# Automated three-dimensional measurement method of *in situ* fish with a stereo camera


Jun Han, Akira Asada
Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, JAPAN
Hideyuki Takahashi, Kouichi Sawada
National Research Institute of Fisheries Engineering, Fisheries Research Agency
Hasaki 7620-7, Kamisu, Ibaraki, 314-0408, JAPAN



***Abstract*-An acoustic-optical integrated measurement deep-tow system called J-QUEST (the Japanese quantitative echo sounder and stereo-video camera system), consisting of a quantitative echo sounder and two synchronized cameras, has been developed for the observation of mesopelagic fishes such as lanternfish by the Fisheries Research Agency, Japan. The quantitative echo sounder is designed for measuring fish target strength and assessing the fish abundance, and the optical cameras are aimed for identifying fish species, measuring fish tilt angle distribution, counting and sizing fish that passing through the camera views. In order to obtain the distributions of fish length and tilt angle, a commercial-release optical image processing software has been used so far. However, the measurement, which has to be carried out frame by frame manually, is time-consuming and labor intensive. This paper presents a fully automated three-dimensional method to measure *in situ* fish using image pairs taken by the two cameras.**


## I. Introduction

Acoustic method is considered to be the most effective quantitative way to estimate the biomass in the oceans. To apply the acoustic assessing method, it is necessary to clarify two variables that greatly affect the presumption accuracy. One is fish species identification, and another is the mean target strength. In order to solve this problem, the Fisheries Research Agency developed an acoustic-optical integrated measurement deep-tow system called J-QUEST (the Japanese quantitative echo sounder and stereo-video camera system) [1]. The system consists of a quantitative echo sounder and a stereo-video camera, as shown in Fig. 1. The quantitative echo sounder is designed for measuring fish target strength and assessing the fish abundance, and the optical cameras are aimed for identifying fish species, measuring fish tilt angle distribution, counting and sizing fish that passing through the camera views.

In order to obtain the distributions of fish length and tilt angle, a commercial release optical image processing software has been used so far. However, the measurement, which has to be carried out frame by frame manually by selecting the fish head and tail with mouse clicks, is time-consuming and labor intensive, and its accuracy is inadequate because the fish length is calculated by the distance of the straight line between the fish head and fish tail. This paper presents a fully automated three-dimensional method to measure in situ fish using image pairs taken by the two cameras.

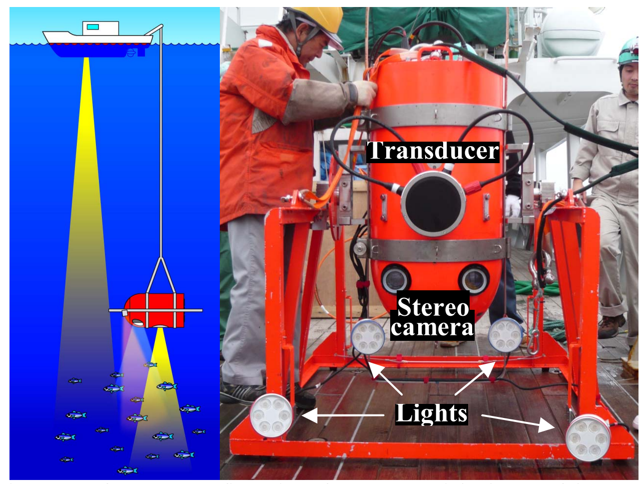



Figure 1. Left: system configuration of J-QUEST; Right: layout of the stereo system.

## II. Materials and Methods

### A. Camera calibration

It is necessary to obtain the camera parameters by calibration when a camera is used for a measuring system. We employed the direct linear transformation (DLT) method [2]

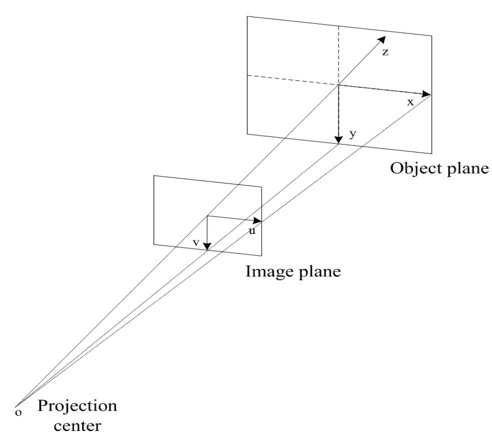



Figure 2. The pinhole camera model in positive position.


