SECOND CYCLE, 30 CREDITS

Development of a Level-0 Geoprocessing Platform for a Multispectral Remote Sensing Payload

SERGIO SANTIAGO BERNABEU PEÑALBA
Development of a Level-0 Geoprocessing Platform for a Multispectral Remote Sensing Payload

Sergio Santiago Bernabeu Peñalba

Abstract—This thesis presented an overview of the development of a geolocating algorithm as part of a geoprocessor for raw satellite imagery. This algorithm was devised for and limited by the specifications of a state-of-the-art multispectral telescope designed by Aistech Space, hosted onboard the Guardian spacecraft, which will observe Earth through the visible, near infrared, and thermal infrared bands of the electromagnetic spectrum. The geolocation algorithm presented here is composed of the combination of two models. The first is a physical model, which makes use of spacecraft telemetry and external satellite-tracking data to approximate the geographical center of a sensed scene. Secondly, an optical model obtains a reference Landsat image based on the timestamp and approximated location of the sensed scene and utilizes image processing techniques to pinpoint a more precise geographical location of the sensed scene within acceptable limits. This performance was achieved in 77% of the cases considered. To conclude, a roadmap of the subsequent development topics and their relevance was laid out.

Index Terms—Infrared imagery, VIS, NIR, LWIR, geoprocessor, geolocation, quaternions, Aistech Space, image processing, ORB, RANSAC, Sentinel, Landsat.


I. INTRODUCTION

Satellite imagery has grown from an unimaginable concept a century ago to a key piece of technology that shapes our everyday life. It guides our decision process on what to wear tomorrow and what to take on a week-long trip, it has radically changed how to orient oneself in a new city and how society follows and keeps track of events happening across the globe. The range of satellite imagery that can be obtained and applied to society’s daily needs has grown over the last decades, with many specialized infrared-based observation programs appearing over the last years, following the rise of CubeSats and the newspace era.

All of this satellite imagery generated and downlinked to Earth has to be processed prior to its use; this series of processing stages are performed by a geoprocessor. On top
of the interpretation and correction of the data contained in the sensed image itself, the area it captures over Earth’s surface must be determined; this task is performed by a geolocator, a design for which is presented in this thesis.

The Customizable Aistech Geoprocessing Engine (CAGE) is a proof-of-concept algorithm that was developed to show the viability of an in-house solution to the geolocation and overall satellite imagery geoprocessing needs of Aistech Space. The company launched Guardian, its first multispectral Earth-observation satellite, in May, 2022, as a first step towards a constellation of 20 satellites aimed at providing a steady stream of high resolution imagery. The geolocator was composed of two models: a physical one and an optical one. A layout of CAGE is presented in Figure 1; it was designed as a modular processor so that stages could be easily added as new developments are made to it. The physical model made use of Guardian’s telemetry to approximate a sensed scene’s location on Earth’s surface through a purely geometrical approach; this was successfully achieved with an error of $44.0 \pm 36.8$ m at 99.9% tolerance. Then, the optical model performs a coregistration between a scene sensed by Guardian and a reference image retrieved from the Landsat catalogue. Images from the catalogue were filtered based on the output of the physical model and the sensed scene’s timestamp. To register the images, computer vision methods were implemented to extract keypoints (ORB) and to perform image matching. Once a valid coregistration was found, the appropriate transformation was added to the sensed scene’s metadata, and deemed ready for further downstream processing. The functionality of the optical model was measured for a test dataset in terms of the Root Mean Squared Error (RMSE) and the Mean Corner Error (MCE). This model resulted in a 77% success rate, which could be increased up to 8% through minimal operator input. Lastly, a series of further development points were presented to indicate future possible improvements of the geolocator.

This thesis is structured as follows. Section I provides a background to the different topics of this thesis and the reasoning behind this project. Section II offers an overview of related past work on satellite imagery processing, geographical information systems, orbital mechanics, and cloud detection. Sections III and IV present the methodology followed for the physical and optical models, respectively. Section V elaborates on the methodology behind the error quantification for Sections III and IV. Section VI presents the test cases used to evaluate the techniques suggested and the results, and Section VII comments on these, as well as other external topics that may relate to the thesis on a non-technical level, like the social impact of geoprocessors. Lastly, Section VIII provides a summary of the thesis and its results, and Section IX touches on the possible refining and complementary developments of the presented work.

A. Remote Sensing

Remote sensing was historically defined as “the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by the device that is not in contact with the object [...] under investigation” [1]. Remote sensing is present in our daily life in endless formats; your eyes, the infrared camera that your phone uses to recognize you, the bell that indicates someone entered a store, doors that automatically open for you, traffic cameras, along with other examples. However, in a contemporary context, the meaning of remote sensing has evolved exclusively to that of data sensing through the use of high-altitude platforms, whether they may be balloons, aircraft or spacecraft [2].

Space-based remote sensing has grown hand-in-hand with the exploration of space; public programs such as SPOT, Landsat, and Copernicus have enabled the general public to access current, global imagery of Earth, and private ventures such as Planet, Maxar, ICEYE, Aistech Space, along with many other companies, represent the expansion of the commercial uses of remote sensing. The wide accessibility of satellite imagery is a fundamental part of this thesis, as it is used as a reference to prove the performance of the algorithms presented.

Fig. 1. Structure of the Customizable Aistech Geoprocessing Engine.
B. Geographic Information Systems and Geoprocessors

Space-based platforms have served as a constant source of information about Earth’s resources and geographical characteristics for decades. Through thousands of sensors, data pertaining Earth’s land and water masses, as well as the atmosphere, is constantly measured and divulged. This data is compiled into computer-based databases of spatial information, denominated Geographical Information Systems (GIS). [3]

GIS were first conceived in the 1850s during a cholera outbreak in the English city of Newcastle, where a doctor used what now is spatial analysis to pinpoint the source of the outbreak at a contaminated water pump. Over the following century, there was not much development in GIS, as mapping was paper-based. In the 1950s, GIS started gaining relevance as maps aided decision-making processes for land development, traffic routing, and the location of prominent features. With the rise of computation since the 1960s, GIS have grown exponentially in their capabilities, and are now commonly used in almost every industry. [4]

GIS are systems capable of mapping and analyzing geographic data to simplify the diverse and extensive amount of information that can be connected to a location, e.g. soil composition, soil hardness, distance to closest parks, solar radiation per building, surface composition, air quality, surface temperature, etc. GIS are a way to more efficiently map and analyze data and establish geographical correlations, bettering information management and decision making. A geoprocessor is a set of tools that allow for the addition and manipulation of data in these databases. [5]

This way, a geoprocessor is essential to the processing of satellite imagery. They prepare raw data as captured by the sensor into a product that can be distributed to users. Similarly, it is not uncommon to integrate data sensed by different sensors to improve the final product obtained from its own spacecraft’s observations [6]; especially between the Sentinel and Landsat spacecraft families [7–9].

