

A Multiple Video Camera System for 3D Tracking of Farmed Fry in an Aquaculture Tank

Koji Abe,^{1*} Shinichiro Kuroda,² and Hitoshi Habe¹

¹Faculty of Science and Engineering, Kindai University,
3-4-1 Kowakae, Higashiosaka, Osaka 577-8502, Japan

²Graduate School of Science and Engineering, Kindai University,
3-4-1 Kowakae, Higashiosaka, Osaka 577-8502, Japan

(Received March 30, 2020; accepted June 17, 2020)

Keywords: image analysis, 3D tracking, farmed fry, visualization, aquaculture

A system for the 3D tracking of underwater farmed fry using action cameras in multiple directions is presented. In a real environment, owing to light reflection and the small size of fry relative to the size of the tank, blind spots and unclear figures often appear in pictures taken by a camera. The proposed system continuously monitors the fry's figure clearly and avoids blind spots by using multiple cameras. The proposed method changes the combination of multiple viewing points from frame to frame in a movie depending on the fry's location to maintain clear fry figures and obtains the 3D coordinates of the location by the direct linear transformation. The regions of fry are sequentially obtained by taking the difference between the background image and each frame where fry appear. By taking movies of some farmed fry, we examined the tracking performance by comparing the 3D coordinates of the locations obtained by the proposed method with the correct coordinates. The mean and the standard deviation of the distance between the two points were 2.619 and 1.333 cm, respectively, i.e., it was confirmed that the proposed method can correctly capture the locations of fry because the distance was smaller than the body length of the fry.

1. Introduction

In the research field of fisheries science, to enhance the quality of aquaculture, the observation of fish in real time has an important role. For example, movements of several free-swimming fish are observed in water tanks under different environments, and their growth is compared to examine differences in the response of the fish to the environments.⁽¹⁾ Also, to develop better fish feed, different test feeds are provided in different water tanks under the same environment, and differences in the response of the fish to the feeds and the movement of fish are observed for a long time.⁽²⁾ However, continuous observation for a long time is difficult for researchers, and a system for tracking fish could be a useful tool because it could analyze patterns of fish movement automatically and find characteristics of the movement in different situations on behalf of human eyes. Moreover, unlike in normal recording, since researchers

*Corresponding author: e-mail: koji@info.kindai.ac.jp
<https://doi.org/10.18494/SAM.2020.2890>

could record movements quantitatively, they could quantitatively compare differences in the movements of fish.

It is possible to track fish by various means if there are no conditions on the recording or sensing environment. However, most tracking systems depend on the environment. For example, a method of tracking fish by directly attaching sensors onto their body cannot be applied to small fish such as fry or larvae.⁽³⁾ Although methods of tracking fish using sound navigation and ranging (sonar) stereo cameras have been proposed as contactless tracking tools,^(4–6) all of them can only reproduce the locus of fish movement in 2D space. Although there are reports on a system for tracking fish in 3D space,^(7–10) the performance of the system presented by Yoneyama *et al.*⁽⁷⁾ has not been evaluated for small fish such as fry, and the system can track fish using only two cameras in the case when fish swim in only a small part ($60 \times 60 \times 60 \text{ cm}^3$) of a water tank. Hence, the system cannot track all the fish in a tank and cannot track fish while viewing the whole tank. In other previously reported systems,^(8–10) the target fish to be tracked must be captured by a pair of cameras mounted from the top of the tank and pointing downward under an ideal environment, where the water depth must be small to catch the figure of the fish clearly by the cameras and there must be no blind spots of the cameras. This means that the tracking could not work when one or both of the cameras lose the figure, and the tracking performance has not been examined in the case when the figure is unclear. Under a real aquaculture environment, owing to light halation and the small size of fry relative to the side of the aquaculture tank, the conditions are unsuitable for tracking using cameras owing to blind spots and low-contrast views. When the figure of an object is unclear in images, the 3D coordinates of the object cannot be precisely obtained. Therefore, it would be difficult to apply the existing 3D tracking methods under such a real aquaculture environment.

In this paper, with the aim of developing a tool for quantitatively analyzing fish movement as discussed above, a 3D tracking system for free-swimming small fish is proposed. Considering that the proposed system is expected to be applied in a real aquaculture environment where halation appears on the water surface and the figures of fish sometimes become unclear, the tank is monitored from multiple directions using six action cameras to avoid blind spots and maintain clear figures of the fry. A few small fish can be simultaneously tracked by altering the combination of multiple viewing points from frame to frame in the moving images, thus reproducing the movement of the fish in a virtual 3D space.

2. Recording Environment

Cylindrical transparent tanks are often employed in fisheries research using indoor water tanks. Hence, it is difficult to mount cameras for tracking fish outside the tank owing to curved-surface refraction. To take a movie of the whole cylindrical tank avoiding such refraction, it is necessary to mount cameras at several positions inside the tank. Figure 1(a) shows a cylindrical tank regularly used in research on aquaculture at the Faculty of Agriculture in Kindai University. Figure 1(b) shows the configuration of the proposed recording system, where six action cameras (GoPro HERO5 Black) record the tank together and all of them are synchronized. The viewing angle is set as the wide-angle mode. The recorded movie is 24-bit

