

Automated Counting of Zebrafish: An Image Processing Approach

Ehsan Noshahri, Cristiano V. M. Araujo, and Alvaro Rodriguez

Department of Computer Science, University of A Coruña, Campus de Elviña, 15071 A Coruña, Spain

Center for Information and Communications Technology Research (CITIC), University of A Coruña, Campus de Elviña, 15071 A Coruña, Spain

Department of Ecology and Coastal Management, Institute of Marine Sciences of Andalusia, Campus Río San Pedro, 11519 Puerto Real, Spain

Correspondence: ehsan.noshahri@udc.es

DOI: <https://doi.org/10.17979/spu.23.c36>

Abstract: Accurate quantification of fish populations is a critical task in aquaculture and behavioral research. In this work, we present a lightweight image processing pipeline for the automatic counting of zebrafish (*Danio rerio*) in a multi-compartment aquatic system. The approach combines background subtraction, morphological refinement, and watershed segmentation to estimate fish counts directly from raw video without annotated training data. The method enables non-invasive and reproducible counting while remaining computationally efficient and interpretable. Although challenges remain under occlusion and limited visibility, the method reduces reliance on manual observation and demonstrates that classical image-processing techniques can provide efficient, interpretable, and accessible solutions for early-stage behavioral experiments.

1 Introduction

Fish counting is an important and necessary process in applications such as fish farming and must be carried out quickly, reliably, and efficiently (Hernández-Ontiveros et al., 2018; Zion, 2012). Traditionally, fish are counted manually, a process that is labor-intensive, time-consuming, and often stressful or harmful to the animals. These drawbacks have motivated a growing interest in automated counting systems that are accurate, efficient, and non-invasive (Cui et al., 2025).

Computer vision has recently emerged as a powerful alternative for automating fish counting. By analyzing images or videos captured in aquaculture settings, vision-based methods reduce the need for human intervention, minimize stress to the fish, and improve counting precision (Cui et al., 2025). These systems have been successfully applied in tasks such as object classification, detection, segmentation, and tracking across a wide range of applications, making them a strong candidate for use in aquaculture monitoring and fish counting (Fan and Liu, 2013; Labuguen et al., 2012).

Despite these advances, fish counting remains an open and challenging problem in real-world conditions. Factors such as rapid fish movement, overlapping individuals, and variable lighting conditions in aquatic environments complicate detection and segmentation (Cui et al., 2025). Previous works have applied both traditional machine learning and deep learning to fish counting. Geometric features such as contour, area, skeleton, and endpoints of the fish body have been used in machine learning approaches, which require intricate handcrafted feature extraction and often lack generalization (Zhang et al., 2023). More flexible models such as deep neural networks have shown improved performance through detection-based (Li et al.,

2017) and density map regression approaches (Song et al., 2020), but they typically require large annotated datasets, which are difficult to obtain in fish counting applications.

In the context of behavioral studies with zebrafish (*Danio rerio*), automatic counting plays an especially important role. Zebrafish are widely used in large-scale genetic screening, and for studying human neurological disorders, drug addiction, social anxiety disorders, and more (Pedersen et al., 2020). In this study, we design an experimental framework using image-processing methods for the counting of zebrafish in controlled environments. Based on the zebrafish footage in an specific experimental setups, namely the Heterogeneous Multi-Habitat Assay System (HeMHAS) (Araújo et al., 2018; Salvatierra et al., 2025), we propose a lightweight automatic zebrafish counting method based on motion detection, morphological refinement, and segmentation. Our approach prioritizes computational efficiency and interpretability. While it cannot fully overcome limitations such as occlusion or poor visibility, it provides a practical, transparent, and reproducible solution suitable for early-stage experiments and exploratory behavioral studies.

2 Materials and Methods

2.1 Dataset and Experimental Conditions

The videos used in this study were recorded in the Heterogeneous Multi-Habitat Assay System (HeMHAS) (Araújo et al., 2018; Salvatierra et al., 2025), an ecotoxicological test environment consisting of a configurable grid of compartments connected by controllable gates. Unlike traditional forced-exposure systems, HeMHAS allows organisms to move freely among compartments, which can be arranged to simulate contamination gradients and more natural conditions.