A geoprocessor can have any level of complexity, including tools such as a scene and layer visualization, geolocation, geometric and radiometric correction, orthorectification, scene mosaicing, scene sharpening, data extraction and management, and so forth. CAGE is a geolocating algorithm that intends to lay the foundation for a proprietary geoprocessor for Aistech Space.

C. Infrared Imagery

Infrared imagery offers a range of information that visible imagery is not able to provide. This region of the electromagnetic spectrum opens up a myriad of applications such as smoke and fire source tracking, the study of cloud structures, the survey of vegetation health, determination of snow cover and ice extent, surface classification, albedo determination, ocean chlorophyll, and aerosol properties, among many others. [6, 10, 11]

Even though infrared sensors were some of the first to go up to space in the 1960s [11], it was not until 1991 that the first Earth-observation infrared telescope was deployed; the Upper Atmosphere Research Satellite mission investigated, among other things, the concentration of aerosol particles and temperature profiles of the atmosphere through the observation of the infrared spectrum below 12 µm [12].

While this range comprises a relatively small subregion of the infrared region of the electromagnetic spectrum, as can be seen in Figure 2. The far-infrared (FIR) region, composed by those frequencies in the realm of 15 to 1 000 µm, allows for less radiation to pass the atmosphere, and is hence not commonly used in Earth-observation science; Figure 3 presents the atmospheric transmission through different wavelengths. Hence, remote sensing in the infrared spectrum is mainly performed through frequencies lower than 15 µm. The thermal infrared (TIR) region is often referred to as long-wavelength infrared (LWIR), the near-infrared (NIR) region is often shortened to the 0.7 µm to 1.4 µm range, and the 1.4 µm to 3 µm range is referred to as short-wavelength infrared (SWIR). [1, 10, 13]

Commercial observation of the Earth is a market that has grown alongside the rise of CubeSats, which opened up the door for non-governmental programs to access space at a much lower cost and accelerated the miniaturization of space components. Specifically, Earth observation based on TIR bands is currently growing, as this smaller market’s observation technology has historically been developed at a lower pace than technologies based on the observation of the visible spectral bands (VIS). Aistech Space is one of few companies developing state-of-the-art, multispectral, space-based, remote sensing platforms (albeit focused on thermal imaging); it shares the market with active constellations being developed by Orora Technologies and ConstellIR, and upcoming projects by Hydrosat and the Albedo Space Corporation.

![Fig. 2. The infrared region of the electromagnetic spectrum. (credit: NASA)](image)

![Fig. 3. Wavelengths observed by different Landsat sensors and atmospheric transmission of wavelengths. (credit: NASA)](image)
D. Aistech Space and Thesis Background

Aistech Space is a multinational geospatial intelligence company based in Spain that utilizes satellite imagery to provide a steady stream of medium-resolution imagery of Earth’s resources to aid its clients’ decision-making processes and augment quality of life on Earth. Aistech Space has been developing its own Earth observation constellation of small satellites since its establishment in 2015. Currently it operates two satellites, which carry demonstrator payloads for the aircraft- and ship-tracking Automatic Dependent Surveillance Broadcast system and the Internet of Things. Later this year, Aistech Space will deploy its first thermally-enabled satellite, Guardian. Guardian will be launched into a Sun-Synchronous Orbit (SSO) and carry a state-of-the-art multispectral telescope (MST) that will empower Aistech Space to expand in its fire and water applications: crop monitoring and risk analysis, land use assessment, algae and fish farm monitoring, and deforestation and fire hazard evaluation, among others.

As Guardian is Aistech Space’s first telescope-bearing satellite, the processing platform that takes the scene captured by the telescope in its most basic state, also called a Level-0 processing level, and transforms it into its final product is still under development. The commissioning of this project roots in this need for the determination of the geolocation of Guardian’s imagery, which is considered to belong to a the first processing step for a Landsat scene [14], taking in the raw (Level-0) imagery and generating Level-1 products. Levels of processing are not standardized, and hence defined differently and to diverse levels of detail [14, 15].

While there are commercial geoprocesors for purchase that meet the expected operational and computational requirements of Guardian, there is a desire at Aistech Space for an in-house solution. This is motivated by two arguments: a financial one and a technological one. On the one hand, the companies approached by Aistech Space that offer comprehensive, on-demand, Level-0 scene geoprocessing services offered slightly different services that, in essence, surpassed CAGE’s geometric correction capabilities by also performing band alignment and radiometric and terrain corrections, and more; their prices ranged in the order of $50 000 annual subscription costs for the processing of imagery generated during nominal operations, or up to $150 000 for a stand-alone software suite to be operated from Aistech Space’s premises. While Guardian would initially only generate enough imagery for the cost to be significantly lower, these externally-provided solutions represent a substantial financial commitment in the long-term strategy. For this, Aistech Space seeks the development of an proprietary solution as an investment for its own future missions, as a multi-year development of a geoprocessor could be argued to be financially and technologically more beneficial.

This thesis pertains specifically to the development of a geolocator for CAGE. Its main goal is to present and demonstrate a method to determine a geographic location for a satellite image using telemetry data and the complement of this through image processing techniques; the first part consists of a physical model, and the second of the matching of the image captured, in its most unprocessed state, with reference imagery captured and processed by other space-based platforms.

E. The Customizable Aistech Geoprocessing Engine

CAGE is a geoprocessing unit for the satellite imagery to be received from the Guardian satellite at Aistech Space. Within it, the first tool developed was the one presented in this thesis: the geolocator. CAGE was designed as a modular system to allow for new methodologies and improvements to be easily added through future work; this is motivated by previous work [16] making an emphasis on the efficacy of multi-stage processors.

To build this geolocator, three main sections were considered:

- an orbital mechanics section, where a physical model was used to provide an initial estimation of a scene’s geolocation;
- an image processing section, where an optical model used reference imagery generated by alternative satellite platforms more precisely determine the geolocation approximated by the physical model;
- an uncertainty assessment section, where the performance of the methodologies was presented and their accuracy and precision quantified. This established the efficiency of CAGE.

Firstly, an approximation of the geolocation was determined utilizing a physical model; knowing the location and orientation of the satellite in space from the telemetry, and finding the intersection between its optical axis and Earth’s surface, a geolocation was approximated through orbital geometry. The location and orientation of the spacecraft are downlinked as part of the telemetry and captured scene’s metadata.

