Advances in Computer, Signals and Systems (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2024.080117
ISSN 2371-8838 Vol. 8 Num. 1

Research on Largemouth Bass Target Recognition and Tracking Utilizing the Optical Flow Approach

Chen Murui[1,a,*], Zhou Zheng[1,b], Tang Wenkai[1,c], Lai Junlin[1,d]

[1]College of Marine Engineering, Jiangsu Ocean University, Lianyungang, Jiangsu, China
[a]2021221303@jou.edu.cn, [b]2021221330@jou.edu.cn, [c]2022220315@jou.edu.cn, [d]2022221703@jou.edu.cn

Keywords: Optical Flow, Largemouth Bass, Target Identification, Motion Tracking, Camera Calibration, Binocular Vision, Domestication Experiment

Abstract: This research focuses on the application of the optical flow method for target identification and tracking of largemouth bass in underwater environments. The optical flow field, representing pixel movement on a two-dimensional plane, is utilized to analyze grayscale variations in image sequences. The study employs the basic variational optical flow model and introduces the Horn-Schunck (HS) algorithm to address challenges such as the aperture problem. Additionally, the paper explores camera selection and calibration, emphasizing the importance of accurate calibration parameters for binocular vision systems. A domestication experiment is designed for California black bass, incorporating sound stimuli and optical flow-based motion tracking. The study concludes with an evaluation of fish domestication effects using motion trajectory and speed analysis.

1. Introduction

The motion of objects in space can be described using a motion field, which, when projected onto a two-dimensional plane, is manifested through the different expressions of grayscale distribution in an image sequence. Optical flow is the instantaneous velocity of pixel movement on a two-dimensional plane, and the motion field in space is represented as the optical flow field on the image plane. The optical flow field is a collection of optical flows for each pixel in the image, forming a two-dimensional vector field that reflects the grayscale variations of all pixels in the image. Ideally, the optical flow field is a dense displacement vector field, and based on this displacement vector field, all points in the first image can be mapped to their corresponding positions in the second image. Optical flow is widely applied in computer vision fields such as target recognition, object tracking, motion estimation, motion segmentation, anomaly detection, video frame interpolation, obstacle detection and avoidance, and assisted driving.

The basic variational optical flow calculation method relies on two fundamental assumptions: the brightness constancy assumption and the small motion assumption. Assuming that at time t, the grayscale value of a point on the image is I(x, y), and after a time interval, it moves to a new position (x', y'), with the new grayscale value denoted as I'(x', y'). According to one of the fundamental assumptions, the brightness constancy assumption, the grayscale values at the initial and final positions of the point are equal:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \tag{1}$$

Expanding the right side of equation (1) using Taylor series yields:

$$I(x, y, t) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt + \varepsilon \tag{2}$$

Here, ε represents the second-order infinitesimal term, which can be ignored under the small motion assumption. Substituting equation (2) into equation (1) and dividing both sides by dt, we get:

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t}\frac{dt}{dt} = 0 \tag{3}$$

Let u and v represent the velocity vectors of optical flow along the x and y directions, respectively:

$$u = \frac{dx}{dt}, v = \frac{dy}{dt} \tag{4}$$

Let $I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t}$ denote the partial derivatives of pixel intensity in the image along the x and y directions. Therefore, equation (2) can be written as:

$$I_x u + I_y v + I_t = 0 \tag{5}$$

Equation (5) is the well-known basic constraint equation for optical flow. However, this equation contains two unknowns, u and v, while there is only one equation, leading to the aperture problem in optical flow estimation.

To address the aperture problem and obtain a solution for optical flow, additional constraints must be introduced. Traditional optical flow algorithms fall into two major categories based on the type of constraints introduced: local methods and global methods[1], represented by the Lucas-Kanade (LK) and Horn-Schunck (HS) optical flow algorithms, respectively[2]. LK-type algorithms rely on parameter registration within a local window, resulting in sparse optical flow. HS-type algorithms, on the other hand, depend on the estimation of a global representation to find a vector field that aligns images and satisfies certain regularity constraints. In 1981, Horn and Schunck conducted research on the aperture problem in the basic optical flow model and proposed the classic Horn-Schunck optical flow algorithm, abbreviated as the HS algorithm. The HS algorithm introduces a global smoothness constraint as an additional constraint, requiring that the optical flow values in the given neighborhood of the current pixel transition smoothly. The HS algorithm transforms the optical flow field's problem into the minimization problem of an objective energy function:

$$E(u, v) = \iint (I_x u + I_y v + I_t)^2 + \lambda^2 (|\nabla u|^2 + |\nabla v|^2) dxdy \tag{6}$$

