3ZG-03

Vision-based 3D Tracking System for Fish Interaction Analysis

Yanghong Zhong, Hiroaki Kawashima, Takashi Matsuyama
Graduate School of Informatics, Kyoto University

1 Introduction

In nature fish collective behaviors governed by the individual-level interactions bring fish many benefits, like energy saving during migration, protection from predators, etc. This has been attracting many researchers to analyze the mechanism of the individual-level interactions in fish groups. The main contribution of this paper is that we propose a vision-based tracking system for a fish group to capture each individual's motion data in 3D space, including position and orientation which are very useful to infer the underlying fish interaction rules.

In our tracking system, a fish group, with a small size around 3 to 10 members, swims in a customized fish tank, as shown in Figure 1. To capture their 3D motion data, we use three synchronized cameras outside the fish tank: one for front-view, one for right-view, and one for top-view.

Tracking in such an environment has a challenge of maintaining the object association due to occlusions among individuals and object mirrors reflected at the fish-tank surfaces. To tackle this challenge, we employ shape and motion priors which are implemented as (1) the 3D fish shape which can be parameterized based on a 3D fish shape model consisting of a midline and elliptical cross sections [1] and (2) mixture particle filter [3] which utilizes current observation and fish motion prior estimated based on historical information. In particular, the calculation of the likelihood of the particle filter requires an efficient forward projection from 3D shape model to 2D image planes considering the effect of light refraction. Therefore, we utilize a pixel-wise varifocal camera model (PVCM) [2], which creates an offline lookup table mapping between sampled points on the inner fish tank surface (in contact with water) and the corresponding projection rays. Based on this lookup table, for each 3D underwater point, its projection ray can be estimated by a quickly converging iteration.

2 Tracking Approach

2.1 3D Fish Shape Model

Our 3D fish shape model is based on a flexible midline which is responsible for shape changing, and fixed elliptical cross sections approximating fish surface. To parameterize them in a local fish coordinate system (FCS), we make three assumptions of fish motion.

Assumption 1 *The midline lies in a plane and only bends laterally.*

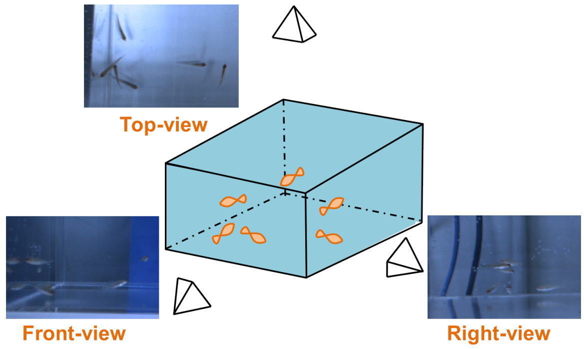

Figure 1: Tracking environment

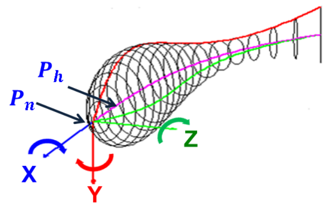

Figure 2: Fish shape model and fish coordinate system Yaw (red); Pitch (green); Roll (blue)

Assumption 2 *The head part of midline does not bend.*

Assumption 3 *Fish yaw and pitch, but do not roll.*

Based on Assumption 2, on the midline, we regard the nose point, $\mathbf{P}_n$, as the origin of FCS and the head part, $\overrightarrow{\mathbf{P}_h\mathbf{P}_n}$, lies on the X-axis. Based on Assumption 1, we assume the Y-axis fish yaw around with is perpendicular to the plane the midline lies in. Then the Z-axis fish pitch around with can be determined as a vector perpendicular to the XY-plane. Finally a FCS is defined as shown in Figure 2. Then we can do the parameterization for the midline and ellipses of our shape model in the FCS. Assume $s \in [0, 1]$, represents the ratio of midline starting at the nose point $\mathbf{P}_n$. Then we use a quadratic Bezier curve with parameters Π_c to model the midline. For every s, there is a corresponding ellipse perpendicular to the midline. We assume its major axis is parallel with the Y-axis, and that its minor axis lies in the XZ-plane and is perpendicular to the tangent line of the midline at the corresponding point. Then we use a cubic function with parameters Π_a to model the curve of major axis, and a quartic function with parameters Π_b to model the curve of minor axis.

As a result, every 3D fish shape can be generated by a parameter set $\{\Pi_c, \Pi_a, \Pi_b\}$. To present a 3D fish shape in the world coordinate system (WCS), we need to know the translation vector $\mathbf{T}_{f2w}$ and rotation matrix $\mathbf{R}_{f2w}$ from FCS to WCS. Suppose in the WCS, the nose position denoted by $\mathbf{r}$, and the heading direction denoted by a unit vector $\mathbf{h}$, are known, then $\mathbf{T}_{f2w} = \mathbf{r}$. Based on Assumption 3, the pitch axis is the cross-product of the vertical $\mathbf{g} = [0\ 1\ 0]^T$ with the heading $\mathbf{h}$, and the yaw axis is the cross-product between the heading and the pitch axis. Then $\mathbf{R}_{f2w} = [\mathbf{h}\ \mathbf{h} \times (\mathbf{g} \times \mathbf{h})\ \mathbf{g} \times \mathbf{h}]$. Eventually every fish in WCS can be modeled by $\{\mathbf{r}, \mathbf{h}, \Pi_c, \Pi_a, \Pi_b\}$.

