Camera-Display System for the Interaction Analysis of Live Fish vs Fish-like Graphics

Hiroaki KAWASHIMA[1,a)] Yu KANECHIKA[1,b)] Takashi MATSUYAMA[1,c)]

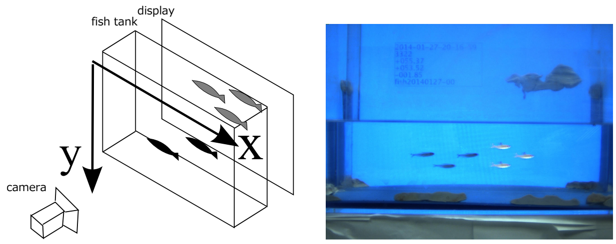
1. Introduction

This paper presents a novel camera-display approach to the analysis and control of fish schools.

Animals' collective behaviors are characterized by the dynamics of individuals and the interaction among them. Even though each individual interacts with its local neighbors based on some local interaction rules, a group of animals shows interesting behaviors as a whole [8], [9]. Among them, the analysis of fish schools is now particularly important since fish are often used for social behavior analyses in genetic and pharmacological studies. In addition, once we successfully understand the behavior of fish schools not only from a macroscopic aspect but also through the detailed modeling of a group, such as how information propagates through the individuals in the group, one will be able to predict and even control the entire group behavior by injecting external inputs from outside to a small number of constituent fish. The idea of controlling a group behavior of "agents" is now attracting scientists in a variety of fields including complex networks and control theory [5], [6].

One of promising approaches to the analysis of fish schools is to provide some stimuli to a fish group and observe the responses. Moving striped patterns are often used to elicit a schooling behavior [2] since the patterns cause fish an optomoter response (OMR), i.e., each fish tries to maintain a fixed visual field. However, the response depends on not only the stimuli but the state of fish. Thus, a real-time feedback control of stimuli becomes important to closely analyze the mechanism of the interaction. In fact, the use of a robotic fish replica was recently proposed to analyze fish behaviors, for example, how fish groups escape from a predator [10].

Our main idea of analyzing and controlling a live fish school is the use of artificial movements of "imitated" fish on a monitor screen. A video-based fish-group control has several advantages compared to existing methods: (i) a video has much more freedom than actual robots in terms that we can generate an arbitrary number of fish as graphical objects and move them in arbitrary patterns, and (ii) a video is suitable for a long-term real-time analysis compared to robots



(a) System configuration (b) Captured image

Fig. 1 Camera-display system. (a) shows the system configuration with the xy-coordinate system. (b) is an example of captured image, where black ones are live fish and the textured ones are fish-like graphics on the monitor screen.

in terms of endurance. The proposed video-based approach therefore enables us to design a variety of experimental settings in order to understand the underlying mechanism of a fish school, for example, via real-time video-based feedback control.

However a fundamental question "whether live fish actually form a school with fish-like graphic objects" is unrevealed in the literature. The objective of this paper is therefore to open up a new arena of computer vision applications by answering the question via the following contributions:

(1) We design a camera-display system with a simple tracking method to evaluate the video-based fish control.
(2) We provide a qualitative analysis to show whether live fish can be synchronized with moving fish-like graphics.

The requirements and design of the proposed system are discussed in Section 2, and the tracking method is shown in Section 3. The analysis of fish synchronization with video stimuli is provided in Section 4 and concluded in Section 5.

2. Camera-Display System for Analysis

The system should be designed to offer appropriate properties for future analyses, and we are particularly interested in the following requirements:

- The system has high fault-tolerance so that the models of fish (e.g., individual dynamics and interaction) can be estimated from long-term collected data using machine learning techniques.
- The system can be used for a variety of fish types with approximately 2-4cm length since (i) it allows us to use compact experimental environments, and (ii) many model fish species (zebrafish, medaka, etc.) in biological experiments have this size.