Secondly, a more precise geolocation was obtained by comparing the sensed scene to imagery from another space-based platform taken in the temporal and spatial proximity to the scene sensed by Guardian. The physical model served as a stepping stone, significantly reducing the computational requirements of the present method, as an accurate (although not necessarily precise) approximation of the geolocation allowed for a reduced assessment of Earth imagery. This optical model determined the final geolocation of a scene sensed by Guardian by comparing it to a reference imagery though image processing techniques.

Lastly, the uncertainty of CAGE was gauged. The main factors considered in the uncertainty of the physical model were the accuracy of Guardian’s timekeeping, Global Positioning System (GPS) location, and the attitude error. This uncertainty had a direct effect on the selection and use of reference imagery. Then, the main contributing factors to the uncertainty of the optical model were considered to be inherent in the algorithm itself. The detection and pairing of non-ideally-distributed keypoints through the imagery...
without operator aid was evaluated in terms of the mean distortion of pixels relative to their expected geolocation. This analysis assumed the uncertainty sources to have a uniform distribution and that Earth is an ellipsoid.

F. Guardian and the Multispectral Telescope

The MST is hosted on a 6U satellite platform provided and assembled by OrbAstro. This is Aistech Space’s third satellite, Guardian, which can be seen in Figure 4.

Aistech Space’s MST is medium-resolution telescope developed by Aistech Space capable of observing five spectral bands of the reflected solar region: visible green, red, and blue, NIR, and LWIR; the specifications of the MST are presented in Table I.

<table>
<thead>
<tr>
<th>Band</th>
<th>Wavelength (µm)</th>
<th>Spatial Resolution (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIS (Blue)</td>
<td>0.40 – 0.55</td>
<td>10</td>
</tr>
<tr>
<td>VIS (Green)</td>
<td>0.48 – 0.58</td>
<td>10</td>
</tr>
<tr>
<td>VIS (Red)</td>
<td>0.58 – 0.70</td>
<td>10</td>
</tr>
<tr>
<td>NIR</td>
<td>0.70 – 0.90</td>
<td>10</td>
</tr>
<tr>
<td>LWIR</td>
<td>8.00 – 11.2</td>
<td>75</td>
</tr>
</tbody>
</table>

The MST is expected to generate 120 scenes per day upon the onset of nominal operations, totalling to almost 44,000 scenes in the first year. This number is expected to increase as the system’s functionality is improved and overlapping margins are decreased.

G. Project Requirements

The operational requirements of CAGE were established in conversation with the team at Aistech Space. Firstly, CAGE shall be computationally efficient enough to process hundreds of scenes per day. Second, CAGE shall be automated and, preferably, free from operator input. Next, CAGE shall only use bands present on all sensors available on Guardian. Lastly, the precision and accuracy of the algorithm was expected to yield a maximum RMSE of one pixel on the final geolocation of scenes processed. As pixel size is the precision of sampling, the geographical precision of the geolocation must be within this margin to be considered valid; this is a standard expressed both in industry and literature [6].

II. Literature Study

Academic literature and scientific divulgation materials served as the main source of information on remote sensing and satellite image processing [1, 3, 6, 11, 17], infrared physics [10, 13] and cloud spectroscopy [18]. Specifications of the main public Earth observation satellites was used to consider the available sources of reference imagery [7, 19–25] and to provide an overview of the most suitable options [15, 19, 20].

The background knowledge required to build a physical model for geolocation determination required of a deeper understanding of the coordinate systems available and the transformations between them [26–28], as well as the geometrical relations between an ellipsoid model of Earth and a spacecraft orbiting it [29, 30]. The location of a spacecraft in space called for technical literature on the tracking of spacecraft in orbit and the determination of location from publicly available tracking data [31–33].

Since the publication of remotely sensed data by the major Earth observation programs in the late 1990s and early 2000s, a wide variety of work has been performed in aims of advancing the ability of automate the processing and interpretation of this data. Considering that the Copernicus database alone uploads the majority of the 16 TB of imagery and related data generated every day by the Sentinel satellites [34], it comes as no surprise that ancillary data is commonly used in the determination and improvement of geolocation algorithms [35, 36].

There is a vast variety of algorithms that may be used in feature detection [37–41] that have been thoroughly benchmarked against one another [39, 42, 43]. An alternative feature detection methodology has successfully been implemented in the orthorectification process for satellite imagery of similar dimensions to those of Guardian [44], though it must be noted that the yielded accuracy was less than required for Guardian, and the computational requirements considerably higher than the algorithm presented here. Additionally, complementary methods to those that extract and match imagery features are commonly combined to increase the accuracy and precision of the relationships established between two scenes such as Lowe’s Ratio [38], adaptive thresholding [45] or fitting models [46–48].

A key part of satellite image processing is the detection of clouds, which has proven to be one of the biggest challenges in the field. An abundance of methodologies have been published regarding cloud and cloud shadow detection, although each is tailored to the sensors available on the target satellite. This yields a series of algorithms that change approach depending on the sensor’s bands and resolution; as methods that work with low- or high-resolution imagery perform poorly for medium-resolution [49]. For this, a wide variety of approaches, not all of which aligned with the MST’s specifications, were surveyed.

The use of thermal infrared imagery and spectral bands similar to those of Guardian’s is exemplified by [50], which presents a cloud detection algorithm of Landsat imagery.
though the analysis and use of the VIS-Red, NIR, and two TIR bands. Through a statistical analysis of each kind of land cover, a distribution of surface reflectance values was approximated and used to determine the presence of thin and thick clouds; thresholding of surface reflectance, brightness temperature, and derived indexes such as the Ratio Vegetation Index and the Normalized Differenced Vegetation Index are used to calculate a Cloud Detection Index. The thresholds are manually determined through human-machine interaction, biasing the algorithm and limiting its operational envelope to certain imaging times and regions.

Moreover, it is worth noting the concept introduced in [51], where Poisson matting is used to delineate a cloud’s area in a TIR scene. This unique approach presents an parametrized, edge-detecting algorithm that would place the focus of cloud detection in infrared imagery, rather than visible imagery. Similarly, [52] lays out a cloud and cloud shadow detection methodology based on Brightness Temperature differences derived from TIR and FIR data. However, as infrared-based remote sensing is a less mature field than visible in terms of active space-based sensors and literature, methodologies that use an alternative set of spectral bands must be considered.

