**Enhancing Water Body Detection with Satellite Image Feature Engineering**

**Team Members:**

[Laxmi Narayana Mangilipally](https://unt.instructure.com/groups/171341/users/324768) - 11605829

[Pranitha Pothuguntla - 11656849](https://unt.instructure.com/groups/171341/users/347874)

Aashritha Sandiri - 11712554

[Shasank Reddy Ganta](https://unt.instructure.com/groups/171341/users/323004) - 11605399

<https://github.com/Pothugunt1a/Group-5---Project-/upload>

**1. Introduction**:

Water bodies play a critical role in various ecological, societal, and economic aspects, making their accurate detection and monitoring essential. Leveraging the capabilities of the Sentinel-2 satellite, our team aims to enhance water body detection through advanced feature engineering techniques applied to satellite imagery. This project holds significant promise for applications in water resource management, disaster response, climate change monitoring, and ecosystem assessment.

Feature engineering plays a critical role in enhancing the performance of water body detection algorithms. By extracting and transforming information from satellite imagery, feature engineering can effectively capture the unique characteristics of water bodies and distinguish them from other land-cover types. This report explores various feature engineering techniques that can be employed to improve water body detection from satellite imagery.

Water indices are a widely used class of features for water body detection. These indices exploit the specific spectral properties of water, such as its high reflectance in the near-infrared (NIR) band and its low reflectance in the green band. By combining multiple bands, water indices can effectively highlight water bodies and suppress the influence of other land-cover types.

Commonly used water indices include:

1. Normalized Difference Water Index (NDWI)

2. Modified Normalized Difference Water Index (MNDWI)

3. Automated Water Extraction Index (AWEI)

**2. Goals and Objectives:**

**2.1 Motivation:**

The motivation for using a Sentinel-2 satellite to detect water bodies is that it provides numerous advantages over other methods, such as:

* **High spatial resolution:** These Sentinel-2 images can be used to detect small and narrow water bodies and have a spatial resolution of 10 meters.
* **Wide spectral range:** The Sentinel-2 satellite has a wide spectral range, covering wavelengths from the visible (blue, green, and red) to the near-infrared (NIR) and shortwave infrared (SWIR). This also allows for the detection of water bodies in a variety of conditions, such as the presence of clouds, shadows, and vegetation.
* **High temporal resolution:** Sentinel-2 satellites can be used to monitor water bodies over time.
* **Free and open access**: Sentinel-2 data is accessible to a wide range of users as it is freely available to the public.

Sentinel-2 satellite imagery advantages make it an ideal tool for a variety of water-related applications, such as: Monitoring water resources, Mapping water bodies, Studying water dynamics, and Detecting water bodies at risk.

It is well-suited for detecting water bodies in developing countries and many other regions where the traditional methods of water monitoring may be limited or unavailable. It is an affordable and accessible tool for governments, NGOs, and other organizations striving to enhance water conservation and management in these areas since it is freely and publicly accessible.

**2.2 Significance:**

The significance of a collection of water bodies images captured by the Sentinel-2 Satellite, with black and white masks where white represents water and black represents something else but water, generated by calculating the Normalized Water Difference Index (NWDI) with a greater threshold to detect water bodies is that it would provide a high-quality, global dataset of water bodies that could be used for a wide range of applications, including:

* Water resource management: The dataset could be used to monitor water levels, track changes in water bodies over time, and assess water quality. This information could be used to inform water management decisions, such as how to allocate water resources and how to protect water quality.
* Disaster response: The dataset could be used to identify and map flood inundation areas, track the movement of oil spills and other pollutants, and monitor the impact of natural disasters on water bodies. This information could be used to help emergency responders and disaster planners make better decisions about how to respond to and recover from disasters.
* Climate change monitoring: The dataset could be used to monitor changes in the distribution and size of water bodies over time. This information could be used to assess the impacts of climate change on water resources and to develop adaptation strategies.

In addition to these specific applications, the water body dataset and masks could also be used to support research in a variety of fields, such as hydrology, limnology, and ecology. The dataset could also be used to develop new educational and outreach materials on water resources.

The significance of the water body dataset and masks is that it would be a valuable resource for researchers, practitioners, and policymakers working in a variety of fields related to water resources management, disaster response, and climate change monitoring.

Here are some specific examples of how the water body dataset and masks could be used to make a positive impact on the world:

* The dataset could be used to develop early warning systems for floods and other water-related disasters. This could help to save lives and reduce property damage.
* The dataset could be used to improve water management in drought-prone areas. This could help to ensure that people have access to clean water during times of water scarcity.
* The dataset could be used to monitor the impacts of climate change on water resources. This information could be used to develop adaptation strategies to protect water resources and the people who depend on them.

**2.3 Objectives:**

To curate and provide a high-quality, global dataset of satellite imagery for water resources, ensuring precise discrimination between water and non-water pixels. This dataset aims to support diverse applications, including water resource management, disaster response, climate change monitoring, wetland and ecosystem health assessment, and the evaluation of human activities' impact on water bodies.

Dataset Quality Assurance:

Precision in Water Pixel Classification: Develop and maintain a dataset with a focus on accuracy, minimizing misclassification of non-water pixels to ensure reliable identification of water bodies.

Comprehensive Global Coverage: Curate satellite imagery from diverse geographical locations to create a dataset that represents water resources across the globe.

High-Resolution Imagery: Prioritize high-resolution satellite images to enhance the dataset's capability for detailed water body detection.

Application Accessibility:

Wide Range of Applications: Structure the dataset to cater to various applications, including but not limited to water resource management, disaster response, and climate change monitoring.

User-Friendly Access: Implement mechanisms for easy and widespread access to the dataset, promoting its utilization across different domains.

Utilization in Water Resource Management:

Water Availability Monitoring: Enable water resource managers to monitor water availability, assess irrigation needs, and plan for sustainable water usage.

Effective Disaster Response: Support disaster response teams in identifying flooded areas, tracking water movement, and coordinating relief efforts efficiently using the dataset.

Climate Change Research:

Impact Analysis: Facilitate climate change researchers in analyzing the dataset to understand the impacts of climate change on water resources.

Change Detection: Monitor changes in water body extent and characteristics, providing valuable information for informed adaptation strategies.

Ecosystem Health Monitoring:

Wetland and Ecosystem Assessment: Enable the monitoring of wetlands and aquatic ecosystems, aiding in the identification of threats and implementation of conservation measures.

Biodiversity Maintenance: Contribute to the preservation of biodiversity by providing data for the regulation of water flow and pollution filtering in ecosystems.

Assessment of Human Activities:

Impact Evaluation: Use the dataset to assess the impact of human activities, such as agriculture, urbanization, and pollution, on water bodies.