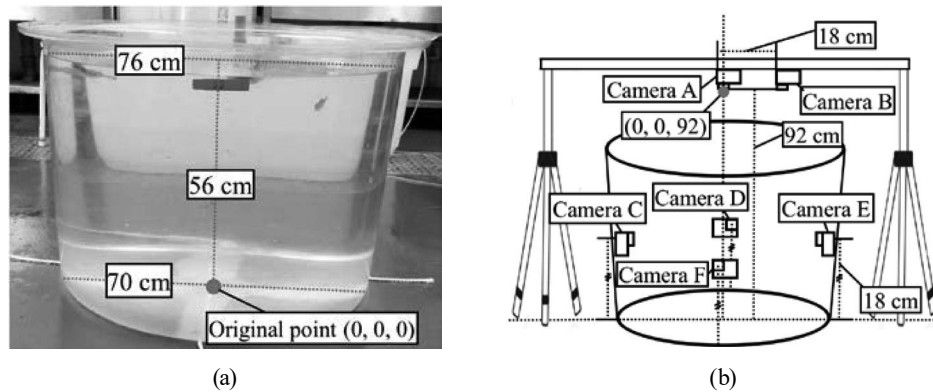


Fig. 1. Water tank and recording environment. (a) Cylindrical tank. (b) Configuration of recording environment.

full color in MP4 format and its frame rate is 60 fps. The image size is 1000×750 pixels (W/H). By mounting the six cameras as shown in Fig. 1(b), it is possible for at least one of the two downward cameras (A, B) and one of the four cameras installed on the side (C–F) to always catch fish. Figure 2 shows a frame in a moving image taken by each camera. As shown in Fig. 1(a), the origin in 3D space $(0, 0, 0)$ is defined as the central point on the tank's bottom. The origin of all frames in the moving image $(0, 0)$ is defined as the location of the pixel at the top left of each frame.

In this research, a water tank where three farmed fry (length: about 3–4 cm) swim freely is prepared for the 3D tracking. Since the recording is conducted in accordance with a real environment used for examinations in fisheries science, three fish can be regarded as a standard number. The tank is not aerated. Tiny floating objects are always present in the water and could become noise in image processing.

3. Proposed Method

3.1 Preprocessing

First, all the frames in moving images captured by the six cameras are converted into 8-bit grayscale images. When tracking multiple fish at the same time, it is difficult to capture the fish promptly in moving images at the six cameras after the tracking has started. Hence, the user next decides the start time for the tracking by selecting the target fish with mouse clicks on the frames obtained from Cameras A and B at the same time. At that time, the user first clicks the multiple targets one by one in the frame for Camera A and then clicks the targets in the frame for Camera B in the same selection order, and the frames become the first frame for the tracking. In addition, the frames for the other four cameras at the same time similarly become the first frame for the tracking. In the two frames for Cameras A and B in which the user has selected the fish, the coordinates of the location for the k th selected fish are defined as (x_{kA}, y_{kA}) and (x_{kB}, y_{kB}) , respectively. In this paper, k is limited to 1, 2, or 3 according to the recording situation.

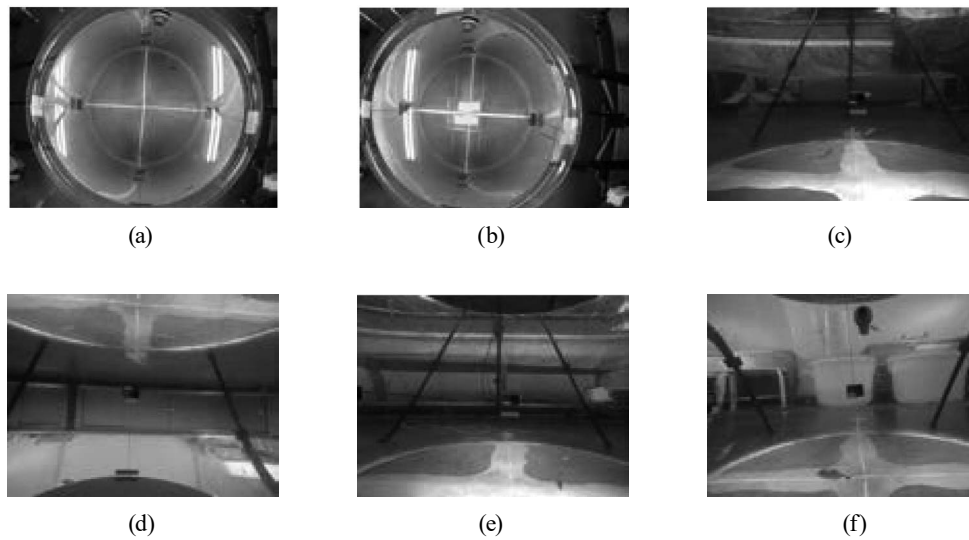


Fig. 2. Frame of moving image taken by each camera. (a) Camera A. (b) Camera B. (c) Camera C. (d) Camera D. (e) Camera E. (f) Camera F.

3.2 Extraction of fish regions

To obtain the coordinates of fish locations in 3D space, the fish regions are extracted from the images obtained by each camera. Although the extraction of objects from an image is generally to find the difference between the image where the object appears and its background image, the background of the water tank is continually changing in a real environment. Hence, the background image is formed by the method of time median filtering.⁽¹¹⁾ The time median filter forms the background image by choosing the median value among the pixel values obtained at every pixel in several frames of a movie. In this research, the set of frames for the filter is introduced by separating a movie into 21 segments at regular intervals and choosing one frame from every segment. Then, if the difference between the pixel value of pixel $b(x, y)$ in the background image and the pixel value at the same location in the n th grayscale frame is more than 8, the pixel in the frame is converted into a white pixel; otherwise, it is converted into a black pixel. In addition, the closing processing of the erosion and dilation is applied to the black-and-white image three times, and the obtained image is defined as the candidate image of the fish region, F_n . F_n is obtained from the movies taken by all six cameras.