[1] Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo, Kyoto 6068501, Japan
[a)] kawashima@i.kyoto-u.ac.jp
[b)] kanetika@vision.kuee.kyoto-u.ac.jp
[c)] tm@i.kyoto-u.ac.jp

Based on the above requirements, we design a camera-display system as simple as possible. The point is that, since visual stimuli are moving objects on a screen, the important information of fish movements to be analyzed is almost parallel to the monitor plane. To analyze the two dimensional trajectories of fish, the system basically requires a single viewpoint. Nevertheless, the overlapping of fish may cause assignment errors. There are several choices to overcome the overlapping issue. One option is to utilize multi-viewpoint cameras, which requires calibration underwater. We however do not take this option since the configuration of the water tank can easily change due to the installation of a separator, etc., during each experiment. Instead, we here focus on using a single color camera with a simple tracking method to know its basic capability for future extensions (e.g., the use of time-of-flight cameras).

The proposed system consists of a fish tank, a camera, and a display monitor attached to one of the wide sides of the tank (Fig. 1 (a)). The monitor presents the movements of several fish-like graphics, where the side views of fish is used. The camera is mounted on the opposite side of the monitor, and it captures images containing both live fish and displayed fish. The spatial mapping of displayed positions on the screen to camera pixels is calibrated with grid points. A separator is installed in the fish tank so that the swimming area of fish can be narrowed.

The direction of the monitor placement is also an important issue. Considering the fact that a fish has a wide visual field in the horizontal direction but the visual field of the vertical direction depends on species, we choose to place the monitor screen not the bottom (or top) but the side of the fish tank. In the tracking and analysis, we use the xy-coordinate system depicted in Fig. 1 (b).

3. Multi-target Tracking with GMM

3.1 Requirements for Motion Analysis

Although the tracking technique is not the main focus of this paper, and in fact multi-target tracking techniques have been vastly studied, we here discuss what is an appropriate method for our particular goal. We first divide the existing tracking techniques into two categories in terms of how prior and observation are combined.

Prediction-oriented methods utilize the framework of sequential Bayesian filtering such as particle filter [4], [7] and unscented Kalman filter [10]. In these tracking methods, the estimation of the state at frame t can be inferred from the observation at time t and the predicted state given from previous frames. Detection-oriented methods, on the other hand, mainly rely on the observation at time t and try to find the best fit of the mixture model (e.g., Gaussian mixture mode, GMM) to the observation. An iterative procedure, such as the expectation-maximization (EM) algorithm or a gradient-based energy minimization method, is used for the fitting in each frame since the solution cannot be found as a closed form in many mixture models [1], [10], [11]. Here, these iterative procedures utilize the result in frame $t-1$, or its one-step prediction, as the initial values at t.

Since our goal is to analyze tracked data rather than to realize robust tracking, and we even consider the use of manual correction as a post processing, it is plausible to rely on the input data as strong as possible. While the balance between data term and prediction (prior) is determined by the parameters of variances in a sequential Bayesian filter, it sometimes suffers from the tuning of these parameters.

Besides, the appearance of every fish in our environment is similar and difficult to be distinguished based on their texture or color (see also Fig. 1 (b)). As a result, shape information becomes an important clue to separate overlapped targets. Indeed, we tested particle filters with several shape models, such as a trained model from actual fish images [1], and found that, for the side view of fish, a simple shape model with ellipses can yield comparable accuracy.

We therefore decided to extend an existing detection-oriented approach with GMM [11] as a simple and convenient tracker. Note also that we do not need to account for the change of the number of fish since we capture the whole area in the tank.