978-1-4244-5222-4/10/$26.00 ©2010 IEEE

for camera calibration because of its simplicity for field uses.

Fig.2 depicts the pinhole camera model in positive position. Based on the DLT method, a world coordinate (x, y, z) can be transformed to image coordinate (u, v) by (1).

$$\begin{pmatrix} u\cdot s \\ v\cdot s \\ s \end{pmatrix} = \begin{pmatrix} L_1 & L_2 & L_3 & L_4 \\ L_5 & L_6 & L_7 & L_8 \\ L_9 & L_{10} & L_{11} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (1)$$

Here $L_{1\text{-}11}$ and $s$ are camera parameters and scale factor respectively. This equation can be reformulated by division to

$$u = \frac{L_1 x + L_2 y + L_3 z + L_4}{L_9 x + L_{10} y + L_{11} z + 1}$$
$$v = \frac{L_5 x + L_6 y + L_7 z + L_8}{L_9 x + L_{10} y + L_{11} z + 1} \quad (2)$$

By using $n$ point pairs, the following over-determined linear equations can be formed.

$$\begin{pmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -u_1x_1 & -u_1y_1 & -u_1z_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -v_1x_1 & -v_1y_1 & -v_1z_1 \\ & & .. & .. & .. & .. & .. & & & & \\ x_n & y_n & z_n & 1 & 0 & 0 & 0 & 0 & -u_nx_n & -u_ny_n & -u_ny_n \\ 0 & 0 & 0 & 0 & x_n & y_n & z_n & 1 & -v_1x_1 & -v_ny_n & -v_nz_n \end{pmatrix} \begin{pmatrix} L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \\ L_8 \\ L_9 \\ L_{10} \\ L_{11} \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \\ \dots \\ u_n \\ v_n \end{pmatrix} \quad (3)$$

For short, it can be rewrote as

$$A \cdot L = B, \quad (4)$$

where $A$ is a rectangular matrix of size $2n \times 11$, $B$ is a column vector of length $2n$, and $L$ is a column vector of length 11 containing the parameters $L_{1\text{-}11}$.

The optimal solution $L$ of these over-determined linear equations can be obtained with the method of least squares according to

$$L = \left(A^T A\right)^{-1} A^T B, \quad (5)$$

where $\left(A^T A\right)^{-1} A^T$ is the Moore Penrose pseudoinverse of $A$. After the camera parameters $L_{1\text{-}11}$ are obtained, the image coordinate ($u$, $v$) for any world point ($x$, $y$, $z$) can be calculated by using (2). By contraries, if we know the image coordinate ($u$, $v$), then the world coordinates ($x$, $y$, $z$) corresponding with the image point can be determined by the following under-determined equations.

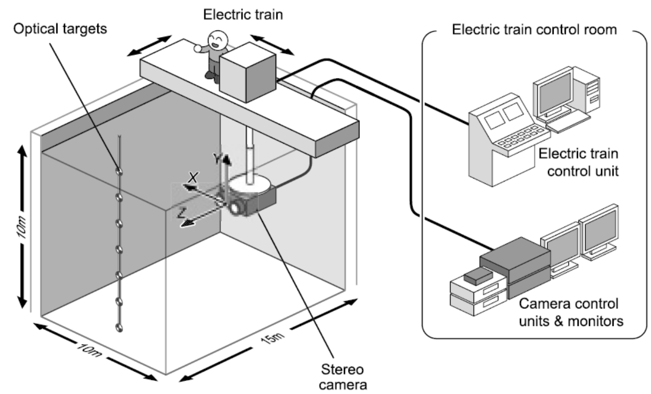



Figure 3. System configuration for the camera calibration conducted in the water tank in the Fisheries Research Agency, Japan. The stereo camera is fixed on the electric train, which can move on the rail over the tank. The optical targets are set on a wire with an interval of 10 cm, and the camera positions are controlled by a PC with 0.1 cm accuracy.

$$\begin{pmatrix} L_9u - L_1 & L_{10}u - L_2 & L_{11}u - L_3 \\ L_9v - L_5 & L_{10}v - L_6 & L_{11}v - L_7 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} L_4 - u \\ L_8 - v \end{pmatrix} \quad (6)$$

For modeling lens distortions, many approaches exist. We use the most famous model (Brown model). According to Brown model, the image offsets $du$ and $dv$ due to lens distortions can be expressed as

$$du = a(L_{12}r^2 + L_{13}r^4 + L_{14}r^6) + L_{15}(r^2 + 2a^2) + L_{16}ab$$
$$dv = b(L_{12}r^2 + L_{13}r^4 + L_{14}r^6) + L_{15}ab + L_{16}(r^2 + 2b^2), \quad (7)$$

where $a = u - u_0$ and $b = v - v_0$. By replacing $u$ with $u+du$ and $v$ with $v+dv$ in (3), we can obtain the radial distortion parameters $L_{12\text{-}14}$ and the tangential distortion parameters $L_{15\text{-}16}$ with iterative calculations.