For the present work, the system was used to conduct a population test in an environment without contaminants. A total of 5 minutes and 35 seconds of video footage, was used as the dataset. The recordings were obtained using a fixed top-view camera positioned above the HeMHAS. The aim of the experiment was to observe the behavior of multiple zebrafish interacting within a shared environment and to quantify their spatial distribution over time.

The arena consisted of a 3×5 grid of compartments with horizontal and vertical gates that could be opened or closed to control connectivity. Each row of the grid was used to replicate the same experiment simultaneously, with separate groups of zebrafish providing three independent repetitions. At the start, all gates were closed while fish were placed in their compartments; once released, only the horizontal gates were opened, allowing movement only along rows.

2.2 Image Processing

In this study, we employ a classical image processing approach rather than deep learning-based methods for detecting and counting zebrafish. The primary reason for this choice is the lack of annotated data, as supervised deep learning models require a large number of precisely labeled frames to train effectively, which is highly time-consuming and labor-intensive to produce. Moreover, generating accurate labels is complicated by factors like fish entering partially hidden regions and occlusions, which make reliable ground truth difficult to obtain.

In contrast, traditional image processing techniques can operate directly on raw videos without any manual annotation, making them well suited for exploratory studies or early-stage experimental setups. These methods are lightweight, interpretable, and computationally efficient, requiring only modest hardware and offering real-time or near real-time performance. Moreover, their behavior can be easily adjusted and tuned through algorithmic parameters rather than retraining a model, providing greater flexibility when experimental conditions or video quality change. While deep learning may offer superior performance in well-labeled, large-scale datasets, the interpretability and low data requirements of classical image processing make it the most practical and transparent choice for this study.

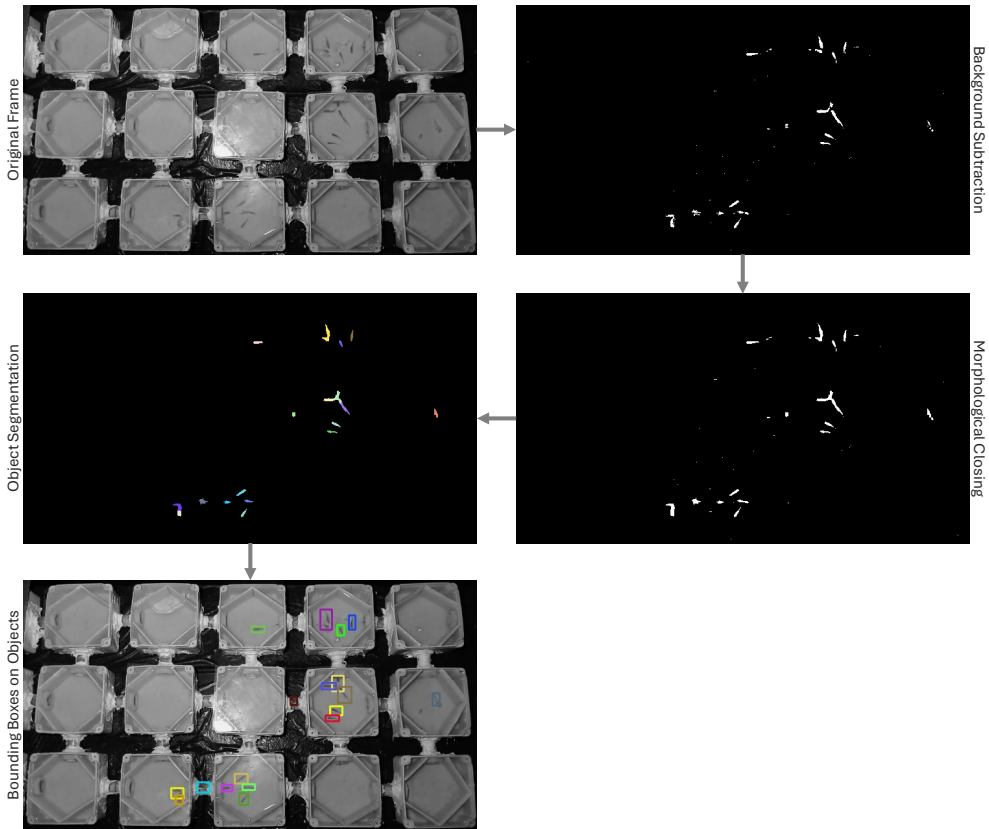


Figure 1: Overall workflow of the proposed image-processing pipeline for zebrafish counting.

