

Memory-Efficient Georeferencing of Segmented Satellite Imagery

Corrado Chiatante , *Graduate Member, IEEE*, Håvard S. Løvås , *Member, IEEE*,
Dennis Langer , *Member, IEEE*, Joseph L. Garrett , *Member, IEEE*, Roger Birkeland , *Member, IEEE*,
Milica Orlandić , *Member, IEEE*, and Tor Arne Johansen , *Senior Member, IEEE*

Abstract—Accurate georeferencing is essential for the prompt and rapid use of Earth Observation (EO) data. Direct georeferencing algorithms, using satellite orbital position and attitude to compute geolocation pixels, are commonly used. However, small satellites employing Commercial Off-The-Shelf (COTS) components can experience significant uncertainties in attitude and position telemetry data. This paper introduces Memory-Efficient Georeferencing of Segmented Satellite Imagery (MEGSSI), an automatic georeferencing pipeline that uses land-sea-cloud segmentation to reliably identify Ground Control Points (GCPs). These are used to correct the satellite attitude for georeferencing by minimizing the reprojection error. The corrected attitude state is then used to re-estimate each pixel geolocation through direct georeferencing. The proposed pipeline is scalable and designed for future implementation onboard, and it achieves a geolocation Root Mean Square Error (RMSE) of 0.81 pixel. This sub-pixel accuracy demonstrates that high-precision georeferencing is attainable despite inherent resource constraints and large initial telemetry errors. The method is validated and demonstrated using the HYPISO-1 and HYPISO-2 satellites which each carry a hyperspectral pushbroom sensor.

Index Terms—Georeferencing, Small-satellite, Machine-vision.

I. INTRODUCTION

Earth Observation (EO) satellites acquire data of the planet's surface. For these data to be operationally used for remote sensing, they must be associated with a unique geolocation. Typically, geometric calculations using orbital positions and orientations, together with the instrument boresight angles and camera parameters, are used to find the geolocation of each pixel in an image. This class of georeferencing algorithms is called direct georeferencing. This type of algorithm requires accurate knowledge about satellite orbital position and body

This work was supported by the Research Council of Norway, in part through the project "Observational Pyramid with Hyperspectral Nano-Satellites for Ocean Science (HYPSCI)" under Grant 325961, in part through the project "A system-of-systems approach to real-time integrated ocean environmental monitoring (SoS)" under Grant 337495, and in part through the project "Real-Time Remote Hyperspectral Imaging on Aerial Vehicles and Small Satellites (ARIEL)" under Grant 333229.

Corrado Chiatante, Roger Birkeland, and Milica Orlandić are with the Department of Electronic Systems at the Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. (e-mail: corrado.chiatante@ntnu.no)

Dennis Langer is with the Department of Marine Technology, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway.

Joseph L. Garrett and Tor Arne Johansen are with the Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), 7034 Trondheim, Norway.

Håvard S. Løvås is with the Norwegian Research Centre (NORCE), 9294 Tromsø, Norway.

attitude during the whole imaging process to reconstruct the geolocation of the acquired image. The advent of satellite miniaturization and other innovations has revolutionized access to space by allowing a larger number of organizations, such as universities and small companies, to design and operate their own satellites. Among those, several small satellite missions have been developed for water monitoring in coastal areas, such as the 3U SeaHawk-HawkEye Cubesat mission [1], or the 6U HYPISO-1 and HYPISO-2 satellites [2]. However, the use of Commercial Off-The-Shelf (COTS) components and pre-made satellite buses often limits the ability of engineers and researchers to control and correct uncertainty in the Attitude Determination and Control System (ADCS) subsystems, leading to inaccurate attitude knowledge of the satellite. Even the most accurate sensors, star trackers, can provide noisy information due to lower sensitivity of their low-cost imaging sensors for small satellites [3]. As a result, geolocation errors are present and must be addressed when using direct georeferencing methods, requiring workarounds to achieve the required geolocation accuracy [4], [5].

State-of-the-art small-satellite platforms can generate large amounts of data, but also typically have limited power and communication bandwidth [6]. Artificial intelligence can be employed onboard satellites to extract meaningful information from raw data, while producing compact outputs such as object, target or anomaly detection, and segmentation algorithms. Techniques such as object detection highly benefit from highly precise geolocation [7]. Investigating and developing a lightweight and robust georeferencing method for global applications is therefore crucial for future satellite onboard implementation of these algorithms.

Automatic georeferencing has evolved significantly in the past years with the introduction of more powerful computer vision techniques. Coastal features have been often employed for georeferencing due to their global availability and the easily recognizable features [5], [8]–[11]. Eugenio et al. [8] proposed an automatic georeferencing approach employing contour matching applied to water bordering regions. However, the method can be less robust than point matching because clouds or noisy data can disrupt the continuity of shoreline contours. Cheng et al. [9] expanded the georeferencing analysis of coastal areas by investigating the use of Scale-Invariant Feature Transform (SIFT) and Maximally Stable Extremal Region (MSER) algorithms to match affine invariant features in the shorelines, while filtering out outliers with Random sample consensus (RANSAC). The method is applied to image pairs

and hence is limited to relative georeferencing. Goncalves et al. [10] introduced an image segmentation approach to reduce image complexity and improve the matching performance. They applied a SIFT-based matching algorithm with outlier removal to perform 2D automatic image registration. Shorelines have been recently employed by Simarro et al. [11] for the co-registration of images captured by Landsat satellites and Sentinel-2 at different time instants. The method performs relative-alignment of image pairs after clustering within their database. As a result, the method presents less flexibility as it focuses on spatial alignments of identified image pairs. Similarly, the AROSICS framework [12] has emerged as a prominent open-source Python library achieving sub-pixel co-registration, and it has therefore been employed by several missions such as PRISMA [13] for small-scale adjustments. However, it requires the availability of full-resolution reference imagery (e.g., Landsat-8/9 or Sentinel-2), leading to significant storage overheads.

Different pipelines have been proposed for automatic orthorectification, including sensor model correction [14], its combined use with a Digital Elevation Model (DEM) [15], and general affine transformations [16]. Kartal et al. [14] investigated the use of SIFT and orthorectification for very high resolution imagery with meter-scale accuracy. The study investigated robustness with variable incidence angle, land cover and surface topography. Hakim et al. [16] developed a methodology for automatic georeferencing of multispectral images for the LAPAN-A3/IPB satellite, employing an affine transformation. The method introduced by Kartal et al. [14] relies on a database of manually georeferenced images to accurately georeference input captures. Similarly, Hakim et al. [16] require input reference images to validate the results. Consequently, both methods are limited by the need for access to an external database composed of previously georeferenced data. These major drawbacks hinder the implementation of a fully-automated georeferencing pipeline, as well as possible onboard implementation. Müller et al. [15] introduced an automated georeferencing approach for optical satellite imagery through the use of reference data and a DEM, while integrating sensor model refinement. Their work achieved relative and absolute geometric accuracies of approximately half a pixel [15]. The study focused on high-resolution satellites, validated with SPOT-4 HRVIR, SPOT-5 HRG, IRS-P6 LISS III, and ALOS AVNIR-2. While effective for on-ground processing, the pipeline relies on high resolution external database for reference image matching and DEM reference model. These requirements are challenging to meet onboard small satellites, where the limited memory capacity makes it impractical to store or scale a global reference database. Other methods have adopted geospatial vector data for georeferencing, employing external databases of shorelines [5] or roads [17], which can limit the size of the database and provide a reference of the global scale. Marsetič et al. [17] proposed a fully automatic georeferencing pipeline using vector roads as geographic reference information. The pipeline involves Ground Control Points (GCPs) extraction, geometric modeling, and orthorectification, achieving a resolution of approximately 1 pixel. The method is based on a reliable and compact database, as roads are

generally persistent features. However, the reported processing time is around 1 hour per experiment, and it may not work in sparsely inhabited areas. In addition, roads can only be effectively detected by high-resolution satellites.

Finally, there is great interest in machine vision methods involving satellite attitude estimation and correction to automatically refine the geolocation of each pixel, exploiting the generation of GCPs for optimizing the satellite state [4], [5], [18]–[20]. In particular, Carozza et al. [18] investigated co-registration of image sequences with extraction and matching of image features along the satellite track, with a focus on small satellites with limited attitude sensor accuracy. Previous works by Sugimoto et al. [19] and Kouyama et al. [4] employ a common methodology for satellite attitude determination through machine vision, using, respectively, a pushbroom and a frame sensor. Both works achieve an increase in geolocation accuracy and orbital estimation of the satellite attitude rotation matrix for roll, pitch, and yaw. Similarly, Zhou et al. [20] addressed geometric errors in wide field-of-view sensors by introducing a joint orientation and polynomial exterior orientation model to compensate external attitude-induced errors—still based on roll, pitch, and yaw—but also accounting for internal geometric distortions. However, [4], [19] rely on base-map images from the Landsat-8 OLI instrument as reference data, thus requiring access to previously georeferenced imagery, while [20] focuses on a large satellite with wide-field of view, with an original error of approximately 3-pixels, unlike smaller satellite where errors of a few to tens of kilometers—which can be equivalent to tens to hundreds of pixels—necessitate much larger attitude corrections. Liu et al. [5] have employed a Non Rigid-Coastline Inflection Point method (NR-CIM) method to correct small-scale geolocation error of a MicroWave Radiation Imager satellite with coarse resolutions on the order of tens of kilometers. However, Coastline Inflection Methods are limited by environmental factors; they perform optimally only with regular coastal geometries and under cloud-free conditions, making them less suitable for diverse global environments and significant cloud cover, commonly encountered by optical imaging satellites [21].

In summary, although several methods exist for image georeferencing, most rely on exhaustive external reference image databases or computationally demanding procedures. The literature review thus underscores the need for an automated georeferencing methodology that performs absolute registration with constrained computational resources, using minimal, stable ground-truth data, without dependence on external reference images, and tailored for onboard execution.

A first step towards onboard georeferencing has been recently introduced by Wang et al. [7], who proposed a novel onboard georeferencing method by combining feature-compression with template matching to reduce the storage needed onboard. The method was tested on the 4-meters resolution Qilu-3 microsatellite. However, the reference dataset was stored onboard for a limited set of nine predefined test areas, as the work was a preliminary onboard experiment [7]. A global implementation would require a substantially larger database, whose size would increase proportionally with the number of target regions. Such scaling introduces considerable

inflexibility and limits feasible system configurations on small satellites.

Our previous work [22] presented a lightweight prototype georeferencing algorithm that leverages segmented hyperspectral data for future onboard implementation, using global shoreline geospatial data as ground-truth. This previous study employs an Iterative Closest Point (ICP) approach to align shoreline point cloud with the generated ground-truth. However, the method does not account for distortions, and its performance is inherently less robust than local feature matching, as its objective is to minimize the alignment error with the global point cloud, rather than focusing on local—and potentially sparse—feature correspondences. In this paper, a novel pipeline is introduced, integrating established machine-vision techniques with an optimization-based state-estimation framework, to explicitly account for geometric distortions and enhance the robustness of the georeferencing process.

As in [4], [5], [19], [20], georeferencing is performed using satellite attitude correction rather than a generic coordinate transformation. The present paper thus seeks to extend, generalize, and increase the accuracy of the approach previously introduced in [22].

Contributions

Existing registration algorithms are based on image databases of previously georeferenced images, human intervention, or computationally intensive algorithms. However, these approaches might not be well-suited for small satellites equipped with pushbroom sensors, which acquire images at varying off-nadir angles and are affected by non-constant errors in ADCS, which can hinder the geolocation accuracy of the data products. This paper addresses the need for an automated, scalable, and computationally efficient georeferencing methodology suitable for global-imagery operations on memory-constrained small-satellite platforms, serving as a foundational step towards future onboard implementation. The proposed approach integrates machine-vision techniques with an optimization-based state-estimation framework, enabling accurate and robust geolocation under the constraints of limited onboard resources. By utilizing a single-channel land-sea-cloud segmentation mask in conjunction with a compact, globally applicable vector ground-truth dataset, the method enables accurate and computationally efficient georeferencing without the overhead of high-resolution raster databases.

This approach leverages shorelines and inland water bodies as the reference features due to their global ubiquity and high compatibility with standard land-sea-cloud segmentation algorithms. While the current approach is validated using sea and lake regions, the framework is inherently modular and extensible. It can be adapted to identify stark contrast between additional feature classes, such as desert boundaries, forests, or urban grids—provided that comparable ground-truth vector maps are available for those environments.

II. METHODOLOGY

A pushbroom instrument acquires one image line at a time along the satellite's ground track, with the sampling rate

determined by the imaging sensor's frame rate and integration time. Each line is composed of multiple samples spanning the across-track swath, as shown in Figure 1. Hyperspectral and multispectral imagers further enhance the data by capturing a few to hundreds of spectral bands per pixel, effectively generating an image for each line that contains both spatial and spectral information. As a result, a three-dimensional data cube is formed, where the across-track spatial samples, the temporally collected along-track lines, and the spectral bands constitute the cube's three axes.

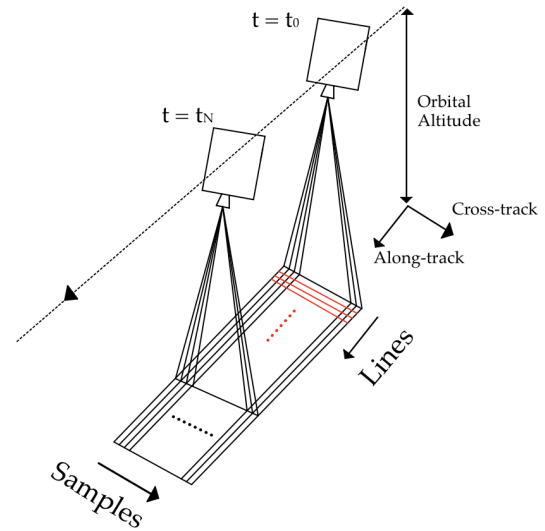


Fig. 1. Pushbroom imaging.

Each image pixel is typically georeferenced using a direct georeferencing method that combines telemetry-derived attitude data, satellite position information, and camera parameters. Small satellites are prone to pointing-knowledge errors, which can significantly degrade the accuracy of direct georeferencing. For example, the UNIFORM-1 microsatellite mission reports an error of up to 50-100 km, highlighting the common challenges associated with achieving precise geolocalization of small satellite imagery [4].

