AUTOMATING LASER POINT DETECTION IN ANTARCTIC BENTHIC IMAGERY TO AID BIODIVERSITY MONITORING

JASMINE NG | MPHIL DATA INTENSIVE SCIENCE | UNIVERSITY OF CAMBRIDGE

1 Motivation and Scientific Justification

Benthic research focuses on seafloor environments and the biodiversity they support. This field has gained increasing importance due to the growing impacts of climate change on benthic environments. To survey benthic environments, cameras are towed along the seafloor capturing benthic imagery. However, the rapid increase in image volume poses significant challenges in this research field. A key bottleneck is the manual identification of laser points (LPs), which are used as spatial references to measure the size of marine organisms and seafloor features. This task is labour-intensive and error-prone, especially under conditions of poor visibility and light scattering in underwater scenes [16].

To address this challenge, Schoening et al. developed the DELPHI method (Detection of Laser Points in Huge Image Collections using Iterative Learning) [17]. DELPHI uses a small set of annotated examples to learn color and spatial features, which are then used to automatically detect LPs in large image datasets. This thesis presents a Python implementation of DELPHI adapted for use by the British Antarctic Survey (BAS). The method is evaluated using precision, recall, and F1-score on two Antarctic transects to assess reproducibility of Schoening et al.'s results. Beyond that, this thesis will explore potential improvements on DELPHI results through cross-validation, parameter tuning, and testing performance under diverse seafloor conditions.

2 Methodology

Baseline: DELPHI Implementation and Evaluation

As the T1 and T2 datasets from Schoening et al.'s study were unavailable, this thesis uses 39 and 61 images from two Antarctic transects, referred to as t1 and t2. These represent hard and soft substrates along the Western Antarctic Peninsula and are broadly comparable to those used by Schoening et al. The images were captured using the OFOBS towed camera system and manually annotated for LPs using LabelMe, with bounding boxes used for training and evaluation.

An overview of the DELPHI method is shown in Figure 1. The training pipeline consists of two main stages. Firstly, colour feature learning uses k-means clustering to learn the typical red, green, and blue (RGB) colours of LPs in the labelled training images. Secondly, spatial layout modelling learns where LPs usually appear in the training images by analysing their coordinates and the typical triangular patterns they form. The colour and spatial information learned

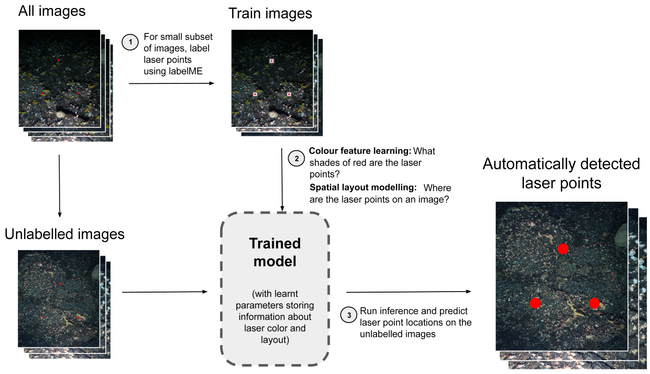

Figure 1: High-level overview of the DELPHI training and testing process

from the training images is then used during inference to predict the most likely LP locations in unlabelled images.

During inference, candidate pixels in unlabelled images are first selected by thresholding based on colour similarity to the learnt RGB values. Secondly, because LPs are assumed to be connected regions of red pixels, morphological opening is applied to remove noise in the form of small, unconnected pixel regions. The third step involves applying a master mask, which restricts the search to regions where LPs were observed in the training images. Finally, red regions most likely to be LPs are identified, and the top three are selected based on how well they match the typical triangular configurations. Performance is evaluated using precision, recall, and accuracy. A detection is considered correct if the predicted point lies within δ_1 (radius) = 25 of a ground truth LP.

Extensions: Improving DELPHI Results

Three extensions were explored to improve DELPHI's performance and robustness. Firstly, Monte Carlo and k-fold cross-validation were applied to address the high variance observed in this thesis's results. Secondly, parameter tuning was conducted using a grid search over key parameters in the learning process, including the background and LP radii (δ_1, δ_2) and the morphological kernel size used for denoising. Thirdly, t1 and t2 were combined to evaluate how using a joint dataset affects DELPHI's performance.