3.2 GMM-Based Tracking with a Dynamic Overlapping Measurement

Given a binary image B_t at frame t, where $B_t(x,y) \in \{0,1\}$ denotes whether the point (x,y) is a foreground pixel, 1, or not, 0, the method tries to fit a mixture of N Gaussian distributions to the sample set $\mathcal{P}_t = \{(x,y) \,|\, B_t(x,y)=1\}$. Here, each Gaussian corresponds to an ellipse-shape target. Namely, Gaussian component $\mathcal{N}(\mu_t^{(i)}, \Sigma_t^{(i)})$ represents target i at frame t, where $\mu_t^{(i)} \in \mathbb{R}^2$ is the mean vector and $\Sigma_t^{(i)} \in \mathbb{R}^{2\times 2}$ is the covariance matrix. These values can be estimated from $\mathcal{P}_t$ via the EM algorithm with initial values predicted from the previous frames, for example, via the Kalman filter [11]. Finally, the sequence of mean vectors, $\{\mu_t^{(i)}\}_t$, is considered as the trajectory of target i.

We here extend this GMM-based tracking by introducing the use of (1) a dynamic prior only when one target has other targets nearby and (2) a size prior for fish.

Let $\mu_{\mathrm{pred},t}^{(i)}$ be the predicted position from the previous frames and $\mu_{\mathrm{ML},t}^{(i)}$ be the original maximum likelihood (ML) estimation in a M-step. To include the predicted information not only for the initial value but also in the M-step of the EM algorithm, we consider the problem of model fitting as a maximum-a-posteriori (MAP) estimation instead of the ML estimation. Suppose the prior has a Gaussian distribution and simply assume that the covariance matrix is proportional to the solution by ML estimation. Then the MAP estimation of the position becomes

$$\mu_{\mathrm{MAP},t}^{(i)} = \alpha_t^{(i)} \mu_{\mathrm{pred},t}^{(i)} + (1-\alpha_t^{(i)}) \mu_{\mathrm{ML},t}^{(i)}, \quad (1)$$

where $\alpha_t^{(i)} \in [0,1]$ is a dynamically adjusted weight; we design the weight so that it becomes a large value when the prediction of target i is close to other targets' predictions.

The question here is how to measure the distance between two ellipse-like objects. While one option is to directly evaluate the overlapped areas of simulated ellipse pairs by changing their radii, an algebraic definition of distance is desirable for lower computational costs. However, the collision analysis of ellipses is difficult in general [3]. We therefore exploit the Bhattacharrya distance between two Gaussians,

$$D_{\mathrm{B}}(i,j,t) = \frac{1}{8}||\mu_t^{(i)} - \mu_t^{(j)}||^2_{\Sigma_t^{(ij)-1}} + \ln \frac{|\Sigma_t^{(ij)}|^{\frac{1}{2}}}{|\Sigma_t^{(i)}|^{\frac{1}{4}}|\Sigma_t^{(j)}|^{\frac{1}{4}}} \quad (2)$$
$$\text{where} \quad \Sigma_t^{(ij)} = (\Sigma_t^{(i)} + \Sigma_t^{(j)})/2,$$

to measure how close two ellipses are. Using this measure, we adjust the weights in Eq.(1) by a sigmoid function

$$\alpha_t^{(i)} = \alpha_{\max}(1 + \exp\{(\min_j D_{\mathrm{B}}(i,j,t) - D_{\mathrm{thres}})/\sigma_\alpha\})^{-1},$$

which takes almost $\alpha_{\max}$ when target i has other targets much closer than D_{thres}, and zero when target i is isolated.

A reasonable choice of D_{thres} is as follows. We first note that, when a target is a filled (i.e., uniformly distributed) ellipse with semi-major axis a, semi-minor axis b, and orientation θ, it can be shown by an integral calculation that these ellipse parameters correspond to the covariance Σ through

$$R(\theta)\,[\mathrm{diag}(a^2, b^2)]\,R(\theta)^T = 4\Sigma, \quad R(\theta) : \text{rotation matrix},$$

and hence $||(x,y)^T - \mu||_{\Sigma^{-1}} = 2$ is the ellipse equation corresponding to $\mathcal{N}(\mu, \Sigma)$. The condition when two ellipses i and j with the same covariance Σ contact each other is given by $||\mu^{(i)} - \mu^{(j)}||_{\Sigma^{-1}} = 4$. Substituting this into the first term of Eq. (2), and noting that the second term is zero for the equal Σ, we have $D_{\mathrm{B}} = 2$, which can be used for D_{thres}.

