

Article A Novel Visual-Range Sea Image Dataset for Sea Horizon Line Detection in Changing Maritime Scenes

Manzoor Ahmed Hashmani^{1,2} and Muhammad Umair^{1,2,*}

- ¹ Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia; manzoor.hashmani@utp.edu.my
- ² High Performance Cloud Computing Center, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia
- * Correspondence: m-umair@outlook.com or muhammad_17008606@utp.edu.my

Abstract: Sea horizon line (SHL) detection plays a pivotal role in the computational performance improvement of computer applications for the maritime environment by dividing the image into sea and sky regions. This division isolates the region of interest and reduces the computational cost of further processing. Testing and performance evaluation of SHL detection methods require a robust image dataset covering the maritime environment's features at different geographical, seasonal, and maritime conditions. However, publicly available maritime image datasets are developed under a limited environment with slight-to-moderate variations in maritime features. This article proposes a novel sea image dataset that fills this gap by incorporating various geographical, seasonal, and maritime features. Across West Malaysia, one offshore and four geographically separated onshore locations were selected. On ten different occasions, field observations were recorded using a visualrange optical sensor and weather station. The data collection experiments were conducted between February 2020 until April 2021. The collected data were preprocessed and SHL images were selected based on their high feature diversity. Manual SHL annotation was applied on images, and a ground truth matrix was generated, which serves as a performance benchmark for SHL detection methods. As a result, the dataset presents 2673 high-definition (1920 \times 1080 pixels) RGB images having a combination of 36 different geographical, seasonal, and maritime features to test and evaluate computer vision-based SHL detection methods.

Keywords: visual image dataset; sea horizon line; computer vision

1. Introduction

A sea horizon line (SHL) is a thin boundary that separates a maritime scene into two major regions, i.e., the sea and the sky. This division helps isolate the area under observation which eventually decreases the computational cost associated with any further image processing task. For this reason, detection of the horizon line is considered a critical preprocessing step for maritime computer applications related to the navigation of unmanned aerial vehicles [1,2] and surface ships [3,4], collision avoidance [5,6], object detection [7] and tracking [8], surveillance [9], distance estimation [10], and video stabilization [11,12].

SHL detection is a challenging task [13]. The presence of abundant features in a maritime scene contributes to its complexity. At the forefront is foreground or background objects and environmental conditions with strong linear features. Floating vegetation, algae, wakes, waterborne vessels, cloud formation near the horizon, sea surface conditions, coastal structures, and background landscape are sources of such strong linear features. Partial occlusion of SHL due to foreground objects can also affect the continuous linear properties of SHL as, generally, SHL is the most distinguishable continuous linear feature in a sea image. In such a case, the horizon line may not run across the full width of an image and its presence is limited to a small region or regions of an image. These conditions can challenge the robustness of projection-based computer vision SHL detection methods as they look



Citation: Hashmani, M.A.; Umair, M. A Novel Visual-Range Sea Image Dataset for Sea Horizon Line Detection in Changing Maritime Scenes. J. Mar. Sci. Eng. 2022, 10, 193. https://doi.org/10.3390/ jmse10020193

Academic Editors: Alberto Alberello and Filippo Nelli

Received: 22 August 2021 Accepted: 29 December 2021 Published: 31 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). for the presence of linear features by applying edge detection and linear transformation methods [11] to identify SHL in an image.

Another challenge for SHL detection methods is varying illumination levels across sea surface, cloud cover and sea states [14]. This creates variation in surface glint and glare levels. In some cases, when strong glint and glare are present, there is little color difference between the sea and sky region. Moreover, at times, sky conditions can change sea water's color. For example, across the sea surface, a scattered cloud cover lets different amounts of sunlight penetrate seawater, consequently creating multiple color bands across its surface. Such color bands are also possible due to sea depth variations and the presence of phytoplankton. The presented scenarios are specifically a challenge for SHL detection methods which attempt to separate sea–sky regions by performing statistical analysis, such as the probability distribution of color across the sea and sky region [14].

A method which attempts to solve the SHL detection problem requires a benchmark image dataset of maritime scenes for testing and performance evaluation purposes. The dataset serves as the sole standard to evaluate the robustness of the method in different maritime scenarios. Researchers have proposed various maritime image datasets [8,11,15–17]; however, limitations in geographical, seasonal, maritime, and day conditions, variation in SHL position and angle in images [8,11,15–17], absence of background objects [11], lowresolution images and presence of compression artifacts [11,15], and lack of annotation scheme [17] in these datasets creates a gap for further improvements. In this paper, we present a novel sea image dataset that caters to the need by incorporating features that are distinguishable due to their geographical, seasonal, and maritime scene diversity. In general, the dataset focuses on computer vision and statistical analysis-based SHL detection methods. In total, the dataset has 2673 high-definition (1920 \times 1080 pixels) RGB images. These images have a combination of 36 different geographical, seasonal, environmental, and maritime features to test and evaluate SHL detection methods. To the best of our knowledge, no other publicly available dataset presents such feature diversity in their images.