Water Quality Monitoring: Track changes in water quality to inform sustainable development practices and promote responsible resource management.

By achieving these objectives, the curated dataset will serve as a valuable resource for a multitude of applications, fostering advancements in water resource understanding, ecosystem conservation, and sustainable development practices on a global scale.

**2.4 Features:**

* Global coverage: The Sentinel-2 satellite provides global coverage, so the dataset would include images of water bodies from all over the world.
* High spatial resolution: The Sentinel-2 satellite has a spatial resolution of 10 meters, which means that the images can be used to identify and map small water bodies.
* Frequent revisit time: The Sentinel-2 satellite revisits the same location every 5 days, which means that the dataset can be used to monitor changes in water bodies over time.
* Accurate water body masks: The water body masks are generated using a robust and accurate method, and they have been validated against reference data.
* User-friendly format: The dataset and masks will be made available in a user-friendly format, with appropriate documentation and metadata.

In addition to these features, the dataset and masks can also be used to retrieve other useful information about water bodies, such as Water body extent, Water body connectivity, and water body type.

**3 Related Work (Background):**

Focuses on enhancing water body detection in satellite imagery through a combination of traditional image processing techniques and advanced feature extraction methods. Initially, the code loads satellite images in different color spaces, such as RGB, HSV, and grayscale, aligning with conventional approaches that utilize spectral indices like NDWI and MNDWI. The integration of the Canny edge detection algorithm underscores the importance of precise boundary delineation, especially for water body edges. Gaussian blur is subsequently applied to reduce noise, and channel-wise analysis of the H, S, and V channels in the HSV spectrum sheds light on the unique characteristics of water bodies in different color spectrums. The introduction of thresholding techniques on individual channels contributes to the generation of feature maps, emphasizing potential water body regions. Moreover, the custom median smoothing function hints at efforts to refine predictions by reducing image noise. Looking forward, the code outlines plans for incorporating K-Means clustering for feature extraction, aligning with contemporary trends in unsupervised learning for extracting meaningful features from satellite imagery. The subsequent classification step indicates the code's potential for real-time prediction, aligning with the overarching project goal of accurate water body detection.

**4 Dataset:**

A collection of water bodies images captured by the Sentinel-2 Satellite. Each image comes with a black and white mask where white represents water and black represents something else but water. The masks were generated by calculating the NWDI (Normalized Water Difference Index) which is frequently used to detect and measure vegetation in satellite images, but a greater threshold was used to detect water bodies.

https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies/data



**5 Detail Design of Features:**

**5.1 Approach**

This Water Body Detection with Satellite Image Feature Engineering project focuses on utilizing advanced feature engineering techniques to accurately detect water bodies in satellite imagery. This involves employing image processing and edge detection algorithms, particularly Canny edge detection, to extract relevant features from the satellite images. The code specifically explores the HSV (Hue, Saturation, Value) color space, analyzing each channel separately and combining the results to enhance water body detection.

**Analysis:**

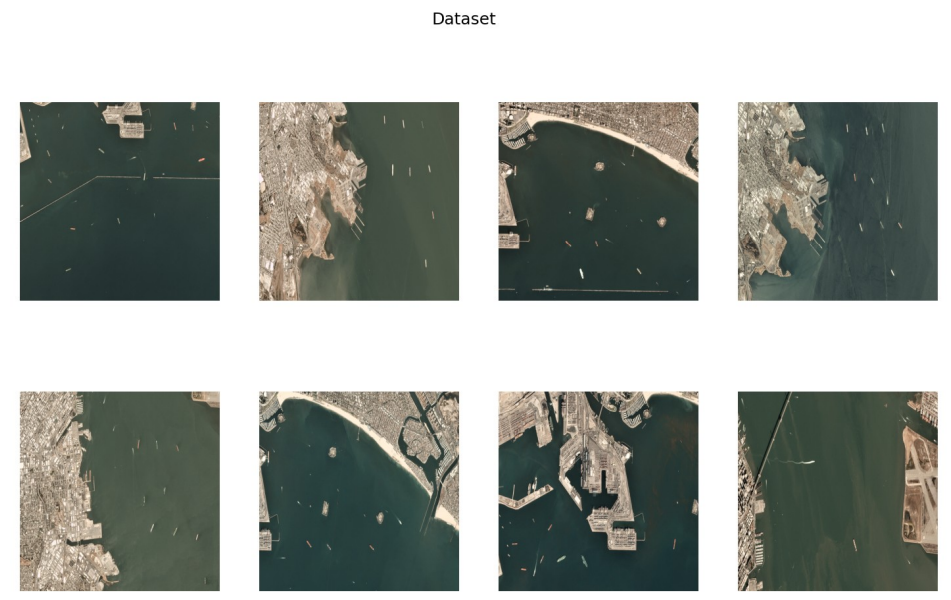
The analysis phase of this code involves assessing the effectiveness of the feature engineering techniques in accurately identifying water bodies in satellite imagery. The analysis includes visualizations of the original images, Canny edge-detected images, individual channels in the HSV spectrum, and thresholded images. The results of these visualizations contribute to the evaluation of how well the applied techniques highlight water bodies and enhance their visibility in the processed images.

**6 Implementation:**

The implementation phase of the code involves the following key steps:

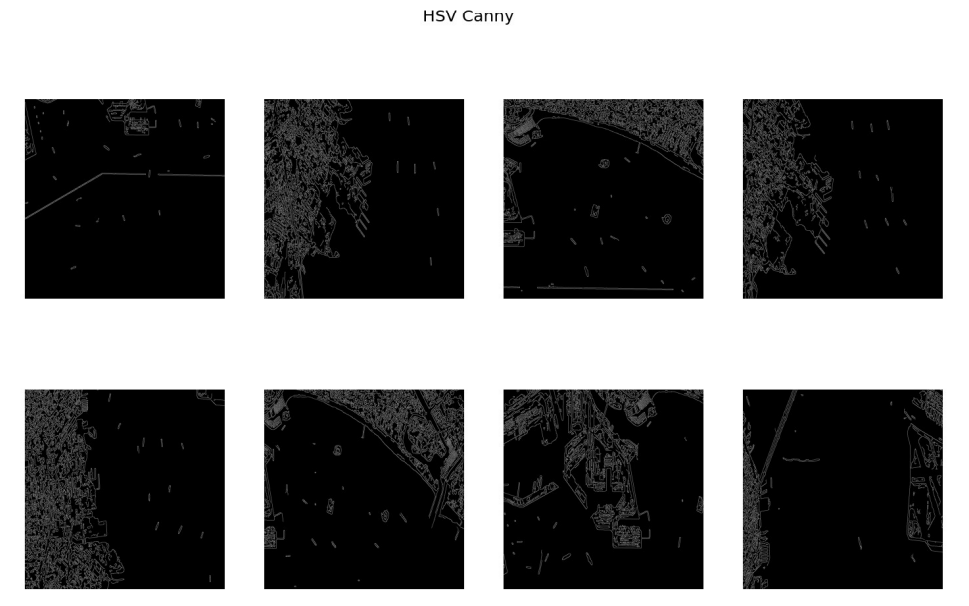
**6.1.1 Loading and Visualization:**

Loaded the dataset of satellite images, converts them to the HSV color space, and visualizes the original images along with their corresponding HSV representations.