As a prior for fish size, we set the upper bound λ_{ub} and lower bound λ_{lb} on the eigenvalues of Σ so that the major and minor axis lengths of the detected ellipses are bounded. We also set the upper bound on trace(Σ), which avoids to fit large ellipses with the both axis lengths are close to λ_{ub}.

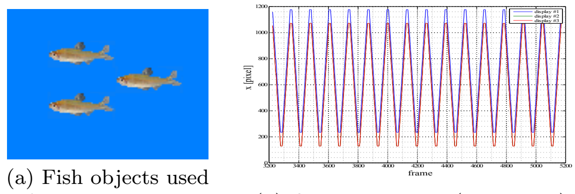
4. Analysis of the Synchronization between Fish Graphics and Live Fish

We conducted an experiment on how live fish respond to and are synchronized with the periodic fish motion generated on a monitor screen by using the proposed system.

4.1 Methods

Three rummy-nose tetras (*Hemigrammus bleheri*) with approximately 3cm size were installed in an acrylic fish tank (length 35cm × height 30cm × width 20cm), where the species was chosen because of the tight schooling behavior. From preliminary experiments, they tend to respond to visual stimuli from a display monitor (1920 × 1080 [pixel]) [*1] while they tend not to move when the number of individuals was less than three.

A captured texture image of the same species was used for fish graphics. The textured object was copied to generate a group with three individuals arranged in a triangle formation (Fig. 2 (a)). A periodic constant-velocity motion with a cycle length T_v in x-coordinate was used for the moving pattern of the fish objects on the screen, where the trajectories of their centroids are shown in Fig. 2 (b). The cycle length T_v depends on each session, and the number of cycles was 15. The motion includes 0.75 [sec] intervals at the peaks, and the direction of each object texture was flipped when they change the moving direction. A blue color was used for the background of the graphics, which was easily distinguished from the foreground objects. The monitor was turned on before each session, while the fish graphic objects appear and disappear suddenly in the beginning and the end of stimuli. The stimuli were started after 3 to 4 minutes each session began, and the lengths of the stimuli were approximately 2 minutes, which vary depending on the period T_v of the stimuli. Additional 4 to 5-minute data after the stimuli disappeared in each session were also used.



(a) Fish objects used for visual stimuli (b) Objects' motion (x position)

Fig. 2 Visual stimuli. (a) shows the arrangement of graphical objects used in dynamic visual stimuli. (b) is an example of the movements of three fish objects. Note that two of them with the same x positions are overlapped.

The movements of live fish were captured by a color camera (1296 × 964 [pixel]) [*2] in 15 fps. Three sessions of experiments were conducted, where $T_v = 7.5$ [sec] for session 1 and 3, and $T_v = 8.7$ [sec] for session 2 were used. In session 1 and 2, three individuals were randomly selected from other fish tank, while in session 2 and 3, the same set of individuals were used. From the captured image sequences, the 2-d trajectories of three fish were obtained semi-automatically: the movements of fish were first tracked automatically by the method described in the previous section and then corrected manually so that the analysis in this section does not contain assignment errors. Each sequence had 9000 frames and the number of errors (i.e., the occurrence of switching of fish IDs) was 31 times in 27000 frames (0.11 [%]); it was in an acceptable range for human to detect manually. Finally, the sequences of three fish ($i = 1, 2, 3$) were obtained in each session $s \in \{1, 2, 3\}$.

4.2 Measurements for synchronization analysis

The sequence of x-coordinate positions of each fish was used in the analysis since the movements in this direction were dominant compared to those in y direction. In particular, we evaluate how the degree of synchronization can be significantly increased during the presentation of the visual stimuli (in what follows, this interval is denoted as "on"), compared to before or after the presentation (these intervals are denoted as "before" and "after", respectively).