1.1. Related Work

For this study, we have reviewed the maritime image datasets which are publicly available for download and used for SHL detection problem. These datasets are the buoy dataset [11], Mar-DCT [15], Marine Obstacle Detection Dataset (MODD) v1.0 [16], IPATCH Dataset [17], and Singapore Maritime Dataset [8]. In this section, we have presented those dataset's objectives, attributes and identified their shortcomings.

The buoy dataset [11] is a low-resolution visible-range color video dataset captured from a buoy-mounted forward-looking camera. The dataset is primarily developed to test a ship detection and tracking system in a maritime environment. It offers videos at 25 frames per second at 800×600 pixels. The lack of diversity in the maritime environment and weather conditions, absence of coastline and infrastructures, and presence of compression artifacts are some of the issues that make this dataset less suitable for evaluating modern SHL detection methods.

The Mar-DCT dataset [15] is a collection of visible-range color and infrared range video datasets. The published dataset focused on a boat classification, detection and tracking problem. A subset of this dataset has been used for SHL detection research [18]. The publicly available dataset consists of low and variable-resolution videos in the range between 352×288 to 704×576 pixels. The frame rate also varies between 10 to 25 frames per second. In total, 410 s of video are available. These videos are captured from a fixed platform, and occasional zoom, pan, and tilt operations have been performed. The dataset offers some variations in time of the day, illumination level, and background. However, due to its low resolution and monotonous environment features, the dataset does not cover the wide range of features present in different maritime environments.

The Marine Obstacle Detection Dataset [16] in essence was developed to focus on obstacle detection issues in a maritime environment. The dataset offers 4454 images captured from a near-to-sea-surface moving camera at a standard definition, i.e., 640×480 pixels. A limited variety of maritime scenes with changing backgrounds are present in the dataset. Since the dataset is created using a small, unmanned surface vehicle (USV) thus its geographical coverage is extremely limited. Additionally, the dataset does not offer great variations in camera height and movement which is essential for capturing an important feature, i.e., variation in angle and elevation of the horizon line. Moreover, similar sea conditions, background features, and lack of environmental noise such as fog or haze limit the features of the dataset.

The IPATCH dataset [17] fundamentally addresses the need for computer vision research in maritime surveillance. The videos are recorded from a tug ship in a coastal area using multiple visible and thermal sensors. The visible-range videos are recorded using a five-megapixel camera, and thermal range videos are recorded at 640×480 pixels and 640×512 pixels. This dataset is entirely focused on object detection and tracking, and no annotation scheme is given to address SHL detection problem. We find it unsuitable for SHL detection research for reasons such as the absence of a detailed annotation scheme, lack of versatility in geographical, environmental, maritime conditions, and dataset accessibility issues.

The Singapore Maritime Dataset [8] is a collection of visible and near-infrared band videos captured under geographically limited, moderate weather condition variations and a high maritime traffic environment. The dataset consists of 19,734 high-definition images, i.e., 1920×1080 pixels, extracted from captured videos. The dataset is focused on the presentation of high marine traffic scenes for SHL detection, object detection, and tracking. Since the dataset is created in a limited geographical boundary, uniformity in the features across different videos is present.

Table 1 presents a summary of the discussed dataset's targeted domain and image attributes.

Dataset Targeted Domain		Video Type	Video Quality	Image Source
Buoy Dataset [11]	Ship detection and tracking	Visible range	800×600 pixels	Video
Mar-DCT [15]	Maritime surveillance	Visible and infrared range	352×288 pixels, 704×576 pixels	Video
Marine Obstacle Detection Dataset [16]	Obstacle detection in maritime environment, Sea-Sky segmentation	Visible range	640 imes480 pixels	Video
IPATCH [17]	Maritime surveillance	Visible and thermal range	5 mega pixels, 640 \times 480 pixels, 640 \times 512 pixels	Video
Singapore Maritime Dataset [8]	Maritime object and sea horizon line detection	Visible and near-infrared range	1920 \times 1080 pixels	Video

Table 1. Discussed dataset's targeted domains and image attributes.