The proposed pipeline follows a sequential procedure, beginning with background subtraction, followed by refinement of the foreground masks, and concluding with instance segmentation. The overall workflow is illustrated in Figure 1 and described in detail in the following subsections.

Background Subtraction

Background subtraction is a motion-based segmentation technique that separates moving objects (“foreground”) from a relatively static scene (“background”) in video. For each pixel location, the algorithm keeps track of how that pixel typically appears over time, and compares incoming frames against this record. Pixels whose current intensity (or color) deviates beyond a threshold are labeled as foreground, those consistent with the model are labeled as background. Because it operates online and per pixel, background subtraction is well suited to fixed-camera recordings where objects of interest move against a stable arena.

In this work we use the K-Nearest Neighbor (KNN) background subtractor (Zivkovic and van der Heijden, 2006), in which each pixel in a frame is classified as background if at least a certain number of its K most recent samples in the previous frames lie within a sufficiently small distance of the current value; otherwise, it is labeled as foreground. This is a straightforward nonparametric, sample-based method in which each pixel’s background model is continuously updated over time. By comparing the current pixel value with its recent history, the algorithm adapts to changes and separates moving foreground objects from the static background. As

an additional step in the background subtraction process, we apply the Otsu's thresholding method (Gonzalez and Woods, 2018) to obtain the binary mask of the foreground objects, which is required by the next stages of our algorithm.

Morphological Operations

While background subtraction can provide an approximate mask of the foreground objects, the resulting mask often contains imperfections such as small holes inside the detected regions or noisy protrusions along their boundaries. To address these issues, we apply a morphological closing operation, which consists of two sequential steps: dilation followed by erosion (Gonzalez and Woods, 2018; Pratt, 2007).

Dilation expands the boundaries of the foreground regions by adding pixels around their edges, which helps to close small gaps and holes inside the detected objects, thus making each region more solid and connected. Erosion is then applied to shrink the boundaries slightly, removing the excess pixels that were added during dilation and smoothing the object edges.

This process fills small holes and connects narrow gaps without significantly changing the overall size or shape of the detected objects, resulting in a cleaner and more coherent mask for subsequent analysis.

Instance Segmentation and the Watershed Algorithm

The watershed algorithm is an object segmentation method that relies on principles of topography and hydrology. It begins by interpreting the grayscale image as a 3D topographic surface (an "elevation" map), where the x and y coordinates represent pixel locations and the z coordinate (height) corresponds to pixel intensity. Dark pixels with lower values correspond to valleys, while brighter pixels correspond to ridges. Based on a set of seed points or markers representing the objects of interest, the algorithm floods the valleys starting from the lowest levels (local minima). Each marker has its own source of water and is assigned a unique label.

As the water from different valleys rises and begins to meet, "dams" are built to prevent merging, forming the watershed lines that define the boundaries between objects. Once the entire image is flooded, these lines determine the final segmented regions, and all pixels connected to the same marker share the same label (Pratt, 2007), representing distinct object segments in the image.

In practice, markers for the watershed algorithm can be selected manually or automatically using operations such as the distance transform. The distance transform of a region of foreground pixels within a background of zeros assigns to each foreground pixel its Euclidean distance to the nearest background pixel. The local maxima in this distance map, which represent the points farthest from the background, can then be used as seed points for the watershed algorithm, as they are most likely located at the center of each object (Gonzalez and Woods, 2018).

2.3 Evaluation Metrics

The primary objective of the proposed methodology is to estimate the number of zebrafish present in the environment at each point in time. To assess performance in the absence of a reliable baseline, we compute descriptive statistics of the detected counts over the entire recording, including the mean, median, standard deviation, minimum, and maximum values.