In [53], a threshold segmentation approach based on the spectral features of three VIS bands and one NIR band is introduced. A guided filtering is performed by means of derived indexes to produce a more precise cloud mask, rather than the buffer zone implemented by other algorithms. This algorithm yields an average efficiency of 97% in cloud and cloud shadow detection. Similarly, [54] presents a study of the efficiency of Otsu’s method [55], an adaptive thresholding technique that relies on the distribution of pixel values of a scene, when applied to cloud detection. Reference [56] proposes a scheme to detect clouds through a simplified procedure where the ratio of radiance values in the VIS and NIR between a cloudless reference scene and the target scene is used to determine the presence of clouds. Then, [57] presents a cloud and shadow masking algorithm that uses the VIS-Red and NIR bands of the MODIS instrument aboard the National Aeronautics and Space Administration’s (NASA) Terra/Aqua satellites and image processing techniques to distinguish between cloudy and cloudless regions. Here, rather than exclusively assessing the pixel data in each band, cloud shadows were detected by comparing darker regions with the shape of detected clouds.

Furthermore, a notable methodology is presented in [49], where a cloud detection algorithm for 10 m resolution imagery that makes use of bands in the VIS-Green, VIS-Red, NIR, and SWIR spectral bands is presented. While only the first three are shared with Guardian, the methodology was deemed relevant to the development of cloud detection software for its similarity to the Guardian-compatible approach presented in [50]. This lack of thermal band poses different challenges for the methodology, which make use of the bands to classify pixels as water, vegetation, cloud or cloud shadow. This algorithm establishes thresholds through a set of training images and requires of human-machine interaction. A 50 m buffer is added to each cloud pixel as a conservative cloud detection measure. Reference [16] introduces a neural network cloud detection algorithm for 30 m resolution Landsat imagery that uses the VIS, NIR, and SWIR bands and demonstrates how spatially-aware post-processing increases classification accuracy.

Independent cloud and cloud shadow detection research has been performed through a diversity of methodologies, from the use of image processing techniques [51, 53, 54, 57] to the use of radiance and top-of-atmosphere (ToA) reflectance readings [49, 50, 52, 56, 57], among others.

The majority of the presented work makes use of ancillary satellite imagery for testing and verification that allows for a secondary classification. Algorithms used by Earth Observation programs, especially those most developed, can be classified by their use of:

- neural networks and machine learning: SPARCS [16], CD-FCNN, s2cloudless, and InterSSIM, and the algorithms presented in [54] and [58].
- multispectral imagery: ATCOR, Fmask, FORCE, IdePix, LaSRC, MAJA, and Sen2Cor (The European Space Agency’s (ESA) cloud detection algorithm for Sentinel-2), and the algorithms presented in [49, 50, 57].
- multitemporal imagery: InterSSIM, MAJA, and the algorithm presented in [35].

From the aforementioned algorithms, those composing the leading algorithms, used partly by Earth Observation programs, are introduced and benchmarked in [25]. Each algorithm presents its own strengths and weaknesses, but overall accuracy ranges from 75% to 95%, where thin clouds were the hardest cloud type to correctly detect and categorize.

It must be noted that a significant portion of the established cloud detection methodology requires of manual threshold settings for different terrains and some degree of human-machine interaction [49, 54, 57].

**III. ORBITAL MECHANICS**

This section covers the methodology behind the physical model used to perform the initial approximation of a scene’s geolocation. Here, from a scene’s timestamp and the satellite’s GPS location and attitude at said time, which are derived from the telemetry, the coordinates on Earth’s surface of the region that the MST captured were determined.

The Guardian satellite will orbit Earth 500 km over its surface at around 7.62 km/s. For a geolocation to be considered precise enough, its RMSE must be smaller than a pixel’s size; in the case of MST, this would be at most 10 m. Bearing in mind the Attitude Determination and Control Subsystem (ADCS) error alone, which is an order of magnitude larger than the pixel size of Guardian, it was expected for this method to not be precise enough.

Nonetheless, the use of a physical model was considered precise enough to combine with other methodologies [17, 29] since, even with an imprecision in the order of hundreds of meters, it would significantly reduce computation time; being able to approximate the geolocation to within a kilometer of its true value would allow to reduce the search for reference
imagery in Section IV to a single scene, instead of performing a wider search through imagery available in the area within line-of-sight of the satellite. Considering the extensive amount of newly sensed data uploaded everyday to databases like Copernicus, where single scenes can be comprised of over 1 GB of data, and that Guardian is estimated to produce up to 120 scenes per day, the search for imagery must be constrained to minimize the computation time and hardware requirements of the geolocator. Therefore, the physical model presented was used to provide a first approximation of a scene’s geolocation, which later was bettered through an image processing algorithm.

The layout of the physical model follows the structure of previous work [29]: determining the location of the spacecraft, its optical vector, and its intersection with Earth, which was approximated as an ellipsoid. Hence, the estimation of the geolocation presented is a purely geometrical approach. Knowing the position and attitude of Guardian, a location and vector expressing the instantaneous field of view of the MST was expressed, and the point where this vector intersected the Earth ellipsoid was determined. The error resultant from this algorithm was rooted in the uncertainty of the GPS location, Guardian’s watchdog and attitude, and the assumption that Earth is an ellipsoid.

In this section, the physical model was introduced. Firstly, a background on coordinate systems was provided in Section III-A. With this foundation, the determination of a spacecraft in space was presented in Section III-B. Next, the orientation of the MST was determined in Section III-C and its intercept with Earth was resolved in Section III-D, both of which follow the layout of previous work [29, 30]. The structure of the physical model is visualized through Figure 5.

A. Coordinate Systems

Earth-Centered Inertial (ECI) or Earth-Centered Earth-Fixed (ECEF) coordinate systems, both of which are Cartesian coordinate systems, were used to describe the geometry of the physical model. The ECI is a non-rotating coordinate system, fixed in celestial bodies. Specifically, its origin is fixed at Earth’s center of mass, and:

- the x-axis points towards the first point of the Aries constellation (the Vernal Equinox).
- the y-axis points 90° eastward of the x-axis, completing the right-handed frame of reference.
- the z-axis points North, along Earth’s rotation axis.

The reference frame used within ECI is the True Equator, Mean Equinox (TEME), which is a common frame of reference in spacecraft dynamics. It is the framework in which the position and velocity vectors are expressed when a Two-Line Element (TLE) is propagated through an orbital model such as the Simplified General Perturbations #4 (SGP4), an analytical method for determining the ephemerides of an Earth-orbiting spacecraft based on general perturbation theory [31–33].

Then, the ECEF is a rotating coordinate system, fixed to the Earth. Similarly to the ECI, its origin is located at Earth’s center of mass, and its axis, shown along with those of the ECI coordinate system in Figure 6, are those such that:

- the x-axis is in the equatorial plane and crosses the prime meridian (the Greenwich Meridian).
- the y-axis points 90° eastward of the x-axis, completing the right-handed frame of reference.
- the z-axis points North, along Earth’s rotation axis.