When Fig. 3(a) is the first grayscale frame of a movie taken by Camera A, F_1 for Fig. 3(a) becomes the image shown in Fig. 3(b). As shown in Fig. 3(b), since many noises appear in the candidate images taken by Cameras A and B, the noises are removed as follows. For example, when $n = 1$, i.e., the original frame is Fig. 3(a) and F_1 is Fig. 3(b), the image of the fish region, R_1 , is obtained by extracting only white regions that include the pixel (x_{kA}, y_{kA}) or (x_{kB}, y_{kB}) . Figure 3(c) shows the image R_1 obtained from F_1 . R_n at the n th frame except for the first frame is obtained as follows. For example, when $n = 2$, the image E_1 is obtained by applying the dilation to R_1 just once, and if a white pixel in F_2 is a white pixel at the same location in E_1 , the white pixel in F_2 is extracted as a pixel in R_2 . Figure 4(a) shows the result of the dilation

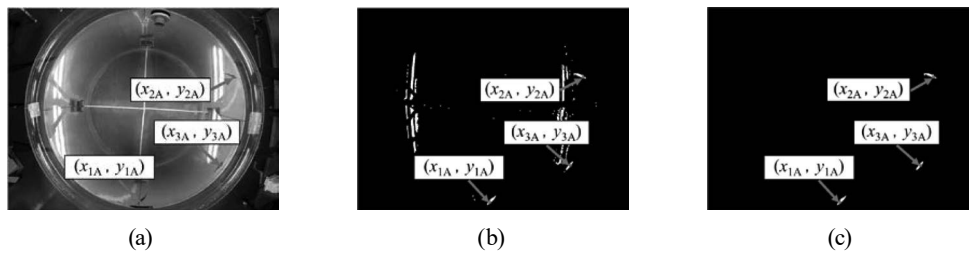


Fig. 3. Extraction of fish regions for the first frame in the movie taken by Camera A. (a) First frame (taken by Camera A). (b) Candidate image F_1 . (c) Image of fish region R_1 .

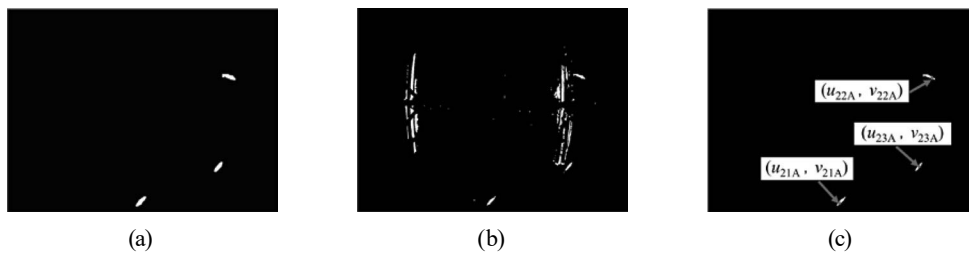


Fig. 4. Extraction of fish regions for the second frame in the movie taken by Camera A. (a) Image E_1 . (b) Candidate image F_2 . (c) Image of fish region, R_2 .

applied to the image in Fig. 3(c), Fig. 4(b) shows F_2 obtained from the next frame of the image shown in Fig. 3(a), and Fig. 4(c) shows R_2 for F_2 shown in Fig. 4(b). After $n = 2$, this extraction is conducted throughout the movie. As shown in Fig. 4(c), since the fish regions hardly move between two consecutive frames when taking a movie with a high frame rate of 60 fps, R_n could be obtained correctly by using R_{n-1} as a query. The initial location of the fish is (x_{kA}, y_{kA}) or (x_{kB}, y_{kB}) , and the tracking point of the fish location in the image taken by Camera i (i represents one of A–F) after $n = 2$ becomes the centroid point of each fish region (u_{nki}, v_{nki}) .

Although halation areas appear in Fig. 4(b), the locations of halation are considerably different between Cameras A and B as shown in Figs. 2(a) and 2(b), i.e., one of them can always capture the fish without including a halation area. On the other hand, the other four cameras (Cameras C–F) do not always capture fish at the start time of the tracking. As shown above, R_n cannot be obtained without R_{n-1} where fish regions appear. Hence, regarding the movies taken by these four cameras, R_n starts to be obtained from the frame after that in which the fish regions first appear.

3.3 Distortion correction of 2D images

As shown above, since all the cameras take a movie with the viewing angle in the wide-angle mode, the 2D image is always distorted in the camera calibration and the distortion brings errors to the conversion from 2D to 3D coordinates. Hence, before the conversion, the distortion in the 2D images is corrected by the methods of “undistort” and “undistortPoints” in the OpenCV library.⁽¹²⁾ These methods need at least 10 images of a chessboard taken from different directions. Figure 5 shows the distortion correction for one of the chessboard images,

where Fig. 5(b) shows the result of correcting the image shown in Fig. 5(a). Table 1 shows the difference between the coordinates at sample points before and after the correction of an image (1000 × 750 pixels). As shown in Table 1, there is hardly any difference at the central point in the image but a large difference at the points located far from the central point, such as the original point (0, 0) or the top right corner (999, 749).

3.4 Conversion from 2D to 3D coordinates by DLT

The 3D coordinates of fish locations are obtained by direct linear transformation (DLT)^(13,14) using the coordinates of the centroid points of the fish regions in images taken by multiple cameras at the same time. In the camera calibration for the DLT, m control points must be obtained in advance, and the vertices of a cubic frame at a fixed position in real space are often used as the points. The 3D coordinates (X_a, Y_a, Z_a) in real space and the 2D coordinates (u_{ia}, v_{ia}) at control point a (a represents 1, 2, ..., or m) in the image taken by Camera i are substituted into Eq. (1), where the 11 weight factors used for converting 2D coordinates into 3D coordinates are obtained as P_{i1} – P_{i11} .