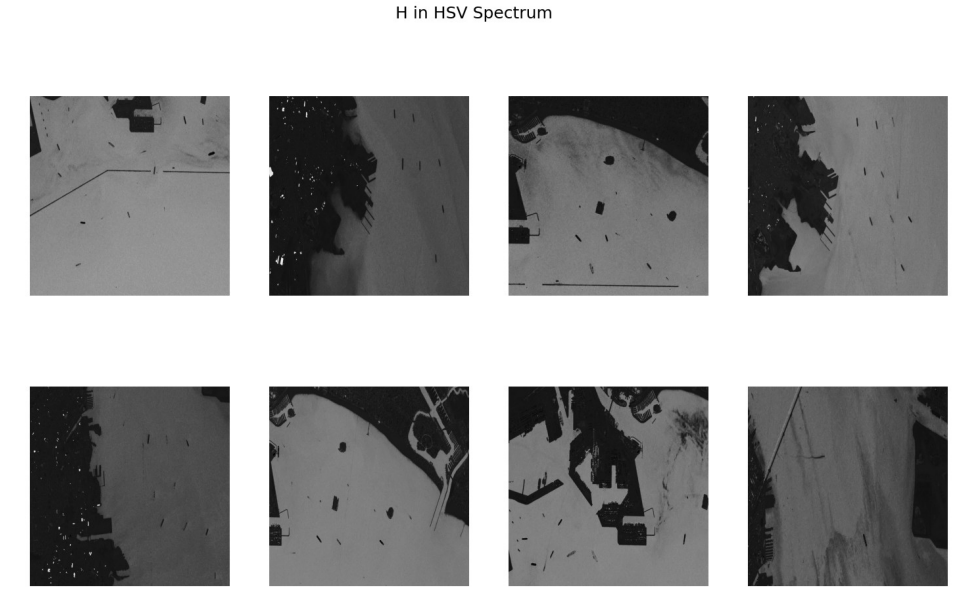
**6.1.2 Canny Edge Detection:**

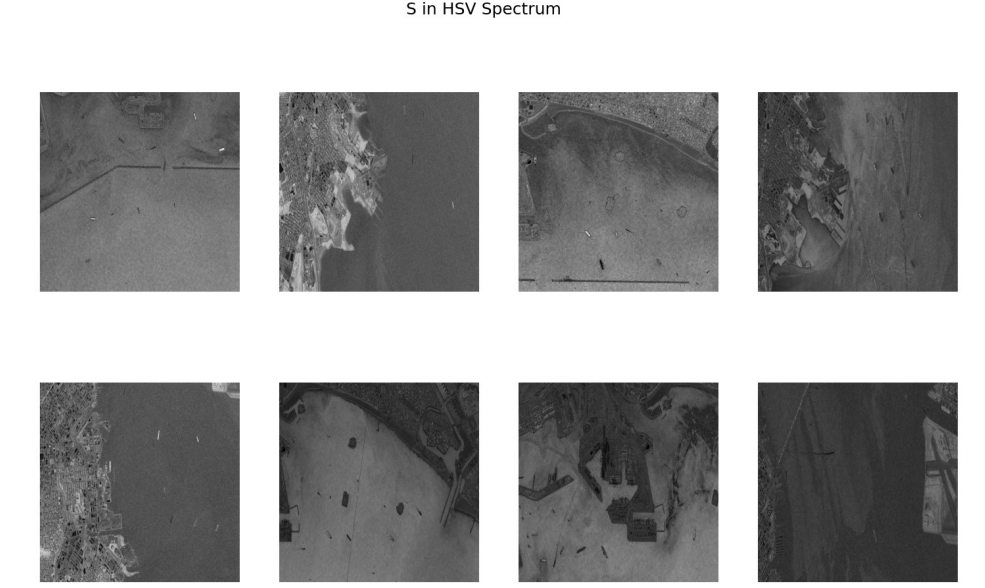
Canny edge detection is applied to the HSV images to emphasize edges and potentially highlight boundaries of water bodies.