Where:

$$|\nabla u|^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 \qquad |\nabla v|^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 \tag{7}$$

In the objective energy function formula (6), there are two components: the data term, (Ixu + Iyv

+ It)^2 derived from the optical flow basic equation (5), and the smoothness term, $(|\nabla u|^2 + |\nabla v|^2)$ from the HS algorithm's global smoothness constraint, where λ>0 is the regularization parameter, and a larger value of λλ results in a smoother optical flow field. By solving the minimization problem of equation (6), iteratively computed optical flow values can be obtained.

Let:

$$L(u,v,u_x,u_y,v_x,v_y) = (I_x u + I_y v + I_t)^2 + \lambda^2(|\nabla u|^2 + |\nabla v|^2) \tag{8}$$

Then, the to-be-solved u and v can be represented as:

$$\begin{aligned} u,v &= \arg\min_{u,v} E(u,v,u_x,u_y,v_x,v_y) \\ &= \arg\min_{u,v} \iint L(u,v,u_x,u_y,v_x,v_y)dxdy \end{aligned} \tag{9}$$

By solving the corresponding Euler-Lagrange equations, u and v can be obtained:

$$\begin{aligned} \frac{\partial L}{\partial u} - \frac{\partial}{\partial x}\frac{dL}{du_x} + \frac{\partial}{\partial y}\frac{dL}{du_y} = 0 \\ \frac{\partial L}{\partial v} - \frac{\partial}{\partial x}\frac{dL}{dv_x} + \frac{\partial}{\partial y}\frac{dL}{dv_y} = 0 \end{aligned} \tag{10}$$

For discrete digital images, the Laplacian operator can be approximated as:

$$\begin{aligned} I_x(I_x u + I_y v + I_t) - \lambda^2 \Delta u = 0 \\ I_y(I_x u + I_y v + I_t) - \lambda^2 \Delta u = 0 \end{aligned} \tag{11}$$

After the approximation, equation (10) becomes:

$$\begin{aligned} ({I_x}^2 + \lambda^2)u + I_x I_y v = \lambda^2 \bar{u} - I_x I_t \\ ({I_y}^2 + \lambda^2)v + I_x I_y u = \lambda^2 \bar{v} - I_y I_t \end{aligned} \tag{12}$$

Since u and v are unknown, an iterative method can be used for solving:

$$\begin{aligned} u^{k+1} = \bar{u}^k - \frac{I_x(I_x \bar{u}^k + I_y \bar{v}^k + I_t)}{\lambda^2 + {I_x}^2 + {I_y}^2} \\ v^{k+1} = \bar{v}^k - \frac{I_y(I_x \bar{u}^k + I_y \bar{v}^k + I_t)}{\lambda^2 + {I_x}^2 + {I_y}^2} \end{aligned} \tag{13}$$

Where the initial values for u and v means are set to 0, and $\bar{u}^0 = 0$, $\bar{v}^0 = 0$ .The stopping criterion for iteration is:

$$|d_{k+1} - d_k| < th \qquad d = I_x u + I_y v + I_t \tag{14}$$

According to formulas (12) and (13), the final u and v values can be calculated.

2. Camera Selection and Calibration

Camera selection and calibration are crucial steps in the fields of computer vision and machine vision [3]. This process involves choosing sensors, lenses, shutters, and addressing calibration

issues such as distortion and intrinsic/extrinsic parameters. It ensures the accuracy of images, laying the foundation for computer vision [4].

