Experiments and actions on the pilot projects 3 until 30.12.2023

Sea State Estimation: A Computer Vision Approach

FIDIT

"CHALLENGE! Create a model for recognizing the state of the sea and detecting small objects that will work autonomously in real time in order to help with navigation and increase the safety of navigation.

HOW? Collect data from real scenes, select appropriate scenes, mark them and prepare them for machine and deep learning in order to be able to learn a computer vision model for recognizing the state of the sea and for detecting small objects at sea.

Form a set for model learning and testing, select appropriate architectures in order to use deep learning and knowledge transfer to create customized models for recognizing the state of the sea and detecting objects.

Research steps:

- Analysis of scientific papers and existing image databases related to the analysis of the state of the sea.
- The existing scales for distinguishing the state of the sea and the basic elements and features that discriminate a certain state of the sea were investigated, and a comparison was made between different interpretations.
- Preliminary data acquisition: Data was collected from real scenes from the ship during sailing and at anchor from different positions, with different shooting and camera settings and in different ways in order to be able to choose the setup that best suits the task.
- Preliminary testing of recorded materials; protocol and setup corrections
- Data acquisition protocol (camera setup, data recording protocol, recording conditions) is defined.
- Data (video) collection from real scenes from the ship during sailing and at anchor
- Determining the area of the image that will be taken into account when learning the model (near the ship, without ship return waves, random, on the horizon, on the edge of the horizon line) and how the data will be labeled.
- Data pre-processing and purification
- Creating a training set and testing set in such a way that the classes are balanced and are equally represented in the set for training and testing.
- Analysis of representative state-of-the-art deep neural networks for the classification and object detection of different architectures that achieve the best results on the benchmark datasets.
- Selection of representative state-of-the-art deep neural networks for experimental training for sea state recognition

- Training and building several computer visions models for recognizing the state of the sea on prepared training sets.
- Evaluation of the model performance in the lab on the test dataset according to standard metrics and choosing the best model.

WHY? For assistance in navigation, assistance in assessing the state of the sea, which is determined manually, and increasing safety, providing additional information about small objects that can be found at sea, such as small boats, buoys, buoys, etc. objects that are not visible on the radar in order to increase the safety of navigation, avoid collisions and enable undisturbed navigation.

FINAL RESULTS:

- Original database ready to train supervised machine learning models for recognizing 8 sea state classes according to Bf 1 to 8, and a database for recognizing small objects on the sea surface.
- Protocol for data acquisition,
- Proposal of image sampling strategies in the case of texture recognition or small object detection and analysis of the results according to the sampling strategy,
- Deep learning model for sea state recognition
- Deep convolutional network model for small object detection at sea

INNO2MARE PROJECT GOALS: Assistance in sea state recognition and detection of small objects at the sea surface.

Detailed explanation of the activities:

Dataset collection

To train the models for sea state recognition, as part of this research study, an extensive set of data that includes the target sea states had to be first collected and then annotated and prepared for machine learning. The collected raw data encompasses 119 video recordings, which collectively constitute 39.61 gigabytes of high-definition visual data. These recordings were made in 11 distinct geographic locations, ensuring a broad coverage of maritime environments. Utilizing a stationary camera with a resolution of 3840x2160, the footage was captured without human intervention and saved using the HEVC codec. The positioning of the camera on the left bridge wing of the vessel afforded a panoramic view of the sea and included segments of the ship, the wake, and the distant horizon. The height at which the recordings were captured varied with the vessel's loading condition: the bridge wing stood at 38.12 meters above sea level while in cargo, and 40.32 meters when in ballast.

Given the inherent challenges in documenting sea states above 6-7, the vessel of choice for this study was a large ocean-going ship, specifically selected for its operational capabilities in the harsh winter conditions and rough seas. This selection was aimed at maximizing the probability of capturing recordings of the higher sea states. The video recordings were conducted on an 89,432 DWT and 173,998 m3 liquefied natural gas transport vessel, akin to the vessels from a 174km3 series previously constructed by Korean shipbuilders.

The exact geographic locations of the recordings are shown in Figure 1.