**6.1.3 Median Smoothing:**

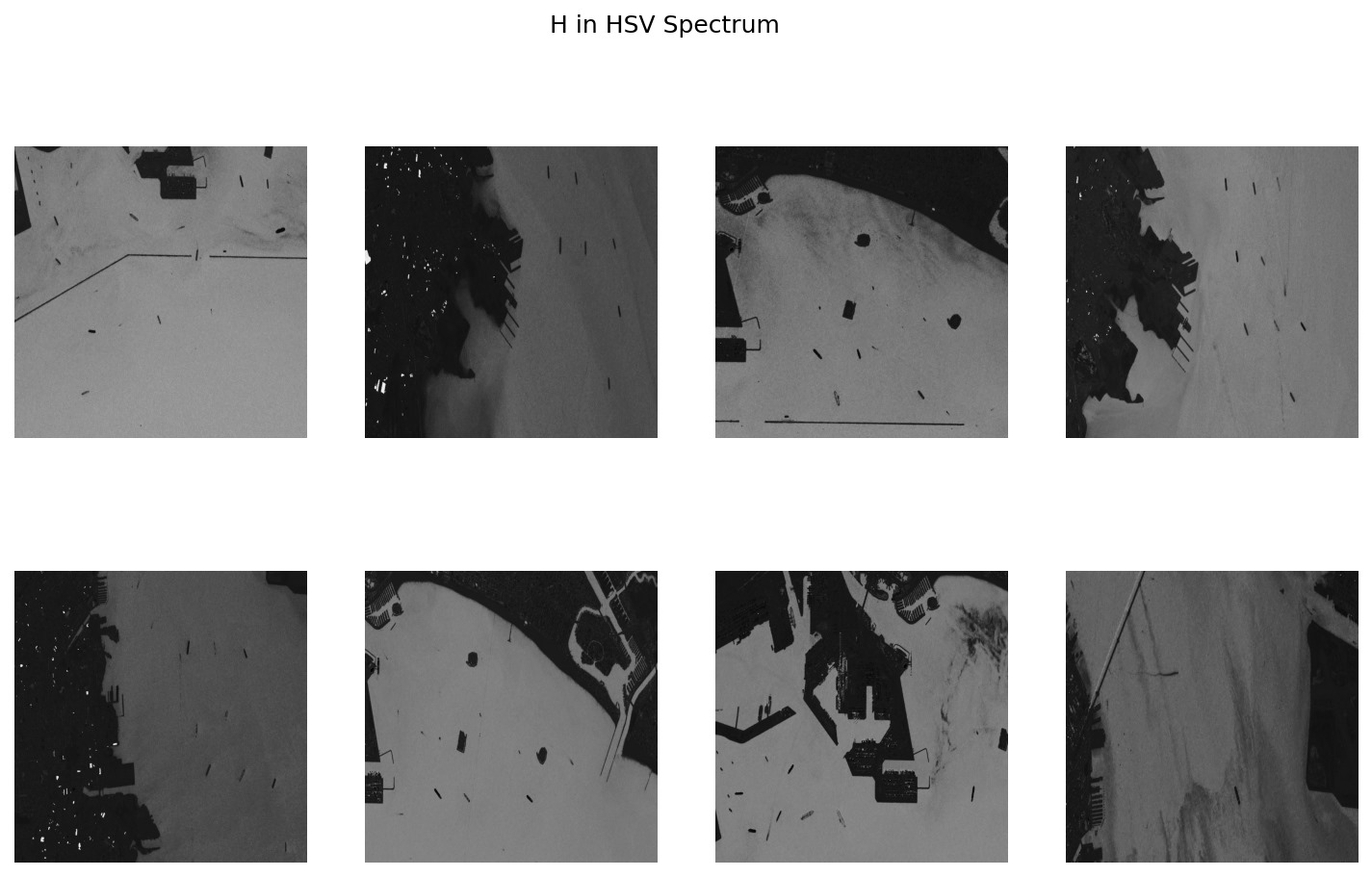
Median smoothing is applied to individual channels (H, S, V) in the HSV spectrum. This step aims to reduce noise and enhance the clarity of features.

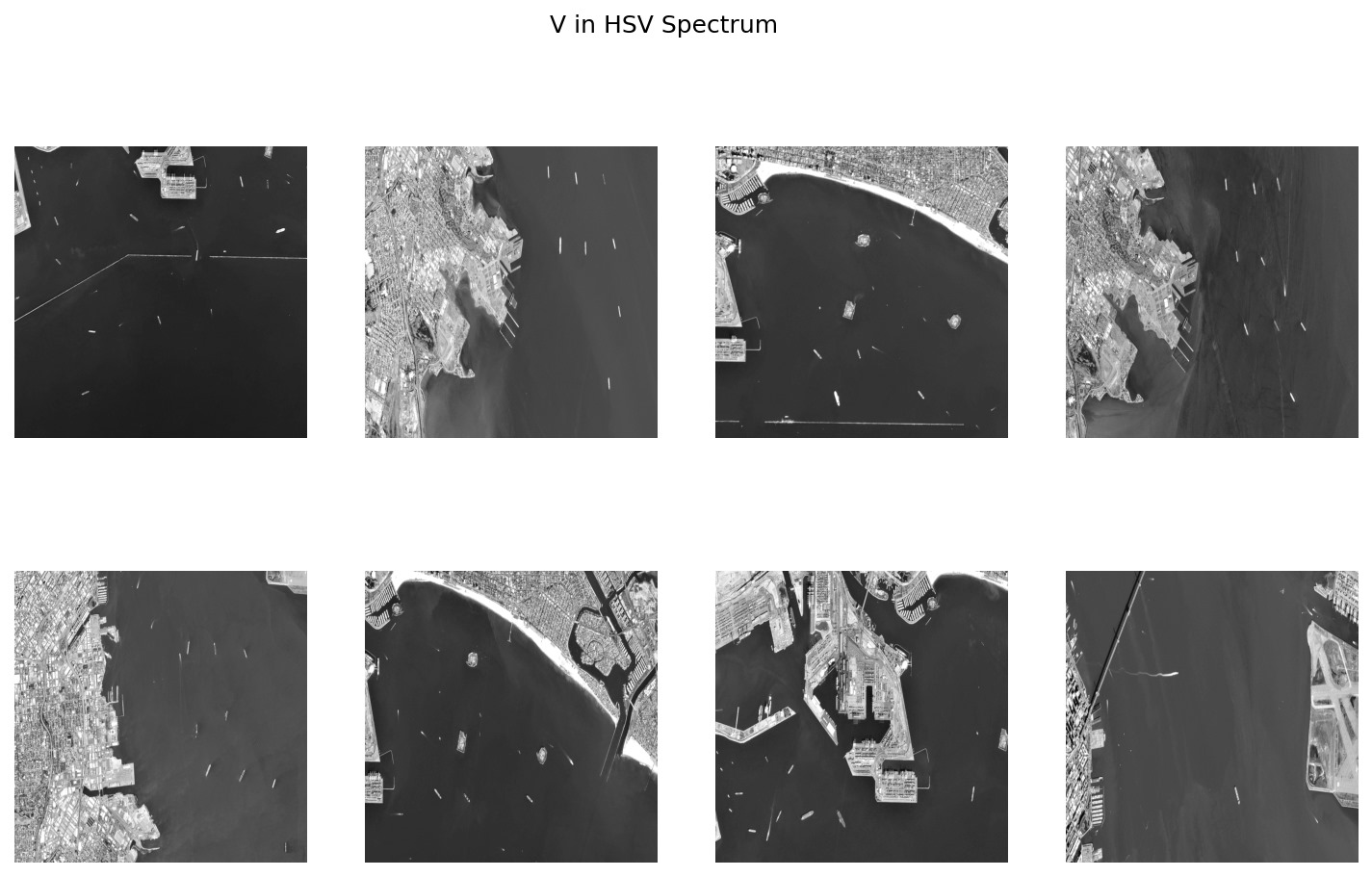
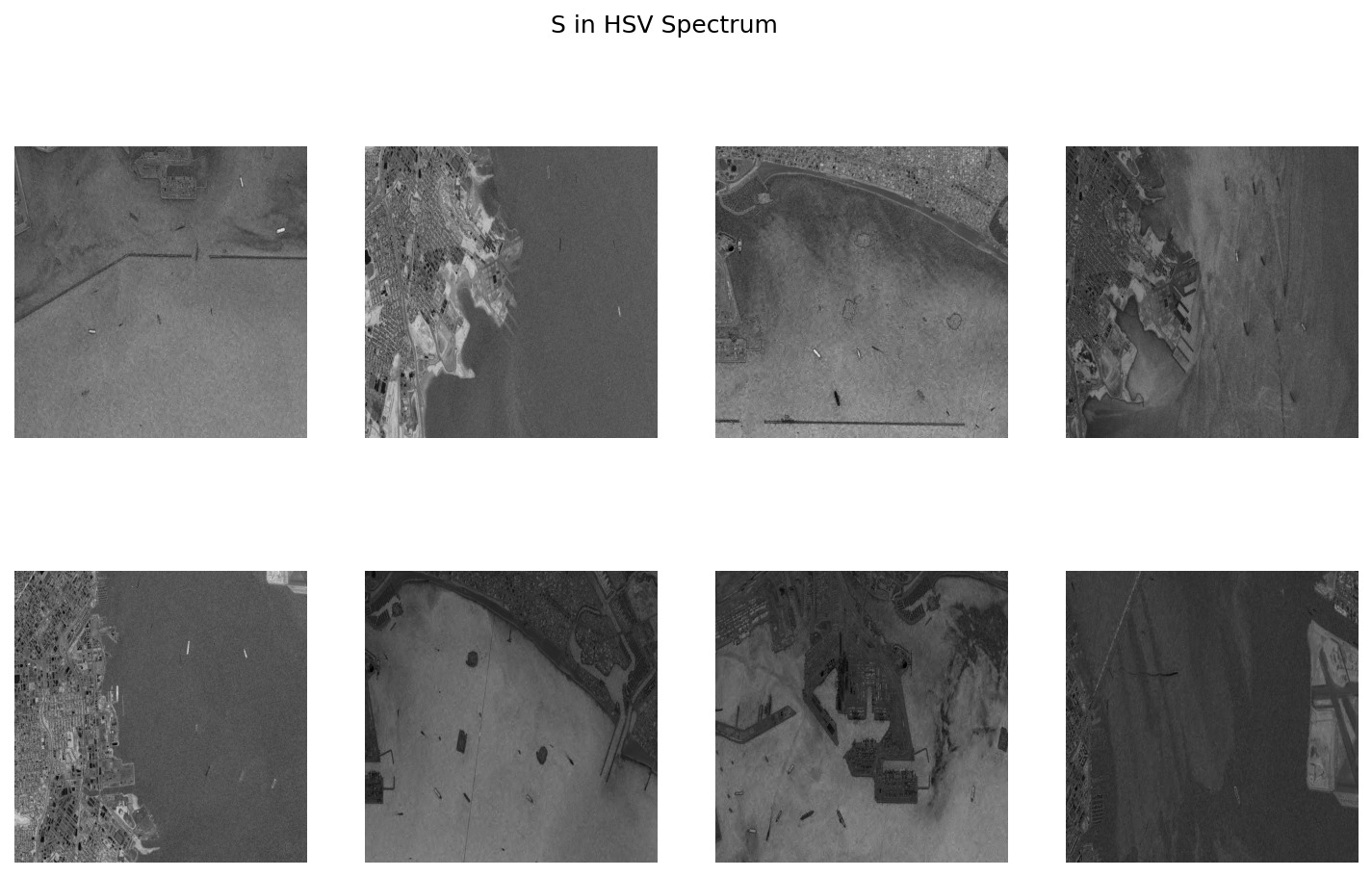