2.1 Camera Selection

In visual depth measurement, there are monocular and binocular visual measurements. Monocular cameras can only capture two-dimensional images and cannot directly provide depth information for objects. This limitation becomes evident in applications requiring 3D scene reconstruction, stereo vision analysis, or depth perception [5]. In captured images, the real position of objects in the camera coordinate system is ambiguous, making it challenging to uniquely determine object depth. This challenge hinders obtaining accurate spatial information in certain applications, such as tracking and determining the positions of fish in this experiment. In contrast, binocular cameras simulate the human binocular vision system, perceiving object depth and stereo effects through the differences between left and right eyes. This stereo vision enhances the understanding of object shapes and structures, improving accuracy in target detection and tracking in complex environments. Binocular cameras can acquire stereo information [6], i.e., depth information. By utilizing the disparity between the left and right cameras (the offset of corresponding points in the images), the distance from objects to the camera can be calculated, providing more precise three-dimensional information [7]. This capability is particularly significant in the fields of robotic navigation and industrial automation. To meet the requirements of underwater images in this experiment, a binocular camera is chosen to enhance the accuracy of the vision system in underwater environments, allowing for distance and size measurements of underwater targets using the binocular vision system.

2.2 Binocular camera calibration

Camera calibration refers to the process of solving the camera's intrinsic and extrinsic parameters based on the transformation relationships between corresponding coordinate points in four coordinate systems related to the camera. The accuracy of camera calibration parameters directly affects the accuracy of binocular measurements [8]. To ensure the accuracy of calibration parameters and the stability of the calibration algorithm, the Zhang Zhengyou chessboard calibration method is adopted. The Zhang method requires synchronous shooting of plane chessboard calibration boards at different poses using only the left and right cameras during the calibration process [9], eliminating the need to pre-obtain motion parameters, etc.

The hardware for this calibration experiment includes a laptop and a binocular camera. The laptop's GPU model is NVIDIA GeForce RTX 3060, equipped with 16GB of memory. The binocular camera is configured with an image resolution of 1280×720 and a baseline length of 60mm, as shown in Figure 1. The system software environment includes Opencv, Matlab 2021a, and Windows 10.

Figure 1: Binocular Camera

Using the binocular camera to capture images of the chessboard calibration board for both left and right eyes. Subsequently, Matlab is utilized to perform single-camera calibration and binocular calibration on the collected calibration images. Based on the obtained camera intrinsic and extrinsic parameters, stereo rectification is applied to the left and right eye images, aligning them to the same plane.

The Zhang Zhengyou calibration method is employed to calibrate the binocular camera in the air. The experiment utilizes a chessboard calibration board with a grid size of 297mm×210mm. Different poses of the calibration board's left and right eye images are captured using the binocular camera. 200 clear images are selected from the captured images for left and right calibration purposes. In Matlab, 20 pairs of selected left and right calibration images are sequentially loaded, as shown in Figure 2. Starting from the top-left corner of the chessboard grid image used for calibration, the four vertices of the chessboard grid are selected clockwise, completing the extraction of the chessboard grid's corner points, as illustrated in Figure 3.