As a reference for comparison, we also report the maximum visually observed count as the ground truth, even though not all frames necessarily contain this number of detected objects. Based on this reference, we further compute the relative accuracy, defined as the average number of counted fish divided by their actual values (assumed here to be the maximum observable fish in the environment), providing an intuitive metric for evaluating the overall performance of the algorithm. In addition to these quantitative measures, we conduct a visual review of the

Environment	Median	Mean	Std	Min	Max	Max _{ground-truth}	Acc _{relative}
Top	6	6.45	1.34	3	35	9	71.67%
Middle	7	6.65	1.50	1	15	9	73.88%
Bottom	7	7.32	1.47	2	13	10	73.20%

Table 1: Evaluation metrics of zebrafish counts across the three environments.

recordings to gain a qualitative understanding of the method’s behavior and limitations.

3 Results

The evaluation metrics for the three environments (top, middle, bottom rows) in the recorded footage from the population experiment are summarized in Table 1. In addition, Figure 2 presents the histogram of fish counts across frames, illustrating their distribution. As observed, the relative accuracy of the algorithm in counting fish across the environments reaches a minimum of 71.67%. While this means that less than 29% of the fish remain undetected (assuming all observable fish were visible in every frame), it is worth discussing the factors that contributed to these missed detections. The main cause lies in the environmental imperfections of the experimental arena, which make it impossible to reliably detect all fish even when they are present in the videos. The most problematic areas are the connecting gates between compartments, including the openings of the gates (inside the compartments), the pipe-like transparent bodies connecting two environments, and the boundary of these connectors sealed with adhesive material to prevent water leakage. As observed, the opening of the compartments is not transparent and is the same color as the compartments, therefore any fish entering or exiting it disappears from view. Furthermore, the adhesive often blocks the transparent connectors partially, and in many cases almost completely, further limiting the detectability of the objects inside.

In addition to these environmental limitations, there are also algorithm-related errors. A major limitation is the algorithm’s inability to separate fish that are occluded by others or that swim in very close proximity. In such cases, their segmentation masks merge, leading the watershed algorithm to treat them as a single object rather than distinct fish. Despite these challenges, the speed and simplicity of the method make it a practical choice for applications with limited labeled data, particularly in early-stage experiments where interpretability and low computational cost are more critical than absolute accuracy.

4 Conclusion

In this study, we introduced a lightweight and transparent image-processing pipeline for automatic zebrafish counting in a multi-compartment experimental setup. By combining background subtraction, morphological refinement, and watershed-based instance segmentation, the approach enables non-invasive estimation of fish counts directly from raw video without requiring annotated training data, making it particularly suitable for early-stage experiments and exploratory behavioral assays.

The results demonstrate that while the method provides useful estimates, its accuracy is limited by environmental factors such as blind spots in the arena and algorithmic challenges like object occlusion.

Future work should integrate tracking to handle occlusions, explore hybrid methods that combine classical image processing with machine learning, and improve experimental design to reduce visibility constraints. Overall, the proposed framework offers a practical and accessible foundation for automated zebrafish counting and supports the development of more robust

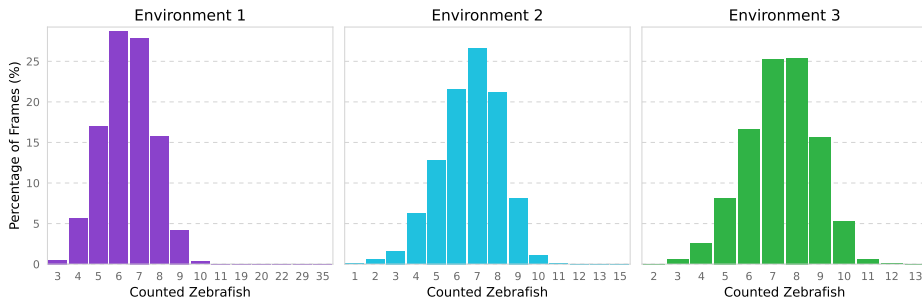


Figure 2: Histogram of zebrafish counts across frames.

and scalable behavioral analysis tools.