The ECEF coordinate system has a subset of reference frames specific to the definition of Earth’s surface coordinates. The most relevant here are the World Geodetic System of 1984 (WGS84) and the International Terrestrial Reference Frame (ITRF); both are considered to be equal with an error of a few cm at most. WGS84 is a geodetic reference frame and the global datum used by the GPS system, while the ITRF is a geocentric Cartesian reference frame. WGS84 and ITRF are updated regularly, as the Earth is under constant deformation due to tectonic movements and other factor and therefore not a rigid body. [28]
B. Spatial Location of a Satellite

Two approaches were taken to determine Guardian’s location in space at the time a scene was captured through MST. The main and most straightforward method was to use the GPS coordinates provided by the spacecraft’s own GPS receiver. Alternatively, a location can be derived from a TLE if GPS data is unavailable.

1) Location Derived from Global Positioning System

Data: GPS coordinates are obtained at regular intervals and downlinked in the telemetry with the measurement timestamp. For two GPS measurements \( p_1 \) and \( p_2 \) made sometime \( t_{GPS_1} \) and \( t_{GPS_2} \) before and after the scene is captured, respectively; a linear interpolation was performed such that

\[
p = p_1 + \frac{t_G - t_{GPS_1}}{t_{GPS_2} - t_{GPS_1}}(p_2 - p_1),
\]

where \( p \) and \( t_G \) are the position vector expressed in ITRF and timestamp of Guardian at the time the scene was captured, respectively.

The GPS coordinates were expressed in geodetic coordinates; that is, latitude, longitude, and altitude. These were easily converted to Cartesian ITRF coordinates through the International Astronomical Union’s software for astronomical computing, which is included in \textit{astropy}, a community-developed Python library for astronomy [59–61]. This transformation is a purely geometrical operation, as explained by Vallado [62] and shown in Figure 7; for a given ellipsoid, any set of geodetic coordinates has a definite set of equivalent ITRF coordinates. The coordinates were interpolated in their Cartesian form.

2) Location Derived from Two-Line Element Sets: The United States Space Surveillance Network tracks every satellite that orbits Earth on a daily basis, generating TLEs that identify and describe each object’s orbit through radar and optical observations. These TLEs are compiled by the North American Aerospace Defense Command (NORAD) and publicly available through Space-Track [31]. Each object is identified through its own NORAD ID.

A TLE is composed of two 69-character lines of data that express the trajectory of a satellite. TLEs can be used together with NORAD’s SGP4 orbit propagator to determine the position and velocity of the associated object at a given time. Their accuracy decreases as the propagation time moves away from the TLE’s epoch, and have an error range in the order of one kilometer for an object in Guardian’s orbit [63]. Further detail on TLEs can be found online, as they are a public, standardized practice [31,33]. The imprecision and rapid decline in accuracy of TLEs led to considering this a secondary methodology, implemented in CAGE as a redundancy measure.

The position and velocity vectors were determined by propagating a TLE to a specific time through SGP4. The conversion between ECI and ECEF coordinates has been well-researched and is implemented in the \textit{astropy} Python library, which is based in Vallado’s work on the United States Department of Defense’s Space-Track project [32].

C. Calculation of the Optical Vector

The determination of the optical vector, the direction in which the MST is pointing, consisted of reference frame transformations and orientation additions. The orientation of the MST relative to the spacecraft, which is a constant value, was added to the orientation of the spacecraft relative to its local reference frame (LRF), which is expressed by the attitude data downlinked within the telemetry. This orientation was transformed to ITRF coordinates for coherence with Guardian’s position vector. The optical vector in ITRF coordinates, \( d_i \), was expressed as

\[
d_i = M A^T d_s,
\]

where \( M \) is the transformation matrix between the nadir-pointing coordinates and ITRF, \( A \) is spacecraft attitude rotation matrix, which indicates the transformation from nadir-pointing coordinates to the the spacecraft LRF (hence the transpose of this matrix was required to go from the spacecraft LRF to nadir-pointing coordinates), and \( d_s \) is the optical vector of the spacecraft in its LRF. The spacecraft attitude rotation matrix \( A \) is a known function of the LRF used.

The spacecraft LRF employed throughout this algorithm is that of Vehicle Velocity, Local Horizontal. In this frame of reference:

- the x-axis points towards the velocity vector. A TLE was required to establish this axis.
- the y-axis points in the negative direction of the orbit normal.
- the z-axis points in the negative radial direction, towards nadir. It is along the same line as the position vector \( p \).

Spacecraft orientation can be expressed in many ways. Two of these were used in this thesis: Direction Cosine Matrices and Euler Parameters. A Direction Cosine Matrix expresses spacecraft orientation through a \( 3 \times 3 \) matrix whose nine components represent the angles between ITRF and the LRF. Then, spacecraft orientation is most popularly represented through quaternions (formally denominated Euler Parameters), as they provide a redundant characterization of attitude without singularities, unlike the intuitive Euler Angles (yaw, pitch, roll). Quaternions are expressed as a 4-element vector, \( \mathbf{q} = (\beta_0, \beta_1, \beta_2, \beta_3) \), composed by the
principal rotation axis and the angle of rotation around it. A thorough introduction to attitude expression was deemed out of the scope of this thesis; the reader can refer to [26] for further information.

The pointing vector of the MST relative to the spacecraft’s LRF, \( \beta_M \), has a fixed value set in the design and assembly of Guardian. The ADCS subsystem provides the spacecraft’s orientation relative to its LRF in quaternions, \( \beta_G \). In accordance with spacecraft dynamics literature [26], these two sets of quaternions can be added up such that

\[
d_s = \begin{bmatrix}
\beta_{s0}^2 + \beta_{s1}^2 - \beta_{s2}^2 - \beta_{s3}^2 \\
2(\beta_{s1} \beta_{s2} - \beta_{s0} \beta_{s3}) \\
2(\beta_{s1} \beta_{s3} + \beta_{s0} \beta_{s2})
\end{bmatrix}
\begin{bmatrix}
\beta_{M0} \\
\beta_{M1} \\
\beta_{M2}
\end{bmatrix} + \begin{bmatrix}
2(\beta_{s1} \beta_{s2} + \beta_{s0} \beta_{s3}) \\
\beta_{s0}^2 - \beta_{s1}^2 + \beta_{s2}^2 - \beta_{s3}^2 \\
2(\beta_{s2} \beta_{s3} - \beta_{s0} \beta_{s1})
\end{bmatrix}
\begin{bmatrix}
\beta_{M0} \\
\beta_{M1} \\
\beta_{M3}
\end{bmatrix} + \begin{bmatrix}
\beta_{s0}^2 - \beta_{s1}^2 - \beta_{s2}^2 + \beta_{s3}^2
\end{bmatrix}
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}
\] (4)

where \( d_s \) is the optical vector; it indicates the pointing of MST relative to the LRF. \( d_s \) was then transformed to Cartesian coordinates in the LRF through a Direction Cosine Matrix following the relation denoted by Equation (4) [26].