3 Key Findings

This thesis found that pecision, recall, and F1-score improved with increasing training size as shown in Figure 2, consistent with findings by Schoening et al. Performance plateaued at 19 training images for t1 and 22 for t2, reaching F1-scores of 0.58 and 0.90, respectively. In contrast,

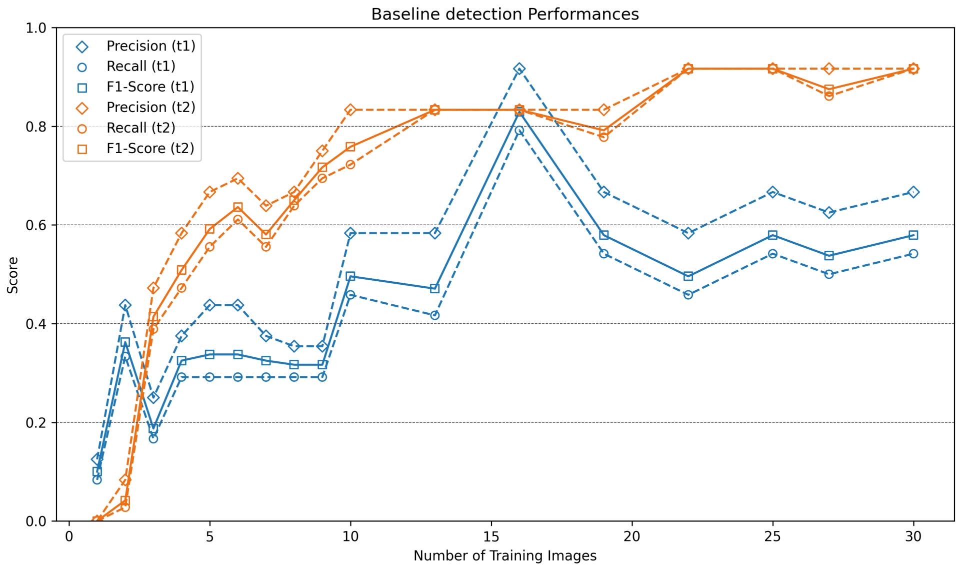

Figure 2: DELPHI detection performance on t1 and t2 with increasing training data size

Schoening et al. reported performance saturation at 13 images.

Interestingly, performance trends were reversed: DELPHI performed better on the soft substrate transect (t2), whereas Schoening et al. reported better results on the hard substrate (T1). The BAS dataset also showed greater variability across training sizes, likely due to the smaller test sample.

Analysis of both successful and failed detections, along with k-means clustering results, suggested that t1 images were more difficult to classify due to noisy backgrounds and dim lighting. Specifically, the morphological filtering step in the DELPHI pipeline was insufficient to remove complex background noise in t1 images, and k-means clustering revealed poor colour separation between LP and background pixels.

Monte Carlo and k-fold cross-validation confirmed the superior performance of t2 over t1. Parameter tuning showed that the default values $\delta_1 = 25$ and $\delta_2 = 3$ were already near optimal and required no further adjustment. Finally, combining t1 and t2 for joint training led to better performance than using t2 alone, achieving an F1 score of 0.80, although more training images were needed to reach saturation. While further analysis of successful and failed cases is needed to fully understand this result, it suggests a potential strategy for improving classification on lower-quality data like t1 by merging it with higher-quality data such as t2 for DELPHI training and detection.

4 Futurework

This study identified key challenges in detecting LPs under low lighting and textured backgrounds, especially in the t1 dataset. Future work could apply image enhancement techniques such as histogram stretching, red channel boosting, and z-score normalization to improve contrast and reduce noise. More advanced approaches like dehazing, convolutional neural networks (CNNs), and support vector machines (SVMs) may offer adaptive solutions for modelling LP features under varying conditions.

5 Conclusion and Research Impact

This thesis replicated the DELPHI method by implementing a semi-automatic LP detection pipeline in Python and evaluating it on BAS imagery, achieving results comparable to Schoening et al. The study highlighted how substrate type and lighting conditions significantly impact detection performance. Despite these challenges, DELPHI shows strong potential to reduce manual annotation and support scalable analysis of benthic imagery, accelerating image processing and enabling faster insights for conservation and climate adaptation.

References

For full references, please refer to the main paper. In this summary, references are cited using the same numbering as the main paper.

LaTeX word count: 999 words