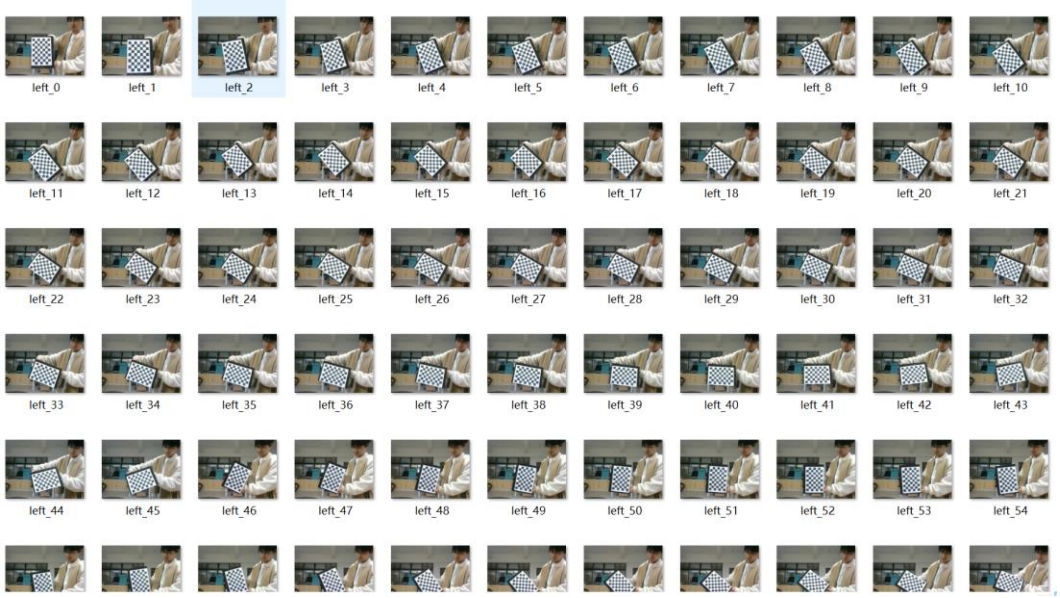

Figure 2: Left and Right Calibration Images

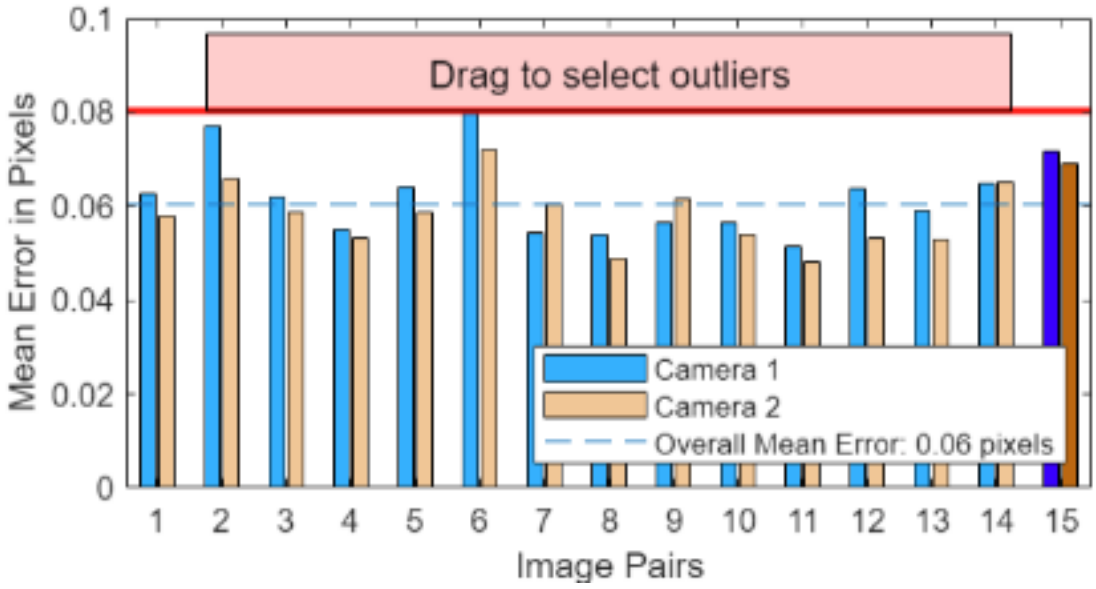

Figure 3: Corner Points Extraction

Table 1: Binocular Camera Calibration Parameters

Parameter	Left camera	Right camera
Inner parameter matrix	$\begin{pmatrix} 497.7420 & -0.1908 & 290.0690 \\ 0 & 497.8283 & 253.8442 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 503.8711 & -0.4289 & 327.0387 \\ 0 & 503.8097 & 253.6183 \\ 0 & 0 & 1 \end{pmatrix}$
Radial distortion coefficient	$[0.0001 \ \ -0.2984 \ \ 2.1756]$	$[0.0098 \ \ -0.3215 \ \ 1.9743]$
Tangential distortion coefficient	$[-0.0018 \ \ 0.0006]$	$[-0.0019 \ \ 0.0012]$
Selection Matrix	$\begin{pmatrix} 1 & -0.0002 & -0.0034 \\ 0.0002 & 1 & -0.0013 \\ 0.0034 & 0.0013 & 1 \end{pmatrix}$	
Translation vector	$[-60.0158 \ \ 0.0089 \ \ 1.3643]$	

Using the single-camera calibration results for the left and right cameras, perform the binocular calibration for the camera. The obtained binocular camera intrinsic matrix, distortion parameters, rotation matrix, and translation vector are shown in Table 1.

After completing the calibration of the binocular camera, I will utilize the obtained camera intrinsic and extrinsic parameters to perform stereo rectification on the left and right images captured by the binocular camera. It can be observed that the images after stereo rectification are coplanar and aligned, as shown in Figure 4.