Figure 2 presents the Memory-Efficient Georeferencing of Segmented Satellite Imagery (MEGSSI) operational pipeline proposed in this work. The methodology is developed for the pushbroom imaging scenario, and is subsequently applied to the HYPSONO mission as a representative case study [2]. The proposed methodology is divided into two sequential stages:

- **Stage 1: Feature Detection and Matching:** The land-sea-cloud segmented image, initially georeferenced using direct methods, is compared with a ground-truth of the corresponding area, allowing machine vision algorithms to identify matching features around the land-sea boundary and identify a set of GCPs, which relate an image pixel to the geolocation.
- **Stage 2: Optimization and Geolocation correction:** An optimization methodology is employed to estimate and correct the satellite attitude state by minimizing the GCPs reprojection error, defined as the projected difference between the actual points identified in the image and the predicted geolocation obtained through the direct

georeferencing approach. The corrected attitude of the satellite is then used to re-estimate the geolocation of the image pixels using the direct georeferencing algorithm.

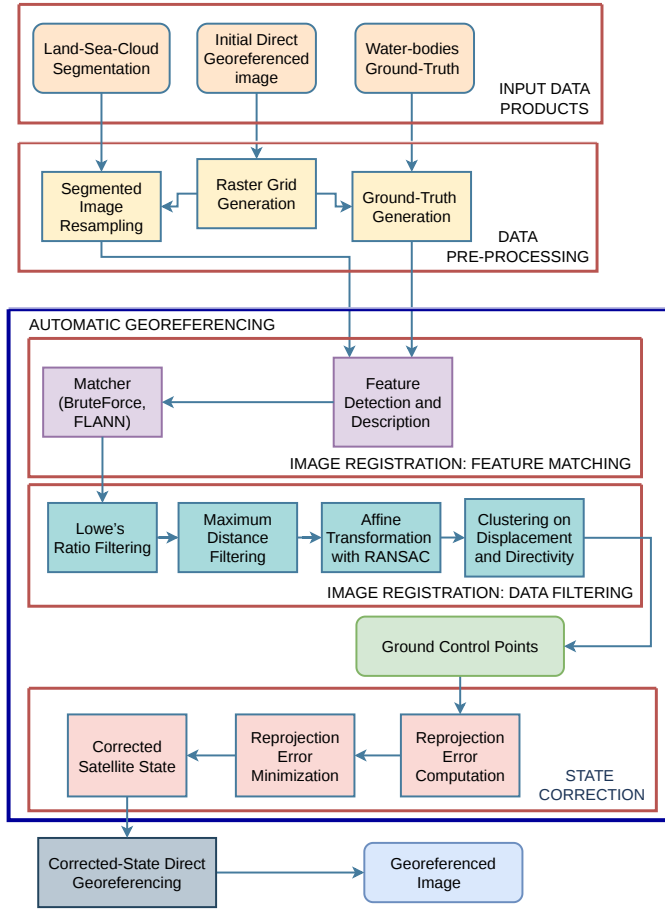


Fig. 2. MEGSSI methodology for Automatic Georeferencing.

A. Input Data Products

The MEGSSI georeferencing pipeline relies on a defined set of input data products. The methodology applied in this work is outlined in the following sections.

1) *Land-Sea-Cloud Segmentation*: Hyperspectral and multispectral imaging enable the identification of features within captured scenes by analysing the information contained in their associated data cubes. Many satellite missions employ land-sea-cloud segmentation data products for applications, such as land masking, cloud screening, and scene classification [23]. By comparing the land-sea-cloud segmented image with a ground-truth of the same area, machine vision algorithms can detect matching features and estimate a transformation that refines the georeferencing accuracy through the detection of GCPs.

2) *Initial Direct Georeferenced Image*: A direct georeferencing method is used to provide an initial estimate of the image location without relying on GCPs. The method uses measurements from the satellite's ancillary sensors, orbital position, ADCS data, and the corresponding time stamps. This dataset is processed to project the viewing geometry onto an Earth model and thus compute geographic coordinates for

each image pixel. The ground location of a pixel, $\mathbf{p}^{ECEF}$, is defined as the intersection point of an Earth ellipsoid model and a ray drawn from the satellite camera position, $\mathbf{x}^{ECEF}$, in the pixel view direction, $\mathbf{v}^{ECEF}$, such that $\mathbf{p}^{ECEF} = \mathbf{x}^{ECEF} + t'\mathbf{v}^{ECEF}$, where t' is the distance to the intersection point. Due to uncertainties in COTS telemetry and timing, errors propagate into these coordinates; if these errors are irregular, georeferencing accuracy becomes unpredictable, and needs to be corrected for each capture.

3) *Water-bodies Ground-Truth*: The availability of a segmentation model enables the development of machine vision methodologies that refine its estimation by comparing it against ground-truth reference map. Shoreline databases, such as Global Self-consistent, Hierarchical, High-resolution Shorelines (GSHHS) [24] and OpenStreetMap [25], can be used to generate the ground-truth dataset identifying sea, land, and inland freshwater features like lakes and rivers. The GSHHS database, introduced by Wessel et al. [24], provides a flexible representation of geographic features through hierarchical accuracy layers, while allowing for the precise identification of shoreline boundaries for both seas and inland water bodies. The data is provided in geospatial vector format, which allows for efficient spatial computation. The locations of water bodies are defined in latitude and longitude according to the World Geodetic System 1984 (WGS84) standard. Five resolution levels are available, each with approximately 80% data volume reduction relative to the next-higher level [24]. Table I shows the database sizes for the three highest-resolution databases - FULL, HIGH, and INTERMEDIATE - all of which achieve sub-kilometer spatial resolution. The RMSE of the GSHHS model was measured at approximately 100 meters on average, when compared against high-resolution (2 m) MAXAR imagery [26]. In particular, when tested across multiple locations with variable tidal conditions and time, the RMSE varied between 53–171 m, with median errors of 24–124 m [26]. This spread reflects the variability across diverse coastal sites compared in that study, which can be partly attributed to tidal state and seasonal water-level change.

Pelich et al. [27] found that OpenStreetMap provides shoreline data with a resolution of the order of 10 meters, comparable to the spatial resolution of Sentinel-1. This makes OpenStreetMap suitable for extending the proposed approach to higher-resolution satellite imagery, at the cost of increased data size and processing requirements. The size reported for OpenStreetMap in Table I corresponds to a dataset created by filtering out lakes smaller than 1 km² to reduce computational complexity.

While vector-based maps are memory-efficient, an alternative approach would be to use global raster products. However, according to the ESA WorldCover project, annual global cloudless composites containing four bands (RGB and NIR) at their native 10 m spatial resolution require approximately 8.5 TB of storage [28]. Although downsampling this dataset to 100 m resolution—matching the RMSE of the GSHHS database—reduces the memory requirement to approximately 85 GB, this is still significantly larger than the memory footprint of a vector-based shoreline reference map.

In addition, using raster reference images limits flexibility

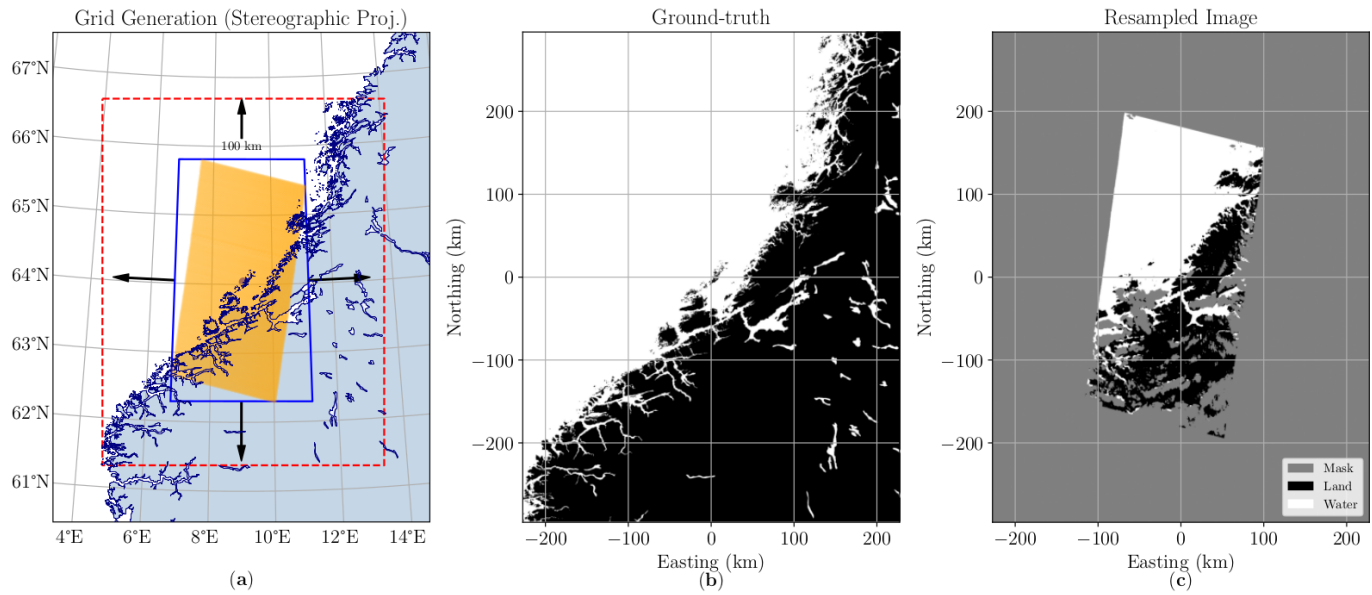


Fig. 3. Raster grid and ground-truth generation for the HYPSON-1 Trondheim capture on September 11th, 2024, at 09:27:22 UTC, with a resolution of 300 meters and omni-directional margin extension of 100 km.

TABLE I
GLOBAL DATA SIZE SPECIFICATIONS FOR GEOSPATIAL VECTOR
BOUNDARIES AND RASTER LAND MOSAIC

Resolution	Sea (L1)	Lakes (L2)
GSHHS Full (f)	177.66 MB	15.61 MB
GSHHS High (h)	46.65 MB	5.27 MB
GSHHS Intermediate (i)	10.13 MB	2.61 MB
OpenStreetMap Lakes and Sea data: ~ 3956 MB		
RGBNIR Land Mosaic resampled at 100 m resolution: ~ 85 GB		

for data manipulation. In contrast, geospatial vector data can be rasterized at arbitrary spatial resolutions, enabling the reference data to be consistently matched to the resolution and sampling grid of the satellite imagery, facilitating consistent resampling of both ground-truth and satellite captures.

B. Data Pre-Processing

1) *Raster Grid Generation*: Remote sensing data are commonly represented by discrete raster systems, which map individual observations onto shared geolocation grids [29]. This representation enables data comparison and fusion, while facilitating data processing. Within the proposed application, generating a common raster grid is a fundamental prerequisite for the MEGSSI georeferencing methodology, as it provides a shared spatial framework for the re-projection of both observed payload data, and reference ground-truth map. For each acquisition, the grid is generated by enclosing the directly georeferenced sensor footprint and extending it with a user-defined margin, as shown in Figure 3(a). This margin ensures that, even in cases of high direct-georeferencing errors, the relevant spatial features remain within the search region, allowing them to be reliably located and used to identify common features for the subsequent alignment process. To ensure geometric fidelity over large geographic extents, a Local Stereographic Projection centered on the image mid-pixel is employed. This projection enables the generation of a grid that preserves

angles and distances even across several hundred kilometers. The raster grid is generated using the area definition method provided by PYRESAMPLE [30]. The parameter settings used to generate the grid within the framework are reported in Table II. The user can select the resolution of the grid and specify a margin extension in kilometers in all directions. The resulting output is a grid with equally spaced pixels.

TABLE II
PARAMETERS FOR THE RASTER GRID GENERATION

Parameter	Information
Projection	Local Stereographic Projection
Resolution	User input [m]
Center	Midpoint of the image
Extension	User input [km]

2) *Ground-Truth Generation and Segmentation Resampling*: The proposed georeferencing method relies on the representation of the two independent information layers on a common raster grid: a reference ground-truth map and a resampled, segmented satellite image.

The ground-truth reference layer is generated by mapping vector-based water-body polygons, comprising both lake and sea shoreline data, onto the common raster grid. After being transformed from WGS84 to the stereographic reference frame, the polygons are overlaid on the raster grid data by generating a binary mask using the geometry mask function from RASTERIO [31]. By iterating over all intersecting polygons, a binary 2-dimensional reference map is produced, classifying each pixel as land or water. The resulting ground-truth image is shown in Figure 3(b).

A satellite-observation layer is simultaneously created by resampling the directly georeferenced segmented image onto the same raster grid. This step maps the land-sea-cloud segmented results onto the grid. In this work, nearest-neighbour interpolation is selected due to its simplicity and computational efficiency.

Figure 3(c) shows the resulting resampled data for the capture in the mid-Norway HYPSON-1 on September 11th, 2024, at 09:27:22 UTC. Land pixels are shown in white, while seas and lakes are in black. Areas not covered by the image, including cloud-obscured regions, are used to form an exclusion mask. Specifically, this mask combines the cloud class from the land-sea-cloud segmentation with a footprint mask marking grid points outside the satellite swath. Pixels belonging to either of these regions are excluded from further processing and passed to the feature detector as a binary mask, so that no keypoints are generated in these areas. As a consequence of using a stereographic reference frame, the axes in Figure 3(b) and Figure 3(c) are given as distances in kilometers relative to the center of the raster grid. Each pixel in the generated raster grid is associated with a pixel index, stereographic distance, and geolocation in latitude and longitude through the reference coordinate transformation.

Automatic Georeferencing

The automatic georeferencing method proposed in this paper consists of two main stages: image registration and state correction. First, image registration is used to identify corresponding features between satellite observations and reference data, enabling the generation of GCPs that relate image pixels to geographic locations. These GCPs are subsequently used in a state-correction step to refine the estimated satellite orientation.

C. Image Registration - Feature Matching

Image registration is a well-established class of algorithms in remote sensing and is commonly used to align different images into a common coordinate system. The present research focuses on a class of algorithms that have been employed for more than a decade in the field. Given the objective of developing a method that is scalable and potentially open-source, the investigation is restricted to algorithms available in the OPENCV 3.0 library [32] that are not protected by patents. This ensures that the proposed methodology can be reproduced, extended, and deployed without licensing restrictions. Unlike recent deep-learning-based feature matching approaches, these classical methods benefit from mature, readily available implementations. Furthermore, several of the underlying algorithms have already been demonstrated in hardware-accelerated and Field Programmable Gate Array (FPGA)-based implementations, making them attractive candidates for future onboard deployment on resource-constrained satellite platforms.

Image registration typically consists of three main stages: feature detection and description, feature matching, and outlier rejection through filtering [33]. The first step includes the detection of features, such as corners or edges, and the description of each feature point with respect to the characteristics of the local image, making the feature highly distinctive [34]. In the second step, feature-matching is performed by identifying common features between two images. Various matching algorithms can be employed, depending on the chosen feature-description method or the distance metric used for the point correlation [33].