1.2. Problem Statement

Many SHL detection methods have been proposed in the literature. They may be categorized into four major domains, i.e., (i) computer vision, (ii) statistical analysis, (iii) machine learning, and (iv) hybrid approaches. Computer vision-based methods usually detect the edges in a targeted image. They then apply line detection algorithms such as the Hough transform, Radon transform, or least-squares regression methods to identify SHL [18]. A statistical analysis approach identifies SHL using techniques such as probability distribution of color to find SHL [14]. An approach based on machine learning applies machine learning models such as SVM, J48, or Naïve Bayes classifiers to classify pixels which belong to the SHL [19]. A combination of the mentioned approaches is categorized as a hybrid

approach. For example, a combination of projection method and statistical analysis to determine SHL [11].

Regardless of their category, the testing and performance evaluation of SHL detection methods requires a well-defined image dataset of diversified scenes [8,20]. For this purpose, a challenging feature matrix can be constructed for quality and suitability evaluation of the SHL image dataset. Such a matrix is presented in Table 2.

Table 2. A challenging feature matrix for SHL detection.

Challenging Feature for SHL Detection	Potential Source	
Background objects [21]	City skyline, Hills, Irregular coastline	
Blur horizon [22]	Haze, Fog, Illumination, Sensor noise	
False linear feature or Edge [3,18,20]	Small or large objects on horizon, Sea waves, Floating vegetation, Edges of cloud, Glint	
Image resolution [23]	Low resolution of optical sensor	
Inseparable color distribution between sea and sky region [20], Absence of line features [13,20]	Haze	
Fewer number of pixel attributes [19]	Small dataset	
Large area of occlusion [14,20]	Foreground ships or objects	

Researchers have proposed various maritime image datasets [8,11,15–17]. These datasets are created from short video clips captured within the vicinity of a specific geographical location at different resolutions and frames per second. This way, a few seconds of a video clip recorded at 60 frames per second can generate more than a thousand images. To its benefit, this approach yields a large volume of an image sequence, but on the downside, the dynamics of foreground or background features do not change much in the resultant images. Thus, similar features are repeated in thousands of images in the dataset, which limits the feature diversity of the dataset. To judge the performance accuracy of the SHL detection method, it is important that it be evaluated under dynamic conditions (i.e., diversity of features) [3]. However, to the best of our knowledge, publicly available SHL image datasets are limited in feature diversity as they are created from a large number of images extracted from a source video, have limited geographical coverage, and do not cover various wave forms (i.e., sea state).

To address this problem, in this paper, we propose the development of a sea horizon line image dataset called Manzoor-Umair: Sea Image Dataset (MU-SID), which is created using still images covering a wider range of environmental, weather, maritime, and sea state conditions at five geographically distinct locations across the eastern and western coasts of West Malaysia.

The prominent contribution of this dataset is the addition of the following features diversity to the existing work.

1. Geographical diversity: extensive feature diversities related to foreground and background objects, sea states, and maritime traffic are subject to geographical location and location of observation platforms. The proposed dataset presents SHL images collected under two scenarios (i.e., offshore, and onshore) at five geographically separated locations across the eastern and western coasts of West Malaysia, covering two separate water bodies (i.e., the Straits of Malacca and the South China Sea). As a result, variations in foreground and background objects and maritime traffic are captured. Additionally, by collecting data from two different bodies of the sea, five sea states (i.e., from sea state 0 to sea state 4) are present in SHL images. To the best of our knowledge, no other sea horizon line image dataset covers five geographical locations and two bodies of sea in its studies. 2. Temporal diversity: features such as illumination levels, glint, glare, haze, fog, and cloud cover change across the day and season. If image collection experiments last for different times of the day and across all seasons of the year, a wide range of such features can be incorporated into a dataset. As a result of ten data collection field trips targeting separate weather and seasonal conditions, the proposed dataset covers various times of the day (i.e., late morning, noon, afternoon, evening, and late evening) and yearly tropical seasons (i.e., summer and monsoon).

The paper is organized as follows: The "Materials and Methods" Section explains the steps taken to capture and construct the proposed dataset. The "Dataset Creation" Section explains the steps taken to preprocess the raw images, the selection of final images and their annotation. Section "Dataset Description and Statistics" discusses the features and statistical properties of the dataset. The "Discussion" Section compares the presented work with previous studies and highlights its novelty. In Section "Applications of MU-SID", we have identified potential application use of the dataset, and finally, the "Conclusion and Future Work" Section presents the concluding remarks and directions for future work.