\[
M_y = \begin{bmatrix}
t_x/t \\
t_y/t \\
t_z/t
\end{bmatrix}
\] (8)

Lastly, the last component of \( M \) is determined from the other two components as

\[
M_z = M_y \times M_x,
\] (9)

leading to the nadir-to-ITRF transformation matrix

\[
M = \begin{bmatrix}
M_x & M_y & M_z
\end{bmatrix}.
\] (10)

This way, from the position vector \( p \), the velocity vector \( v \), and Guardian’s attitude \( \beta_G \) at a given time \( \tau_G \), the optical vector of the satellite in ITRF, \( d_i \), was determined through Equation (2).

---

**TABLE II**

<table>
<thead>
<tr>
<th>Constant</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_m )</td>
<td>Mean Earth radius</td>
<td>6371 km</td>
</tr>
<tr>
<td>( R_e )</td>
<td>Equatorial Earth radius</td>
<td>6378.137 km</td>
</tr>
<tr>
<td>( f )</td>
<td>Earth ellipsoid flattening factor</td>
<td>1 / 298.25722356</td>
</tr>
</tbody>
</table>

Then, the transformation matrix \( M \) provided the conversion from spacecraft’s LRF to ITRF, and was determined through a vector-based general method [30] that allows for non-nadir pointing of a sensor. While this is a closed-form and exact solution, the assumption of an ellipsoidal Earth induces error into the algorithm; to maintain consistency across the algorithm, WGS84 was the ellipsoid model used throughout this thesis. Previous work [29,30] using this methodology shows that the calculations have an accuracy of, approximately, 0.3 arcseconds for a 700 km orbit. This is equivalent to an error in the order of 1 m. Conservatively, this was assumed to be similar for the 500 km orbit of Guardian, and was considered the expected error when verifying the algorithm’s functionality.

The matrix \( M \) was obtained by determining its three components individually. Knowing the position and velocity vectors \( p \) and \( v \), respectively, of the satellite, the z-component of \( M \) can be determined such that

\[
M_z = \begin{bmatrix}
-\frac{p_y f'}{\sqrt{p_x^2 + f'^2 (p_x^2 + p_y^2)}} \\
-\frac{p_x f'}{\sqrt{p_x^2 + f'^2 (p_x^2 + p_y^2)}} \\
\frac{f'}{\sqrt{p_1^2 + f'^2 (p_x^2 + p_y^2)}}
\end{bmatrix}
\] (5)

where

\[
f' = \frac{R_m (1 - f)^2 + |p| - R_m}{|p|}
\] (6)

and \( R_m \) is the mean Earth radius and \( f \) is the Earth ellipsoid flattening factor. The relevant constants used throughout the thesis are presented in Table II.

Then, the vector perpendicular to the orbital plane, \( T \), was determined through

\[
t = M_z \times v
\] (7)

\[
M_y = \begin{bmatrix}
t_x/t \\
t_y/t \\
t_z/t
\end{bmatrix}
\] (8)

---

Fig. 8. Orbital cross-section showing the nadir vector \( d_i \), pixel location vector \( g \), spacecraft position vector \( p \). The green ellipsoid, of flattening factor \( f \), represents Earth, and the orange ellipsoid represents the satellite’s orbit.
D. Determination of Earth Intercept

Having the position and optical vectors, \( d_i \) and \( p_i \), of the MST, the corresponding intercept with the Earth ellipsoid was determined through geometry. The relation between them is presented in Figure 8. The ground intercept position \( g \) can be expressed as

\[
g = \left( \frac{p_x + D d_{ix}}{p_y + D d_{iy}}, \frac{p_z + D d_{iz}}{p_z + D d_{iz}} \right),
\]

where \( D \) is a scaling factor for the unitary MST optical vector and, effectively, the distance to the ellipsoid. To find \( D \), \( g \) can be equated to the definition of the Earth ellipsoid. An ellipsoid is defined such that

\[
\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.
\]

Equation (14) was solved through the quadratic formula, yielding two results, as the optical vector can intercept the Earth ellipsoid twice. The lowest value represents the first intersect, and was hence chosen as the appropriate solution.

Lastly, the Earth Intercept location was converted to geodetic latitude and longitude, as this format is more intuitive and standard to the expression of locations on Earth’s surface. Since these coordinates are on the ellipsoid, the latitude was calculated per

\[
\theta_d = \arctan \left( \frac{g_z}{(1 - f)\sqrt{g_x^2 + g_y^2}} \right),
\]

and the longitude as

\[
\phi_d = \arctan \left( \frac{g_y}{g_x} \right).
\]

IV. IMAGE PROCESSING & IMAGE MATCHING

The next step in the geolocation determination process was to match the sensed image with a reference image obtained from publicly available ancillary data. This process is visualized in Figure 9 and flows as follows: firstly, the reference image was retrieved from an existing database of satellite imagery; the selection of database and retrieval methodology are discussed in Section IV-A. Each scene was analyzed to find identifying patterns (keypoints) through the algorithm presented in Section IV-B. Next, the sensed image was located within the reference scene through the methodology in Section IV-C; this section covers feature matching and the filtering of invalid matches (outliers). Once the transformation between the sensed and reference scenes has been completed, the final form of the transformation matrix that relates each pixel in the sensed scene with a geographical coordinate is presented in Section IV-D. Additionally, there are two pre-processing stages: image enhancement and cloud detection; these are presented in Section IV-E and IV-F, respectively. Cloud detection is a crucial stage of the pre-processing of satellite imagery.

The 3-dimensionality of satellite imagery and its effects belongs outside the scope of the geolocation of a satellite image and, consequently, this thesis. The correction of the warping and deformation of a satellite image caused by the
Satellite imagery is commonly associated with four kinds of cloud detection methods based on the detection of Ground Control Points (GCP) are commonly used in medium-to-high resolution satellite imagery. The methods involve an alteration of the image to degrade its quality by means of resampling or pixel merging [17]. Scene comparison methods based in the detection of Ground Control Points (GCP) are commonly used in medium-to-high resolution satellites such as Landsat, and yield satisfactory results [6].