Three algorithms that have been studied and compared in previous work [33], SIFT [34], Oriented FAST and Rotated BRIEF (ORB) [35], and Binary Robust Invariant Scalable Keypoints (BRISK) [36], are implemented and evaluated. SIFT is included due to its long legacy in the field, whereas ORB and BRISK were developed as alternatives that aim to balance accuracy with high computational efficiency. SIFT employs floating-point descriptors for the identification of features, whereas ORB and BRISK use binary-string descriptors [33]. As a result, SIFT descriptors are typically matched using Euclidean distance, and binary descriptors using Hamming distance. The configuration parameters used for the corresponding OpenCV methods are summarized in Table III.

TABLE III
COMPARISON OF FEATURE MATCHING METHODS

Detector/Descriptor	Matcher	Distance Metric
SIFT	FLANN-based Matcher	Euclidean distance (L2 norm)
ORB	BFMatcher	Hamming norm
BRISK	BFMatcher	Hamming norm

Figure 3(b) and Figure 3(c) illustrate the 2D images that will be used for matching the features with SIFT, ORB and BRISK.

The selected machine vision algorithms yield up to thousands of candidate matches with variable accuracy. Although the strongest and most distinctive features are retained, automatically selected and isolated, the resulting match set often contains several false positives, which can degrade the performance of the automatic georeferencing pipeline. To ensure reliability, systematic outlier identification and removal from the dataset are included in the MEGSSI pipeline.

D. Image Registration - Data Filtering

The proposed pipeline implements the following data-filtering steps in sequence:

- 1) Lowe's ratio filtering with uniqueness filtering.
- 2) Maximum geodesic-distance filtering.
- 3) Affine transformation using RANSAC.
- 4) Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering on the displacement and directivity.

1) Lowe's Ratio Filtering: Researchers have proposed several algorithms to retain only the desired matches. One of the simplest approaches is Lowe's method [34], which suppresses outliers based on distance metrics when applied to the identified candidate matches. In this work, the Lowe's constant selected for the filtering step is 0.7, based on empirical considerations originally introduced in [34]. To avoid multiple keypoints matching to the same feature, a uniqueness filter removes all matches that reuse a keypoint. Even after this step, the data remain noisy, and therefore additional filtering is required.

2) Maximum Geodesic Distance Filtering: Outliers may occur when strong but misleading local features produce incorrect matches. As a precautionary measure, all matched pairs that imply a correction exceeding a **satellite-specific tolerance** are discarded. Distance between geolocations is computed using the Python package GEOPY [37]. The algorithm assesses

the distance between two geolocations with the geodesic, which is defined as the shortest path between two points on the Earth's surface [37], [38]. For other satellites, this tolerance needs to be tuned based on the satellite's measured georeferencing accuracy and the image size.

3) *Affine Transformation with RANSAC*: A RANSAC-based method is used to reject matched outliers by estimating an affine transformation between two images. This step is implemented using OPENCV function ESTIMATEAFFINE2D [32] with RANSAC. The affine model is selected because it allows for modeling local geometric distortions while remaining computationally efficient, and prior work has demonstrated its validity for describing pushbroom satellite imaging geometry [39]. Furthermore, higher-order polynomial transformations do not necessarily yield improvements in registration accuracy [40]. The affine transformation is used solely for outlier rejection by identifying the dominant set of matches that follow the primary displacement pattern between the images, and not as the final georeferencing model. This is consistent with previous studies that used RANSAC for outlier rejection, before deriving attitude correction based on the remaining GCP inliers [4], [19]. Similarly, MEGSSI passes the filtered GCP set through the attitude correction optimization in Equation (1).

4) *Clustering on the Displacement and Directivity*: To support full automation of the complete processing pipeline, a high level of robustness to outliers is required, particularly for the unsupervised process. Therefore, an additional filtering step in the filtering pipeline is introduced using DBSCAN [41], applied to the displacement and directivity of each matched pair. The DBSCAN algorithm is used to group large spatial datasets containing noise and can be applied to remove outliers that may remain after the previous filtering steps. The algorithm computes the vertices of the identified original and corrected geolocation feature points, enabling calculation of the displacement vectors associated with each pair of matched points. The angular direction of this displacement is used as a measure of directivity. The ground-truth feature coordinates are expressed by (ϕ_t, λ_t) for the longitude λ and latitude ϕ , respectively, while the corresponding coordinates from the directly georeferenced image are represented by (ϕ_o, λ_o) . Figure 4 illustrates the feature point identified in the original image shown in purple, and its corrected predicted geolocation derived from the ground-truth in yellow.

The directivity of the displacement vector in the WGS84 reference system can be defined by computing the displacement between the original and ground-truth coordinates as:

$$\Delta\phi = \phi_o - \phi_t, \quad \Delta\lambda = \lambda_o - \lambda_t.$$

The angle θ of the displacement vector, measured with respect to the positive eastward direction (i.e., increasing longitude), is calculated as:

$$\theta = \text{atan2}(\Delta\phi, \Delta\lambda).$$

As in the maximum distance filtering step, distances are calculated from a pair of geolocation coordinates that are matched together, computing the geodesic distance with the Python package GEOPY [37].

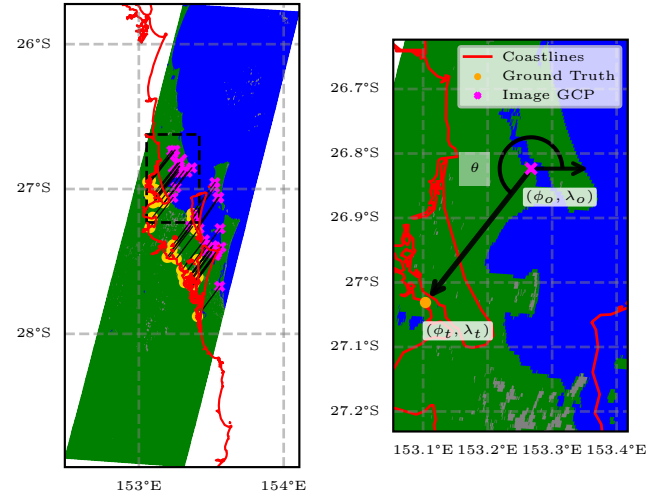


Fig. 4. Displacement and directivity filtering

DBSCAN is applied to the two 1D feature vectors using a minimum cluster size of three samples and the clustering radii specified in Table IV.

TABLE IV
CLUSTERING PARAMETERS

Variable	Clustering Radius	Min. Samples
Angles	User Input [deg]	3
Distances	User Input [km]	3

Only the matches identified as inliers in both clustering stages—angular and spatial—are retained. This dual-stage clustering approach effectively removes outliers exhibiting anomalous displacement direction or magnitude.

E. State Correction

The resulting output of the filtering stage is a refined set of GCPs in which likely outliers have been removed. Each GCP is represented by its pixel coordinate within the raster grid, defined relative to the grid's top-left origin. Latitude and longitude can be extracted in both cases through a simple interpolation using the REGULARGRIDINTERPOLATOR class from the SCIPY library [42].

A GCP is denoted as:

$$\text{GCP} = (\mathbf{p}_{\text{cl}}, (l, s))$$

where

$\mathbf{p}_{\text{cl}}$ is the 3D geocentric coordinate of the shoreline feature (equivalent to latitude, longitude),

(l, s) are the line and sample (row, column) coordinates in the camera frame.

Hence, a linear interpolation between the geolocation grid and the camera frame is required to determine the corresponding *frame* and *line* coordinates in the camera frame for each identified feature. This conversion uses a local linear interpolation within a small 10×10 pixel window around each feature, with a nearest-neighbor fallback for edge cases where sufficient neighboring points may not be available. In this fallback case, the introduced error is bounded by the local pixel spacing.

1) *Reprojection Error Optimization*: From the feature-matching and filtering stages, each GCP is extracted with:

- the feature's coordinates in the Earth-centered, Earth-fixed (ECEF) coordinate system, p_{cl} .
- the corresponding line l and sample s coordinates in the hyperspectral data cube.

A line coordinate, e.g. $l = 12.3$, is converted into a timestamp by linear interpolation of the timestamps of line 12 and line 13. Similarly, the satellite's position and orientation are then interpolated to this timestamp. The position is obtained through linear interpolation, while the orientation is interpolated using spherical linear interpolation (SLERP). This allows expressing the pose of the satellite and pushbroom camera at the Time-of-Passing (ToP), i.e. when the satellite swath passes over the shoreline feature. Using this pose at ToP, the feature's geographical coordinates are reprojected into the camera frame, producing the 2D image-frame point $\mathbf{x}_r = [x_r, y_r]^\top$. In the 2D image plane ($z = 1$) at ToP, the image observation is solely defined by the sample coordinate, s , focal length, f , and principal point $w/2$. The normalized image coordinate is then expressed as $\mathbf{x}_{im} = [x_{im}, 0]^\top = [(s - w/2)/f, 0]^\top$. In Figure 5, the image plane is illustrated with the scanline indicated as the green dotted line and $\mathbf{x}_{im}$ indicated as the green point on the image plane. The red point represents the 2D reprojection of the shoreline feature, $\mathbf{x}_r$. Consequently, the reprojection error for a shoreline feature is then the vector $[x_{im} - x_r, -y_r]^\top$ shown as the blue vector in the figure.

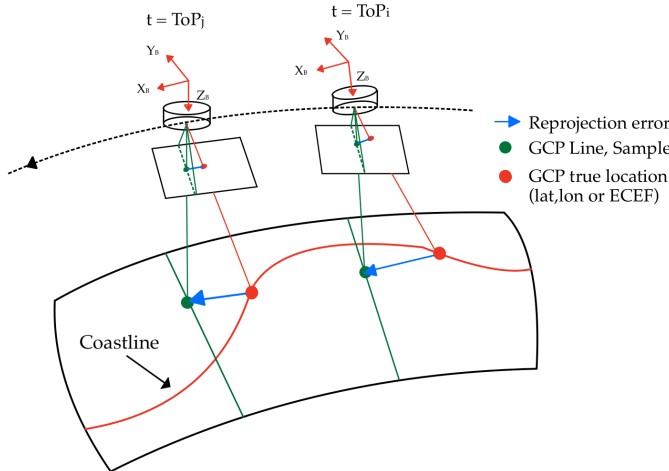


Fig. 5. Representation of attitude error estimation for two GCPs.

In the MEGSSI method, the reprojection errors (and registration errors in general) are attributed to relatively small errors in the ADCS orientation between successive recordings. This assumption is motivated by the fact that the orbital position of small satellites is typically known with substantially higher accuracy than the pointing direction. Moreover, due to the large distance between the satellite and the observed scene, even small attitude errors can result in significant ground displacements. Consequently, only attitude is corrected within the proposed framework. However, these errors in the ADCS orientation are assumed to be sufficiently stationary within each data cube to be corrected by a constant rotation. The

orientation error is expressed by a local rotation matrix, $\Delta\mathbf{R}$, such that the corrected total orientation in ECEF is $\mathbf{R}_{corr} = \Delta\mathbf{R} \cdot \mathbf{R}_{est}$, where $\mathbf{R}_{est}$ denotes the ADCS estimate of the orientation of the satellite imager in ECEF. The latter, therefore, expresses both attitude and boresight calibration orientation. Since rotation matrices belong to the special orthogonal group, they require exactly 3 parameters to be fully defined. Here, Euler angles $\Delta\boldsymbol{\theta}$ are used to represent $\Delta\mathbf{R}$. Assuming Gaussian distributed reprojection errors, the maximum-likelihood estimator for $\Delta\mathbf{R}$ - and thus $\mathbf{R}_{corr}$ - is found by minimizing the sum of squared reprojection errors over all features, as follows:

$$\min_{\Delta\boldsymbol{\theta}} \sum_{i=1}^N \left\| \mathbf{x}_{im,i} - \pi_z \left(\mathbf{R}_{corr} \cdot [\mathbf{p}_{cam,i} - \mathbf{p}_{cl,i}] \right) \right\|^2 \quad (1)$$

where $\mathbf{p}_{cam}$ is the ECEF position of the hyperspectral camera based on the satellite telemetry, i and N are the index and total number of features, respectively. The function π_z is a projection function, normalizing the vector by the z coordinate, so that $\pi_z([v_x, v_y, v_z]) = [v_x/v_z, v_y/v_z]$. The reprojection error can be expressed in pixels by multiplying the resulting normalized coordinates by the focal length f , since the camera model used is a pinhole model with centred principal points.

Note that $\mathbf{R}_{est}$ and $\mathbf{p}_{cam}$ are typically calculated from Earth-centered Inertial (ECI) positions and orientations, provided by the ADCS, while the optimization uses ECEF. The rotation matrix from ECI to ECEF, $\mathbf{R}_{ECI,i}^{ECEF}$, is derived at each feature's ToP, t_i .

The reprojection error for two shoreline features is illustrated in Figure 5. The optimization is implemented using SCIPY's least squares optimization [42] with the objective function defined in Equation (1), initialized at $\Delta\boldsymbol{\theta} = [0, 0, 0]^T$ to employ the raw ADCS telemetry as an initial guess for the orientation. The cost function ensures that the final estimation of the satellite state is driven by the geometric alignment of the identified GCPs. The optimization then returns the parameter $\Delta\boldsymbol{\theta}$ (or $\Delta\mathbf{R}$) which minimizes the sum of squared reprojection errors over all shoreline features, using the algorithm default implementation tolerance (10^{-8}). The numerical stability is supported by the low dimensionality of the state space (3 rotational parameters). However, overfitting might theoretically occur in scenarios with limited spatial distribution, such as a capture containing a single small island with very few features. Using all the identified GCPs across the entire image greatly limits this problem, as a wider geometric distribution provides better constraints for the overall rotation.

F. Corrected-State Direct Georeferencing

Using the updated orientation matrix $\mathbf{R}_{corr}$, the known satellite orbital position $\mathbf{x}^{ECEF}$, and the camera model, each pixel in the image can be geolocated through the direct georeferencing algorithm.

III. HYPSON MISSION USE CASE

The proposed MEGSSI pipeline has been evaluated using data from the HYPSON mission. The HYPSON-1 satellite is a

6U CubeSat that has been acquiring hyperspectral imagery since its launch in January 2022. It is equipped with a pushbroom hyperspectral imaging payload [2] and is operated by the SmallSat Lab at Norwegian University of Science and Technology (NTNU), which is also responsible for data processing and management. The Hyperspectral imager observes the Earth's surface in the visible region of the radiance spectra, with 120 spectral bands spanning approximately 400 to 800 nm [2]. In August 2024, HYPISO-2, another CubeSat equipped with a payload configuration closely matching that of HYPISO-1, was launched, extending the observational capabilities of the mission. Owing to the similarity between the two platforms, the georeferencing pipeline developed for HYPISO-1 was successfully extended and deployed to HYPISO-2. The HYPISO-1 satellite typically captures and downlinks up to six images per day. Upon successful optimization of HYPISO-2 software, the downlink capacity is expected to increase to approximately 50 images per day.