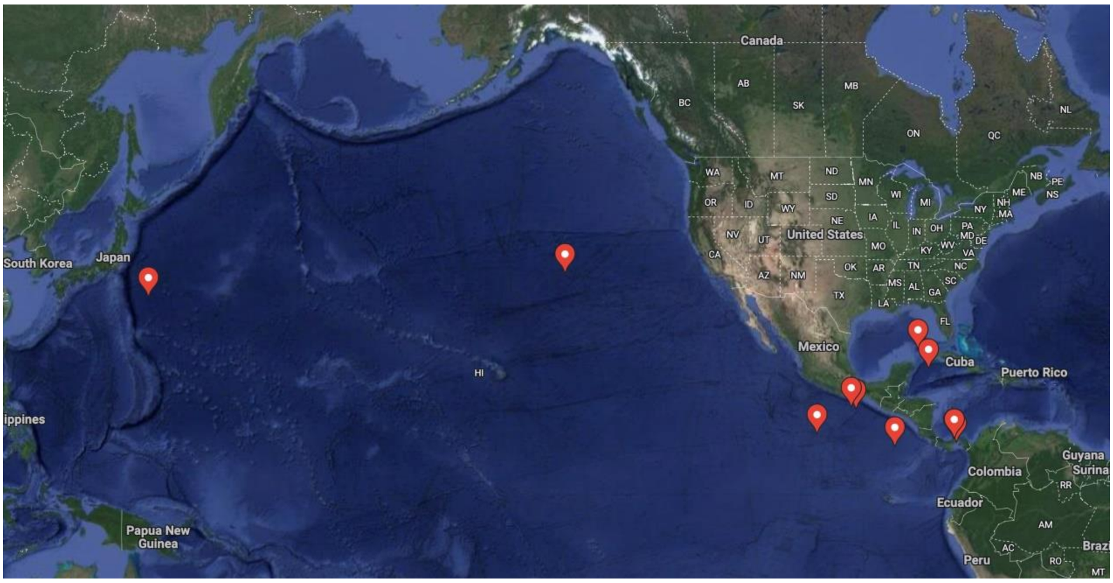

Figure 1. Map of sea state image acquisition locations.

Figure 2 illustrates two representative frames of the video footage from which crops were extracted for subsequent analysis. These frames were selected to demonstrate the type of raw visual data obtained from the stationary camera mounted on the vessel during navigation or when anchored. The image showcases the specific section of the maritime environment that was consistently captured across the various recordings. In subsequent processing, individual crops were extracted from such frames to ensure that the further analysis would be based on consistent and relevant segments of the visual data, focusing on key features of the sea state, and excluding e.g., most of the sky or ship's wake.

Figure 2. Sample frames for maritime video dataset for sea state analysis.

For each recorded video, the sea state in Beaufort scale was estimated by human experts, as would be during regular ship operations. The estimated sea state values according to the Beaufort scale were compared with the measured wind speed using the ship's anemometer, and they mostly matched. The basic data about the collected videos for each sea state states spanning from 1 to 8 on the Beaufort scale is shown in Table 1.

Table 1. Statistics of collected videos for sea states from 1-8 Bft.

Sea state	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
No. sessions	6	7	7	7	7	4	3	3
No. frames	19.173	24.101	14.030	33.063	46.303	16.282	15.094	6.393
Total duration (m:s)	5:20	6:42	3:54	9:11	12:52	4:31	4:12	1:47

Dataset preparation

The basic prerequisite for the implementation of this research is the definition of a customized dataset intended for the development and validation of machine learning models aimed at sea state analysis that is formed from the collection of video recordings. From the collected raw video material, a customized data set for training and testing the model was constructed. The videos were first divided into training and test subsets to prevent further processing the frames from the same video are used for both training and testing the models.

The initial collection of videos was found to be imbalanced concerning the total number of frames, with certain classes containing more data than others. To form the balanced dataset from the initial uneven quantity of data for each sea state, and to distill the video data into a form more suitable for analysis, a systematic sampling approach was employed. This process involved the selection of frames at regular intervals from the recorded footage so that the number of frames for each sea state is about 750. From each chosen frame, a segment with dimensions of 331x331 pixels was extracted from the bottom-left corner of the frame (Fig. 3). These segments were then incorporated into the training or test subsets, depending on whether they were extracted from videos in the training or testing subsets. This dataset was named UNIRI-SeaState-LL. An illustrative set of samples representing each sea state in the resulting dataset is shown in Figure 5.