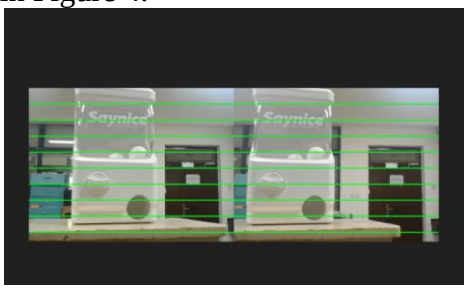

Figure 4: Stereoscopic Rectification of Binocular Cameras

3. Domestication Experiment Design

This chapter outlines the domestication scheme for California black bass primarily based on sound, providing a reliable theoretical foundation and technical support for the practical application of fish domestication equipment.

3.1 Selection of Fish Species

California black bass and Nile tilapia are economically significant exotic aquaculture species in China. As of 2022, the top three exotic fish species with the highest annual production in China are Nile tilapia (Oreochromis niloticus), California black bass (Micropterus nigricans), and Leiocassis longirostris [10]. Nile tilapia is mainly cultivated in regions with higher temperatures, such as Guangdong, Guangxi, and Hainan, with an adaptable temperature range between 15 ℃ and 37 ℃, and the optimal growth temperature between 23 ℃ and 31 ℃. Its main disadvantage is its

intolerance to low temperatures and cold weather, primarily cultivated in pond-intensive systems.

California black bass (Micropterus nigricans) [11], also known as largemouth bass, belongs to the Perciformes order, Centrarchidae family, with the scientific name Micropterus nigricans. Originally from the Mississippi River system in California, USA, it is a valuable carnivorous fish known for its delicious meat, disease resistance, rapid growth, easy capture, and broad temperature adaptability. California black bass exhibits strong adaptability to aquaculture environments and is suitable for cultivation in freshwater ponds. Currently, largemouth bass cultivation has high production, widespread scale, and good economic benefits, earning it the reputation of one of the most successful foreign species in China's aquaculture industry [12]. In this experiment, black bass is chosen as the experimental fish species.

Experimental equipment includes domestication devices, tanks, binocular cameras, computers, etc. The binocular cameras are used to observe and record the process and results of the black bass domestication experiment. They can be remotely activated through a computer and the footage stored on the computer. The specific experimental layout is illustrated in Figure 5, and the experimental site arrangement is depicted in Figure 6.

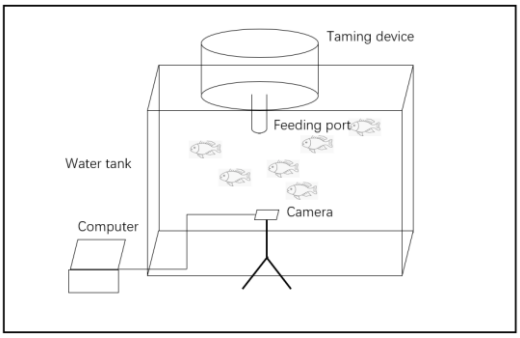

Figure 5: Experimental Layout Schematic

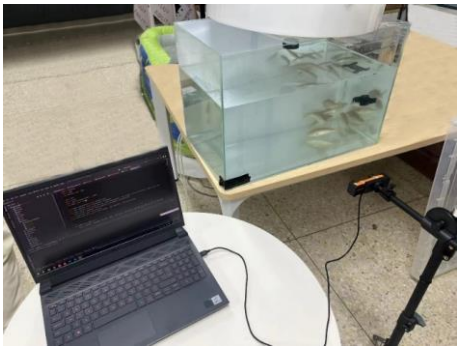

Figure 6: Experimental Site Arrangement

Domestication Method: The domestication period is set to 80 seconds. During the initial 20 seconds, the fish's normal swimming speed is recorded. Subsequently, a 20-second audio playback is initiated, and the fish's reactions are observed. At the 40-second mark, feeding is commenced, followed by the use of binocular cameras to record fish activities. The remaining time is dedicated to continuous observation of fish aggregation and dwell time, as shown in Figure 7. The experimental fish are divided into two groups: an experimental group and a control group, each consisting of ten bass, totaling twenty bass. Based on the feeding time of bass, two domestication experiments are conducted daily at 10:00 and 22:00. Upon observation, the domestication process for bass is relatively slow, leading to a 14-day duration for the domestication experiment.