**6.1.4 Visualization of Channels:**

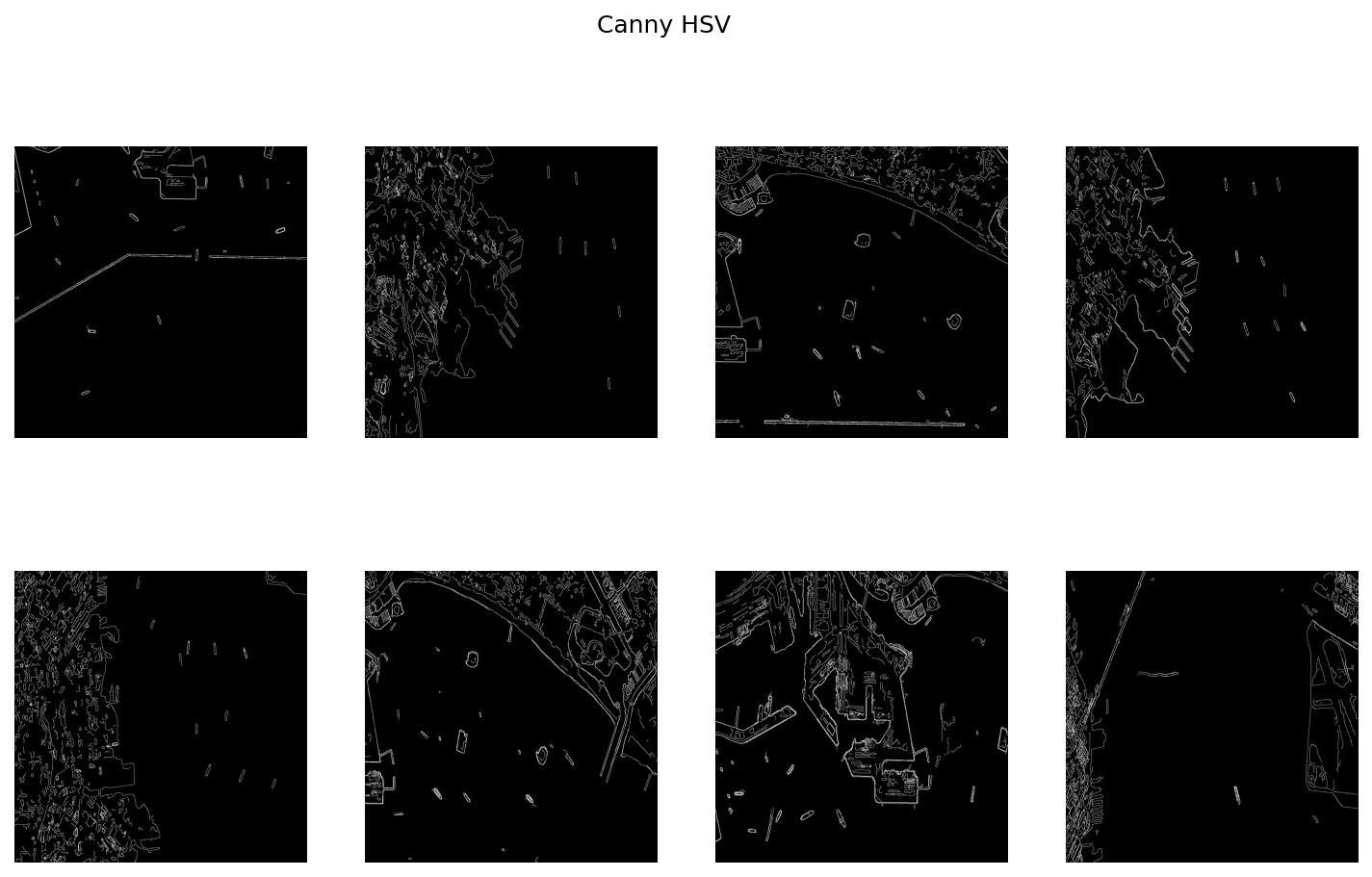
This further visualizes the individual channels (H, S, V) after applying Gaussian blur and Canny edge detection to each channel separately.





