Deep Learning for 6-DoF Underwater Pose Estimation

Mari Hovem Leonhardsen
marihl@stud.ntnu.no

NTNU
Norwegian University of Science and Technology

Abstract

This thesis investigates the opportunity of applying deep learning techniques for underwater localization systems. In particular, the application of a convolutional neural network (CNN) to estimate the relative position and attitude, together referred to as the six degrees of freedom (6-DoF) pose, between a camera and a fixed object based on imagery input has been evaluated. This is intended to serve as a complementary high-precision localization system for an ROV equipped with a camera.

Background and Motivation

The demand for subsea inspection, maintenance and repair (IMR) operations is expected to increase in the coming years. ROVs are crucial for the execution of IMR operations, and an increased level of autonomy in ROV operations is desired to reduce cost and risk. Situation awareness is a necessity in autonomous operations, and localization systems are therefore essential. Localization refers to an object's understanding of its position and attitude relative to the surroundings. Performing manipulation tasks on subsea installments require extremely precise maneuvering of the ROV, and the need for highly accurate localization systems is therefore adressed.

State of the art methods for estimating the relative pose between an underwater vehicle and a fixed point in the environment, are based on computer vision (CV) approaches and imagery input. These methods rely on artificial markers that must be placed in the environment, requiring additional infrastructure on the subsea installment subject to intervention.

Recent advances within the field of image processing with deep learning techniques have resulted in promising methods for visual pose estimation. Convolutional neural networks (CNNs) have been successfully applied to regress 6-DoF pose based on one single image. Such methods do not require artificial landmarks placed in the environment. To the author's best knowledge, such deep learning methods have not yet been applied to the underwater localization problem.

Objective

The objective of this thesis is to investigate, identify and implement a deep learning method for estimating the relative 6-DoF pose between and underwater vehicle and a fixed object. The suggested method shall be assessed on underwater datasets.

Acknowledgements

I would like to thank my supervisor Professor Ingrid Schjolberg for assistance and advice during the work of this thesis. I would also like to thank Postdoctoral Fellow Mikkel Cornelius Nielsen, for giving me advice regarding the method in this work and for assisting the laboratory activities.

Method

MODEL

The deep learning architecture PoseNet [1] was chosen for implementation. This is a 23 layer deep convolutional neural network, regressing relative pose from one image. The network outputs a pose vector $\mathbf{p} \in \mathbb{R}^7$ composed of the 3D camera position $\mathbf{x} \in \mathbb{R}^3$ and the orientation represented by the quaternion $\mathbf{q} \in \mathbb{R}^4$. The network was trained with stochastic gradient descent and an Adam optimizer, minimizing the modified mean squared error loss function seen in Equation 1. The term β is a scaling factor balancing the position and orientation errors. The optimal value for this factor was identified with grid search.

$$Loss = ||\hat{\mathbf{x}} - \mathbf{x}||_2 + \beta \left|\left| \hat{\mathbf{q}} - \frac{\mathbf{q}}{||\mathbf{q}||} \right|\right|_2 \quad (1)$$

The CNN was implemented with Python and the open source machine learning framework Tensorflow. The model was trained on a GeForce GTX 1080 Ti GPU.

DATASET

Real-world datasets were obtained with experiments in the MC-laboratory at NTNU. Underwater datasets containing images of a submerged target object taken from several camera poses were produced, where the relative pose between the camera and the object serves as the label for each image. This was achieved with the ROV model bluerov2 equipped with a camera. A plastic ring imitating a subsea valve was used as the target object. The laboratory setup is seen to the left **, and the bluerov2 in the MC-lab basin is seen to the right.

Figure 1: Setup in MC-lab (left) and sample image from dataset (right)

The positioning system Qualisys is available in the MC-lab, measuring 6-DoF position and orientation of objects in motion with millimeter accuracy. The relative poses serving as the labels of the images in the dataset are calculated from these measurements. Several datasets with varying light conditions and backgrounds were produced.

Results and Conclusion

The network was trained and tested in several iterations, with training sets in the range of 6-11K images. The following results describe the network performance in the case where the test data consist of images similar to the sample right in Figure 1. Histograms of the 6-DoF errors are seen in Figure 2, and it is observed from the plots that the errors are approximately normally distributed around zero in each DoF. Outliers of large magnitude does however occur, which is likely related to noise in the pose data from the Qualisys system. A total position and a total orientation error were also measured, being the Euclidean distance of the position errors and the relative angle between the true and predicted quaternion, $\mathbf{q}$ and $\hat{\mathbf{q}}$. The median of these errors were evaluated to 19 mm and 0.4 degrees in the case.