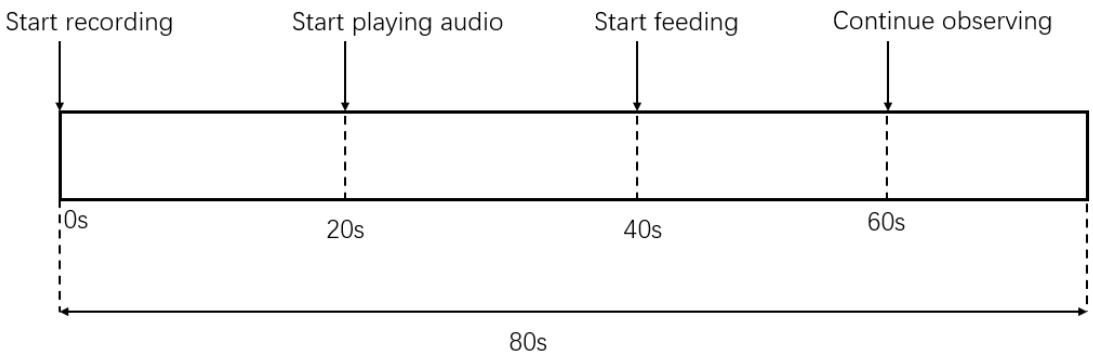

Figure 7: Domestication Time Settings

4. Target Detection System Design

This experiment is developed based on the OpenCV open-source computer vision library (cv2), utilizing the Python programming language and the PyCharm integrated development environment. The goal is to design a motion target detection system based on the HS dense optical flow method. The system's establishment and processing steps are as follows:

(1) Video Reading and Initialization Initially, the motion video of fish is read from a specified path, obtaining information such as the video's frame rate and total frames. This section uses fish in a calm state as an example. To facilitate subsequent processing, an output video object is initialized. This object will save the processed video results and define parameters needed for video processing, such as data smoothing filter window size and distance threshold.

(2) Optical Flow Calculation This step employs the HS dense optical flow method, calculating the optical flow field by comparing pixel intensity changes between adjacent frames in the video. The OpenCV function cv2.calcOpticalFlowFarneback is used in the code, returning the displacement vector for each pixel. By converting the amplitude and direction of the optical flow field into an image, a grayscale image representing inter-frame motion is obtained.

(3) Motion Contour Extraction By setting an appropriate threshold, the optical flow field is converted into a binary image where white pixels represent areas of motion. Contour detection functions are then applied to find contours in the binary image. Area filtering is employed to exclude motion regions that are too small or too large, resulting in processed outcomes as shown in Figure 8.

(4) Fish Position Tracking this step primarily involves processing the filtered contours, calculating the centroid position of the fish. By comparing the current frame's fish centroid with the average centroid position from previous frames, simple motion filtering is applied to ensure tracking accuracy. Ultimately, the fish's motion trajectory in the video is obtained. Fish positions are marked by drawing rectangular boxes, and data for each frame's fish centroid is saved, as depicted in Figure 9.

Figure 8: Extraction of Motion Contours

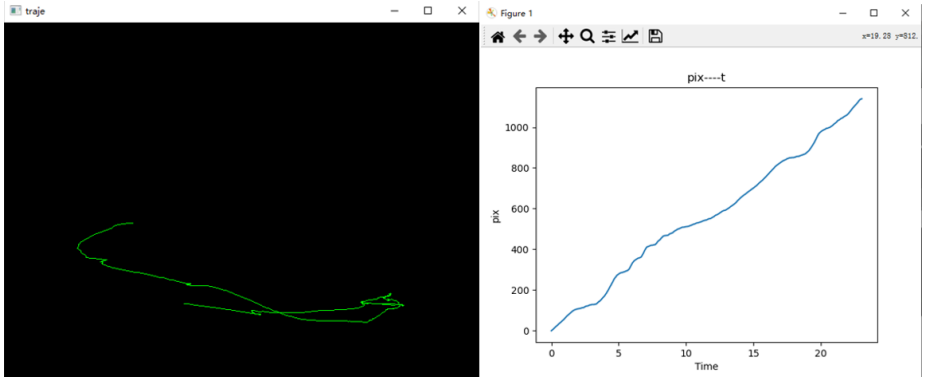

Figure 9: Tracking Results of Fish Positions

5. Evaluation of Fish Domestication Effects

In the preceding sections, fish motion targets in the videos were detected, and data on fish movement were obtained. The main focus of this section is how to effectively and accurately analyze this data. The core task is to analyze the collected data by deriving the fish's motion trajectory and speed information, thereby delving into a detailed analysis of the effectiveness of fish domestication.