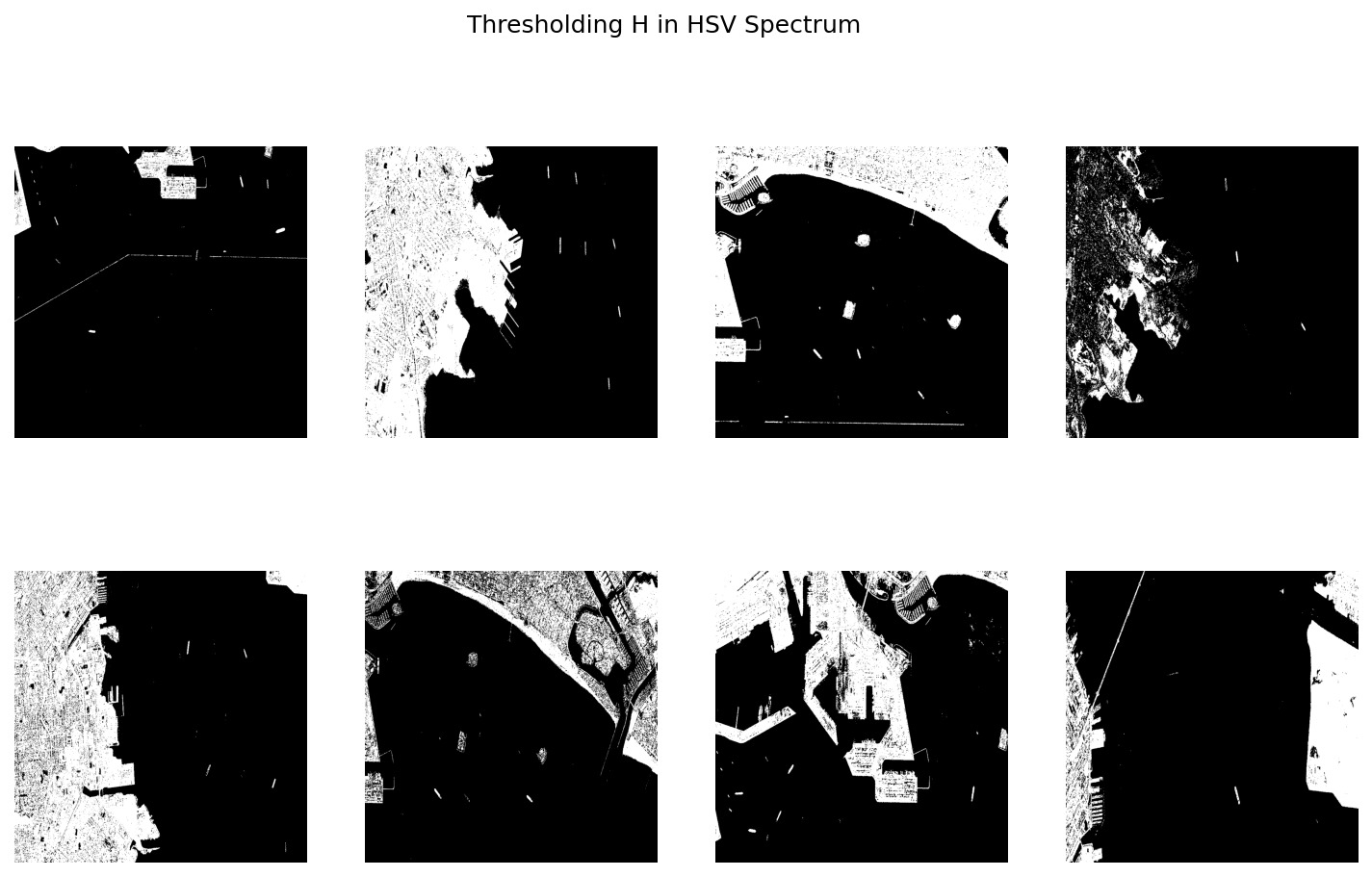
**6.1.5 Combining Canny Edges:**

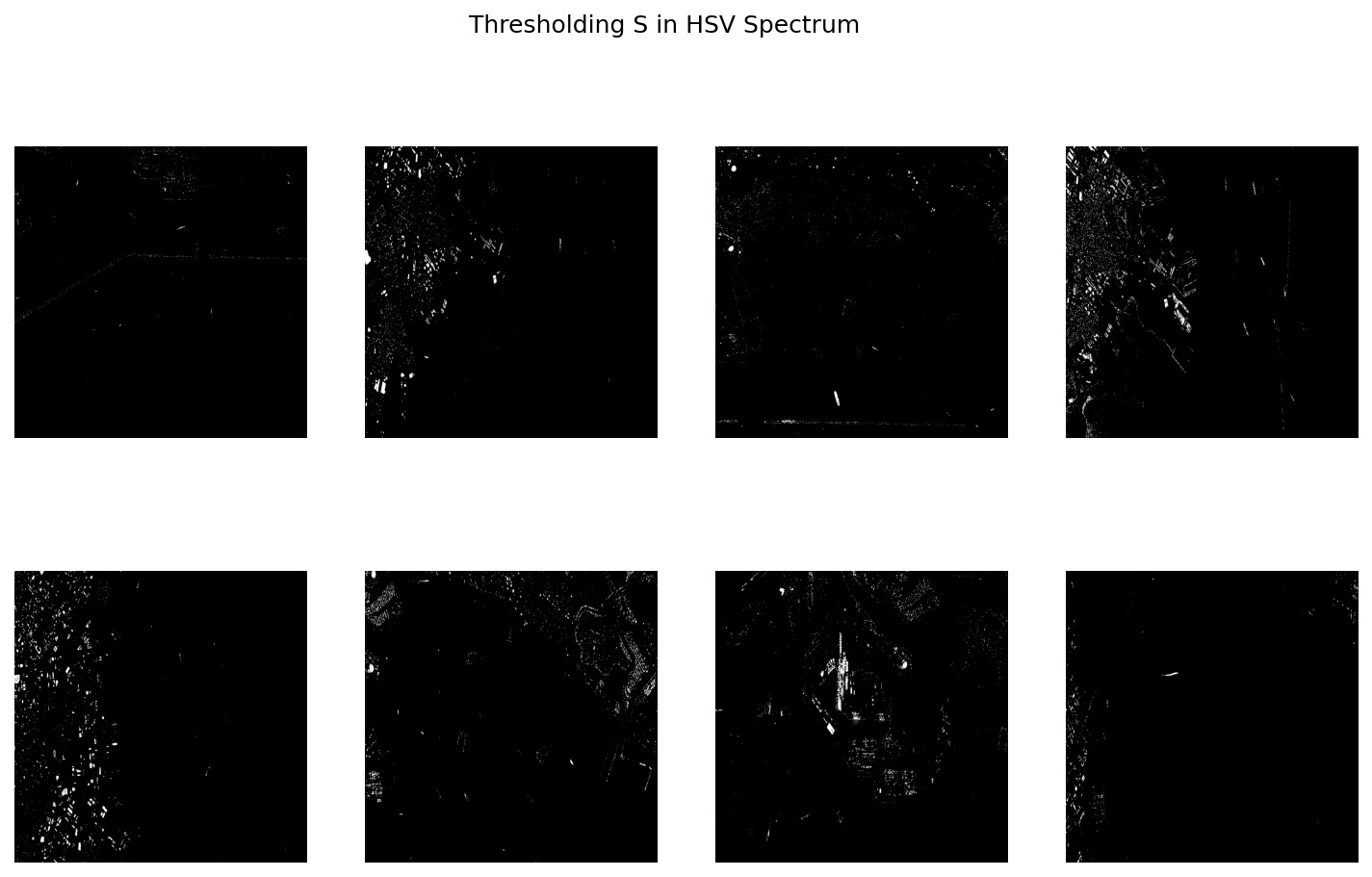
The Canny edges obtained from individual channels are combined to create a composite image that potentially enhances the detection of water bodies.