2. Materials and Methods

- 2.1. Sensors Detail
- 2.1.1. Optical Sensor

The MU-SID is primarily focused on the testing and performance evaluation of visiblerange sea horizon line detection methods. In similar maritime visual-range image dataset development studies, a single or an array of high-definition visual-range cameras were successfully employed to capture the source videos [8,11,15–17]. A similar approach is adopted in this study as well, by employing a Nikon D3400 visual-range digital camera. The camera's CMOS image sensor has 24.2 million effective pixels and a high ISO sensitivity of between 100 and 25,600. The shutter speed of the camera is 1/4000 to 30 s. The captured image is saved in standard RGB color space and has a maximum dimension of 6000×4000 pixels at 300 dpi.

2.1.2. Weather Sensor

To report precipitation and prevailing sea state (based on the Beaufort wind force scale and existing wind conditions) in the proposed dataset, the experiment requires in situ reading of weather parameters. To achieve this objective, a weather station can be deployed at an observation site. For this study, a Vantage Vue wireless weather station is selected, primarily for its ability to record wind speeds as low as 1.7 knots per hour and at a smaller interval of 2.5 s. Additionally, the weather station can record rain, temperature, humidity, and atmospheric pressure data at an interval of 1 min. At each observational site, the setup of the weather station requires updating values such as time, date, longitude, latitude, and the weather sensor's height above sea level. Figure 1 depicts the setups of both sensors during onshore and offshore data collection experiments.



Figure 1. Cont.



(c)



Figure 1. (a) Offshore optical sensor setup; (b) onshore optical sensor setup; (c) offshore weather station setup; (d) onshore optical sensor setup.

2.2. Experimental Design and Considerations

2.2.1. Geographical Diversity

The selection of an observation site is a major factor that affects the feature dynamics in an SHL image. For example, observation sites near a strait or a nearly enclosed sea are subject to less wind–water interaction, resulting in relatively calm sea conditions. However, a site facing a partially enclosed or open sea exhibits dynamic sea conditions due to the larger surface area of the water being exposed to wind. Keeping this fact in consideration, five geographically distinct locations across West Malaysia were selected for data collection experiments. To capture open sea and calm sea state conditions, two locations (one offshore, one onshore) were selected in the Straits of Malacca, while the remaining three locations (all onshore) were facing the South China Sea to capture dynamic sea state conditions. Further details on these locations are presented in Figure 2 and Table 3.



Figure 2. (a) Data collection sites across West Malaysia, offshore site is highlighted by blue marker; (b) ship route during open sea data collection experiment.

Nature of Location	Name	Geographical Coordinates	Corresponding Sea	Number of Observations
Offshore	Open sea off the coast of George Town	5.69 N, 100.08 E	Straits of Malacca	Two
Onshore	Coast of Batu Ferringhi	5.46 N, 100.24 E	Straits of Malacca	Four
Onshore	Coast of Tanjung Balau	1.61 N, 104.26 E	South China Sea	One
Onshore	Coast of Kuala Nerus	5.37 N, 103.12 E	South China Sea	Two
Onshore	Coast of Kampung Kempadang	3.75 N, 103.32 E	South China Sea	One

2.2.2. Temporal Diversity

Meteorological conditions change across the year. These changes are the sources of various challenging weather and environmental features such as illumination levels, cloud cover, precipitation, fog, and haze, etc. To cover a fair proportion of these features over the year, the field data collection experiments in this study were planned and executed at different times of the day (i.e., late morning, noon, afternoon, evening, and late evening) and from February 2020 to April 2021.

2.2.3. Maritime Scene Diversity

One of the challenging scenarios in an SHL image is the presence of false linear features, which can be mistakenly identified as SHL or part of an SHL. The sources of these false linear features are foreground objects such as maritime traffic, wakes, floating vegetation, or background objects such as hills, city skylines, etc. To capture these false features, two data collection experiments were conducted on a ship which travelled 45 miles into the open sea. During these experiments, the city skyline, hills, and a variety of maritime traffic and wakes were captured in images. Additionally, four experiments were conducted at an onshore location (5.46 N, 100.24 E) during which small boats, wakes, hills, ships on the horizon, and cloud formation near the horizon were captured.

2.2.4. Observation Platforms and Sensors' Placement

In an open sea, sea waves cause a movement of six degrees of freedom. On a small boat or ferry, this movement is noticeable. However, a large ship generally remains stable under moderate sea conditions. For accurate readings, it is important to place the weather station on a stable surface. Therefore, a large cruise ship was selected for open sea data collection experiments. The ship is 160.3 m long and 22.8 m wide, with a gross tonnage of 15,653. As depicted in Figure 2b, the ship goes 45 km into the open sea. At its mid-journey position (5.69 N, 100.08 E), the reported sea depth is 50 m. The weather station was mounted on the forward deck of the ship, where minimal interference due to the movement of the ship was observed. For the onshore data collection experiments, sites having platforms stretching into the sea (5.46 N, 100.24 E and 1.61 N, 104.26 E) or flat surfaces next to the shoreline (5.37 N, 103.12 E and 3.75 N, 103.32 E) were chosen to install a weather station and optical sensors.