To avoid the decrease in processing efficiency caused by fish obscuring each other, individual video tracking and data collection and calculation were conducted for a single largemouth bass, resulting in the motion distance line chart shown in Figure 10.

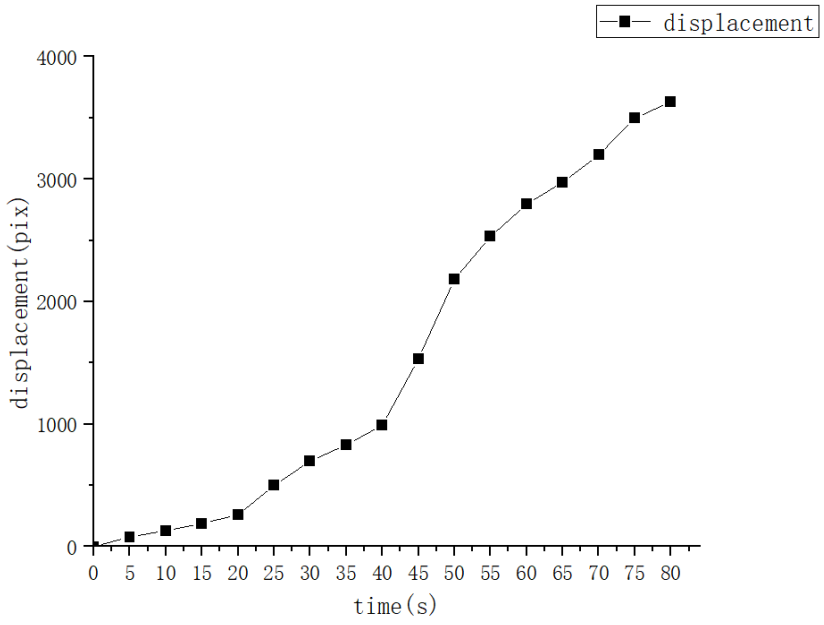

Figure 10: Motion Distance of Largemouth Bass

To enhance the experiment's reliability, the same method was applied to 10 largemouth bass. The average values were calculated, yielding Figure 11 for calculating the group's movement speed.

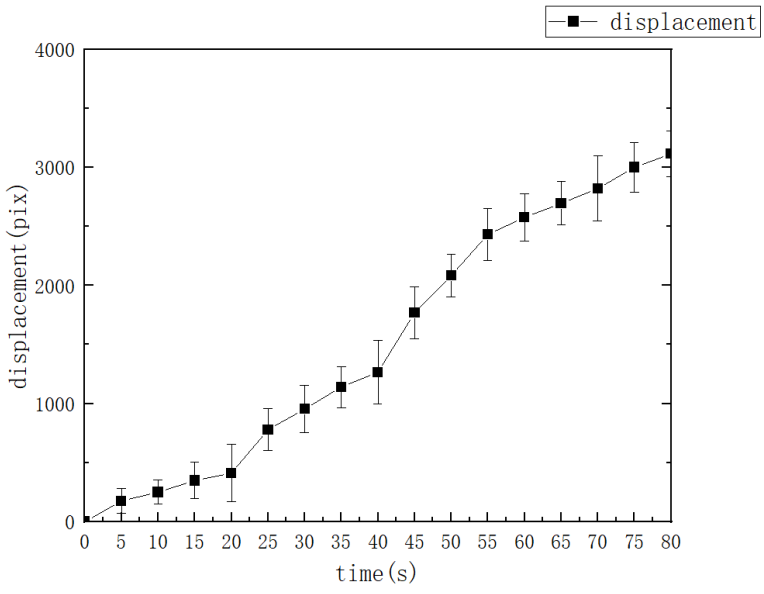

Figure 11: Group Motion Distance of Largemouth Bass

From the graph, it is evident that there is no significant change in fish movement from 0 to 20 seconds, with an average swimming speed of 20.6 pixels per second. At the 20-second mark, when a continuous square wave sound is played, the fish's movement distance gradually increases, and the average speed reaches 52.55 pixels per second. This indicates that the 300Hz square wave sound has a notable attracting effect on largemouth bass, demonstrating a good domestication effect. It suggests that after domestication, largemouth bass have developed a conditioned reflex to the square wave sound, exhibiting active behavior immediately upon hearing the sound. When feeding begins at 40 seconds, fish movement reaches its most active point, lasting for about 15 seconds. Subsequently, the amount of movement decreases gradually, and after 75 seconds, it returns to a calm state, with an average swimming speed of 24.81 pixels per second at this point.