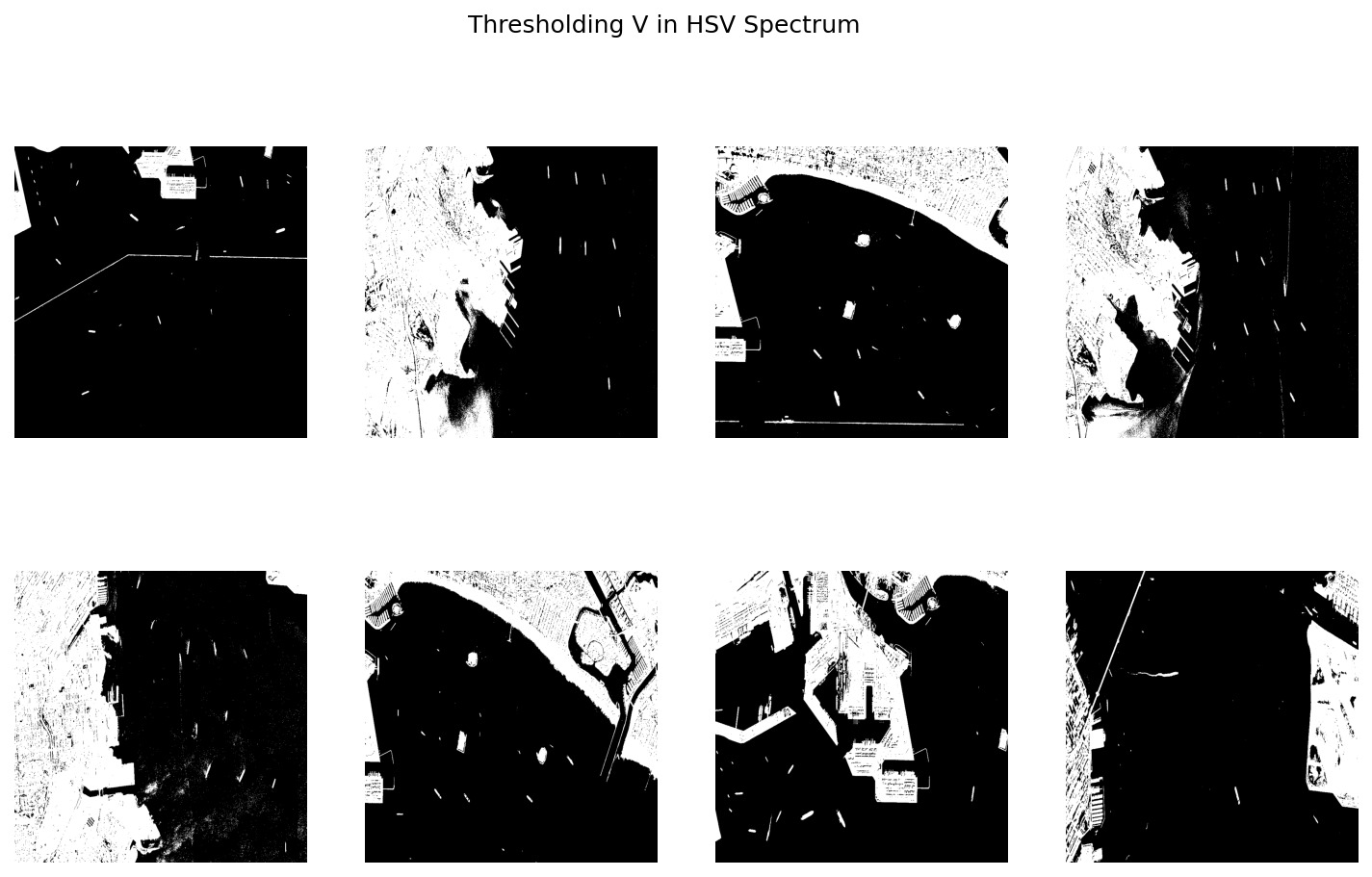


**6.1.6 Thresholding:**

Thresholding is applied to individual channels to create binary masks highlighting specific intensity ranges associated with potential water bodies.