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_{i1}X_1 & -u_{im}Y_1 & -u_{i1}Z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_m & Y_m & Z_m & 1 & 0 & 0 & 0 & 0 & -u_{im}X_m & -u_{im}Y_m & -u_{i1}Z_m \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_{i1}X_1 & -v_{i1}Y_1 & -v_{i1}Z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & X_m & Y_m & Z_m & 1 & -v_{im}X_m & -v_{im}Y_m & -v_{im}Z_m \end{bmatrix} \begin{bmatrix} P_{i1} \\ \vdots \\ P_{i11} \end{bmatrix} = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{im} \\ v_{i1} \\ \vdots \\ v_{im} \end{bmatrix} \quad (1)$$

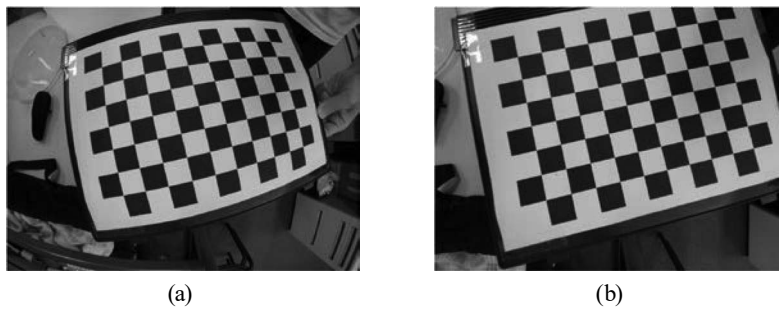


Fig. 5. Chessboard used in the distortion correction. (a) Before distortion correction. (b) After distortion correction.

Table 1
Difference between 2D coordinates at sample points in an image before and after the distortion correction.

	Coordinates in the original image	Coordinates after the correction	Difference
Original point	(0, 0)	(-205, -151)	(205, 151)
	(300, 225)	(280, 210)	(20, 15)
Central point	(500, 375)	(499, 375)	(1, 0)
	(700, 525)	(717, 539)	(17, 14)
	(999, 749)	(1198, 900)	(199, 151)

In the DLT, a point (u_i, v_i) in the 2D image taken by Camera i is represented by Eqs. (2) and (3). To show the relationship between the coordinates (u_i, v_i) and the location in 3D space corresponding to the coordinates (X, Y, Z) , Eq. (4) is obtained from Eqs. (2) and (3).

$$u_i = \frac{XP_{i1} + YP_{i2} + ZP_{i3} + P_{i4}}{XP_{i9} + YP_{i10} + ZP_{i11} + 1} \quad (2)$$

$$v_i = \frac{XP_{i5} + YP_{i6} + ZP_{i7} + P_{i8}}{XP_{i9} + YP_{i10} + ZP_{i11} + 1} \quad (3)$$

$$\begin{pmatrix} P_{i1} - u_i P_{i9} & P_{i2} - u_i P_{i10} & P_{i3} - u_i P_{i11} \\ P_{i5} - v_i P_{i9} & P_{i6} - v_i P_{i10} & P_{i7} - v_i P_{i11} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} u_i \\ v_i \end{pmatrix} \quad (4)$$

For example, when a point in the 2D images taken by Cameras A and B is represented as (u_A, v_A) and (u_B, v_B) , respectively, the relationship between the coordinates of the point in 3D space (X_{AB}, Y_{AB}, Z_{AB}) and the two 2D coordinates is represented by rewriting Eq. (4) as follows.

$$\begin{pmatrix} P_{A1} - u_A P_{A9} & P_{A2} - u_A P_{A10} & P_{A3} - u_A P_{A11} \\ P_{A5} - v_A P_{A9} & P_{A6} - v_A P_{A10} & P_{A7} - v_A P_{A11} \\ P_{B1} - u_B P_{B9} & P_{B2} - u_B P_{B10} & P_{B3} - u_B P_{B11} \\ P_{B5} - u_B P_{B9} & P_{B6} - u_B P_{B10} & P_{B7} - u_B P_{B11} \end{pmatrix} \begin{pmatrix} X_{AB} \\ Y_{AB} \\ Z_{AB} \end{pmatrix} = \begin{pmatrix} u_A - P_{A4} \\ v_A - P_{A8} \\ u_B - P_{B4} \\ v_B - P_{B8} \end{pmatrix} \quad (5)$$

By writing the first matrix on the left side as L and the matrix on the right side as R in Eq. (5), Eq. (6) is obtained.

$$L \begin{pmatrix} X_{AB} \\ Y_{AB} \\ Z_{AB} \end{pmatrix} = R \quad (6)$$

Finally, the coordinates of the point in 3D space (X_{AB}, Y_{AB}, Z_{AB}) are obtained by rewriting Eq. (6) as

$$\begin{pmatrix} X_{AB} \\ Y_{AB} \\ Z_{AB} \end{pmatrix} = (L^T L)^{-1} (L^T R). \quad (7)$$

In this research, a transparent cubic frame with a side of 15 cm and 15-cm-long wires were prepared in order to define the 16 control points as shown in Fig. 6.

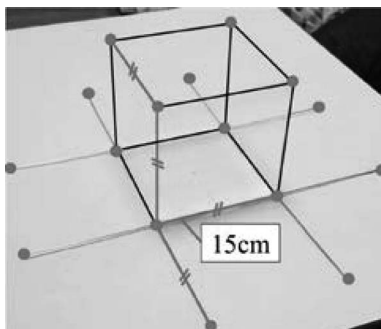


Fig. 6. 16 control points prepared for the calibration.