Let $f_v = 1/T_v$ [Hz] and ϕ_v [rad] be the major fre-

[*1] Iiyama ProLite T2735MSC

[*2] Point Grey Research, Chameleon CMLN-13S2C

quency and the phase of the fish object motion at session $s \in \{1,2,3\}$. Since these f_v and ϕ_v are known, the degree of the synchronization between the motion of the live fish and that of the presented stimuli can be evaluated by observing how both the frequency and its phase of live fish become close to f_v and ϕ_v, respectively. Specifically, to compare the degree of synchronization in two intervals, we define the following measurement. Let $\{x_t^{(i)}\}_t$ be the trajectory of fish i in x direction. The phase information of the live fish i can be evaluated by the cross-correlation of sinusoidal function and sequence $\{x_t^{(i)}\}_t$. We therefore define the following value in the interval $I_k = [b_k, e_k]$ with length $|I_k| = e_k - b_k + 1$:

$$\gamma_k^{(i)} = \frac{1}{|I_k|} \sum_{t=b_k}^{e_k} x_t^{(i)} \cos(2\pi f_v(T_d t) + \phi_v), \qquad (3)$$

where $T_d = 1/15$ [sec] is the sampling period of sequence $\{x_t^{(i)}\}_t$, and $k \in \{\text{before}, \text{on}, \text{after}\}$; session index, s, is omitted in the notations. Note that the value is normalized based on the length of the interval.

Hypothesis: The value $\gamma_k^{(i)}$ is significantly larger under the visual stimuli. We used Welch's t-test by assuming $\{c_t^{(i)} \mid t \in I_k\}$ is a normally distributed sample set, where $c_t^{(i)} := x_t^{(i)} \cos(2\pi f_v(T_d t) + \phi_v)$. In each session, we used two types of null hypotheses $\gamma_{\text{on}} = \gamma_k$, $k \in \{\text{before}, \text{after}\}$.

4.3 Results

We first qualitatively observe how the amplitude of frequency changes over time. Figure 3 shows the spectrogram (the short-term Fourier analysis) of the velocity of single fish motion in session 3. The frequency has strong peak around $f_v = 1/7.5$ [Hz] during the graphic objects' movement was presented ($I_{\text{on}} = [3584, 5259]$). However, in other session, such as in session 2 ($f_v = 1/8.7$ [Hz]), the fish motion took almost the same frequency as the presented stimuli during all the period in the session, which suggests the amplitude of the frequency is not adequate to closely analyze the fish response.

To quantitatively verify the synchronization between live fish and stimuli, we calculated the correlation $\gamma_k^{(i)}$ defined in Eq. (3) for each session $s \in \{1,2,3\}$, as shown in Fig. 4. From the Welch's t-test (5% significance), we found that $\gamma_{\text{on}}^{(i)}$ was significantly larger than $\gamma_{\text{before}}^{(i)}$ or $\gamma_{\text{after}}^{(i)}$ in 15 out of 18 (= 3 sessions × 3 individuals × 2 interval pairs).

While the number of experiments is still limited, these preliminary results indicate that it is possible to analyze and control live fish using fish-like graphic objects on a monitor screen. From a detailed analysis of each individual, we observed some fish did not only follow but tried to take a lead of the fish-like graphics. This causes the loss of synchronization and eventually makes the γ_{on} smaller. Meanwhile, this behavior indicates that a real-time feedback control needs to be incorporated.

5. Conclusion

We proposed a camera-display approach to the interaction analysis of fish schools, where we utilized the movements

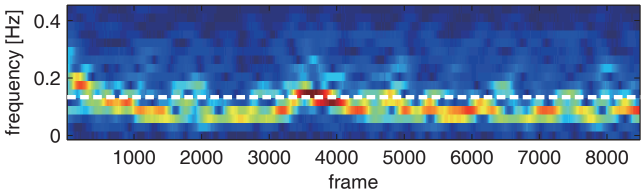


Fig. 3 Spectrogram of fish velocity (session 3, individual $i = 2$). The frequency of stimuli, $f_v = 1/7.5$ [Hz], is depicted by a white dashed line, where $I_{\text{on}} = [3584, 5259]$ [frame].