6. Conclusions

This paper focuses on the selection and calibration of the camera, with camera calibration being a crucial step in the fields of computer vision and machine vision. Calibration in this step addresses distortion and intrinsic-extrinsic parameter issues, ensuring image accuracy and providing a foundation for computer vision. Developed using the OpenCV open-source computer vision library (cv2) and implemented in the Python programming language, this system employs the PyCharm integrated development environment. The design incorporates a motion target detection system based on the HS dense optical flow method.

The system extracts motion contours and calculates optical flow from the imported fish domestication videos. This process enables fish identification and tracking, utilizing computer vision techniques. The feasibility and accuracy of the system are validated through domestication experiments, confirming its effectiveness.

Acknowledgements

I would like to express my sincere gratitude to my classmates who provided invaluable assistance and support throughout the course of this research. Their collaboration and willingness to share insights greatly enriched the development of this paper.I am deeply thankful for the encouragement and guidance provided by my teachers. Their expertise, constructive feedback, and dedication to fostering academic growth have been instrumental in shaping the direction and quality of this work. I also extend my appreciation to myself for the diligence, perseverance, and hard work invested in this research endeavor. The personal commitment and effort dedicated to this project have been crucial in achieving its successful completion.

This journey would not have been possible without the collective contributions and support of those around me. I am truly grateful for the collaborative and encouraging environment that has fueled the progress of this research.

References

[1] Anthwal S, Ganotra D. An overview of optical flow-based approaches for motion segmentation [J]. The Imaging Science Journal, 2019, 67(5):284-294.

[2] Wang Y, Li Y, Wang J, et al. Sparse optical flow outliers elimination method based on Borda stochastic neighborhood graph [J]. Machine Learning: Science and Technology, 2024, 5(1).

[3] Lei L ,Hao C, Minze T , et al. Research on 3D reconstruction technology based on laser measurement[J].Journal of the Brazilian Society of Mechanical Sciences and Engineering,2023,45(6).

[4] Matsuoka R, Fukue K, Sone M, et al. A Study on Selection of Unknown Parameters of Image Distortion Model in Calibration of Non-metric Digital Camera [J].Journal of the Japan society of photogrammetry and remote sensing, 2005, 44(1):6-25.

[5] Stadlinger B, Grunert K, Sumner W R. The Perception of Depth and Form in Fine Arts, Video Games and

Medicine.[J].International journal of computerized dentistry,2024,0(0):0-0.
[6] Siyu C, Chao M, Chao L, et al. Feature selection based on the self-calibration of binocular camera extrinsic parameters[J].International Journal of Wavelets, Multiresolution and Information Processing, 2024, 22(01).
[7] Zhang J, Xie H, Zhang L, et al. Information Extraction and Three-Dimensional Contour Reconstruction of Vehicle Target Based on Multiple Different Pitch-Angle Observation Circular Synthetic Aperture Radar Data [J].Remote Sensing, 2024, 16(2).
[8] Xiaojing Z ,Yong H ,Chong W , et al. Binocular vision measurement for large-scale weakly textured ship hull plates using feature points encoding method[J].Measurement,2023,221.
[9] Zapiain O D M D, Tran A, Moore W N, et al. Calibration of thermal spray microstructure simulations using Bayesian optimization[J].Computational Materials Science,2024,235112845 .
[10] Hao W, Weidong H, Yulu S, et al. Calibration error study of binocular stereo camera[J]. Journal of Physics: Conference Series, 2022, 2338(1).
[11] T. J T, A. M S, A. L R, et al. Effects of Different Feeding Regimes on Growth Rates and Fatty Acid Composition of Largemouth Bass Micropterus nigricans at High Water Temperatures[J].Animals,2022,12(20):2797-2797.
[12] Xu C, Liu Y, Pei Z. Research on Legal Risk Identification, Causes and Remedies for Prevention and Control in China's Aquaculture Industry[J]. Fishes. 2023, 8(11).