2.3. Data Collection

2.3.1. Image Data Collection

It is important that the captured images present optimal variations in geographical, temporal, and maritime features. The study covers (i) offshore and (ii) onshore SHL image collection experiments. Thus, separate methods are followed to capture images. For offshore image collections, the images are captured from the bridge-roof, upper and lower decks of the ship. Additional variation is achieved in combination with the bow, starboard, and port side of the ship. By following this protocol, variations in height, illumination, background objects, and an overall 270-degree view of SHL are captured from the ship.

Since a variety of viewpoints are not available from an onshore observation site, sensor zoom, pitch, yaw, and roll are more extensively applied during onshore experiments. Regardless of the nature of the observation point, the optical sensor is either mounted on a tripod or held in hand while capturing the images. In total, 3184 super high-definition standalone images at 6000×4000 pixels are captured. These images have a spatial dot density of 300 dpi, a bit depth of 24, and an sRGB color representation.

2.3.2. Weather Data Collection

Weather data is collected at each site to monitor precipitation and wind. For every experiment, the weather station is setup prior to the image capturing experiment. The standard setup relays wind data at an interval of 2.5 s. Additionally, every minute, weather parameters are internally stored in a CSV format. A half-hourly average wind speed

8 of 15

is calculated and mapped on the Beaufort wind force scale [24] to estimate the current sea state.

3. Dataset Creation

3.1. Image Selection

As a first step toward dataset creation, a subset of images with unique maritime features is manually selected from raw images using the following two stages. Firstly, location-wise images are clustered based on their time signatures. Secondly, an image that represents the maximum feature diversity in its cluster is manually selected and removed from the cluster. Step two is repeated if necessary. This way 1299 images are selected from a total of 3184 images.

3.2. Sub-Image Extraction and Selection

Since the raw images are captured at a super high definition of 6000×4000 pixels, thus from a single image, it is possible to extract multiple sub-images of high definition at 1920×1080 pixels using a sliding window. By applying this approach, nine sub-images are extracted from each raw image. An extracted image is named by coupling source raw image's name and a postfix value mentioning the extracted image's order in the extraction sequence. For example, a sub-image named "DSC_0035_3.JPG" indicates that it's the third sub-image extracted from a raw image named "DSC_0035.JPG". As a result, a pool of sub-images is available for further pre-processing. At this stage, sub-images with high blur, undesirable artifacts and visually missing SHL (e.g., Figure 3b,c,e,f,h,i) are removed. The remaining sub-images are used to construct the proposed dataset. An example of the sub-image extraction and selected SHL images are presented in Figure 3.



(a)



(**b**)









Figure 3. (a) Source raw image; (b–j) Extracted sub-images; (d,g,j) Selected images with visible SHL.

3.3. Ground Truth Scheme

The ground truth (GT) value identifies the actual position of an object of interest in an image. To create GT values for the sea horizon line, we have referred and modified the GT scheme presented by Prasad et al. [8]. The modified GT scheme is depicted in Figure 4. Here, P1(x1,y1), and P2(x2,y2) are the starting and ending coordinates of sea horizon line. *M* is the middle point of the sea horizon line. *D* is the distance from the starting point of normal to *M*. The angle of sea horizon line with reference to the normal passing through *M* is denoted with θ .



Figure 4. Ground truth marking scheme.

3.4. Image Annotation

The final set of images are annotated manually using the desktop version of Super-Annotate. The annotation serves two purposes; firstly, it identifies SHL; secondly, it is used as a reference for performance evaluation of an SHL detection algorithm. For each image, the sea horizon line is manually drawn and labeled to the best of our ability. The annotation software generates a JSON file from which the starting and ending points of SHL are retrieved, and the proposed GT values are calculated. The final GT values and its corresponding image file name is presented in a single CSV file. Figure 5 shows the manual annotation process using SuperAnnotate software.



Figure 5. Image annotation using SuperAnnotate software.

4. Dataset Description and Statistics

The SHL dataset presented in this work consist of 2673 high-definition (1920×1080 pixels) RGB images. The dataset offers 36 different features. A combination of these features such as different day, weather, and environmental conditions, sea states, linear features, partial occlusion, foreground and background objects, and artifacts present a challenge for sea horizon line detection algorithms. These 36 features are categorically described in Table 4.

Table 4. Features list of proposed dataset.