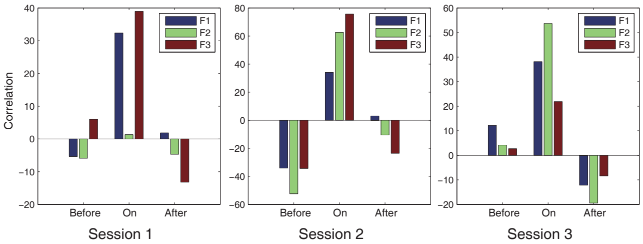


Fig. 4 Synchronization analysis. Each bar denotes the correlation $\gamma_k^{(i)}$ in each interval $k \in \{\text{before}, \text{on}, \text{after}\}$. F1, F2, and F3 denote fish IDs.

of fish-like graphics on the monitor and tracked fish motion patterns with a simple tracking algorithm. The experiments have shown that the movements of live fish can be significantly synchronized with the stimuli and the response can be quantified by the proposed framework. This results thus afford possibility for novel collective behavior analyses using video-based feedback control in future study.

Acknowledgment: This work was supported by JSPS KAKENHI Grant Numbers26540084.

References

[1] Cremers, D.: Dynamical Statistical Shape Priors for Level Set-Based Tracking, *IEEE Trans on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 8, pp. 1262–1273 (2006).

[2] Imada, H., Hoki, M., Suehiro, Y., Okuyama, T., Kurabayashi, D., Shimada, A., Naruse, K., Takeda, H., Kubo, T. and Takeuchi, H.: Coordinated and Cohesive Movement of Two Small Conspecific Fish Induced by Eliciting a Simultaneous Optomotor Response, *PLoS ONE*, Vol. 5, No. 6 (2010).

[3] Jia, X., Choi, Y.-K., Mourrain, B. and Wang, W.: An Algebraic Approach to Continuous Collision Detection for Ellipsoids, *Computer Aided Geometric Design*, Vol. 28, No. 3, pp. 164–176 (2011).

[4] Khan, Z., Balch, T. and Dellaert, F.: An MCMC-based Particle Filter for Tracking Multiple Interacting Targets, *ECCV*, pp. 279–290 (2004).

[5] Liu, Y.-Y., Slotine, J.-J. and A.-L.Barabasi: Controllability of complex networks, *Nature*, Vol. 473, No. 7346, pp. 167–173 (2011).

[6] Mesbahi, M. and Egerstedt, M.: *Graph Theoretic Methods in Multiagent Networks*, Princeton University Press (2010).

[7] Okuma, K., Taleghani, A., Freitas, N. D., Little, J. J. and Lowe, D. G.: A Boosted Particle Filter : Multitarget Detection and Tracking, *ECCV*, pp. 28–39 (2004).

[8] Partridge, B. L.: Internal Dynamics and the Interrelations of Fish in Schools, *Journal of Comparative Physiology A*, No. 144, pp. 313–325 (1981).

[9] Reynolds, C. W.: Flocks, herds, and schools: A distributed behavioral model, *Computer Graphics*, Vol. 21, No. 4, pp. 25–34 (1987).

[10] Swain, D. T., Couzin, I. D. and Leonard, N. E.: Real-Time Feedback-Controlled Robotic Fish for Behavioral Experiments with Fish Schools, *Proceedings of the IEEE*, Vol. 100, No. 1, pp. 150–163 (2012).

[11] Xiong, G., Feng, C. and Ji, L.: Dynamical Gaussian mixture model for tracking elliptical living objects, *Pattern Recognition Letters*, Vol. 27, pp. 838–842 (2006).