The camera calibration experiment was conducted in the water tank (L10m×W10m×D15m) in the Fisheries Research Agency. The camera was fixed on the electric train, which can move on the rail over the tank. Usually a checkerboard is used for camera calibrations. As it cannot be easily fixed in the water, we specially made optical rig shown in Fig. 3 for

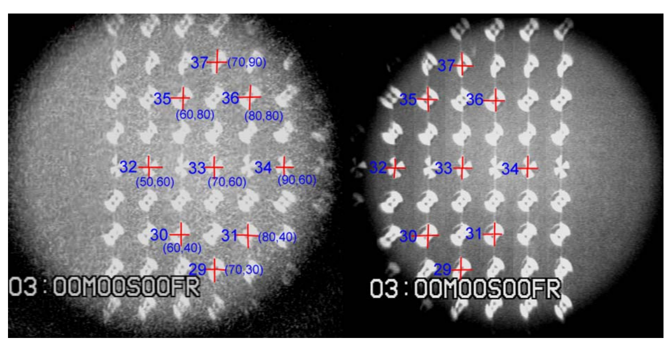



Figure 4. The combined image examples taken at a distance of 3 m. The points marked with "+" were used in the calibration calculations.

underwater use. The optical targets are set on a wire with an interval of 10 cm. The camera moved in the direction of the front, back, left and right, and its positions are controlled by a PC with 0.1cm accuracy. On each position the images are taken and recorded for 10 seconds.

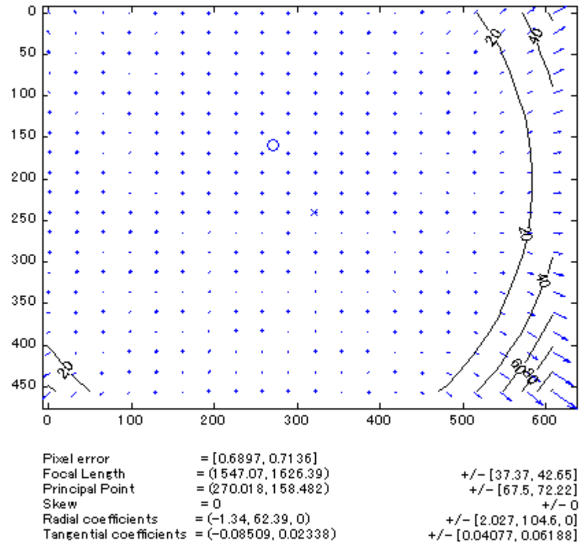



Figure 5. Left camera distortion model.

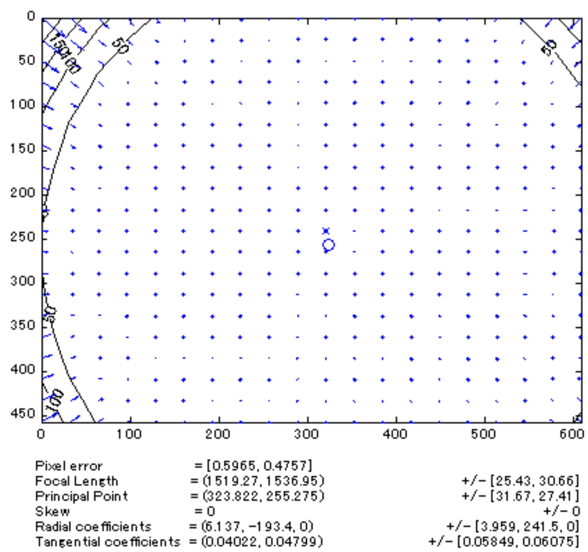



Figure 6. Right camera distortion model.