Day Conditions	Weather Conditions	Sea States	Environmental Conditions	False Linear Features	Occlusion	Presence of Object	Artifacts
 Late morning; Noon; Afternoon; Evening; Early night 	 6. Sunny; 7. Partially cloudy; 8. Mostly cloudy; 9. Overcast; 10. Drizzle 	11. Flat surface; 12. Ripples; 13. Small wavelets; 14. Large wavelets with few whitecaps 15. Small waves with frequent whitecaps	 16. Haze; 17. Mild-to-strong solar glare; 18. Mild-to-strong glint; 19. Variation in sea surface color; 20. Smooth color transition between sea and sky region 	21. Wakes; 22. Foam; 23. Sea waves; 24. Linear color bands on sea surface; 25. Linear cloud formation at horizon	26. Partial occlusion by ship; 27. Partial occlusion by other objects	28. Hills; 29. City skyline; 30. Shoreline; 31. Boats and ships	32. Variable-sea to sky area proportion; 32. Variable-sea skyline angles; 34. Blur; 35. Occasional optical sensor noise; 36. Grainy noise

The SHL is captured from different heights (2 to 20 m from sea level) and from different angles thus a wide diversity in P1(x1,y1), and P2(x2,y2), SHL's relative position (*D*) in the image, and angle (θ) is present in the dataset. The details of these attributes are given in Table 5.

Table 5. Sea horizon line statistics.

	Values in Pixels					Degree
	x1	у1	<i>x</i> 2	<i>y</i> 2	D	θ
Minimum	0	0	254	0	14	53.6
Maximum	1637	1080	1920	1080	1054	125.8

The dataset also offers a fair distribution of the occurrence of SHLs in 4 equally distributed cross-sections of the images; thus, a variable sea-to-sky ratio is present. For the interest of our reader, Table 6 represent SHL's frequencies of occurrence in a specific cross-section of images along with their percentages.

Table 6. Presence of sea horizon line in equally distributed cross-sections of images.

Image Cross-Section	Frequency of SHL Occurrence	Percentage of SHL Occurrence
Тор	756	28.3%
Upper Middle	699	26.2%
Lower Middle	593	22.2%
Bottom	625	23.4%

A wide deviation in SHL's angle is essential to address the need for testing the SHL detection algorithm in moving conditions. The proposed dataset offers SHL angles in the range of 53 to 126 degrees. The overall distribution of SHL angles is presented in Table 7.

Table 7. Sea horizon line angle distribution.

Degree Greater than	Degree Equal to or	Frequency of SHLs	Percentage of SHLs
	Less than	in Given Range	in Given Range
53	90	1576	59%
90	126	1097	41%

As mentioned earlier, the images in the MU-SID dataset have a combination of 36 features. Details on those features are mentioned in Table 4. In Figure 6, we present some of the samples from MU-SID which shows a combination of features, such as overcast and shoreline (Figure 6a), strong glare and sea state two (Figure 6b), haze and hilly terrain (Figure 6c), partially cloudy evening (Figure 6d), background city skyline (Figure 6e), sea state three, sea waves, and foam (Figure 6f), haze, foreground object and city skyline (Figure 6g), SHL angle and position variation (Figure 6h), late evening and uniform sea-cloud color (Figure 6i), small objects on the horizon (Figure 6j), early night conditions (Figure 6k), partial occlusion of SHL by close-to-camera object (Figure 6l), strong glint (Figure 6m), sea state four (Figure 6n), and drizzle, wake and foreground object (Figure 6o).



Figure 6. Sample images from MU-SID; (**a**–**o**) features such as different level of illumination, sky conditions, strong glint and glare, foreground and background objects, false linear features, sea states, partial occlusion, haze, variation in SHL angle and position in different cross-sections of images are visible.

5. Discussion

In this paper, a novel visual-range sea image dataset for sea horizon line detection is proposed. The dataset contributes towards the existing work by incorporating a new approach to SHL image dataset creation. Rather than using a standard method of creating an SHL image dataset by extracting similar-features image-frames from short video clips, this study focused on capturing high-definition still images of SHLs at different times of the day and across the seasons, resulting in visible changes in maritime scenery. Additionally, to the best of our knowledge, this is the first study that covers five geographically distinct locations and two separate water bodies, i.e., the Straits of Malacca and the South China Sea, resulting in a variety of sea and weather conditions. Moreover, by utilizing a weather station to report wind parameters to estimate sea state while capturing the SHL images, the dataset reports a novel feature, i.e., the sea state in its images. Overall, the study reports the presence of 36 distinct features.

6. Applications of MU-SID

The detection of SHL is the first step in various maritime computer applications. The MU-SID contributes towards the addition of important features by incorporating geographical, seasonal, and maritime diversities along with a wide range of challenging scenarios for testing and evaluating such applications. Some of the use cases for MU-SID are presented in this section.