A. Input Data Products for HYPISO

1) *Land-Sea-Cloud Segmentation Map*: Justo et al. [23] developed a computationally efficient image segmentation method based on deep neural networks, called the 1D-Justo-LiuNet, which labels each pixel into land, sea, and cloud classes. The algorithm has been available onboard the HYPISO-1 satellite since May 2024 [43], and it is also implemented in the ground-processing pipeline. For this work, both execution modes are applicable, provided that the segmentation output is available to the ground processing pipeline for automatic georeferencing. However, the flexibility of the implementation enables both onboard and ground-based solutions to be considered. An overall segmentation accuracy of 0.93 across all classes is reported in [23]. An example of the algorithm output is illustrated in Figure 6, represented by the capture taken over mid-Norway on September 11th, 2024: Figure 6(a) shows the RGB representation, while Figure 6(b) displays the corresponding land-sea-cloud segmentation. Similarly, a new land-sea-cloud segmentation model is employed for HYPISO-2 to extend the usability to both satellites, obtaining comparable results [44].

2) *Direct Georeferencing for HYPISO*: Each pixel in the acquired imagery is georeferenced through a direct georeferencing method, based on the attitude telemetry and Two-Line-Element (TLE) position data. The pixel view direction $\mathbf{v}^{CAM}$ is computed in the camera frame using the pixel view angle φ through a camera model. In this work, the simple pinhole camera model is used. Bore-sight calibration rotates the view direction from the camera to the body frame. Using satellite attitude telemetry, the body-frame view vector is rotated to the ECI frame, and, finally, from ECI to ECEF. This chain of rotations may be expressed as:

$$\mathbf{v}^{ECEF} = \mathbf{R}_{ECI}^{ECEF} \mathbf{R}_{BODY}^{ECI} \mathbf{R}_{CAM}^{BODY} \mathbf{v}^{CAM}(\varphi). \quad (2)$$

The same pinhole camera model used to compute the observed normalized coordinate $\mathbf{x}_{im}$ in Equation (1) is also used to define $\mathbf{v}^{CAM}$ in Equation (2).

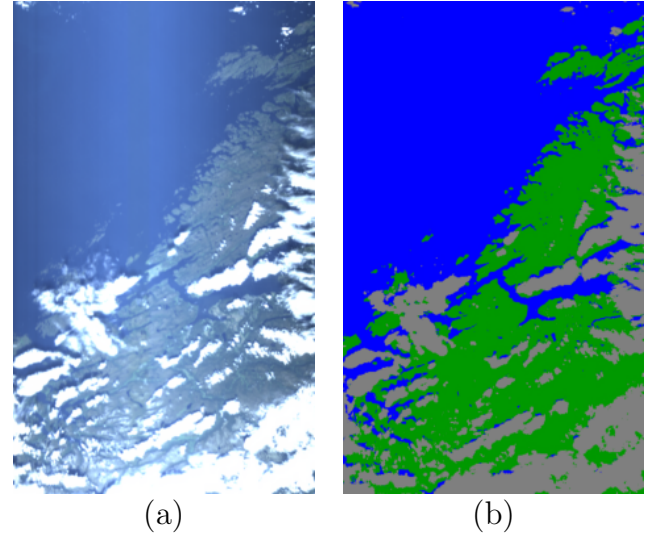


Fig. 6. Capture taken by HYPISO-1 over mid-Norway / Trondheim on September 11th, 2024, where (a) is the RGB composite, and (b) is the segmented output.

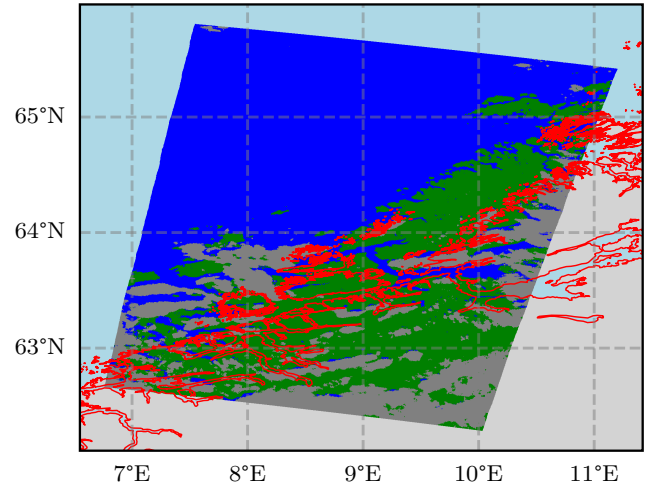


Fig. 7. Direct Georeferencing of the capture from Figure 6 using raw telemetry, and compared to shorelines ground-truth in red.

The corresponding satellite position is given as:

$$\mathbf{x}^{ECEF} = \mathbf{R}_{ECI}^{ECEF} \mathbf{x}^{ECI}. \quad (3)$$

HYPISO satellite telemetry provides satellite position in True Equator, Mean Equinox (TEME) frame, which is the ECI reference frame used by TLE orbit propagators, as well as the satellite attitude between the satellite body frame and the TEME frame. The ECEF frame selected for the simulation is the International Terrestrial Reference Frame (ITRF) coordinate system. Because the line-acquisition rate and the telemetry-logging rate differ on the HYPISO satellites, the satellite telemetry is interpolated to the line-acquisition timestamps. The Python package ASTROPY [45] is used for the ECI to ECEF transformation. Figure 7 shows an example of a directly georeferenced capture with a georeferencing error. In the sea-land-cloud raster image, the shoreline is to the west

of the ground-truth shoreline marked in red, indicating that the georeferencing error is directed westward.

3) *Ground-Truth Model*: Figure 8 depicts a zoomed-in portion of a HYPSON-1 capture over British Columbia, Canada. While the GSHHS FULL and HIGH models produce barely distinguishable results, the reduced resolution of the INTERMEDIATE level is clearly visible. This level is therefore discarded, as it does not align with the Ground Sample Distance (GSD) of the HYPSON satellites. Although the FULL model incurs a significant increase in data size compared to the HIGH model, the overhead is considered acceptable under the current configuration due to the improved accuracy it provides.

This accuracy has been reported in the range of 53–171 m RMSE, with median errors between 24 and 124 m depending on location, acquisition time, and tidal conditions [26]. While tidal state and seasonal water-level variations contribute to these errors, their impact is not expected to be a dominant source of uncertainty for HYPSON imagery, as shoreline-reference uncertainty is generally smaller than or comparable to the effective resolution relevant to this study. It may, however, become significant in some areas with large intertidal zones, especially for across-track evaluations, where HYPSON's native resolution is finer.

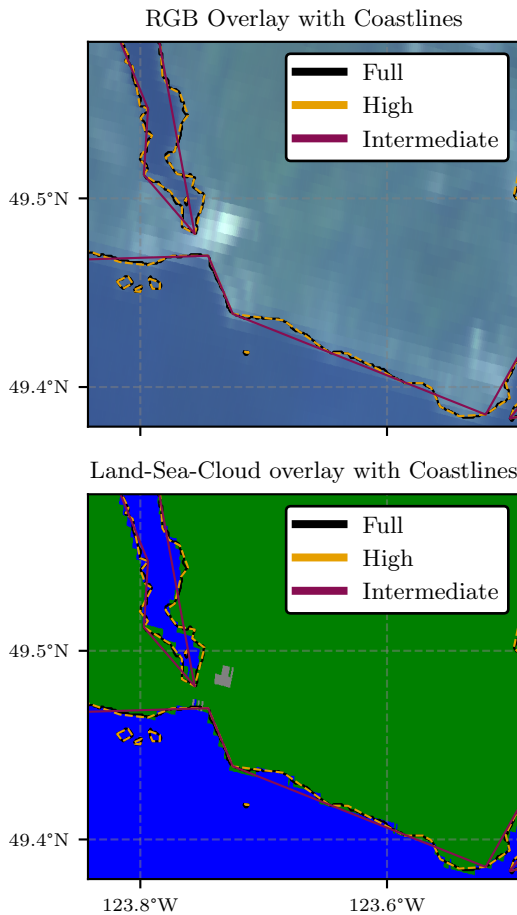


Fig. 8. Close-up of shorelines with different Ground-truth model resolution.

B. Data Pre-Processing: Raster Grid Generation

1) *Raster Grid Generation*: A user may define the spatial extent of the area according to the imaging properties of the sensor. In this case study, an omnidirectional extension of 100 km in both latitude and longitude was used to accommodate the HYPSON image footprint, which spans approximately 70 km across-track, and between 350–500 km along-track. In addition, 100 km is set as the maximum initial georeferencing error that can be corrected by the pipeline, consistent with the maximum geodesic distance threshold applied during feature filtering (Section II-D2). The raster grid resolution was set to 300 m per pixel, matching the spatial resolution of HYPSON [2], while maintaining fine shoreline details for feature matching. This choice is supported by the analysis reported in the empirical comparison between different grid resolution, reported in Table V, where the captures reported in Table VIII were processed using different grid resolutions and assessed in terms of RMSE against manually generated GCPs and computational runtime. The results showed that a finer grid resolution of 200 m offered no advantage over 300 m, yielding both slightly higher RMSE and longer processing time. In contrast, coarser resolutions of 600 m and 1000 m resulted in substantial degradation in georeferencing accuracy, as reflected by higher RMSE values.

TABLE V
RMSE IN PIXEL AND MEAN RUNTIME PER GRID RESOLUTION.

Grid Resolution	RMSE (pix)	Mean Runtime (s)
200 m × 200 m	0.88	15.16
300 m × 300 m	0.81	12.22
600 m × 600 m	1.24	9.14
1000 m × 1000 m	2.69	8.04

The properties of the raster-grid generation are reported in Table VI.

TABLE VI
GENERATION OF RASTER GRID FOR HYPSON

Parameter	Information
Projection	Local Stereographic Projection
Resolution	300 m × 300 m
Center	Midpoint of the image
Extension	Original image area + 100 km buffer

2) *Ground-Truth Generation and Segmentation Resampling*: The generation of ground-truth data and the resampling on the same grid have been illustrated previously in Figure 3, where image (a) on the left shows the original image extent in orange, overlaid on the shoreline and lake ground-truth for a capture over mid-Norway. After applying the spatial extension, the raster grid boundaries are indicated by the red dashed polygon. Both the reference ground-truth and the segmented satellite imagery are then projected onto this common frame for georeferencing, as shown in Figure 3(b) and (c).

C. Image Registration - Feature Matching

For the HYPSON mission, GCPs are generated through automatic georeferencing process by identifying correspondences between the satellite imagery and the ground-truth model. Either SIFT, ORB, or BRISK can be selected by the user to match the distinctive coastal and lake features on the generated raster grid.

D. Image Registration - Data Filtering

As described in Section II-D2, a maximum-distance filtering may be applied to remove all GCPs that result in excessively large corrections. In this study, a maximum geodesic distance of **100 kilometers** is selected as a precautionary measure to avoid retaining potentially erroneous matches. This threshold value is consistent with the area extension introduced in Table VI. As shown in Table XIV, the observed initial georeferencing error across the tested captures ranges from 7.8 to 97 km, so the maximum geodesic distance is set to consider even the observations with the most extreme initial errors. DBSCAN clustering is then applied to the two 1D feature vectors, distance and directivity, using a minimum sample size of three and clustering radii specified in Table VII. To assess the robustness of these parameter selections, a sensitivity analysis was performed by varying the Angle parameter over $\{1, 3, 5, 10\}$ degrees and Distance parameter over $\{1, 5, 10, 20\}$ km across all 15 captures using ORB. The resulting RMSE values showed only minor variation, indicating that the default parameter choice (3° , 5 km) is not particularly sensitive. The value of the DBSCAN filtering stage was further evaluated by disabling it and re-running the experiments. Although the overall results changed only marginally, DBSCAN proved to be a useful low-cost validation step in two captures for which BRISK failed. Without DBSCAN, these captures produced georeferencing results with an RMSE of approximately 20 px, despite the matches passing the subsequent RANSAC affine-transformation step. This indicates that DBSCAN successfully identified inconsistencies that were not rejected by RANSAC, thereby improving the robustness of the registration pipeline.

The resulting filtered feature matches for a HYPSON-1 capture are shown in Figure 9, where the reference ground truth is shown on the left and the projected satellite data on the right, with matching features connected by a yellow line.

TABLE VII
CLUSTERING PARAMETERS FOR HYPSON

Variable	Clustering Radius	Min. Samples
Angles	3 deg	3
Distances	5 km	3

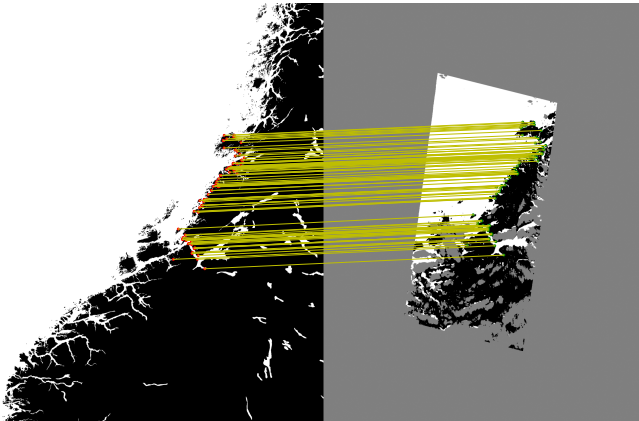


Fig. 9. Matching results after filtering for the Trondheim capture on September 11th, 2024 taken by HYPSON-1.

The filtering steps and grid parameters were empirically calibrated to HYPSON based on its georeferencing error characteristics and the ablation studies presented above. Applying MEGSSI to a different satellite platform would require repeating the calibration procedure, taking into account platform-specific characteristics such as attitude-errors and spatial resolution.

E. State Correction

The satellite attitude is corrected by optimizing the reprojection error for the identified GCPs. This correction can be applied to the HYPSON satellites, and, using Equation (2), the corrected camera-to-ECEF rotation is given by:

$$\mathbf{R}_{\text{CAM}}^{\text{ECEF}} = \mathbf{R}_{\text{corr}} = \Delta \mathbf{R} \cdot \mathbf{R}_{\text{est}}, \quad (4)$$

where

$$\mathbf{R}_{\text{est}} = \mathbf{R}_{\text{ECI}}^{\text{ECEF}} \cdot \mathbf{R}_{\text{BODY}}^{\text{ECI}} \cdot \mathbf{R}_{\text{CAM}}^{\text{BODY}} \quad (5)$$

F. Corrected-State Direct Georeferencing

Once the attitude correction is applied, the final direct georeferencing is carried out. This last step projects each pixel of the datacube into the updated geolocation using the corrected-state, providing the final georeferenced product for further analysis.