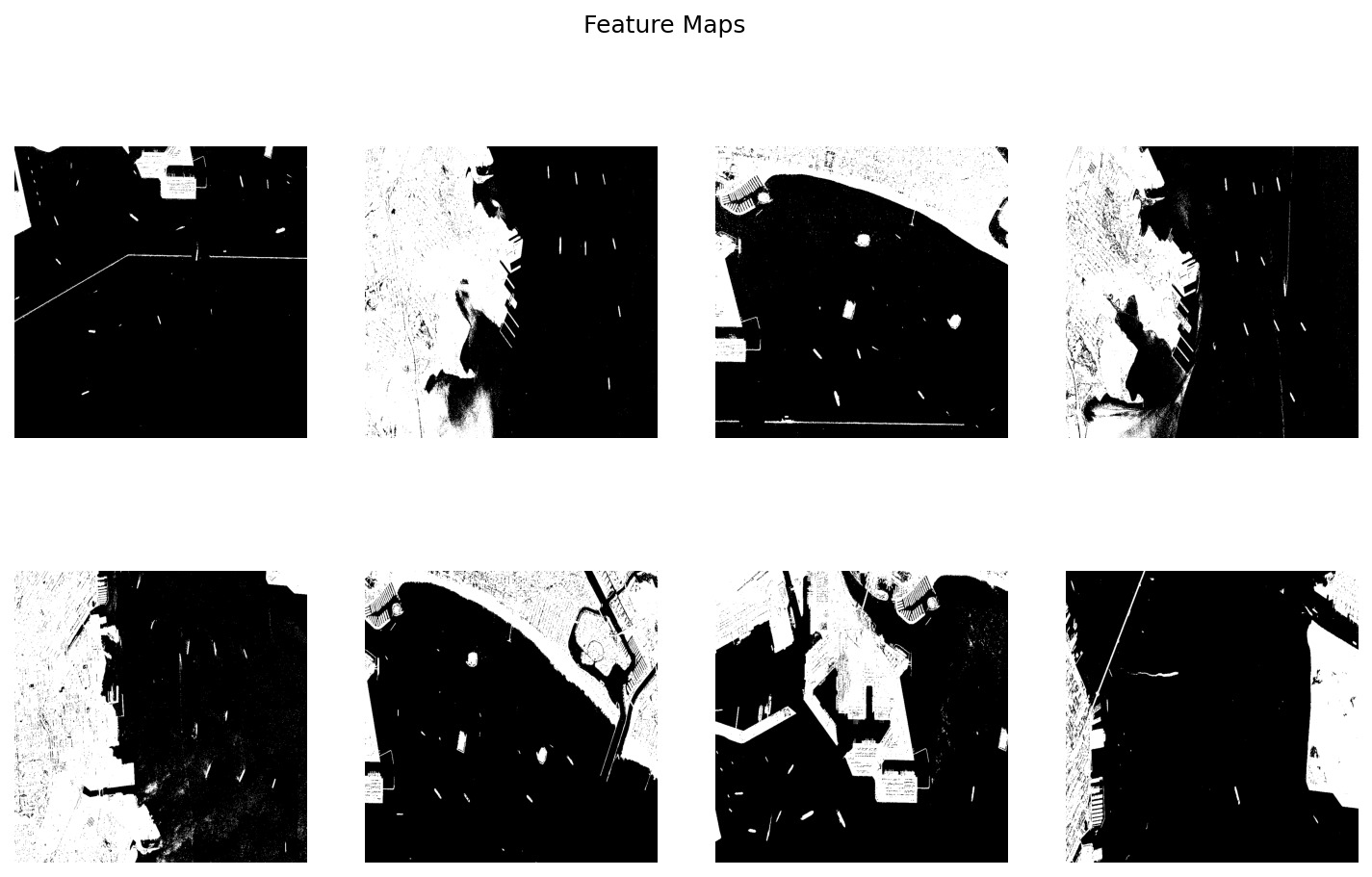






**6.1.7 Feature Maps:**

The final step involves creating feature maps based on the thresholding results, consolidating information from multiple channels to identify regions likely to represent water bodies.

****

**7 Preliminary Results:** The preliminary results indicate that the code is successful in applying various image processing techniques to emphasize potential water bodies in satellite imagery. The visualization of feature maps and thresholded images provides insights into the effectiveness of the applied methods in highlighting regions that may correspond to water bodies.

**8 Project Management:**

**8.1 Implementation Status Report:**

**8.1.1 Work Completed:**

**•Responsibilities and Contribution**

| Student | Task and Contribution |
| --- | --- |
| Laxmi Narayana Mangilipally | Worked on Canny methodology and implemented same in documentation  (25%) |
| Aashritha Sandiri | Data Selection, Data visualisation, and documented in report.  (25%) |
| Pranitha Pothuguntla | Worked on Gaussian blurring, Canny Edge detection and documented in report  (25%) |
| Shasank Reddy Ganta | Feature extraction and documentation  (25%) |

**8.1.2 Work to be Completed:**

**Description:**

**Step 1: K-Means Clustering for Feature Extraction**

**K-Means Clustering:**

K-Means is an unsupervised machine learning algorithm used for clustering similar data points into groups, called clusters. In your case, you're using K-Means to extract features. The algorithm partitions the data into 'k' clusters, where each data point belongs to the cluster with the nearest mean. These clusters can be considered as features representing certain characteristics of the data.

**Feature Extraction:**

After applying K-Means, each data point will be assigned to a cluster, and you can use the cluster assignments as features for your classification model. These features represent the underlying structure or patterns in the data that the clustering algorithm has identified.

**Step 2: Classification for Real-Time Prediction**

**Classification Technique:**

Now that you have extracted features using K-Means, the next step is to build a classification model. This model will be trained on a labeled dataset where each data point is associated with a class label (e.g., water body or non-water body). Common classification algorithms include decision trees, support vector machines, or neural networks.

**Real-Time Prediction:**

In a real-time scenario, your model will take the K-Means-derived features as input and predict whether a given area contains a water body or not. The model will generalize from the training data to make predictions on new, unseen data.

**•Responsibilities and Contribution**

| Student | Task and Contribution |
| --- | --- |
| Laxmi Narayana Mangilipally | Model training and implemented same in documentation  (25%) |
| Aashritha Sandiri | Prediction Analysis, Data visualisation, and documented in report.  (25%) |
| Pranitha Pothuguntla | Model Selectionand documented in report  (25%) |
| Shasank Reddy Ganta | Performance testing, analysis and documentation  (25%) |

**9 References:**

https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies/data

<https://ieeexplore.ieee.org/document/9498442>

<https://www.researchgate.net/publication/371498940_Deep_Learning_for_Automatic_Extraction_of_Water_Bodies_Using_Satellite_Imagery>

<https://www.mdpi.com/2072-4292/10/4/554>

https://orbit.dtu.dk/en/publications/towards-automatic-monitoring-and-mapping-of-benthic-vegetation-us

<https://www.mdpi.com/2072-4292/14/6/1429>

https://www.mdpi.com/2073-4433/10/11/725