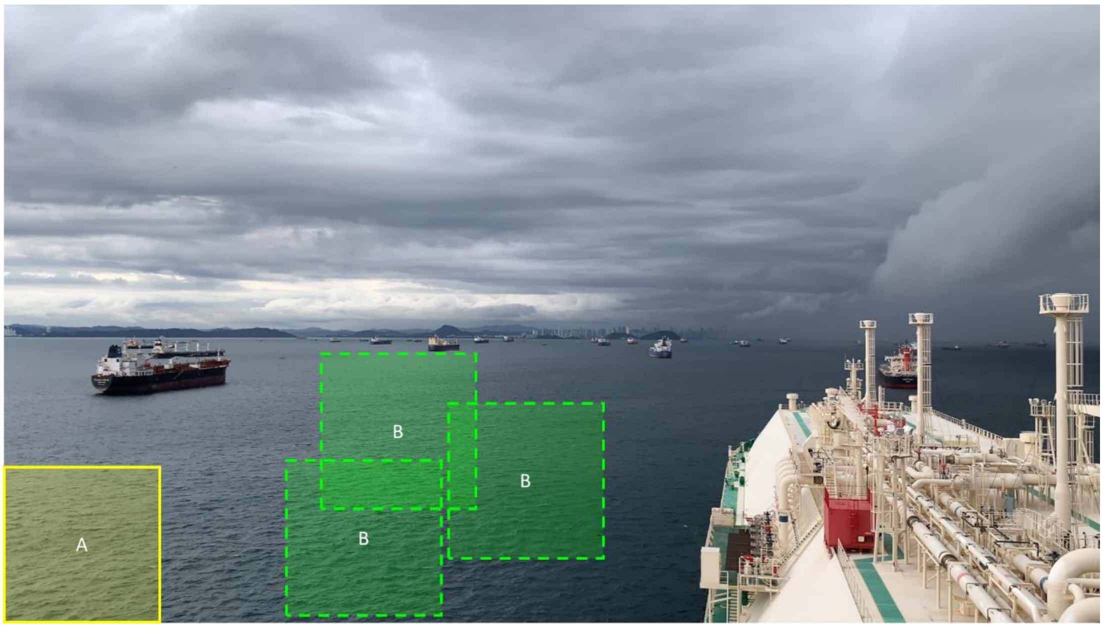

Figure 3. The size and location of cropped portion of initial frame included in the dataset UNIRI-SeaState-LL (Yellow frame marked A). Green frames marked B represent some possible locations of crops included in the dataset UNIRI-SeaState-R.

The rationale behind the frame size and sampling frequency is directly correlated with the model input size and the height from which the sea was recorded. In the development of the dataset, we have also explored a variant where the sample is taken from a random position to the left of the ship's hull (Fig. 3). This variant of the dataset is named UNIRI-SeaState-R.

Figure 4. Examples of extracted sea state images. Top row: states from 1 to 4 Bft. Bottom row: states from 5 to 8 Bft.

Table 2. Class Distribution Balance Across Different Cropped Image Datasets

Dataset	Classes								Total images
	1	2	3	4	5	6	7	8	
UNIRI-SeaState-R	750	750	737	750	750	744	710	766	**5.957**
UNIRI-SeaState-LL	750	736	750	750	750	758	706	770	**5.970**

Table 2 details the resultant class distribution for both datasets. It provides a numerical overview of how images were equitably distributed across the classes to facilitate a balanced representation in line with the Beaufort scale's defined sea states.

Augmentation techniques:

During training, the following data augmentations were used:

- *Random cropping*: The input image size for all models was 224x224 pixels, while the images in the dataset were sized 331x331 pixels. The 331x331 images are themselves either random crops from the initial whole video frame, or crop from the fixed bottom-left position. This augmentation was always applied.

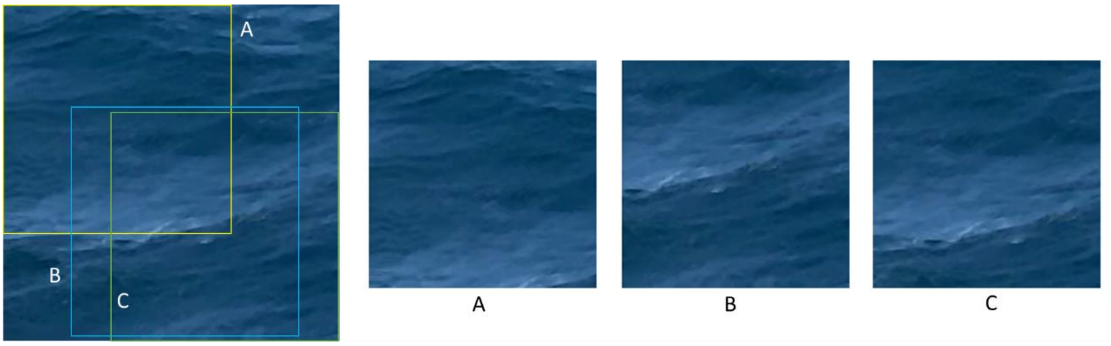

Figure 5. Example of cropping data augmentation. A, B and C show three random crop regions and corresponding cropped images.

- *Motion blur*: A horizontal motion blur was applied to the batch with 50% probability by convolving the image with a horizontal 7x7 motion blur kernel. This was chosen as similar blur was often observed in the original images due to camera panning motion.
- *Horizontal flip*: images were flipped horizontally with 50% probability.
- *Random brightness and contrast adjustments*: Brightness was randomly scaled with a factor in the range [-0.2..2] and contrast was adjusted with a random factor in the range of [0.5..1.5]
- *Random rotation*: images were randomly rotated in range of angles from -0.2pi to 0.2pi radians.
- *Random grayscale conversion:* Images were converted to grayscale with 20% probability.