3.5 Detection of fish locations in images taken by Cameras C–F

In the recording environment, when a fish is hidden by halation or remains near the bottom of the tank, sometimes one of Cameras A and B, which view the fish from above, cannot capture the fish region. To keep obtaining the 3D coordinates of a fish location in the movies, the fish regions viewed from above are matched to those viewed from the side. That is, when fish regions disappear in an image taken by one of the two overhead cameras (Camera A or B), the images taken by the other cameras not used to obtain the 3D coordinates (Cameras C–F) are used to obtain the 3D coordinates as follows. First, the k th fish selected in the images taken by Cameras A and B at the beginning is identified in the images taken by the other cameras. For example, by substituting the 3D coordinates (X_{AB}, Y_{AB}, Z_{AB}) obtained from the images taken by Cameras A and B to (X, Y, Z) into Eqs. (2) and (3) and also $P_{C1}-P_{C11}$, $P_{D1}-P_{D11}$, $P_{E1}-P_{E11}$, and $P_{F1}-P_{F11}$, the 2D coordinates (u_C, v_C) , (u_D, v_D) , (u_E, v_E) , and (u_F, v_F) are obtained from (X_{AB}, Y_{AB}, Z_{AB}) . After that, if the obtained point (u_i, v_i) in the image taken by Camera i is included in a white region of the candidate image F_n , the white region is regarded as the area of the fish located at (X_{AB}, Y_{AB}, Z_{AB}) . The fish locations in 2D images taken by Cameras C–F are first obtained from the second frame in the movie, then found continuously until the fish disappears from the frame, and the location is obtained again when the fish appears in the frame again. Therefore, if a fish disappears in the image taken by Camera A or B, the tracking can be conducted by using the 2D coordinates of the fish location in the image taken by at least one of the other four cameras. Thus, in this 3D tracking of fish by multiple cameras, the location of a fish in 3D space is renewed in each frame by repeatedly increasing and decreasing the number of viewing points.

4. Interface and Results of the 3D Tracking for Farmed Fry

In the proposed method, the user has to input two sets of information by clicking the mouse. The first is the location of every control point in 2D images. Since there are six cameras and 16 control points in this system, the user inputs the locations of the 16 points in an image taken by each camera in the same order. Once this calibration of the tank is finished, the user does

not have to input these points again. The other information is the number of fish to be tracked and their locations in the first frame by clicking the mouse as shown in Sect. 3.1. The proposed system equips an interface for inputting the two sets of information as well as another interface to display the 3D tracking of the fish. Figure 7 shows the interface used to display the 3D tracking and an example of tracking three fish, where Figs. 7(a)–7(c) are the tracking results at the starting time $t = 0$, $t = 2$, and $t = 4$ (s), respectively. The user can freely change the viewing point of the 3D space by dragging the mouse on the interface. To examine the performance of the proposed system, a movie of 5 min 50 s was taken at each of the six viewing points.

During the first 2 min 30 s (9000 frames) after the starting time determined by the user in the movie, the percentages of frames used for the DLT at each of the viewing points were examined. Table 2 shows the percentage of frames. About 80% of all the frames from Cameras A and B were used for the DLT and about 30% of the frames from the other four cameras were used to assist the DLT.

5. Discussion

The tracking performance could be examined precisely if the correct location for each fish were obtained in every frame by attaching a sensor onto the fish body. However, since the target fish were fry, a sensor could not be attached onto the fish body. In addition, the fry used in experiments in fisheries research must not receive any stimulation from outside because a stimulation may affect their movement and introduce uncertainty in experimental results. In research on computer vision, to obtain the correct coordinates of a target without sensors, a standard in the space such as the ground truth is often introduced.⁽¹⁵⁾ However, it would be difficult to introduce a standard in the space because the target is underwater and very small relative to the tank. In other reports on fish tracking,^(16,17) the correct tracking trajectory was drawn manually while monitoring the target, and experimental results were compared with the trajectory. Thus, since it is difficult to examine the fish tracking performance quantitatively,

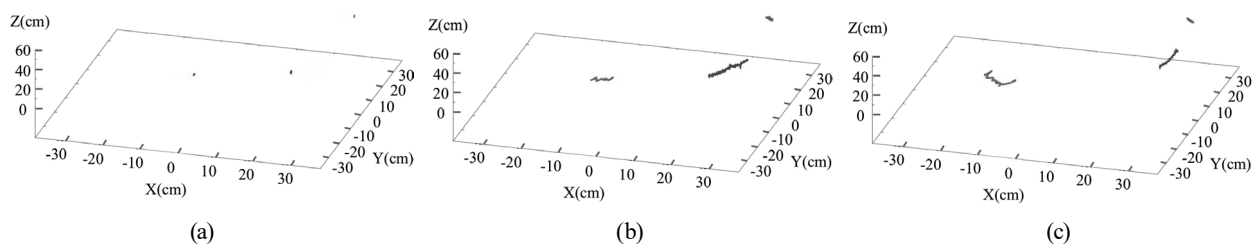


Fig. 7. Situation of 3D tracking from 0 to 4 s for three fish. (a) $t = 0$ (s). (b) $t = 2$ (s). (c) $t = 4$ (s).

Table 2

Percentage of frames used for DLT during the recording time of 150 s (9000 frames) vs each camera.

Viewpoint	Camera A	Camera B	Camera C	Camera D	Camera E	Camera F
Frames from the movie used in DLT (%)	80.3	86.7	37.2	33.5	34.1	31.5

the performance of the proposed 3D tracking was examined by using the locations of the control points, the tank size, and other factors, as shown below.

