Fish Detection and Tracking Using Kinect v2 Sensor

Abstract

Development of deep-sea observatories has led to increased demand of automatic fish detection and tracking. Detecting fishes, estimating fish number, and their tracking can help the marine biologists to understand underwater environment and nature of marine animals. Due to the bleak body deformation of fish, that results in their complex motion and multiple occlusions, robust fish tracking from video image sequences is a highly challenging task. This thesis explores the different techniques used for automatic fish detection and tracking followed by providing an application of the same for marine researchers in determining the critical velocity of fish. The thesis discusses the steps involved in the process, the challenges encountered, and different ways to proceed in the direction. The work entailed though focuses on fish detection and tracking but the same can be generalized as an application of object detection and tracking.

From the last few years, RGB-D cameras are widely used by researchers in various fields. Their reasonable cost and the ability to estimate distances at a high frame rate have made these sensors recommendable for applications in gaming accessories, robotics, computer vision, etc. In addition to colour, these sensors also provide depth information. Aspects like the stability, accuracy, and reliability of depth-sensing cameras like Kinect v2 must also be considered before using the device for applications like that of 3D space modelling. In this thesis, an analysis of the error in the depth measurement as well as calculation of Depth Entropy given by Kinect v2 sensor in different mediums viz. air, glass and water has been done. We have validated our findings using the theories of optics. The findings from error analysis were used to make an error compensation model which can correct depth at each pixel of the image. The error analysis and error

compensation model, proposed in our work, will help in improving the accuracy of present and future depth sensing devices.

Our thesis also presents Kinect v2 sensor based fish detection and tracking. The work builds on our knowledge of error analysis and error compensation to generate a 3-D depth map of aquarium.

Fish detection and tracking in a constrained environment (aquarium) aids in the research of marine biologists to understand and study the behaviour of fishes. The use of commodity hardware devices, like Microsoft Kinect, for detection and tracking is economical for the researchers and helps to automate the study of fishes round the clock without any requirement of manual intervention. In our work, we have calibrated depth and RGB images to detect and track multiple fishes present in each frame of a recorded video. A multi-tiered methodology involving classical clustering, classification, and image processing techniques is proposed to overcome the problems associated with fish detection and tracking, and multiple fishes were tracked using centroid of the fish as central component.

Apart from utilizing the traditional machine learning (ML) algorithms for fish detection and tracking, we explored the domain of deep neural network for multi-fish detection and tracking. The data set provided by Kinect v2 was cleaned, aligned, and calibrated to make it suitable for object detection and tracking. We have utilised YOLOv5s model and proposed three different variants of the model namely YOLOv5-WAM, YOLOv5-RGBD, and YOLOv5-SpecialFeatures.

YOLOv5-WAM (With Attention Mechanism) used Transformer learning based attention mechanism to detect smaller objects like fishes. YOLOv5-RGBD incorporates 4 channel information i.e. 3 colour channels (R, G, B) and one Depth channel (D), to perform object detection on RGBD or multiple channel datasets. YOLOv5-SpecialFeatures utilised custom features derived from traditional feature detection techniques to experiment on amalgamation of custom features with deep learning model i.e. YOLOv5. YOLOv5-WAM and YOLOv5-RGBD can be adapted for use on devices having low compute availability, making them suitable for real time object detection and tracking using low-cost experimental setup. An object detection and tracking pipeline is also designed to detect and track underwater objects (fishes in our work), belonging to multiple classes in 3 dimensions.

Finally, the insights gained from fish detection were utilized to train a prediction model for predicting Critical Velocity (Ucrit) of fishes. Ucrit is one component which is calculated by marine researchers to identify the swimming capacity of fish. Traditional methods of computing Ucrit are human intensive, involve elaborate setup, and are time-taking. Hence, in this thesis, a model is trained by utilising the existing Ucrit data to predict Ucrit of fishes for which detection and tracking pipeline was developed previously.