5 Acknowledgements

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101034261. The study is carried out within the framework of the R&D grant (Ref. PID2021-126289OA-I00) and *BeingHavior* project (Ref. PID2022-137402OB-I00), both funded by MCIN/AEI/10.13039/501100011033/ and by the European Regional Development Fund (ERDF), a way of making Europe. The study is also part of *ConBio+* project (Ref. LINCG23005) funded by the Spanish National Research Council (CSIC).

The authors would also like to express their gratitude to Paola Scarcia and Fernando de la Torre at the National University of Luján in Argentina for recording and sharing the videos that served as the dataset for this study.

Bibliography

- C. Araújo, D. Roque, J. Blasco, R. Ribeiro, M. Moreira-Santos, A. Toribio, E. Aguirre, and S. Barro. Stress-driven emigration in complex field scenarios of habitat disturbance: The heterogeneous multi-habitat assay system (hemhas). *The Science of the total environment*, 644: 31–36, 07 2018. doi: 10.1016/j.scitotenv.2018.06.336.
- M. Cui, X. Liu, H. Liu, J. Zhao, and D. Li. Fish tracking, counting, and behaviour analysis in digital aquaculture: A comprehensive survey. *Reviews in Aquaculture*, 17, 01 2025. doi: 10.1111/raq.13001.
- L. Fan and Y. Liu. Automate fry counting using computer vision and multi-class least squares support vector machine. *Aquaculture*, s 380–383:91–98, 03 2013. doi: 10.1016/j.aquaculture.2012.10.016.
- R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Pearson Education, 2018. ISBN 9780133356724.
- J. Hernández-Ontiveros, E. Inzunza Gonzalez, E. García-Guerrero, O. Lopez-Bonilla, S. Infante, J. Cárdenas-Valdez, and E. Tlelo-Cuautle. Development and implementation of a fish counter by using an embedded system. *Computers and Electronics in Agriculture*, 145:53–62, 02 2018. doi: 10.1016/j.compag.2017.12.023.

- R. Labuguen, E. Volante, A. Causo, R. Bayot, G. Peren, R. Macaraig, N. Libatique, and G. Tangonan. Automated fish fry counting and schooling behavior analysis using computer vision. 03 2012. doi: 10.1109/CSPA.2012.6194729.
- X. Li, Y. Tang, and T. Gao. Deep but lightweight neural networks for fish detection. In *OCEANS 2017 - Aberdeen*, pages 1–5, 2017. doi: 10.1109/OCEANSE.2017.8084961.
- M. Pedersen, J. Haurum, S. Bengtson, and T. Moeslund. 3d-zef: A 3d zebrafish tracking benchmark dataset. 06 2020. doi: 10.1109/CVPR42600.2020.00250.
- W. K. Pratt. *Digital image processing*. Wiley-interscience, 2007. ISBN 9780471767770.
- D. Salvatierra, M. A. Islam, M. D. P. Gonzalez Muñoz, J. Blasco, and C. Araújo. The heterogeneous multi-habitat assay system (hemhas): A non-forced ecotoxicology test system to study contamination-driven habitat selection behavior from landscape and stress ecology perspectives. *Environmental Pollution*, 368:125818, 02 2025. doi: 10.1016/j.envpol.2025.125818.
- Z. Song, X. Yang, Y. Wang, Z. Zhao, J. Liu, L. Yang, C. Sun, and C. Zhou. Automatic fish population counting by machine vision and a hybrid deep neural network model. *Animals*, 10: 364, 02 2020. doi: 10.3390/ani10020364.
- H. Zhang, W. Li, Y. Qi, H. Liu, and Z. Li. Dynamic fry counting based on multi-object tracking and one-stage detection. *Computers and Electronics in Agriculture*, 209:107871, 06 2023. doi: 10.1016/j.compag.2023.107871.
- B. Zion. The use of computer vision technologies in aquaculture – a review. *Computers and Electronics in Agriculture*, 88:125–132, 10 2012. doi: 10.1016/j.compag.2012.07.010.
- Z. Zivkovic and F. van der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780, 2006. ISSN 0167-8655. doi: <https://doi.org/10.1016/j.patrec.2005.11.005>. URL <https://www.sciencedirect.com/science/article/pii/S0167865505003521>.