To examine the difference between the location of the fish in 3D space obtained by the proposed method and the correct point, it is necessary to obtain the correct point (X, Y, Z) for the estimated point of the target fish in a frame of a movie. However, the correct point can only be obtained at limited locations in the recording environment. The correct point is obtained in two ways that involve measuring the 3D coordinates manually. In one way, the correct coordinates of a fish location are obtained by measuring the fish location at the height of the side cameras. First, the correct Z coordinate of the fish location becomes 18 cm, which is equal to the height of the side cameras. Second, the correct X and Y coordinates of the fish location are obtained as follows. In the image of the transparent cubic frame taken by Camera A, the point on each of the two vertical sides of the frame at a height of 18 cm from the bottom of the tank is chosen by clicking the mouse manually. In the view from Camera A, the cubic frame appears thinner toward the bottom of the tank. In fact, the coordinates of the two points were $(-7.5, -7.5, 18.0)$ and $(7.5, -7.5, 18.0)$. The 2D coordinates of the corresponding two points in the image taken by Camera A were $(447, 427)$ and $(552, 430)$, respectively. The correct X and Y coordinates can be obtained by measuring the distance between the fish location and the two points. The other way of obtaining the correct coordinates of the fish location is to measure them when the fish is located at the water surface. Since the height of the surface is 50 cm, the correct Z coordinate becomes 50 cm. Next, the location of fish in the image taken by Camera A is chosen by clicking the mouse manually, and the correct X and Y coordinates are obtained by measuring the distance between the clicked point and the central point of the image $(500, 375)$ and considering the ratio between the distance and the real diameter of the tank at the water surface. The difference between the 3D coordinates of the fish location obtained by the proposed method and the correct coordinates was examined. Ten points were selected from the movie of three free-swimming fish to obtain the correct coordinates. Table 3 shows the results of the measurement for the 10 points. From Table 3, at two points near the bottom of the tank where the fry's figure was unclear and at the other eight points near the water surface where halation appears, it was confirmed that the proposed method can correctly obtain the 3D coordinates of the fry. The mean and the standard deviation of the difference for the 10 points were 2.619 and 1.333 cm, respectively, and the maximum difference was less than 4.7 cm.

Table 3
Difference between the coordinates obtained by the proposed method and the correct coordinates.

Point No.	Obtained 3D coordinates (cm)	Correct coordinates (cm)	Difference (cm)
1	(25.467, -2.297, 17.568)	(28.194, -4.166, 15.971)	3.671
2	(-3.316, -0.696, 15.181)	(2.083, -0.694, 15.971)	1.455
3	(-23.000, -23.854, 52.530)	(-24.376, -22.598, 50.000)	3.141
4	(33.488, -8.256, 50.640)	(32.999, -9.130, 50.000)	1.188
5	(1.958, -34.653, 47.469)	(0.314, 31.102, 50.000)	4.660
6	(33.488, -8.256, 50.640)	(32.047, 1.181, 50.000)	2.098
7	(-13.128, 31.467, 50.180)	(-11.732, 30.000, 50.000)	2.025
8	(-22.789, 20.301, 49.832)	(-21.496, 19.842, 50.000)	1.382
9	(16.883, -6.558, 47.039)	(15.433, -6.771, 50.000)	3.303
10	(33.461, 12.854, 47.915)	(32.441, 10.551, 50.000)	3.269

Next, we examined how the proposed tracking was conducted in a time sequence. Regarding an arbitrary time in the movie as the starting point of the tracking, the direction that the fish moved in 3D space is compared with that observed by human eyes. Three starting points for a fish were chosen and the difference was examined in the X , Y , and Z directions, respectively. Tables 4–6 show the change in the location for each of the three points in a time sequence, where the 3D coordinates obtained by the proposed method and the difference in fish location from the coordinates before 30 frames are shown in the first frame, the 30th frame, the 60th frame, and the 90th frame in the movie. In addition, the increase or decrease in the change in the X , Y , and Z directions was judged visually and their results are shown in the “Increase/Decrease” column in Tables 4–6. “No change” means that the observer could not observe a change in that direction. If the change in the location obtained by the proposed method is almost the same as the observation in the “Increase/Decrease” column in each direction of 3D space, “Same” is shown in the tables. Otherwise, “Different” is shown. As shown in Tables 4–6, the direction of the 3D tracking was basically the same as that observed by human eyes at the three points.

Throughout the movie recorded by the proposed system, the tracking was lost only once. Figure 8 shows the frames captured by Cameras A and B at that time and the frames after 20 and 40 frames, where (a) and (d) are the frames at that time, (b) and (e) are the frames 20 frames later, (c) and (f) are the frames 40 frames later, and the figure of the fish appears in the

Table 4

Comparison between direction of fish movement and observation by human eyes (starting point 1).

Frame No.	Obtained 3D coordinates (cm)	Difference in fish location from location before 30 frames (cm)	Increase/Decrease			Same or Different
			X axis	Y axis	Z axis	
0	(−3.208, −0.450, 16.182)	—	—	—	—	—
30	(−1.282, −1.956, 14.335)	(1.926, −1.506, −1.847)	Increase	Decrease	No change	Same
60	(0.350, −3.506, 15.999)	(3.558, −1.550, 1.664)	Increase	Decrease	No change	Same
120	(5.390, −6.668, 16.537)	(8.598, −3.162, 0.538)	Increase	Decrease	No change	Same

Table 5

Comparison between direction of fish movement and observation by human eyes (starting point 2).

Frame No.	Obtained 3D coordinates (cm)	Difference in fish location from location before 30 frames (cm)	Increase/Decrease (visually)			Same or Different
			X axis	Y axis	Z axis	
0	(−29.777, −30.934, 37.425)	—	—	—	—	—
30	(−32.959, −4.267, 37.462)	(−3.182, 3.333, −0.047)	Decrease	Increase	No change	Same
60	(−37.145, −36.831, 37.472)	(−4.186, −2.564, −0.183)	Decrease	Decrease	No change	Same
120	(−29.538, −38.225, 37.472)	(7.607, −1.394, 0.355)	Increase	Decrease	No change	Same

Table 6

Comparison between direction of fish movement and observation by human eyes (starting point 3).