6.1. Sea Horizon Line Detection Applications

With a combination of 36 different features in its images, the MU-SID is a challenging dataset for SHL detection methods. Characteristics such as variation in daylight and sky conditions, presence of five sea states, foreground and background objects, false linear features, the smooth color gradient in between sea and sky regions, partial occlusion of SHL, environmental noise, and artifacts makes it a dataset to test the robustness of SHL detection methods.

6.2. Coastal Surveillance Applications

Coastal environments are usually cluttered with features such as maritime traffic, city skyline, complex terrain, and seashore. The presence of these features in a test dataset helps in the performance evaluation of coastal surveillance applications. The proposed dataset offers these features with variations in maritime traffic, illumination levels, different background conditions and sea states.

6.3. Navigational Awareness Applications

Nowadays, unmanned aerial and surface vehicles, small and large boats, and ships rely on fully or partially automated navigational systems. These systems examine image frames from a video feed and suggest navigational corrections accordingly. In such scenarios, the horizon line is considered as a reference for navigational awareness [1,3,4,6,16]. For these applications, variation in optical sensor height from sea surface can be a feature of interest. The images presented in MU-SID are captured from different heights (i.e., from the top deck of a ship to 2 m above sea level). Additionally, different zoom, pitch, yaw, and roll settings are applied to the optical sensor to represent maritime scenes better.

7. Conclusions and Future Work

Detection of SHLs is pivotal for any image processing task related to the maritime scene. However, to test an SHL detection method, an image must represent a wide range of maritime features. In this paper, we have presented the Manzoor-Umair: Sea Image Dataset (MU-SID), which offers versatility in image feature set. Maritime scenes from one open-sea and four on-shore locations across West Malaysia were captured during February 2020–April 2021. The images were captured at different heights from the sea surface, time of the day, and by applying various camera zoom, pitch, yaw, and roll settings. As a result, the dataset offers a combination of 36 features of interest. Some of the unique features of this dataset, which distinguish it from previously presented datasets, are maritime scenes that present diverse geographical, seasonal, and sea state conditions. The dataset consists of a total of 2673 high-definition RGB images at 1920 imes 1080 pixels. These images are manually annotated, and ground truth values for starting and ending points of the sea horizon line, the distance between the starting point of normal drawn to the mid-point of the SHL, and the SHL's angle with respect to the normal are given to evaluate the performance of SHL detection methods. Potential application domains which can use this dataset for their testing and performance evaluation purpose are also mentioned.

For future work, we recommend two directions for feature diversity improvement; firstly, more dynamic sea state conditions can be added. This addition will increase the complexity arising from much rough sea states where big waves, spray, foam, and water splashes occlude the sea horizon line, and SHL no longer exhibits a linear property across the horizon; secondly, open sea conditions with fog and light-to-heavy rain can be considered for weather related feature enhancement of the dataset.

Author Contributions: M.A.H. supervised the data collection field trips, managed the cross-country travels, helped in data collection procedures, provided onsite technical support and guidance, proposed the structure of the manuscript, and reviewed it; M.U. planned the offshore and coastal travels,

14 of 15

designed, and executed the experiments, processed the data, created the dataset, and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Yayasan Universiti Teknologi PETRONAS—Fundamental Research Grant (YUTP-FRG-2019, grant number 015LC0-158).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Manzoor Umair: Sea Image Dataset (MU-SID) can be downloaded from https://www.kaggle.com/umairatwork/manzoorumair-sea-image-dataset-musid (accessed on 1 December 2021).