IV. RESULTS

To date, the HYPSON mission has collected more than 5000 Hyperspectral Imaging (HSI) captures during its four years of operations. From this database, a dataset of images is selected to evaluate the performance of the proposed georeferencing pipeline and to assess the accuracy of the methodology. The dataset is selected to test the pipeline for heterogeneous areas and under various conditions, including coastal regions from different parts of the world, as well as several lakes. The geographical distribution of the selected captures is shown in Figure 10. The dataset includes 10 captures taken by HYPSON-1, and 5 captures by HYPSON-2, listed with the relative acquisition time in Table VIII. The captures contain varying cloud cover to assess robustness, as well as instances where the direct georeferencing exhibits substantial errors - such as the Dardanelles capture, where the initial georeferencing error reaches approximately 97 km. Moreover, the Lake Ellesmere image is an agile capture performed through dynamic pointing, an imaging mode where the target is observed following a pre-defined ADCS sequence [46].

The three machine vision algorithms - SIFT, ORB, and BRISK - are compared by visualizing the number of matches at various stages, the total number of successful captures, and the number of generated GCPs for those successful captures.

Table IX summarizes the average number of matches obtained across all captures for the different feature matching approaches, together with the subsequent GCPs filtering steps. The Lowe's ratio test substantially decreases the average number of matches, demonstrating its effectiveness in removing outliers. The two subsequent steps, Max distance and RANSAC filtering, further eliminate the majority of remaining

TABLE VIII
SUMMARY OF TESTED HYPISO CAPTURES

HYPISO-1					
No.	Name	Lat	Lon	Time	Comment
1	Aquawatchmoreton	-27.29	153.26	2024-12-26 23:14	–
2	Blanca	-39.15	-61.72	2024-12-14 13:50	–
3	Hypernetestuary	45.54	-1.04	2025-04-10 10:39	–
4	Mariamadre	21.63	-106.59	2025-02-01 17:13	–
5	Mvco	41.33	-70.57	2025-02-24 15:13	–
6	Tampa	27.74	-82.58	2025-02-15 15:40	–
7	Vancouver	49.11	-123.12	2024-08-29 18:15	–
8	Trondheim	63.50	10.39	2024-09-11 09:27	Cloudy
9	Dardanelles	40.31	26.61	2025-04-13 08:51	Huge Error
10	Barbados	13.16	-59.55	2025-03-04 13:54	–
HYPISO-2					
No.	Name	Lat	Lon	Time	Comment
11	Durban	-29.92	31.06	2025-04-13 08:23	–
12	Tharthar	33.87	43.32	2025-04-13 08:05	Lake
13	Dubai	25.11	55.18	2025-04-20 07:09	–
14	Lakeellesmere	-43.78	172.45	2025-04-25 22:22	Agile Capture
15	Socheongcho	37.42	124.74	2025-03-20 02:37	–

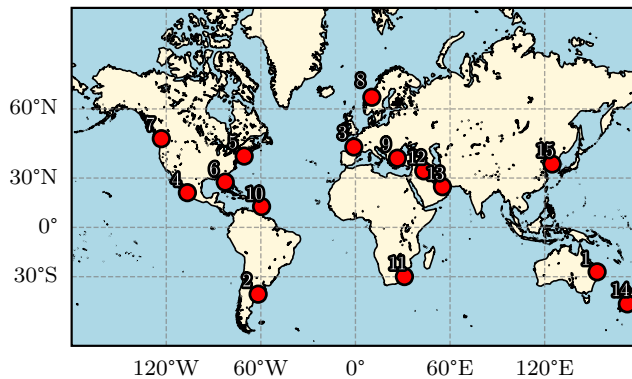


Fig. 10. Location of the Targets on Earth surface.

outliers. In most cases, particularly for ORB and BRISK, the DBSCAN distance and directivity filtering is not necessary. However, it also serves as a baseline to confirm the consistency of the results.

Beyond reporting the total number of matches, their spatial distribution is analyzed. To quantify their coverage, each image was divided into 18 cells (6 along-track $\times$ 3 across-track), chosen to obtain approximately square cells given HYPISO's image footprint, and then counting how many cells contain at least one GCP. As 4 captures failed for SIFT and 2 for BRISK due to lack of identified matches, the analysis of all performance parameters is restricted to the 10 captures that were successfully georeferenced by all three methods. The results show that ORB achieves the best spatial spread, occupying the largest fraction of image cells. In addition, it also produces the highest number of GCPs per occupied cell, consistent with it retaining the highest number of matches after all filtering steps.

A. Relative Accuracy with Ground-Truth

Georeferencing performance is evaluated using a pixel-wise accuracy metric, in which each pixel is described by its geographic coordinates — latitude (ϕ), longitude (λ), and its segmentation class label (c). Given the predicted coordinates (ϕ, λ), the proportion of land and sea pixels that correctly align with the ground-truth model at the corresponding locations can be assessed. The metric can be applied to the

TABLE IX
COMPARISON OF SIFT, BRISK, AND ORB ON SELECTED PERFORMANCE METRICS

Metric	ORB	SIFT	BRISK
Success Ratio	15/15	11/15	13/15
Matches	5159	1452	4831
Matches after Lowe's ratio	156	74	121
Matches after Max distance	118	39	76
Matches after RANSAC	97	29	47
Final Matches after DBSCAN	96	20	46
Mean Occupied Cells (%)	37.8	29.4	31.7
Mean GCPs per Occupied Cell	12.73	3.08	6.97

Success ratios are computed on all 15 captures, while all the other parameters are computed on the common 10 captures that were successful for SIFT, ORB, and BRISK for a fair and comparable analysis. The number of matches has been rounded to the nearest integer.

original, erroneous direct georeferencing to assess the accuracy improvement achieved by the proposed pipeline. This approach provides only a relative indication of accuracy, as class imbalance may lead to disproportionately high accuracy scores. Similarly, captures containing extensive cloud cover may show inflated accuracy due to fewer land and sea pixels being available. However, the measure remains suitable for comparing the performance of SIFT, ORB, and BRISK within the context of the proposed MEGSSI method.

Table X presents the results obtained using the land-sea-cloud accuracy metric. Each method achieved the highest accuracy for at least one capture; however, BRISK and SIFT fail in 2 and 4 cases, respectively; these failures occurred during the filtering stages due to an insufficient number of valid matches. In contrast, ORB is successful for all 15 captures. Overall, the ORB algorithm outperforms the other two approaches, particularly SIFT, demonstrating greater robustness to varying conditions, including cloudy scenes, lakes, and images with an initial georeferencing error of almost 100 km.

TABLE X
COMPARISON OF SIFT, ORB, AND BRISK ON RELATIVE ACCURACY PERFORMANCE METRICS (IN PERCENT)

Capture	Direct georef.	ORB	SIFT	BRISK
Aquawatchmoreton	84.84	99.06	99.10	99.06
Blanca	78.62	94.39	Fail	94.09
Hypernetestuary	92.44	97.84	97.97	97.13
Mariamadre	94.92	99.54	98.85	99.39
Mvco	91.49	99.03	98.75	98.99
Tampa	92.02	98.11	97.92	97.62
Vancouver	81.58	97.30	97.23	97.35
Trondheim	84.77	95.15	95.17	94.62
Dardanelles	62.43	97.81	97.74	97.69
Barbados	97.27	99.92	Fail	99.92
Durban	92.45	97.40	Fail	Fail
Tharthar	97.23	98.18	98.14	Fail
Dubai	95.22	98.90	98.78	98.91
Lakeellesmere	81.50	98.10	98.00	97.99
Socheongcho	93.91	98.61	Fail	98.38

To quantify the contribution of each filtering step, a leave-one-stage-out ablation study is performed using ORB. For each iteration, one of the four filtering steps is left out while the other three steps are left unchanged. The results are reported in Table XI. Disabling ratio test leads to a higher number of GCPs, however a worse RMSE is observed with respect to the default error. Similarly, removing the max-distance filter increases the RMSE, indicating that retaining this step mainly affects the georeferencing accuracy more than the total GCPs count. Removing RANSAC and DBSCAN shows negligible

effect when disabled individually, indicating that the two filtering steps are acting on well-filtered GCP sets rather than on independent processing on the initial matches. Indeed, when disabling both ratio test and max-distance filter, while retaining only the latter two stages, the proposed method successfully georeferenced 12 out of 15 captures with a degraded RMSE of 1.41 pixels, considerably worse than the default case. As such, retaining all four filtering steps provides more robustness and also maximizes the accuracy of the identified GCPs.

TABLE XI

RMSE AND MEAN FINAL GCP COUNT FOR THE ABLATION STUDY, REMOVING SINGLE FILTERING STEPS AND DEFAULT PIPELINE.

Configuration	RMSE (pix)	Mean GCPs
Full pipeline (default)	0.81	73.7
No ratio test	0.93	132.3
No max-distance filter	0.89	74.2
No RANSAC	0.81	81.4
No DBSCAN	0.82	74.1

B. Evaluating Georeferencing Using GCPs Subsets

The overall georeferencing accuracy can be evaluated by dividing the set of generated GCPs into two subsets: a training set, used during the optimization process, and an evaluation set, used solely to assess performance. This approach enables estimation of the geodesic distance between the predicted geolocation (ϕ, λ) and the identified ground location $\mathbf{p}_{cl}$, as obtained by the feature-matching algorithms.

The experimental setup is evaluated on a HYPISO-1 capture acquired over the Trondheim area on September 11th, 2024. Optimization and direct georeferencing are performed using different numbers of GCPs. To assess the robustness and estimate the standard deviation of the solution, ten independent trials are conducted for each number of GCPs used for training. The geodesic error is converted from meters to pixels using the HYPISO ground sampling distance. For the selected HYPISO-1 image, the along-track resolution is approximately 590 meters and the across-track resolution 160 meters, reflecting the rectangular pixel footprint. Consequently, errors are expressed in pixels by normalizing with the larger of the two resolutions (590 m).

Given a total number of GCPs N_{GCP} , a subset GCP_{train} , consisting of $GCP_{i,train}$ where $i = 1, 2, \dots, N_{train}$, is selected for training. This training dataset is used within the optimization algorithm. The performance of the algorithm is then independently evaluated using a separate subset, GCP_{eval} , which includes $GCP_{i,eval}$ where $i = 1, 2, \dots, N_{eval}$. The subsets satisfy $N_{GCP} = N_{eval} + N_{train}$. The mean and standard deviation of the RMSE are computed over ten independent runs of the algorithm, each using a randomly selected training subset of GCPs of size N_{train} . The error is calculated as the difference between the latitude and longitude identified by machine vision and the corresponding optimized state-correction geolocation for each GCP. This error assessment is performed for the GCP_{train} and GCP_{eval} subsets, as well as for the complete GCPs dataset. The results are presented in Table XII and Figure 11. The error on the training subset increases as the optimization process handles a larger number

of candidates, resulting in an averaging effect. In contrast, both the evaluation-set error and the overall error decrease as N_{train} increases, along with the reduction of the standard deviation. A similar trend is observed for the maximum error recorded for GCPs across the 10 runs.

The test demonstrates an asymptotic behaviour with respect to N_{train} , indicating that the performance stabilizes beyond a certain training-set size. Moreover, the overall RMSE reaches sub-pixel accuracy after only ten training GCPs used in the attitude optimization correction.

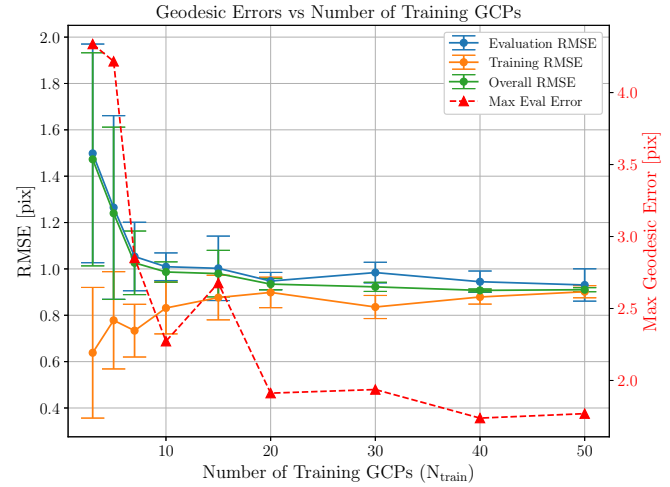


Fig. 11. Error with increasing number of used GCPs.

TABLE XII

EVALUATION OF GEOREFERENCING ACCURACY FOR INCREASING NUMBERS OF TRAINING GCPs

N_{train}	Eval RMSE ± Std [pix]	Train RMSE ± Std [pix]	Overall RMSE ± Std [pix]	Max Error [pix]
3	1.50 ± 0.47	0.64 ± 0.28	1.47 ± 0.46	4.34
5	1.26 ± 0.40	0.78 ± 0.21	1.24 ± 0.37	4.22
7	1.05 ± 0.15	0.73 ± 0.11	1.03 ± 0.14	2.85
10	1.01 ± 0.06	0.83 ± 0.11	0.99 ± 0.04	2.27
15	1.00 ± 0.14	0.88 ± 0.10	0.98 ± 0.10	2.68
20	0.95 ± 0.04	0.90 ± 0.07	0.93 ± 0.02	1.91
30	0.98 ± 0.04	0.84 ± 0.05	0.92 ± 0.02	1.94
40	0.94 ± 0.05	0.88 ± 0.03	0.91 ± 0.01	1.74
50	0.93 ± 0.07	0.90 ± 0.03	0.91 ± 0.01	1.77

The Trondheim capture is also considered to visualize the process of dividing the GCPs into training and evaluation subsets, with $N_{train} = 15$. Figure 12(a) depicts the generated GCPs before correction, together with all residual errors to be corrected. Different symbols are used to indicate the 15 training points and the evaluation points, respectively. Figure 12 (b) shows the output of MEGSSI after correction. Here, residuals are nearly indistinguishable, as the corresponding symbols overlap almost entirely, indicating that GCPs were correctly minimized also on the evaluation set. In addition, this provides qualitative confirmation by comparing the segmented labels with the ground truth, depicted here in black. The slight deviation with ground truth in the lower part of the image can be attributed to the cloud-shadow inducing error in segmentation process, rather than the georeferencing itself.