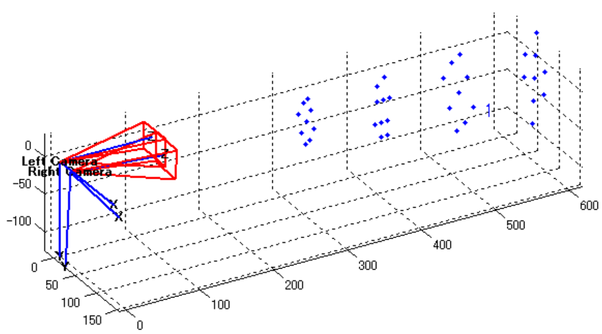



Figure 7. Stereo camera layout and the position data used for calibrations.

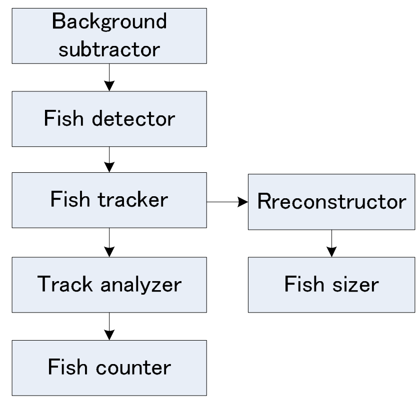



Figure 8. Flowchart for the automated three-dimensional measurement system.

The combined images taken from the distance of 3 m are shown in Fig. 4.

With these calibration image data we performed camera calibrations. The complete distortion model for the left and right cameras are shown in Fig. 5 and Fig. 6 respectively. The stereo camera extrinsic parameters (layout) and the position data used for calibrations are shown in Fig. 7. The mean reprojection errors in $x$, $y$ and $z$ are $0.38 \pm 0.35$, $0.23 \pm 0.53$ and $4.03 \pm 6.28$ cm respectively.

### B. *Automated three-dimensional measurement system*

As shown in Fig. 8, the proposed automated three-dimensional measurement system integrates seven modules: (1) a background subtractor, which estimates and removes the background; (2) a fish detector, which detects and extracts candidate fish regions in the input image utilizing region, edge, and shape information; (3) a fish tracker, which performs prediction and filtering of the fish motion dynamics using a biologically Kalman filter algorithm; (4) a track analyzer, which analyzes and updates an fish track; (5) a fish counter, which oversees the entire tracking history and counts the disappearing fish; (6) a reconstructor, which employs a direct linear transformation (DLT) method to determine the three-dimensional coordinates of a point from two-dimensional views of the point so as to obtain fish model coordinates and (7) a fish sizer, which evaluates the body length by calculating the centerline of the candidate fish. The first five modules process two-dimensional image pairs in pixel respectively.

### C. *Background subtraction*

Background subtraction is frequently carried out to detect moving objects in videos obtained using a static camera. Several methods have been proposed for performing background subtraction [3]-[4]. In all these methods, a background model is developed from the temporal sequences of the frames. In this case, we model the background using the following formula [5]

$$I = \bar{I} + A\sin(\omega t) + \Gamma, \quad (8)$$

where $I$ , $\bar{I}$ , $A$ and $\Gamma$ indicate the image intensity, image mean intensity, the amplitude of the background and the noise level respectively. If the intensity of a pixel is within $\bar{I} - A - \Gamma \leq I \leq \bar{I} + A + \Gamma$, then it is taken as background. $\bar{I}$ and $A$ of the background and foreground are updated using

$$\begin{aligned} \bar{I} &= \frac{(n-1)}{n}\bar{I} + \frac{1}{n}I \\ A &= \frac{(n-1)}{n}A + \frac{1}{n}\sqrt{2(I-\bar{I})^2}. \end{aligned} \tag{9}$$

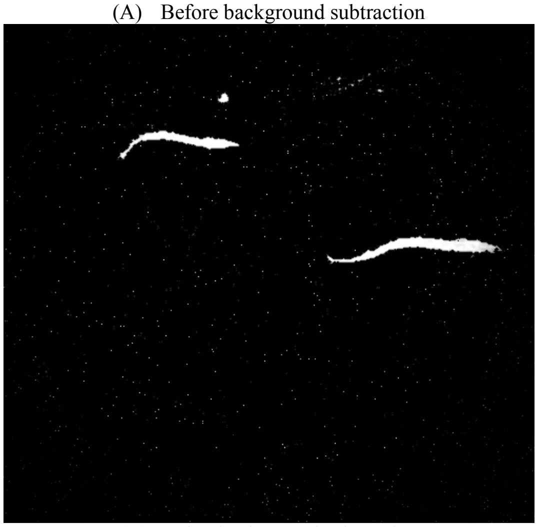
and

$$\begin{aligned} \bar{I} &= I \\ A &= \frac{(m-1)}{m}A + \frac{1}{m}\sqrt{2(I-\bar{I})^2} \end{aligned} \tag{10}$$

respectively, where $m \geq n$ .

