted state at time t + 1 is calculated as: $\hat{X}^{t+1} = X^t + \hat{V}^{t+1}$

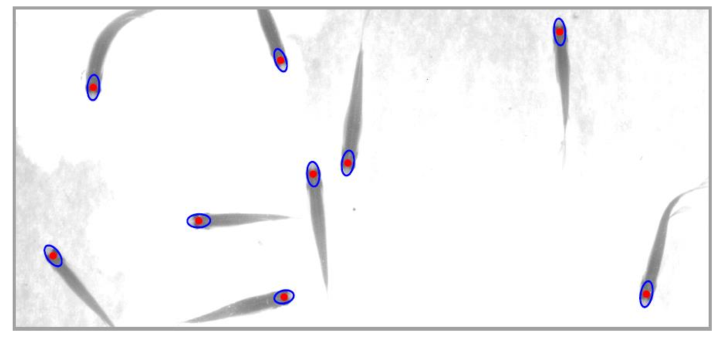
The LSTM network is trained offline. Fish are tracked using conventional Kalman filter. Then the velocity sequences with different lengths are randomly selected from the manually checked tracking results.
We selected totally 50000 velocity sequences of different fish of 8-20 frames in length to be used as training sequences. The LSTM network is trained with Backpropagation Through Time (BPTT) under a matrix-based batch learning paradigm.

Tracking Method



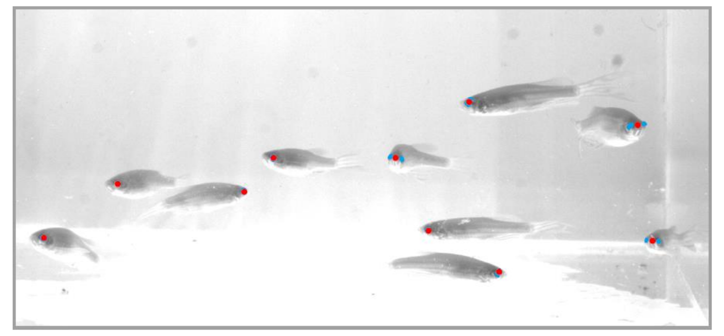
I. Fish detection in master view

Eye-focused scale-space Determinant of Hessian blob detector



II. Fish detection in slave views

Eye-focused Gabor feature detector



Gabor filter $\psi_{\vec{k}}(\vec{x}) = \frac{\vec{k}^2}{\sigma^2} \exp(-\frac{\vec{k}^2 x^2}{2\sigma^2})[\exp(i\vec{k}\vec{x}) - \exp(-\frac{\sigma^2}{2})]$

Gabor feature: 2 frequency levels (v={0, 1}), 4 orientations (u={0, 2, 4, 6}) to generate local descriptions at different scales and orientations.
Image patch: 25*25, Gabor feature dimension: 25*25*8, After dimension reduction: 40

III. 2D Fish tracking in master view

Use Kalman filter, totally two cues are applied to calculate
the weight term to build cross-frame association

1. Motion continuity
measures the distance between the predicted 2D state in master view reprojected from the 3D state predicted by LSTM network and detection

2. Appearance coherency
calculated by Normalized Cross Correlation (NCC) which measures the similarity of fish head image patches determined by predicted 2D state and detection.

IV. Cross-view data association

Associate the 2D tracking results in master view and detection results in two slave views and reconstruct 3D trajectories. Two cues used:

1. Motion continuity
The probability of detection j in slave view v being associated with object i is inversely proportional to Euclidean distance between them.

2. Epipolar constraint
inversely proportional to the Euclidean distance from detection j to the corresponding epipolar line of each object

Experimental Results

Table 1. Description of the 3 data sets

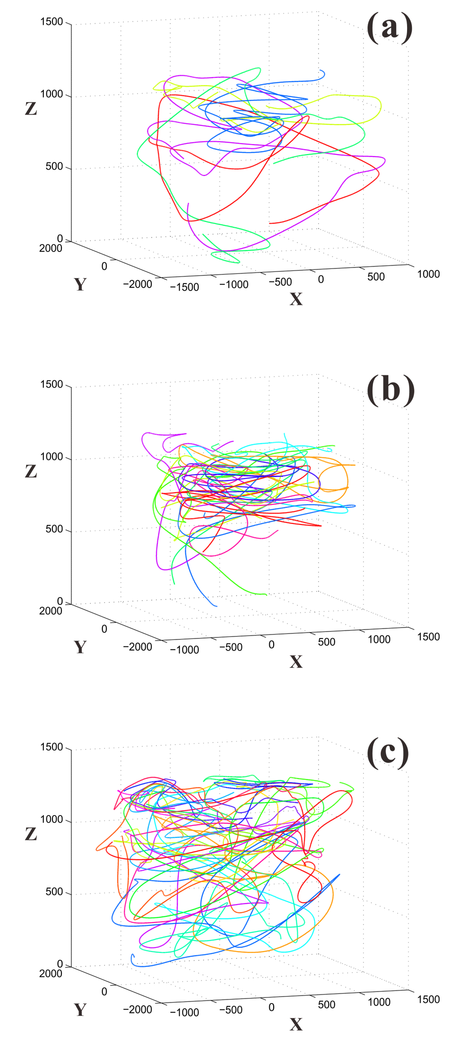
	Group size	OF in top view	OF in side views
V1	5	0.40%	6.26%
V2	10	17.14%	45.21%
V3	20	24.17%	63.33%

OF indicates the occlusion frequency. The video clips are available online at: http://www.cv.fudan.edu.cn/fishtracking3d.htm

Table 2. Performance comparison with other methods

	Method	P	R	F1	Frag	IDS
V1	Ours	**0.977**	**0.992**	**0.984**	**0.9**	**0.7**
	Ours*	0.968	0.987	0.977	1.4	0.9
	Liu *et al.*	0.967	0.975	0.971	2.9	1.1
	idTracker	0.889	0.950	0.918	6.3	1.9
V2	Ours	**0.961**	**0.971**	**0.966**	**3.9**	**0.9**
	Ours*	0.953	0.967	0.960	4.8	1.1
	Liu *et al.*	0.942	0.958	0.950	6.3	3.8
	idTracker	0.833	0.907	0.868	36.9	7.3
V3	Ours	**0.920**	**0.925**	**0.922**	**5.7**	**1.9**
	Ours*	0.913	0.920	0.916	6.2	2.5
	Liu *et al.*	0.812	0.854	0.832	11.2	7.3
	idTracker	0.285	0.436	0.345	122.7	15.0

Ours* denotes the proposed method without LSTM. P, R, F1, Frag and IDS correspond to Precision, Recall, F1-score, Fragments and ID Switches respectively.



(a) V1; (b) V2; (c)V3

Conclusion

- Proposes a reliable 3D fish tracking method using a novel master-slave camera setup.
- The kinematic model is learned with an LSTM network, which predicts the 3D state of fish at each moment.
- Propose to use an innovative master-view-tracking-first strategy. The fish are first tracked in the master view. Cross-view association is then established utilizing motion continuity and epipolar constraint cues.
- Experiments on data sets of different fish densities show that the proposed method is effective and outperforms the state-of-the-art methods.