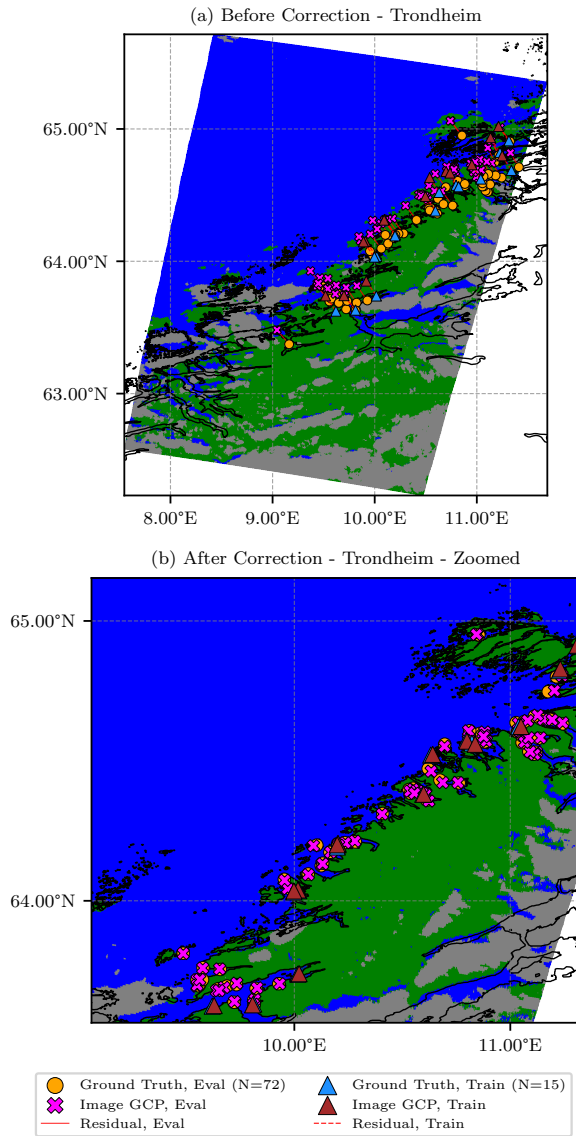


Fig. 12. Before and After applying MEGSSI when using $N = 15$ training points over Trondheim capture.

C. Qualitative Evaluation Across Diverse Sites

To illustrate the performance of MEGSSI across the heterogeneous dataset summarized in Table VIII, three representative captures are presented. The Tharthar capture is shown in Figure 13 to demonstrate the method's applicability to inland water bodies. Lake Habbaniyah and Lake Razzaza, two smaller reservoirs located south of Lake Tharthar, show a visible shoreline mismatch attributable to actual water-level changes associated with the region's ongoing water crisis rather than georeferencing errors. However, this does not affect the overall accuracy, as the extensive shoreline of Tharthar lake provides sufficient geometric information for accurate image registration and georeferencing. In addition, the Dardanelles capture, presented in Figure 14, was selected because it exhibits one of the largest initial georeferencing errors in the dataset, approximately 97 km. Finally, the Barbados capture is reported in Figure 15 to show the performance of the algorithm in a small island setting, where the amount of

available shoreline information is considerably limited. The RGB composites are reported to provide a more intuitive visual reference alongside the shoreline overlay, complementing the segmented representation used in Figure 12.

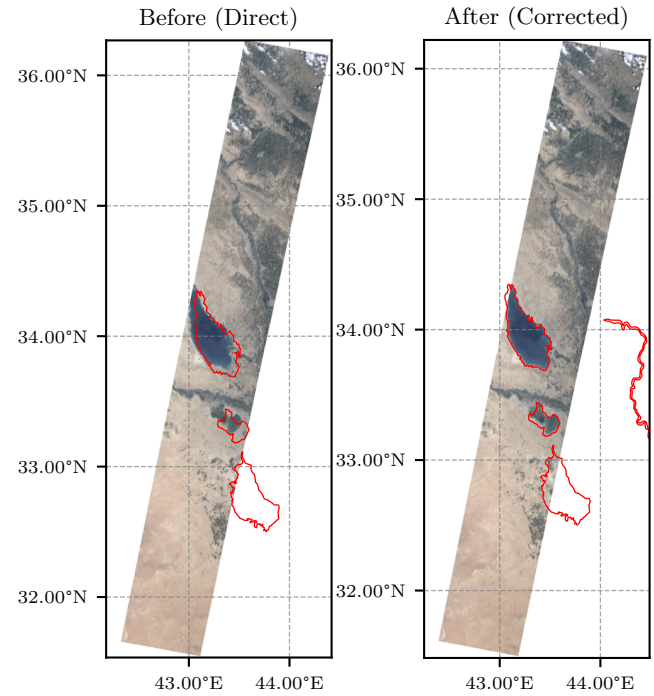


Fig. 13. Before and after applying MEGSSI for the Tharthar capture.

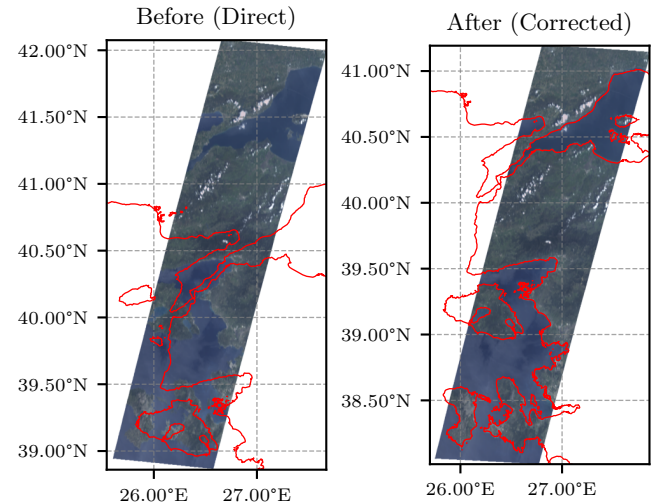


Fig. 14. Before and after applying MEGSSI for the Dardanelles capture.

D. Comparative Analysis against Manually Georeferenced Landmarks

The accuracy and robustness of MEGSSI are evaluated against three established co-registration techniques. First, AROSICS [12] is selected as a baseline, as it has been utilized operationally in hyperspectral missions such as PRISMA [13]. AROSICS is highly effective for ground-based co-registration where initial georeferencing is already approximate, i.e. for

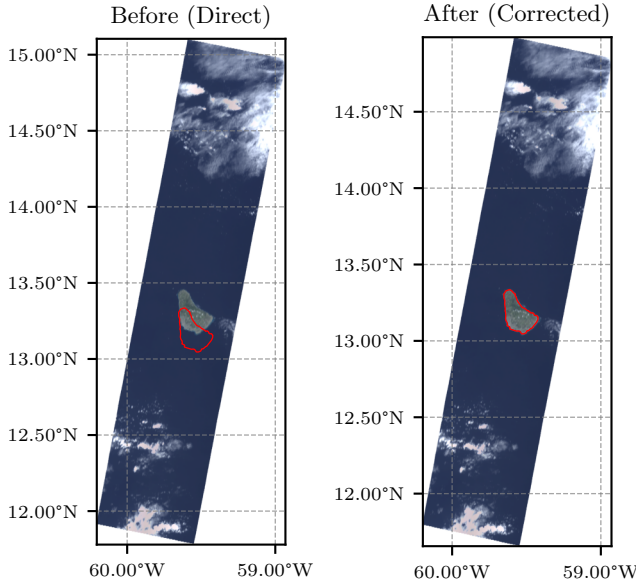


Fig. 15. Before and after applying MEGSSI for the Barbados capture.

small-georeferencing error. When applied to HYPSON data, AROSICS exhibits instability for errors in the order of tens of pixels, and fails to converge entirely when the initial displacement is on the scale of the sensor swath width. Indeed, such large errors result in insufficient geographic overlap with relevant information to achieve matching. Furthermore, the AROSICS global image registration calculates a single global translation for the entire image; for larger errors, this approach cannot account for the significant geometric distortions. A further practical limitation of AROSICS is the memory requirements; in this work, approximately 70-80 MB of Landsat Reference data (for a single band) are utilized for each capture, as reported in Table XIII. This comparison concerns reference-data storage requirements, and not RAM usage.

Beyond the AROSICS baseline, the ICP algorithm [22] and the NR-CIM method [5] are evaluated. Both methods use the same reference data as proposed MEGSSI, as shown in Table XIII. While a rigid-body rotation (as in ICP) is often insufficient for correcting coastal geolocations, the NR-CIM method supports non-rigid transformations. To ensure consistency with the approach in [5], state correction of the HYPSON mission is performed using NR-CIM-identified GCPs before final reprojection. Both AROSICS and NR-CIM frequently fail to converge for datasets with initial geolocation errors ranging from several to tens of kilometers. Consequently, AROSICS has been initialized with the output of the proposed method for sub-pixel refinement, while the NR-CIM method relies on the initial estimate obtained from the ICP algorithm, since both operate directly on point clouds. Global performance of the different methods is assessed through an Aggregated RMSE metric, which is calculated by concatenating the geodesic error residuals from all manually generated GCPs for the entire dataset into a single error vector. This provides a global metric of the overall system precision, defined as $RMSE_{agg} = \sqrt{\frac{1}{N} \sum \epsilon^2}$, where ϵ is the individual error and N is the total number of GCPs. As shown in Table XIV, the proposed

MEGSSI achieves an Aggregated RMSE of 0.81 pixels with a 15/15 success rate. AROSICS, applied here as a post-MEGSSI refinement step rather than an independent baseline, further refines the RMSE to 0.70 pixels when initialized with MEGSSI's output. Given the large initial georeferencing errors in this dataset, initializing AROSICS with MEGSSI output in this way is necessary for a meaningful comparison. Both its accuracy and success rate are therefore conditional on MEGSSI initialization, and the results should be interpreted with this dependency in mind. Furthermore, our approach demonstrates significantly higher accuracy and robustness than the ICP (3.80) and NR-CIM (1.31) methods. While point cloud algorithms struggle with noisy data and complex coastlines, they also face geometric ambiguities in relatively straight coastal segments, as evidenced by the failures in the Durban and HypernetEstuary captures. To consider failures in a success-aware score, we also introduce a Reliability-Adjusted RMSE ($RMSE_{adj} = RMSE_{agg} / \text{Success Rate}$) which penalizes methods in proportion to their failure rate. This parameter is an author-defined heuristic, not a standard statistical error metric, and it does not distinguish between failure reasons, applying the same penalty regardless of the cause. It should therefore be considered as additional information, while the primary accuracy parameters remain the $RMSE_{agg}$ and success rate. Because of the lower success rate, as shown in Table XIV, $RMSE_{adj}$ is raised with respect to $RMSE_{agg}$ to 4.38 and 1.51 pixels for ICP and NR-CIM, respectively, while MEGSSI and AROSICS are unaffected.

Along-track and across-track residuals are defined as the corrected-minus-ground-truth positional error projected onto the local line-index (along-track, direction of satellite motion) and sample-index (across-track) directions, respectively. Decomposing the error between along-track and across-track components, MEGSSI achieves $RMSE_{agg}$ of 463.8 m and 431.8 m, comparable in magnitude and consistent with the isotropic 300 m $\times$ 300 m grid resolution used for feature matching and GCP generation. In native pixel units, this corresponds to approximately 0.75 and 2.78 pixels along-track and across-track, respectively. The larger across-track value is expected, as this finer, anisotropic pixel direction is penalized by resampling onto the coarser, regular 300 m grid—consistent with the ablation study on grid resolution in Table V. It is important to note that the $RMSE_{agg}$ equal to 0.81 is not obtained by combining the 0.75 and 2.78 pixel values above; rather, it derives from normalizing the total, non-decomposed error by the along-track native pixel size.

TABLE XIII
COMPARISON OF MEMORY REQUIREMENTS

Method	Reference Type	Storage	Onboard Scalability
MEGSSI	Vector (GSHHS - Full)	~192 MB (Global)	High (Compact)
AROSICS [12]	Landsat Imagery	~80 MB per capture	Limited (High storage)
ICP [22]	Vector (GSHHS - Full)	~192 MB (Global)	High (Compact)
NR-CIM [5]	Vector (GSHHS - Full)	~192 MB (Global)	High (Compact)

E. MEGSSI Robustness to Cloud Coverage

Robustness to clouds is a key parameter for optical systems, as atmospheric effects degrade the sensor capabilities. To quantify the performance with respect to clouds, four different

TABLE XIV
GEOREFERENCING ACCURACY COMPARISON OF DIFFERENT METHODS
EXPRESSED AS RMSE

Capture	Clouds [%]	Orig. Err. [pix, km]	MEGSSI [pix]	ARS** [pix]	ICP [pix]	NR-CIM [pix]
Aquawatchmoreton	0.9	47.0, 28.1	0.68	0.67	1.16	1.06
Blanca	27.4	45.1, 27.9	0.67	0.78	8.71	2.52
Hypermetestuary	1.7	19.9, 11.9	1.23	0.66	FAIL	FAIL
Mariamadre	0.3	25.0, 14.8	0.74	0.67	1.89	0.81
Mvco	6.0	15.4, 9.3	0.75	0.31	0.58	0.94
Tampa	9.4	19.2, 11.4	0.48	0.51	2.64	0.60
Vancouver	8.6	47.8, 28.1	0.72	0.80	5.46	1.17
Trondheim	20.3	67.1, 40.2	0.93	0.82	3.98	1.65
Dardanelles	4.4	161.6, 97.0	1.17	1.15	3.04	1.38
Barbados	16.7	21.8, 13.2	0.38	0.57	0.69	0.87
Durban	9.3	21.8, 18.7	0.94	0.63	FAIL	FAIL
Tharthar	0.3	10.3, 9.0	0.87	0.38	0.64	0.97
Dubai	2.1	9.0, 7.8	0.53	0.65	0.97	0.94
Lakeellesmere	8.3	30.5, 16.7	1.60	1.57	5.13	1.42
Socheongcho	2.2	9.7, 8.5	0.52	0.41	0.46	0.76
RMSE _{agg} *	-	-	0.81	0.70	3.80	1.31
Success Rate	-	-	15/15	15/15	13/15	13/15
RMSE _{adj}	-	-	0.81	0.70	4.38	1.51

*Aggregated RMSE results exclude failed captures for NR-CIM and ICP. Lakeellesmere is also excluded as it is an agile ADCS capture with non-fixed pointing and is not representative of standard imaging modalities. ARS refers to the AROSIS.

**ARS is applied as a refinement step initialized with MEGSSI output; its values are conditional on successful MEGSSI convergence, not an independent baseline.

capture sites with varying cloud conditions are evaluated, as summarized in Table XV. The results in Figure 16(a) show that high cloud coverage reduces the algorithm's ability to identify a sufficient number of GCPs. However, as long as a small portion of the coastline remains visible, the algorithm can often find a solution, as demonstrated by a successful case despite 80% cloud coverage. For cloud levels below 40%, almost all captures are successfully georeferenced, except for a single Los Angeles capture where the coastline is specifically obscured despite a total scene cloud cover of about 20%. Figure 16(b) shows the RMSE for successful captures, evaluated by splitting the detected GCPs into distinct training and evaluation sets. These results indicate that, even as cloud levels increase, the RMSE remains at or below the sub-pixel level for successfully georeferenced captures.