Acknowledgments: Thanks to Syed Muslim Jameel for his help in local transportation of equipment and materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Hiba, A.; Sántha, L.M.; Zsedrovits, T.; Hajder, L.; Zarandy, A. Onboard Visual Horizon Detection for Unmanned Aerial Systems with Programmable Logic. *Electronics* 2020, *9*, 614. [CrossRef]
- 2. Timotheatos, S.; Piperakis, S.; Trahanias, P. Visual Horizon Line Detection for UAV Navigation. Int. J. Mech. Control 2019, 20, 16.
- 3. Li, F.; Zhang, J.; Sun, W.; Jin, J.; Li, L.; Dai, Y. Sea–sky line detection using gray variation differences in the time domain for unmanned surface vehicles. *Signal Image Video Process.* **2021**, *15*, 7. [CrossRef]
- 4. Wang, B.; Su, Y.; Wan, L. A Sea-Sky Line Detection Method for Unmanned Surface Vehicles Based on Gradient Saliency. *Sensors* **2016**, *16*, 543. [CrossRef] [PubMed]
- McGee, T.G.; Sengupta, R.; Hedrick, K. Obstacle detection for small autonomous aircraft using sky segmentation. In Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 4679–4684.
- De Croon, G.C.H.E.; De Wagter, C.; Remes, B.D.W.; Ruijsink, R. Sky segmentation approach to obstacle avoidance. In Proceedings of the Aerospace Conference, Big Sky, MT, USA, 5–12 March 2011; pp. 1–16.
- Shi, J.; Jin, J.; Zhang, J. Object detection based on saliency and sea-sky line for USV vision. In Proceedings of the IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 14–16 December 2018; pp. 1581–1586.
- 8. Prasad, D.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. Video Processing from Electro-Optical Sensors for Object Detection and Tracking in a Maritime Environment: A Survey. *IEEE Trans. Intell. Transp. Syst.* 2017, *18*, 1993–2016. [CrossRef]
- 9. Petković, M.; Vujović, I.; Kuzmanić, I. An Overview on Horizon Detection Methods in Maritime Video Surveillance. *Trans. Marit. Sci.* 2020, *9*, 106–112. [CrossRef]
- Gladstone, R.; Moshe, Y.; Barel, A.; Shenhav, E. Distance estimation for marine vehicles using a monocular video camera. In Proceedings of the 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 29 August–2 September 2016; pp. 2405–2409.
- 11. Fefilatyev, S.; Goldgof, D.; Shreve, M.; Lembke, C. Detection and tracking of ships in open sea with rapidly moving buoy-mounted camera system. *Ocean Eng.* **2012**, *54*, 1–12. [CrossRef]
- 12. Schwendeman, M.; Thomson, J. A Horizon-Tracking Method for Shipboard Video Stabilization and Rectification. J. Atmos. Ocean. Technol. 2015, 32, 164–176. [CrossRef]
- 13. Jeong, C.; Yang, H.S.; Moon, K. A novel approach for detecting the horizon using a convolutional neural network and multi-scale edge detection. *Multidimens. Syst. Signal Process.* **2019**, *30*, 1187–1204. [CrossRef]
- Liang, D.; Liang, Y. Horizon Detection from Electro-Optical Sensors Under Maritime Environment. *IEEE Trans. Instrum. Meas.* 2020, 69, 45–53. [CrossRef]
- 15. Bloisi, D.D.; Iocchi, L.; Pennisi, A.; Tombolini, L. ARGOS-Venice boat classification. In Proceedings of the 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Karlsruhe, Germany, 25–28 August 2015; pp. 1–6.
- 16. Kristan, M.; Sulic Kenk, V.; Kovacic, S.; Pers, J. Fast Image-Based Obstacle Detection from Unmanned Surface Vehicles. *IEEE Trans. Cybern.* **2016**, *46*, 641–654. [CrossRef]
- Patino, L.; Cane, T.; Vallee, A.; Ferryman, J. PETS 2016: Dataset and challenge. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1240–1247.
- 18. Prasad, D.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. MuSCoWERT: Multi-scale consistence of weighted edge Radon transform for horizon detection in maritime images. *J. Opt. Soc. Am. A* **2016**, *33*, 2491–2500. [CrossRef]
- Sergiy, F.; Volha, S.; Lawrence, O.H.; Dmitry, B.G. Horizon detection using machine learning techniques. In Proceedings of the 5th International Conference on Machine Learning and Applications (ICMLA'06), Orlando, FL, USA, 14–16 December 2006; pp. 17–21.
- 20. Prasad, D.K.; Prasath, C.K.; Rajan, D.; Rachmawati, L.; Rajabaly, E.; Quek, C. Challenges in video based object detection in maritime scenario using computer vision. *arXiv* 2016, arXiv:1608.01079.

- 21. Mou, X.; Shin, B.-S.; Wang, H. Hierarchical RANSAC for accurate horizon detection. In Proceedings of the 2016 24th Mediterranean Conference on Control and Automation (MED), Athens, Greece, 21–24 June 2016; pp. 1158–1163.
- 22. Sun, Y.; Fu, L. Coarse-fine-stitched: A robust maritime horizon line detection method for unmanned surface vehicle applications. *Sensors* **2018**, *18*, 2825. [CrossRef]
- Lu, Y.; Zhang, M.; Zheng, L.; Shen, Y. Dark channel prior principle and morphology based horizon detection method. In Proceedings of the 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Turin, Italy, 22–25 May 2017; pp. 1–5.
- 24. Singleton, F. The Beaufort scale of winds-its relevance, and its use by sailors. Weather 2008, 63, 37-41. [CrossRef]