Frame No.	Obtained 3D coordinates (cm)	Difference in fish location from location before 30 frames (cm)	Increase/Decrease (visually)			Same or Different
			X axis	Y axis	Z axis	
0	(37.832, −9.103, 31.246)	—	—	—	—	—
30	(38.479, −6.999, 30.665)	(0.647, 2.104, −0.581)	No change	Increase	No change	Same
60	(38.707, −5.561, 30.778)	(0.228, 3.542, −0.113)	No change	Increase	No change	Same
120	(38.200, 2.177, 32.246)	(−0.507, 7.738, 1.468)	No change	Increase	No change	Same

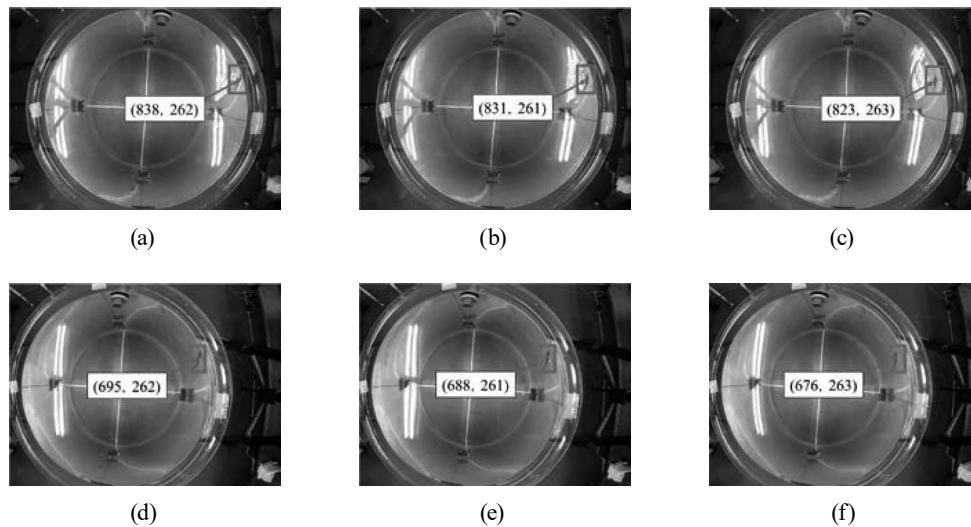


Fig. 8. Sequential frames at the tracking loss at the viewing points of Cameras A and B. (a) Frame at tracking loss (Camera A). (b) 20 frames later (Camera A). (c) 40 frames later (Camera A). (d) Frame at tracking loss (Camera B). (e) 20 frames later (Camera B). (f) 40 frames later (Camera B).

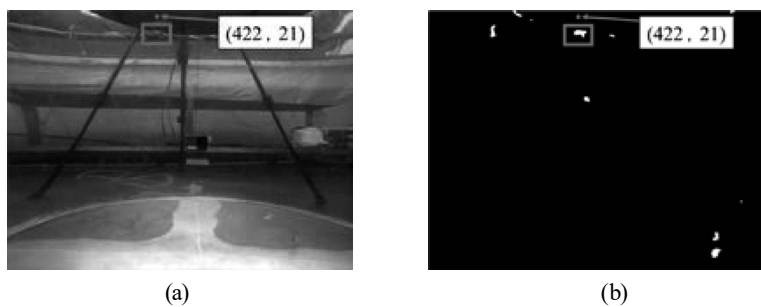


Fig. 9. Frame at the viewing point of Camera C at the tracking loss. (a) Frame at tracking loss. (b) Candidate image of fish regions.

red rectangle. By comparing Figs. 8(a)–8(c) and Figs. 8(d)–8(f), it can be confirmed that the centroid point of the fish region shifted to the left side, although a fish cannot swim to produce such a change. At that time, since a wave appeared on the water surface in the horizontal direction in the frames, the fish region was shifted irregularly in the same direction and the wave caused the tracking loss. The 3D coordinates obtained from the 2D coordinates (838, 262) in Fig. 8(a) and (695, 262) in Fig. 8(d) were (29.367, 11.717, 55.054), i.e., the z -coordinate was above the height of the water tank. Then, although only Camera C captured the same fish among the side viewing points, as shown in Fig. 9(a), the 2D coordinates obtained from the shifted 3D coordinates became (422, 21) in the frame shown in Fig. 9(a), which were out of the fish region, as shown in Fig. 9(b). Thus, the proposed tracking method failed when the location of the fish was irregularly shifted by an external factor in the frames taken by both Cameras A and B. In the future, it will be necessary to cope with such shifts.

6. Conclusions

In this paper, we presented a system for the 3D tracking of multiple small fish at the same time by setting action cameras in different directions under a real aquaculture environment with blind spots and the appearance of unclear figures of fry in the moving images taken by the cameras. The multiple cameras were equipped to continuously record clear figures of the fry and avoid the blind spots caused by halation. In the proposal, after fish regions are obtained by difference processing between the background image and each frame taken by some of the cameras, the 3D coordinates of the location of fish are obtained by the DLT method. The combination of cameras that take an image to obtain the 3D coordinates of the fry changes from frame to frame depending on the location of the fry. The performance of the proposal was examined by tracking three small fish of 3–4 cm size to evaluate the precision of the estimated 3D coordinates by the proposal and the correspondence of the direction between the 3D tracking by the proposed method and by human eyes. From the experimental results, it was confirmed that the mean and the standard deviation of the difference between the estimated point and the correct point were 2.619 and 1.333 cm, respectively. Moreover, the locus of the movement obtained by 3D tracking with the proposal was the same as that observed with human eyes. As future work, it will be necessary to develop a countermeasure for the occurrence of waves on the water surface to prevent tracking loss and to propose a method of automatically detecting fish regions.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant No. 20K11899.