TABLE XV
CLOUD COVERAGE AND GEOREFERENCING SUCCESS RATE OVER
DIFFERENT TEST-AREAS

Location	Cloud Cover [%]			Success Rate
	Min	Max	Mean	
Ariake	0.2	77.2	23.4	16 / 18
Forth Estuary	2.4	95.2	71.6	4 / 16
Los Angeles	0.3	62.2	17.6	14 / 17
Oslo Fjord	6.0	87.9	38.3	14 / 18

*Success defined as registration with ≥ 5 detected GCPs.

F. Validating Constant Rotation-Matrix Assumption

The state correction in Equation (1) assumes a single rotation matrix correction, $\Delta \mathbf{R}$, applied to the entire capture. This assumption is motivated by the operational characteristics of HYPISO, because most HYPISO captures are acquired following fixed-pointing operations, with a prescribed ADCS quaternion sequence lasting approximately one minute (598 image lines), showing that the attitude error remains largely stable throughout the capture, consistent with the results reported in Table XIV. To evaluate the validity of this assumption, an experiment has been carried out on a nominal capture over Ariake, Japan. This scene contains 250 identified GCPs, spread along most of the along-track direction, between lines 131 and 501. Following the methodology described in Section IV-B, the GCPs are divided into a training set and an evaluation

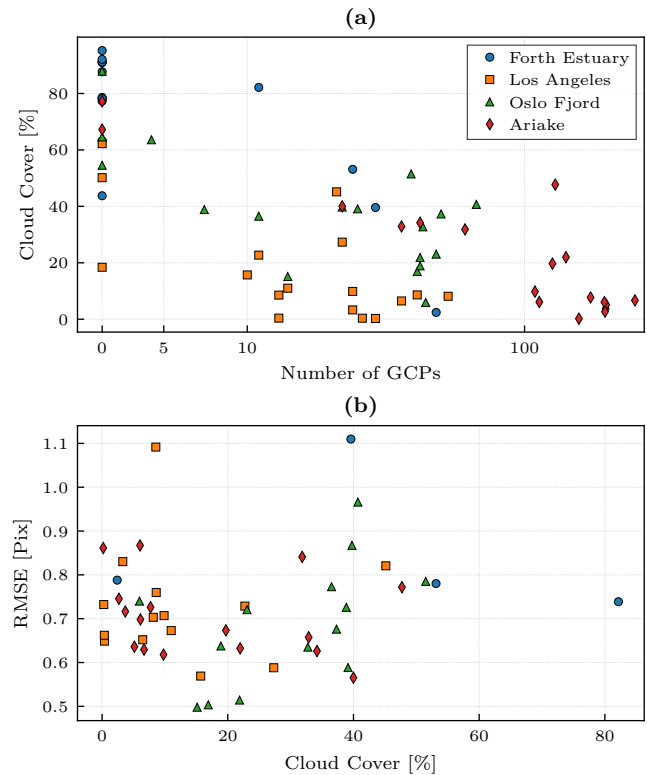


Fig. 16. Impact of cloud coverage on georeferencing performance: (a) number of GCPs detected across all captures, and (b) overall RMSE for successful registrations.

set, with the residual error on the evaluation set computed from the attitude correction and reprojection obtained using the training set. In addition to a random split, in which the GCPs are split in half at random, along-track splits are also considered. The image lines are divided into two, three, or four contiguous segments, with the model trained on the first segment and evaluated on the last (opposite) segment, discarding any intermediate segments for the three- and four-way subdivisions. The resulting evaluation RMSE values are reported in Table XVI. The RMSE rises modestly from the

TABLE XVI
GCP-SPLIT EVALUATION STRATEGIES, ARIAKE CAPTURE

Strategy	Train N	Eval N	Eval RMSE [pix]
Random-Split	125	125	0.63
Half-Split	93	157	0.65
Third-Split	33	112	0.74
Quarter-Split	20	29	1.00

Random-Split to the Quarter-Split configuration, confirming that fitting a correction on one side of the image leads to slightly worse results on the opposite side. However, even this limiting case remains at pixel-level accuracy, supporting the validity of the constant-rotation assumption for nominal HYPISO acquisitions. To further assess the limits of this assumption, the agile Lake Ellesmere capture is considered. Unlike the nominal fixed-pointing acquisition over Ariake, this image was acquired during an agile observation involving considerable attitude variation throughout the capture. Because of the along-track distribution of the identified GCPs, only the Half-Split and the Random-Split configurations could be

evaluated, as reported in Table XVII.

TABLE XVII

GCP-SPLIT EVALUATION STRATEGIES, AGILE LAKE ELLESMERE CAPTURE

Strategy	Train N	Eval N	Eval RMSE [pix]
Random-Split	32	32	1.45
Half-Split	47	17	2.71

Here, the Half-Split forces the rotation fit from one part of the maneuver to extrapolate to a region acquired under a different attitude condition, resulting in reduced accuracy compared to the Random-Split. The latter is representative of the standard correction used by MEGSSI, with an averaging effect across the different parts of the image when GCP are identified. As such, the Random-Split error of 1.45 pixels shows that, even for agile captures, the constant-rotation-matrix assumption remains useful for reducing the georeferencing error. Nevertheless, the increased error observed in the Half-Split case indicates that residual time-varying attitude effects are present. However, a time-varying extension of the correction model – segmenting the image along-track, correcting each segment independently, and interpolating between them – is a natural next step for reaching sub-pixel accuracy in agile captures as well, provided that a sufficient number of reliable GCPs are available within each segment.

G. Failure Case Analysis

Although the proposed MEGSSI has demonstrated robustness across the heterogeneous dataset considered in this work, its success depends on the availability of distinctive and correctly segmented shoreline features. However, several challenging conditions can limit its applicability and prevent the identification of a sufficient number of reliable GCPs. In this section, we report the RGB composite acquired by the HSI camera, together with the corresponding land-sea-cloud segmentation for three representative captures in which MEGSSI failed to identify a sufficient number of matches and was therefore interrupted during the filtering stage (Figure 17). The first example is a capture over Tripoli. In this scene, low-curvature coastlines can greatly limit feature matching, due to the lack of corner-like features and the resulting difficulty in defining unambiguous descriptors and establishing reliable correspondences between the satellite image and the ground-truth reference map. A second challenge is that extensive cloud cover can severely limit the method by obscuring the coastline needed for detection. In the Los Angeles capture, most of the ocean-facing coastline is covered by clouds, and MEGSSI could not identify a sufficient set of GCPs. The final example highlights the impact of segmentation errors. The segmentation errors can propagate into the georeferencing pipeline when the segmentation model provides erroneous class estimates. In particular, a capture acquired over northern Norway during an evening pass exhibited a very low signal-to-noise ratio due to reduced illumination conditions. Under these conditions, the segmentation model produced ambiguous or fragmented land-sea-cloud classifications, as shown in Figure 17. The resulting shoreline representation lacked the consistency required for reliable feature matching, causing the georeferencing process to fail.

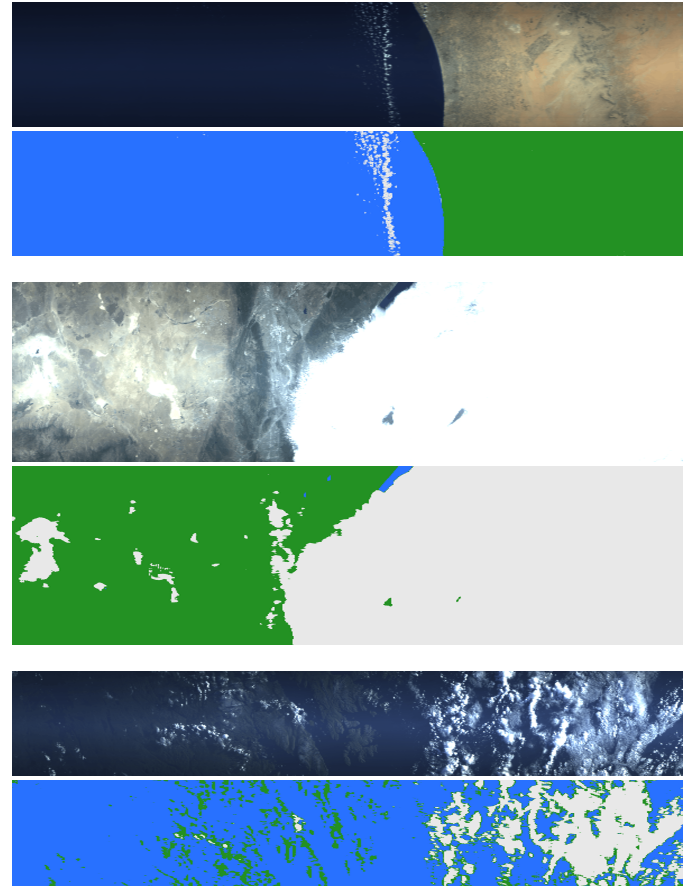


Fig. 17. Failure cases for MEGSSI. From top to bottom: the Tripoli capture, limited by a smooth, low-curvature coastline; and the Los Angeles capture, where extensive cloud cover obscures the coastline; the Nordkapp capture, affected by low illumination during an evening pass. For each capture, the RGB composite is shown above its corresponding land-sea-cloud segmentation.

H. Processing Time

The experiments were conducted on a Dell Precision 3591 laptop with an Intel Core Ultra 7 155H CPU and 32 GB of DDR5 RAM, running Windows 11 Pro (64-bit). The system used a 512 GB NVMe SSD for storage. The algorithms were implemented in Python 3.12.3, and executed using a single thread. The run-time reported in Table XVIII refers to the pipeline that implements ORB.

TABLE XVIII
RUN-TIME OF THE PROCESSING ALGORITHM

Mean time	Max time	Standard Deviation
12.22 sec	15.75 sec	2.24 sec

The On-board Processing Unit (OPU) on HYPISO-1 and HYPISO-2 consists of a PicoZed System-on-Module (SoM) mounting a Zynq XC7Z030-1 System-on-a-Chip (SoC), with dual 667 MHz ARM Cortex-A9 cores and a Kintex-7 FPGA [47]. Although onboard execution is not the focus of this work, the implementation of Python code was tested on the PYNQ-Z1 board [48]. The platform was selected for its computational capabilities comparable to OPU on HYPISO-1 and HYPISO-2, while being suitable for rapid integration and testing through its Python framework. RAM usage is reported for the proposed pipeline only. As reported in Table XIX, the

pipeline successfully georeferenced the Mariamadre capture (see Table VIII) with a peak RAM usage of around 383 MB, well within the 512 MB capacity of the PYNQ-Z1 board. While the current execution time is approximately 4 minutes, a significant reduction in execution time is expected by translating the code into C++, and exploiting FPGA acceleration capabilities.

TABLE XIX
RESOURCE BENCHMARKING: MARIAMADRE SCENE

Platform	Run-Time	Peak RAM Usage
Workstation (WSL2)	9.89 sec	608.35 MB
PYNQ-Z1	242.39 sec	382.82 MB

V. LIMITATIONS

The method has been applied to captures located all over the world, and has been tested for cloudy captures or special imaging modes. However, it is important to recognize and analyze the limitations of the method.

A. Accuracy of the Segmentation

The methodology is segmentation-agnostic, relying on the output of the segmentation algorithm, which identifies land-sea-cloud classes in the image. As a result, low-light conditions, thin clouds, and haze can all affect the proposed methodology. However, since the segmentation map is an input data product, these factors cannot be directly controlled by the georeferencing algorithm. As a result, adapting the methodology requires extensive work in assessing the performance of the selected segmentation algorithm for the satellite in question.

B. Lack of Features

The selected machine vision algorithms - SIFT, ORB, and BRISK - detect local features, such as corners, blobs, and other distinctive features. As a result, during algorithm testing, smooth - almost straight - shorelines proved to be challenging for detecting GCPs due to the low information they carry, often causing the method to fail in this particular scenario. Future work could explore extending the methodology to other methods, such as contour matching, as a fallback solution for the captures where the current method is ineffective.

Similarly, a lack of features in an image might be caused by the scarcity of shorelines in the image or their total absence. Any homogeneous region without lakes or sea, such as deserts or forests, would require another approach for registration: one option is to use the RGB information. As an alternative, a segmentation-approach can still be viable if the implemented segmentation algorithms can extract relevant information from the captures and compare them to geospatial data sources, such as OpenStreetMap. The proposed framework can be extended to any high-contrast geographic boundary present in OpenStreetMap [25], and identifiable by the satellite segmentation model. As an example, many desert areas (*natural=desert*) contain cultivated land (*landuse=farmland*), presenting a strong spectral and geometric contrast, as seen in regions like the Imperial Valley in California. In these arid

environments, the sharp boundaries between irrigated crops and sand can be used for correcting the model. Similarly, in densely populated regions, the transition between urban areas (*boundary=urban*) and the surrounding land cover can be used to perform georeferencing due to the high spectral contrast between the two classes. The core requirement for the proposed methodology is the presence and identification of static, spectrally distinctive features that allow for the extraction of reliable GCPs to provide accurate georeferencing across diverse global environments.

C. State Correction Modelling Limitations

The proposed state correction model employs an Earth ellipsoid model for direct georeferencing, without accounting for terrain elevation through a DEM. While incorporating a DEM could improve accuracy in coastal areas with steep terrain, its inclusion would substantially increase the memory footprint of the ground-truth reference data, conflicting with the memory-efficiency objective of MEGSSI. Moreover, DEM integration is expected to become more relevant for higher-resolution sensors, where terrain-induced distortions and mountain shadows may additionally affect both georeferencing and segmentation performance. As such, projection errors caused by unmodeled terrain profiles are not considered in the present work, but represent a natural direction for future development.

Similarly, the static shoreline reference maps (GSHHS) used for GCP generation do not account for tidal or seasonal water-level variations. This limitation is expected to have only a minor impact for HYPISO, since shoreline-reference uncertainty is generally smaller than or comparable to the effective resolution relevant to this study. However, tidal and seasonal effects should be taken into consideration when adapting the method to higher-resolution sensors, particularly in areas with significant intertidal zones or large seasonal water-level fluctuations.

Second, the reprojection-error optimization in Equation (1) uses unweighted least squares optimization, treating all GCPs as equally reliable. In practice, the reliability of individual GCPs may vary depending on factors such as feature-match quality or segmentation confidence. For example, the Lowe's ratio test computed in feature matching could be used as a proxy for match confidence, although it is currently used only as a filtering criterion. Likewise, confidence estimates from the segmentation model could carry more information for weighting GCPs. However, the current land-sea-cloud segmentation framework does not support confidence levels, as it outputs hard class labels rather than confidence scores, limiting the applicability of such approaches. As such, unweighted least squares optimization was considered for its simplicity, and because it operates on GCPs that were heavily filtered before being ingested by the optimization. Nevertheless, incorporating confidence-based weighting schemes or robust cost functions may further improve resilience to residual outliers and constitutes a promising direction for future research. Accordingly, all statements regarding GCPs identified as reliable, or the MEGSSI method being robust, should be interpreted based on the dataset evaluation presented in this work, rather

than an explicit propagation of segmentation, shoreline, or individual GCP uncertainty.