Example images before and after background subtraction are shown in Fig. 9.


(A) Before background subtraction



(B) After background subtraction

Figure 9. Example images before and after background subtraction.

*D. Fish detector*

A black and white binary image is obtained after background subtraction. Only the white pixels are considered, and the others are treated as background. To eliminate isolated pixels, morphological filtering (dilation and erosion) is performed using a 3 × 3 mask. An inner boundary tracing method is used to trace the edges. The area of the objects is then calculated. The noise, the area of which is obviously smaller than that of the fish, is then removed. After the above steps have been carried out, the remaining areas in the image are considered to be targets. The centroid of an object is calculated by using moments.

*E. Fish tracker*

In order to prevent the multiple counting of fish and to eliminate the echo from fallen leaves, tracking the moving fish is essential for image processing. From this viewpoint, the detection of candidate fish is attempted. Identifying the same candidate fish in subsequent images is a problem. Image-matching techniques are commonly used to solve this type of problem. However, in this case, the echo images of the candidate fish are not clear, because the fish are in motion; therefore, the image-matching techniques are not effective. The prediction and filtering of the motion dynamics of the fish are carried out using the Kalman filter algorithm [6]-[7]. Detail information about Kalman filter algorithm for fish tracking can be found in [8].

*F. Track analyzer and fish counter*

The track analyzer compares the outputs of the fish detector and the fish tracker, and takes one of the following actions: it creates a new track segment, updates an existing track, or terminates a track. For continuing track segments, the track analyzer updates the fish motion state, and predicts its state for the next frame. For new track segments, initial motion states are initialized. The fish counter oversees the entire tracking history, establishes the complete fish trajectories, and counts the disappearing fish, as shown in Fig.10 .

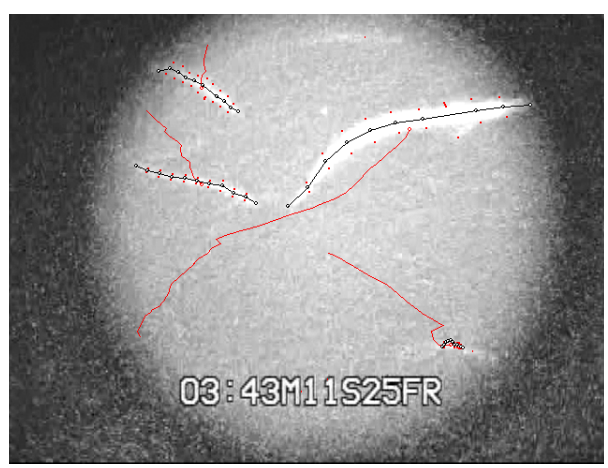
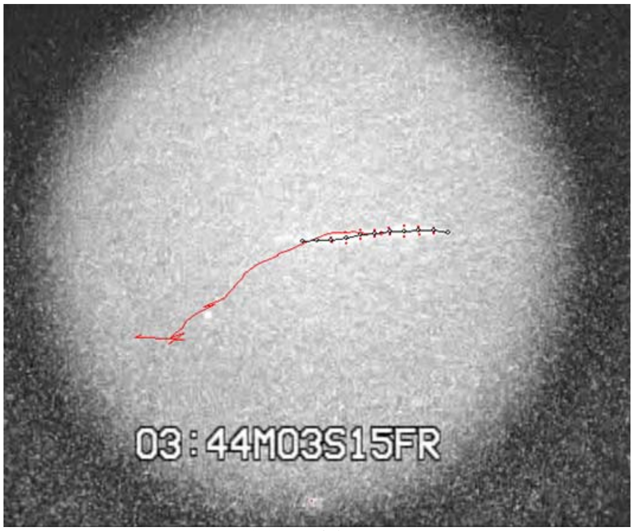
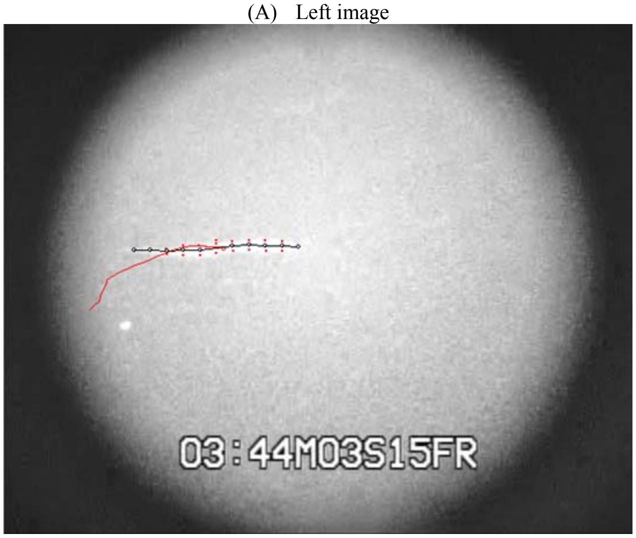



Figure 10. Example of tracking fish trajectories.