References

- 1 A. S. Vila, S. Coughlan, M. D. Guiry, and S. Kraan: *J. Appl. Phycol.* **21** (2009) 617. <https://doi.org/10.1007/s10811-009-9423-z>
- 2 J. S. F. Lee and B. A. Berejikian: *J. Fish Biol.* **72** (2008) 1736. <http://www.blackwell-synergy.com>
- 3 M. Kadota, E. J. White, S. Torisawa, K. Komeyama, and T. Takagi: *PLoS One* **6** (2011) e28241. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3229534/>
- 4 H. Fukuda, S. Torisawa, Y. Sawada, and T. Takagi: *Aquaculture* **305** (2010) 73. <https://www.sciencedirect.com/science/article/abs/pii/S0044848610002723>
- 5 S. Torisawa, T. Takagi, H. Fukuda, Y. Ishibashi, Y. Sawada, T. Okada, S. Miyashita, K. Suzuki, and T. Yamane: *J. Fish Biol.* **71** (2007) 411. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1095-8649.2007.01498.x>
- 6 M. Kadota, S. Torisawa, T. Takagi, and K. Komeyama: *Rev. Fish. Sci. Aquacult.* **77** (2011) 993. <https://link.springer.com/article/10.1007/s12562-011-0415-y>
- 7 K. Yoneyama, S. Kunisawa, H. Sotozono, K. Kotani, A. Imamura, and T. Matuoka: *Math. Phys. Fish. Sci.* **12** (2015) 51 (in Japanese). http://jsmpfs.org/12_2015/Komeyama_2015.pdf
- 8 Z. M. Qian, X. E. Cheng, and Y. Q. Chen: *PLoS One* **10** (2014) 1. <https://doi.org/10.1371/journal.pone.0106506>.
- 9 G. Wu and L. Zeng: *Sci. China, Ser. G* **50** (2007) 779. <https://link.springer.com/article/10.1007/s11433-007-0071-5>
- 10 Using Track3D for Fish: <https://www.noldus.com/track3d/fish> (accessed May 2020).
- 11 K. Miyake, M. Yoneda, H. Hase, M. Sakai, and H. Maruyama: *J. Inst. Image Electr. Eng. Jpn.* **30** (2001) 251 (in Japanese). <https://doi.org/10.11371/iiej.30.251>
- 12 OpenCV 2.4.13.7 Documentation: https://docs.opencv.org/2.4/search.html?q=undistort&check_keywords=yes&area=default (accessed March 2020).

- 13 Y. L. Adbel-Aziz and H. M. Karara: *Photogramm. Eng. Remote Sens.* **81** (2015) 101. <https://www.sciencedirect.com/science/article/pii/S009911215303086>
- 14 F. Gazzani: *J. Biomed. Eng.* **14** (1992) 476. [https://doi.org/10.1016/0141-5425\(92\)90099-7](https://doi.org/10.1016/0141-5425(92)90099-7)
- 15 Q. Wang, F. Chen, W. Xu, and M. H. Yang: *IEEE Trans. Image Process.* **21** (2012) 4454. <http://azadproject.ir/wp-content/uploads/2013/07/Object-Tracking-via-Partial-Least-squares-analysis.pdf>
- 16 D. Sharma and R. K. Singh: *Int. J. Adv. Technol. Eng. Res.* **8** (2018) 31. http://www.ijater.com/Files/21c3a216-8d2b-416a-abfa-edefd90bb836_IJATER_42_05.pdf
- 17 C. Spampinato, S. Palazzo, D. Giordano, I. Kavasidis, F. P. Lin, and Y. T. Lin: *Int. Conf. Computer Vision Theory and Applications* **2** (2012) 409. <https://www.scitepress.org/papers/2012/38666/38666.pdf>

About the Authors



Koji Abe received his B.S. and M.S. degrees from Kogakuin University, Japan, in 1996 and 1998, respectively, and his Ph.D. degree in engineering from Kanazawa University, Japan, in 2001. He was affiliated with the Institute for Image Data Research, University of Northumbria at Newcastle, UK, as an honorary research fellow in 2002. He became an assistant professor in Kanazawa Institute of Technology in 2003, then a lecturer in the Department of Informatics, Faculty of Science and Engineering, Kindai University, in 2006. He has been an associate professor in Kindai University since 2010. His research interests include pattern recognition, medical image processing, CBIR, multimedia databases, and artificial intelligence. He is a member of IEEE, IEICE, IEEJ, and IIEEJ. (koji@info.kindai.ac.jp)



Shinichiro Kuroda received his B.E degree from Kindai University in 2019. He is currently a master's student at the Graduate School of Science and Engineering, Kindai University. His research interests include pattern recognition and image processing. He is a member of IEICE. (shin.96chah108@gmail.com)



Hitoshi Habe received his B.E. and M.E. degrees in electrical engineering and his D. Info. degree in intelligence science and technology from Kyoto University, Japan, in 1997, 1999, and 2006, respectively. After working for Mitsubishi Electric Corporation, Kyoto University, NAIST, and Osaka University, he is now an associate professor at the Department of Informatics, Faculty of Science and Technology, Kindai University, Japan. From 2010 to 2011, he was a visiting researcher at the Department of Engineering, University of Cambridge, UK. His research interests include computer vision, pattern recognition, and image processing. He is a member of IEEE, ACM, IEICE, IPSJ, JSFS, and JSAI. (habe@kindai.ac.jp)