VI. CONCLUSION

The paper has introduced a computationally efficient implementation of a georeferencing pipeline for pushbroom imagers. The MEGSSI methodology has been adapted to the HYPISO mission, for both HYPISO-1 and HYPISO-2, demonstrating consistent sub-pixel accuracy during routine imaging operations, while proving robustness across varying cloud levels. The methodology has been developed to be fast, accurate, and based on limited-size locally available ground-truth, without relying on large online image databases.

Future work will focus on the onboard implementation of the HYPISO satellites and in-orbit testing. However, steps such as resampling, ground-truth generation, or ADCS interpolation remain time-consuming. To address this, an FPGA accelerator may be integrated to exploit the onboard computational capabilities of the HYPISO platform. In addition, deep-learning-based feature matching techniques could be explored to increase matching accuracy and robustness. Beyond the analyzed variation in cloud cover, initial georeferencing error, and water-body type, the evaluation dataset could be further categorized according to coastal terrain types (e.g., steep cliffs vs. mudflats) for future work. Finally, the proposed method could also be adapted to snapshot imagers by reformulating the reprojection error with respect to a single acquisition time, extending the applicability to multiple imaging modalities.

REFERENCES

- [1] P. J. Bresnahan, S. Rivero-Calle, J. Morrison, G. Feldman, A. Holmes, S. Bailey, A. Scott, L. Hong, F. Patt, N. Kuring *et al.*, "High-resolution ocean color imagery from the SeaHawk-HawkEye CubeSat mission," *Sci. Data*, vol. 11, no. 1, p. 1246, 2024.
- [2] M. E. Grøtte, R. Birkeland, E. Honoré-Livermore, S. Bakken, J. L. Garrett, E. F. Prentice, F. Sigernes, M. Orlandić, J. T. Grasdahl, and T. A. Johansen, "Ocean color hyperspectral remote sensing with high resolution and low latency—The HYPISO-1 CubeSat mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2022.
- [3] H. Zhao, M. F. Lembeck, A. Zhuang, R. Shah, and J. Wei, "Real-time convolutional neural network-based star detection and centroiding method for CubeSat star tracker," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 4, pp. 8172–8184, 2025.
- [4] T. Kouyama, A. Kanemura, S. Kato, N. Imamoglu, T. Fukuhara, and R. Nakamura, "Satellite attitude determination and map projection based on robust image matching," *Remote Sens.*, vol. 9, no. 1, 2017.
- [5] J. Liu, W. Li, J. Peng, L. Shen, H. Han, P. Zhang, and L. Yang, "Geolocation error estimation and correction on long-term MWRI data," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9448–9461, 2021.
- [6] B. Chintalapati, A. Precht, S. Hanra, R. Laufer, M. Liwicki, and J. Eickhoff, "Opportunities and challenges of on-board AI-based image recognition for small satellite Earth observation missions," *Adv. Space Res.*, vol. 75, no. 9, pp. 6734–6751, 2025.
- [7] L. Wang, Y. Xiang, Z. Wang, H. You, and Y. Hu, "On-board geometric rectification for micro-satellite based on lightweight feature database," *Remote Sens.*, vol. 15, no. 22, p. 5333, 2023.
- [8] F. Eugenio and F. Marques, "Automatic satellite image georeferencing using a contour-matching approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2869–2880, 2003.
- [9] L. Cheng, L. Tong, Y. Liu, M. Li, and J. Wang, "Automatic registration of coastal remotely sensed imagery by affine invariant feature matching with shoreline constraint," *Mar. Geod.*, vol. 37, no. 1, pp. 32–46, 2014.
- [10] H. Goncalves, L. Corte-Real, and J. A. Goncalves, "Automatic image registration through image segmentation and SIFT," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2589–2600, 2011.
- [11] G. Simarro, D. Calvete, F. Ribas, Y. Castillo, and C. Puig-Polo, "UOrtos: Methodology for co-registration and subpixel georeferencing of satellite imagery for coastal monitoring," *Remote Sens.*, vol. 17, no. 7, 2025.
- [12] D. Scheffler, A. Hollstein, H. Diedrich, K. Segl, and P. Hostert, "AROSICS: An automated and robust open-source image co-registration software for multi-sensor satellite data," *Remote Sens.*, vol. 9, no. 7, 2017.
- [13] S. Cogliati, F. Sarti, L. Chiarantini, M. Cosi, R. Lorusso, E. Lopinto, F. Miglietta, L. Genesio, L. Guanter, A. Damm, S. Pérez-López, D. Scheffler, G. Tagliabue, C. Panigada, U. Rascher, T. Dowling, C. Giardino, and R. Colombo, "The PRISMA imaging spectroscopy mission: overview and first performance analysis," *Remote Sens. Environ.*, vol. 262, p. 112499, 2021.
- [14] H. Kartal, A. Alganci, and E. Sertel, "Automated orthorectification of VHR satellite images by SIFT-based RPC refinement," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 6, 2018.
- [15] R. Müller, T. Krauß, M. Schneider, and P. Reinartz, "Automated georeferencing of optical satellite data with integrated sensor model improvement," *Photogramm. Eng. Remote Sens.*, vol. 78, pp. 61–74, Jan. 2012.
- [16] P. R. Hakim, A. P. S. Jayani, A. Sarah, and W. Hasbi, "Autonomous image georeferencing based on database image matching," in *2018 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES)*, 2018, pp. 1–6.
- [17] A. Marsetič, K. Oštir, and M. K. Fras, "Automatic orthorectification of high-resolution optical satellite images using vector roads," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6035–6047, 2015.
- [18] L. Carozza and A. Bevilacqua, "Error analysis of satellite attitude determination using a vision-based approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, pp. 19–29, 2013.
- [19] R. Sugimoto, T. Kouyama, A. Kanemura, S. Kato, N. Imamoglu, and R. Nakamura, "Automated attitude determination for pushbroom sensors based on robust image matching," *Remote Sens.*, vol. 10, no. 10, 2018.
- [20] S. Zhou, H. Pan, T. Huang, and P. Zhou, "High accuracy georeferencing of GF-6 wide field of view scenes toward analysis ready data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, 2023.
- [21] L. H. Hoffman, *Calculation and accuracy of ERBE scanner measurement locations*. National Aeronautics and Space Administration, 1987, vol. 2670.
- [22] C. Chiatante, D. D. Langer, J. L. Garrett, R. Birkeland, S. Berg, and M. Orlandić, "Onboard hyperspectral classification enables georeferencing," in *2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2023, pp. 1–5.
- [23] J. A. Justo, A. Ghiță, D. Kováč, J. L. Garrett, M.-I. Georgescu, J. Gonzalez-Llorente, R. T. Ionescu, and T. A. Johansen, "Semantic segmentation in satellite hyperspectral imagery by deep learning," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 273–293, 2025.
- [24] P. Wessel and W. H. Smith, "A global, self-consistent, hierarchical, high-resolution shoreline database," *J. Geophys. Res. Solid Earth*, vol. 101, no. B4, pp. 8741–8743, 1996.
- [25] OpenStreetMap contributors, "Planet dump retrieved from <https://planet.osm.org>," 2017. [Online]. Available: <https://www.openstreetmap.org>
- [26] S. D. Mudiyansele, C. Dai, I. M. Howat, E. Larour, and E. Husby, "A global high resolution coastline database from satellite imagery," *Sci. Data*, vol. 12, no. 1, pp. 1–9, 2025.
- [27] R. Pelich, M. Chini, R. Hostache, P. Matgen, and C. López-Martínez, "Coastline detection based on Sentinel-1 time series for ship- and flood-monitoring applications," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1771–1775, 2021.
- [28] ESA WorldCover Consortium, "ESA WorldCover Annual Composites: Data Access," 2020, accessed: 2026-02-22. [Online]. Available: <https://esa-worldcover.org/en/data-access>
- [29] B. Bauer-Marschallinger, D. Sabel, and W. Wagner, "Optimisation of global grids for high-resolution remote sensing data," *Comput. Geosci.*, vol. 72, pp. 84–93, 2014.
- [30] D. Hoese, M. Raspaud, P. Lahtinen, W. Roberts, T. Lavergne *et al.*, "pytroll/pyresample: Version 1.16.0," 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3372769>
- [31] S. Gillies *et al.*, "Rasterio: geospatial raster I/O for Python programmers," Mapbox, 2013–. [Online]. Available: <https://github.com/rasterio/rasterio>
- [32] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
- [33] S. A. K. Tareen and Z. Saleem, "A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK," in *2018 International Conference*

on Computing, Mathematics and Engineering Technologies (iCoMET), 2018, pp. 1–10.

- [34] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.
- [35] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in *2011 International Conference on Computer Vision*, 2011, pp. 2564–2571.
- [36] S. Leutenegger, M. Chli, and R. Y. Siegwart, “BRISK: Binary Robust invariant scalable keypoints,” in *2011 International Conference on Computer Vision*, 2011, pp. 2548–2555.
- [37] G. Developers, “geopy: Geocoding library for Python,” 2024, version 2.4.1. [Online]. Available: <https://github.com/geopy/geopy>
- [38] C. F. Karney, “Algorithms for geodesics,” *J. Geod.*, vol. 87, pp. 43–55, 2013.
- [39] C. Fraser and T. Yamakawa, “Insights into the affine model for high-resolution satellite sensor orientation,” *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 5–6, pp. 275–288, 2004.
- [40] L. Yan, D. P. Roy, H. Zhang, J. Li, and H. Huang, “An automated approach for sub-pixel registration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery,” *Remote Sens.*, vol. 8, no. 6, p. 520, 2016.
- [41] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD’96. AAAI Press, 1996, p. 226–231.
- [42] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat. Methods*, vol. 17, pp. 261–272, 2020.
- [43] J. A. Justo, D. D. Langer, S. Berg, J. Nieke, R. T. Ionescu, P. G. Kjeldsberg, and T. A. Johansen, “Hyperspectral image segmentation for optimal satellite operations: In-orbit deployment of 1D-CNN,” *Remote Sens.*, vol. 17, no. 4, 2025.
- [44] P. Berg, S. Boyle, A. Alinejad, and M. Orlandić, “Wavelet convolutional neural network for hyperspectral image segmentation,” in *2025 15th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2025.
- [45] T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, W. E. Kerzendorf *et al.*, “Astropy: A community Python package for astronomy,” *Astron. Astrophys.*, vol. 558, p. A33, 2013.
- [46] D. D. Langer, J. L. Garrett, B. A. Kristiansen, S. Bakken, S. Berg, R. Birkeland, J. Tommy Gravdahl, T. A. Johansen, and A. J. Sørensen, “Agile maneuvers for push-broom imaging satellites,” *IEEE Trans. Geosci. Remote Sens.*, vol. 63, pp. 1–16, 2025.
- [47] D. D. Langer, M. Orlandić, S. Bakken, R. Birkeland, J. L. Garrett, T. A. Johansen, and A. J. Sørensen, “Robust and reconfigurable on-board processing for a hyperspectral imaging small satellite,” *Remote Sens.*, vol. 15, no. 15, 2023.
- [48] PYNQ, “PYNQ — Python Productivity for AMD Adaptive Computing platforms,” Advanced Micro Devices, Inc., 2024, accessed: 2025-08-11. [Online]. Available: <https://www.pynq.io/>



Corrado Chiatante (Graduate Member, IEEE) received the B.Sc. in Aerospace Engineering and the M. Sc. in Space Engineering from the Politecnico of Milano, Italy. He is currently pursuing a PhD with the Department of Electronic Systems at the Norwegian University of Science and Technology (NTNU). His main interests include the integration of small satellites with other data sources to enable cooperative Earth-observation in a system-of-system framework.



Håvard Sneffjellå Løvås Håvard Sneffjellå Løvås received the M.Sc. degree in marine technology from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2019, and the Ph.D. degree in 2023 with the Centre for Autonomous Marine Operations and Systems (AMOS) on hyperspectral imaging techniques and photogrammetry from underwater and aerial platforms. Since 2024, he has been working in NORCE Research AS, focusing on optical imaging from drones and aircrafts.



missions with hyperspectral imager payloads.

Dennis Langer (Member, IEEE) received the B.Sc. degree in electrical engineering from UiT—The Arctic University of Norway, Tromsø, Norway, in 2016, and the M.Sc. degree in industrial cybernetics from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2019. In 2024, he received the PhD degree from the Department of Marine Technology at NTNU. His research interests include geometric processing of pushbroom satellite data, onboard software for satellite systems, and uncrewed aerial vehicle (UAV)



Joseph L. Garrett (Member, IEEE) received a B.Sc. degree in mathematics and physics from the Ohio State University in 2011, followed by a PhD in physics in 2017 from the University of Maryland – College park. He has worked in the Department of Engineering Cybernetics with the NTNU SmallSat Lab since 2018. His research interests include hyperspectral imaging, image processing, and uncertainty.



Roger Birkeland (Member, IEEE) received the M.Sc. degree in electronic engineering and the Ph.D. degree in satellite communications from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2007 and 2019, respectively. Since 2026, he has been an Associate Professor in the Department of Electronic Systems at NTNU, where he manages various satellite projects. His main interests include small satellite technologies and systems for optical and RF remote sensing as well as communication systems for remote areas.



Milica Orlandić (Member, IEEE) received B.Sc. and M.Sc. degrees in electrical engineering from the University of Montenegro in 2007 and 2009, respectively, and the Ph.D. degree in electronics from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2015. She has been holding the position of Associate Professor in Electronic Systems Engineering at NTNU since 2019. Her research interests include satellite system engineering, remote sensing imaging system design, and high-performance reconfigurable systems.



Tor Arne Johansen (Senior Member, IEEE) received the MSc degree in 1989 and the PhD degree in 1994, both in electrical and computer engineering, from the Norwegian University of Science and Technology, Trondheim, Norway. From 1995 to 1997, he worked at SINTEF as a researcher before he was appointed Associated Professor at the Norwegian University of Science and Technology in Trondheim in 1997 and Professor in 2001. He has published several hundred articles in the areas of control, estimation, optimization, autonomous systems and sensing. In 2002 Johansen co-founded the company Marine Cybernetics AS where he was Vice President until 2008. Prof. Johansen received the 2006 Arch T. Colwell Merit Award of the SAE, and is currently director of the Unmanned Aerial Vehicle Laboratory at NTNU and the SmallSat Laboratory at NTNU. He co-founded the spin-off companies Scout Drone Inspection, UBIQ Aerospace, Zeabuz and SentiSystems.