### A. Reference Imagery

Space-based imagery has seen an exponential growth over the last decades, leading to a multitude of Earth-imaging programs, both public and private. Given the scope and limitations of this project, only public programs were considered as sources of reference imagery. Within these, there is still a long list of possible sources: Sentinel, SPOT, Pleiades, the 30 missions within NASA's Earth Observing System, and the Indian Space Research Organization’s 17 Earth-observation satellites. Bearing in mind that cloud detection is a relevant step in the analysis of satellite imagery, and knowing that some cloud detection methods consist of statistical models and that the efficiency of these models can vary highly with imagery of different resolutions [49], the desired characteristic of a reference image was that it had similar characteristics to those of a Guardian scene. Satellite imagery is commonly associated with four kinds of resolution:

- **Spatial resolution:** equivalent physical distance on the surface to the pixel size of a sensor’s field of view.
- **Spectral resolution:** internal size and number of wavelength intervals that a sensor can measure.
- **Temporal resolution:** revisit time of a sensor or network of sensors over a specific location.
- **Radiometric resolution:** bit depth of an image, expresses the resolution of intensity values that can be measured.

Spectral and radiometric resolution are endemic to the specifications of the MTS, and not directly reflected on the performance of the algorithm, so they were not considered for the selection of a reference imagery source. Temporal resolution was reflected on the revisit time of the different available platforms; though this could be compared to the revisit time of Guardian and their revisit time relative to one another, as Guardian has not yet been deployed this was deemed outside the present breadth of this thesis.

Given that, at the time of submission of this thesis, Guardian had not yet been launched and its functionality verified, the use of ancillary data was considered as a substitute for the verification of CAGE. Therefore, a second source of satellite imagery was needed.

While the Indian Space Research Organization has, similarly to ESA and NASA, a public portal for satellite imagery, the data in their geoportal covers exclusively the territory of India and neighboring areas [24]. Hence, this was disregarded as a source of reference global imagery.

The most modern satellites of the main public Earth Observation programs are presented in Table III, along with the bands that overlap those of the MTS. From these, only those with thermal infrared bands were considered in aims of radiometric similarity. While the ability to compare thermal reference imagery was not essential to the algorithm introduced in this paper, it was considered that it may prove useful in future work. In the two final candidates, Landsat and the Terra and Aqua sister satellites, the radiometric similarity was considered as sources of reference imagery. Within these, only public programs were considered as sources of reference global imagery.

#### Table III: Satellites with Publicly Accessible Data of Specifications Similar to Guardian [19–23]

<table>
<thead>
<tr>
<th>Satellite</th>
<th>VIS Band Number: Wavelength (μm)</th>
<th>NIR Band Number: Wavelength (μm)</th>
<th>LWIR Band Number: Wavelength (μm)</th>
<th>Scene size (km × km)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Guardian</strong></td>
<td>1: 0.40 – 0.55 [10 m]</td>
<td>4: 0.70 – 0.90 [10 m]</td>
<td>5: 8.00 – 11.2 [75 m]</td>
<td>5 × 7</td>
</tr>
<tr>
<td></td>
<td>2: 0.48 – 0.58 [10 m]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3: 0.58 – 0.70 [10 m]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sentinel-2</strong></td>
<td>1: 0.43 – 0.45 [10 m]</td>
<td>5: 0.70 – 0.71 [20 m]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2: 0.46 – 0.53 [10 m]</td>
<td>6: 0.73 – 0.76 [20 m]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3: 0.55 – 0.59 [10 m]</td>
<td>7: 0.77 – 0.79 [20 m]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4: 0.65 – 0.68 [10 m]</td>
<td>8: 0.78 – 0.89 [10 m]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1: 0.43 – 0.45 [30 m]</td>
<td>8a: 0.85 – 0.88 [20 m]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Landsat 8-9</strong></td>
<td>2: 0.45 – 0.52 [30 m]</td>
<td>5: 0.85 – 0.89 [30 m]</td>
<td>10: 10.6 – 11.2 [30 m]</td>
<td>185 × 180</td>
</tr>
<tr>
<td></td>
<td>3: 0.53 – 0.60 [30 m]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4: 0.63 – 0.68 [30 m]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SPOT 6-7</strong></td>
<td>1: 0.45 – 0.52 [1.5 m]</td>
<td>4: 0.76 – 0.89 [1.5 m]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2: 0.53 – 0.59 [1.5 m]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3: 0.63 – 0.70 [1.5 m]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pleiades</strong></td>
<td>0: 0.45 – 0.53 [2 m]</td>
<td>4: 0.76 – 0.89 [1.5 m]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1: 0.51 – 0.59 [2 m]</td>
<td>3: 0.78 – 0.92 [2 m]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2: 0.62 – 0.70 [2 m]</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Terra/Aqua</strong></td>
<td>1: 0.52 – 0.60 [15 m]</td>
<td>3: 0.76 – 0.86 [15 m]</td>
<td>13: 10.25 – 10.95 [90 m]</td>
<td>60 × 60</td>
</tr>
<tr>
<td></td>
<td>2: 0.63 – 0.69 [15 m]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: spatial resolution of each band expressed in parenthesis next to the respective wavelength.
Digital Numbers (DN) that denote the intensity value of a pixel within a certain spectral band. A raster has an origin in its upper left corner, and carries in its metadata an affine, a transformation matrix that relates each pixel to a global datum. The intensity value of a pixel lays within a range established by the bit depth of the image (e.g. between 0 and 255 for 8-bit imagery). While different sensors may use different bit depths (e.g. Sentinel-2 and Landsat have 12- and 16-bit depth of radiometric resolution, respectively), CAGE resampled all imagery to reduce it to 8-bit depth, as the OpenCV-based keypoint detector did not tolerate other bit depths. In view of core functions requiring exclusively 8-bit depth, this resampling was assumed to not have a strong negative influence in results.

A keypoint is a so-called point of interest in an image; it signals a distinctive feature such as a corner or an edge, where the intensity between a pixel and its neighbors changes abruptly. Keypoints are represented by descriptors, vectors that summarize the fundamental characteristics of a keypoint in a binary string. Keypoint descriptors are designed for robustness against image transformations, e.g. translations, scaling, and rotations.