2.2 Mixture Particle Filter

Our tracking framework is based on a Dynamic Bayesian Network (DBN) in which at each time there are two phases: prediction and update. Firstly in the prediction phase, a proposal distribution of target state is predicted based on historical information. Then in the update phase, the current observation is used to create a posterior distribution of target state. To approximate the non-Gaussian distribution of multiple targets' state, we employ the mixture particle filter (MPF). In MPF, the state distribution of each target is approximated by a set of particles. In the prediction phase, the particles responsible for the previous frame posterior distribution of different targets drift and diffuse independently to form a proposal distribution for each target. In the update phase, firstly each particle's likelihood is calculated. Then for a set of particles responsible for the same target, we do the importance sampling by regarding each particle's likelihood as its weight. Finally, a set of unweighted particles are produced to approximate each target's posterior distribution. Then in the next frame, the same process continues.

The key for MPF is to compute the likelihood of each particle represented as $\mathbf{x} = \{\mathbf{r}, \mathbf{h}, \Pi_c, \Pi_a, \Pi_b\}$. To calculate the likelihood of a given $\mathbf{x}$, firstly we reconstruct a 3D fish surface in WCS based on our fish shape model. Then we use PVCM to forward project the generated 3D surface onto each of image planes to get a set of 2D regions. For each projected region, we use the corresponding captured image to calculate the F-score as its likelihood. Finally, the likelihood of each particle is the multiplication of the likelihood of its corresponding 2D projected regions on each of image planes.

3 Experiment

In our experiment, the frame rate of each camera for capturing images was 30 frames per second. By using the proposed tracking approach, we successfully tracked each fish in a group with 6 members. One snapshot of the tracking result is shown in Figure 3. The first row shows 2D projected contours of the 3D fish model in the tracking results. In these three figures, the contours with the same color are produced by the same estimated fish. The second row shows another visualization of our tracking results. The red regions represent the foreground extracted by background subtraction, while the green ones are the regions surrounded by the projected contours which are the same as the ones in the first row. The yellow regions show the overlaps between red regions and green regions. From Figure 3, despite occlusions and object mirrors reflected on the fish tank surface, we can see each fish is successfully tracked.

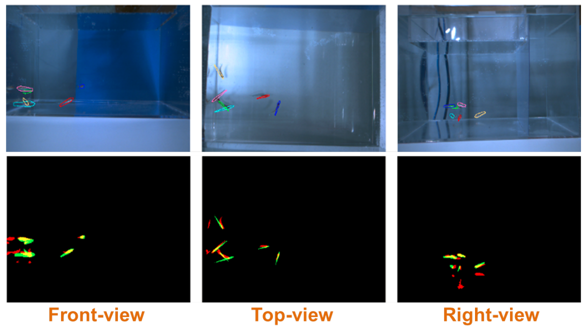

Figure 3: Visualization of tracking results

Since we want to track mainly the position $\mathbf{r}$ and orientation $\mathbf{h}$ of each fish, we did some evaluation for them. Firstly, for the same image sequences, we created the ground truth for $\mathbf{r}$ and $\mathbf{h}$ by manually picking up the projections of the nose point $\mathbf{P}_n$ and the end point of head part $\mathbf{P}_h$ on each image. Then we used backward projection to estimate $\mathbf{P}_n$ and $\mathbf{P}_h$ in WCS. For each fish, we regard the estimated $\mathbf{P}_n$ as the ground truth for $\mathbf{r}$, and the estimated $\frac{\overrightarrow{\mathbf{P}_h\mathbf{P}_n}}{|\overrightarrow{\mathbf{P}_h\mathbf{P}_n}|}$ as the ground truth for $\mathbf{h}$. Then we calculated the average error and standard variance for the tracking results of randomly picked 20 frames. For position, based on the Euclidean distance, the average error is around 4.41 millimeter and the standard variance of the error distance is around 1.61 millimeter. For orientation, based on the angle's cosine value of two vectors, the average error is around 0.05 and the standard variance is around 0.06.

4 Conclusion

Currently in a fish group with 6 members, each individual can be successfully tracked with a relatively high accuracy. However, the running speed is slow. In the future, we will focus on speeding up our tracking system to make it available for online tracking.

Acknowledgment: This work was supported by JSPS KAKENHI Grants 26540084 and 26240023.

References

[1] Sachit Butail and Derek A. Paley. Three-dimensional reconstruction of the fast-start swimming kinematics of densely schooling fish. *CVPR*, 2011.

[2] Ryo Kawahara, Shohei Nobuhara, and Takashi Matsuyama. A pixel-wise varifocal camera model for efficient forward projection and linear extrinsic calibration of underwater cameras with flat housings. In *ICCV Workshop on Underwater Vision*, pages 819–824, 2013.

[3] Jaco Vermaak and Arnaud Doucet. Maintaining multimodality through mixture tracking. In *ICCV*, pages 1110–1116, 2003.