(A) Left image




(B) Right image

Figure 11. The fish centerline extracted in the image pair.

### G. *Reconstructor and fish sizer*

In order to measure fish size accurately, we need to reconstruct the fish. The stereo reconstruction has been widely studied and many algorithms exist, however, most of them are based on stereo image matching. As to the images taken in the underwater, it is difficult to find out the correspondence point pairs between the left and right images. Therefore we contrived a matching method for sizing fish.

First, the fish sizer calculates the minimum bounding box and then extracts the centerline of the candidate fish polygon by connecting the midpoint of equally spaced section lines along the thinner side of the minimum bounding box in the image pair (Fig. 11). According to epipolar geometry [9], a point on the left fish center line must be on the correspondence epipolar line in the right image, so we can determine the correspondence point by calculate the cross point of the right fish center line and the epipolar line in the right image. After the correspondence point pairs are obtained, the three- dimensional positions on the fish center line can be calculated.

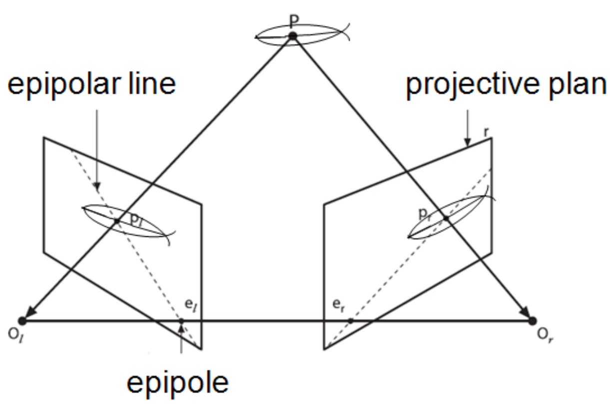



Figure 12. Epipolar geometry for calculating fish center line.

## III. RESULT AND FUTURE WORK

An automated three-dimensional method to measure *in situ* fish using image pairs taken by a stereo camera has been developed. It can be used to calculate the distributions of the fish length and the tilt angle.

As future work, automated fish species identification will be challenged and the applications will be conducted.

## REFERENCES


[1] H. Takahashi, K. Sawada, Y. Takao and K. Abe, "Inspection of the three dimensional measurement accuracy of J-QUEST Stereo TV-camera system," Technical Report of the National Research Institute of Fisheries Engineering, vol. 27, pp. 81-92, 2005 (in Japanese).

[2] Y.I. Adbel-Aziz, H. M. Karara, "Direct linear transformation from comparator coordinates into objects space in close-range photogrammetry," Symposium on Close-Range Photogrammetry, Falls Church, VA, pp.1-18.

[3] C. Zhang, S.C. Chen, M.L. Shyu and S. Peeta, "Adaptive background learning for vehicle detection and spatio-temporal track," ICICS-PCM, pp. 15-18, Singapore, 2003.

[4] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.19(7), pp.780–785, 1997.

[5] S. Morita, K. Yamazawa, M.Terazawa and N.Yokoya, "Networked remote surveillance system using omnidirectional image sensors," *J. of the Institute of Electronics, Information and Communication Engineers (J. IEICE)*, vol.J88-D-II(5), pp.864-875, 2005.

[6] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transaction of the ASME J. of Basic Engineering*, pp.35–45, 1960.

[7] H. W. Soreson, "Least-squares estimation from Gauss to Kalman," *IEEE Spectrum 7*, pp. 63–68, 1970.

[8] J. Han, N. Honda, A. Asada and K. Shibata, "Automated acoustic method for counting and sizing farmed fish during transfer using DIDSON," *Fisheries Science*, vol. 75, pp. 1359–1367, 2009.

[9] H. Richard, *Multiple View Geometry in Computer Vision,* 2nd ed., Cambridge University Press, 2003, pp.65-86.