The CNN model suggested and assessed in the work of this thesis delivers pose estimation of the accuracy presented above. This method eliminate the need for insertion of artificial markers in the environment when estimating the relative pose between a fixed object an underwater vehicle.

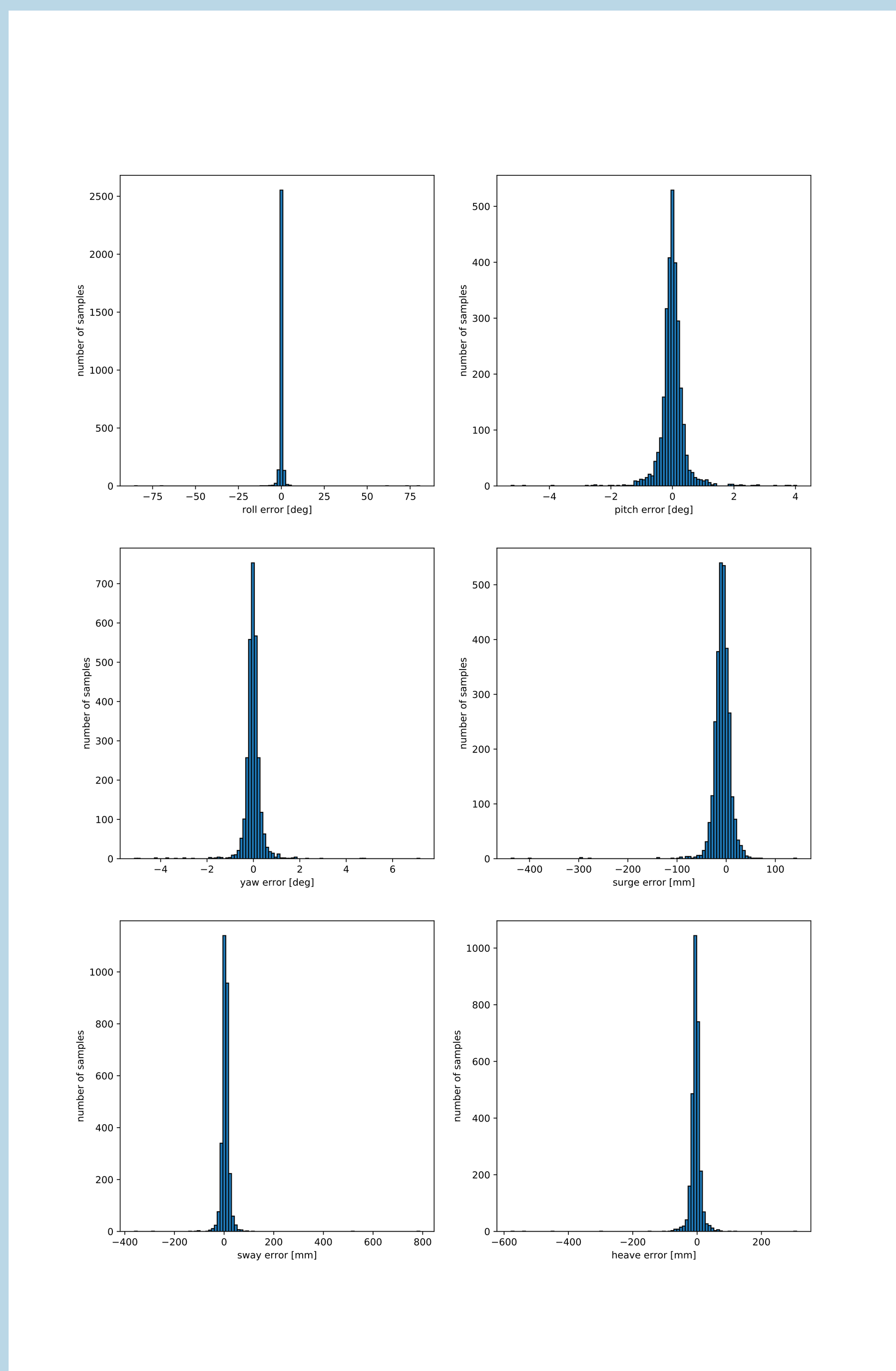

Figure 2: Pose errors in 6-DoF

References

[1] Alex Kendall, Matthew Grimes, and Roberto Cipolla: *Posenet: A convolutional network for real-time 6-dof camera relocalization*, (2015)