The development of a keypoint extractor and a keypoint matcher were considered outside the scope of this project, as there is a multitude of readily-available, open-source algorithms. There are two main kinds of keypoint extractors: feature-based and deep-learning-based. [37]

The most prominent feature-based extractors are SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. SIFT was the first keypoint extraction algorithm, presented in 2004 [38], and along with SURF represents the first generation of keypoint extractors. ORB has been presented as an efficient option to SIFT and SURF [39]; a faster extractor than AKAZE for high-resolution imagery, especially in imagery larger than 640 × 480 pixels [42], which is the case; and as an overall preferable choice to all the other feature-based keypoint extractors presented here in terms of ability to detect high quantity of features and computational efficiency [43]. ORB [39] is an extractor that fuses and expands previously established keypoint detection and feature description algorithms, FAST and BRIEF, respectively [40,41]. Another advantage of ORB is that it is free from the licensing limitations of SIFT and SURF, as it is not patented. However, it has been previously observed that different remote sensing platforms may perform better with different feature-based extractors [67].

FAST classifies a pixel as a corner if there is a set of neighboring pixels which are all either brighter (higher DN) than the intensity of the pixel plus a threshold or darker (lower DN) than the intensity of the pixel minus a threshold, as visualized in Figure 10. From the resulting corners, those with the highest variation (i.e. the sharpest corners) are deemed keypoints. In ORB, FAST is combined with additional features to allow for multi-scale (i.e. multi-resolution) keypoint extraction. BRIEF is a feature descriptor that, through binary tests on a smoothed region of interest of an image, returns a signature for any arbitrary keypoint. A keypoint and its respective descriptor are referred to as a GCP throughout this paper.

B. Keypoint Extraction

Satellite imagery is most commonly formatted as a TIFF, a common file format as it allows to express several layers of an image, each of which can represent a band or mask, in addition to metadata. This format is easily manipulated in Python through GIS-dedicated libraries such as OpenCV [64], rasterio [65], and GDAL [66].

A satellite image is represented as a raster (a matrix) of Digital Numbers (DN) that denote the intensity value perceived by each pixel within a certain spectral band.
Fig. 10. Corner detection through FAST. The highlighted pixels are the ones used in determining the presence of a corner. The arc is indicated by the dashed line passes through 12 contiguous pixels which are brighter than \( p \) by more than the threshold. (credit: E. Rosten and T. Drummond [40])

C. Image Matching

Once keypoints and their descriptors were extracted from the sensed and reference image, they were cross-referenced to find possible matches between them, effectively correlating both images. The matching was performed through a brute force approach, where each descriptor in the first image was compared with all descriptors in the second image, and the Hamming distance between them calculated; the two reference descriptors with the closest Hamming distances were assigned as possible matches to the corresponding sensed descriptor. Hamming distance, a common term in information theory, is defined as the number of positions where the corresponding characters in each string are not equal between two strings of the same length; in this scenario, the two strings considered were a pair of binary descriptors. Nonetheless, while a reference descriptor may have the lowest Hamming distance to a sensed descriptor, this does not entail that one is correlated to the other, so the accuracy of matches needs to be ensured.

To reduce the occurrence of invalid matches, two methods were implemented: Lowe’s ratio and Random Sample Consensus (RANSAC). Lowe’s ratio was theorized by David Lowe in 2004, in the same publication that presented the SIFT feature extractor [38]. This ratio presents a methodology to differentiate valid matches, originated in an actual match between two descriptors, from invalid matches, originated in background noise native to the imagery. Lowe’s ratio states that, for the two closest reference descriptors to a sensed descriptor, their distances must present the relation

\[
\frac{d_{\text{closest}}}{d_{\text{next closest}}} < 0.8
\]

(17)

to consider the closest reference descriptor, of distance \( d_{\text{closest}} \), a valid match. This approach, which rejects all matches for which the closest match is not significantly closer than the second-closest match, was proved to eliminate 90% of all invalid matches at the expense of just 5% of valid matches.

RANSAC [46] is a robust, iterative method used for model fitting. It is capable of extracting an approximation of a trend within a dataset that contains outliers. Given a dataset, RANSAC iterates through randomized groups of data points to create models, choosing as optimal that with the largest amount of inliers. Inliers and outliers are data points at a distance from the model lower and greater than an established threshold, respectively. In the context of image processing, RANSAC is used to determine which matched keypoints best describe the correlation between the sensed and reference scenes. As shown in Figure 11, RANSAC yields a set of GCPs that have similar proportions between the sensed \( x_s, y_s \) pixel coordinates and the reference \( x_r, y_r \) pixel coordinates. Once the occurrence of false matches has been minimized, the precision and accuracy of the algorithm were calculated; this is discussed in Section V.

Regarding this comparison of satellite imagery as two distinct observers looking at the same 2-dimensional space, as exemplified in Figure 12, we can perform an approximation of the relations between each sensed and reference pixel through a homography. A homography is a transformation matrix that expresses the translation, rotation, scaling, and shear deformations incurred by an image as the perspective is changed.

The valid matches output by Lowe’s ratio and RANSAC were used to compute the homography, which expressed the transformation from \( x_s, y_s \) to \( x_r, y_r \) as

\[
\alpha \begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} = H \begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix},
\]

(18)

where \( \alpha \) is a scale factor and \( H \) is the transformation matrix [68]. To determine \( H \), consider that a GCP in the sensed
image can be expressed in terms of the corresponding reference GCP’s pixels as

\[ \alpha x_s = \frac{h_{00} x_r + h_{01} y_r + h_{02}}{h_{20} x_r + h_{21} y_r + h_{22}} \]  

(19)

and

\[ \alpha y_s = \frac{h_{10} x_r + h_{11} y_r + h_{12}}{h_{20} x_r + h_{21} y_r + h_{22}} \]  

(20)

which can be expressed as a matrix in the form of Equation (21). Considering that Equation (21) has nine degrees of freedom, assuming that \( H \) is defined to scale provides an additional constraint reduces the system’s unknowns to eight. By expanding Equation (21) to include \( m \) matches, with \( m \geq 4 \), a single finite solution can be obtained. This way, having computed a homography between the sensed and reference image, Equation (22) was used to determine the homography \( H \). If \( m > 4 \), a Linear Least Squares solution was additionally used.

\[
\begin{bmatrix}
  x_r & y_r & 1 & 0 & 0 & 0 & -x_s x_r & -x_s y_r & -x_s \\
  0 & 0 & 0 & x_r & y_r & 1 & -y_s x_s & -y_s y_r & -y_s
\end{bmatrix}
\begin{bmatrix}
  h_{00} \\
  h_{01} \\
  h_{02} \\
  h_{10} \\
  h_{11} \\
  h_{12} \\
  h_{20} \\
  h_{21} \\
  h_{22}
\end{bmatrix}
= \begin{bmatrix}
  0 \\
  0
\end{bmatrix}
\]

(21)

\[
\begin{bmatrix}
  x_{r_1} & y_{r_1} & 1 & 0 & 0 & 0 & -x_{s_1} x_{r_1} & -x_{s_1} y_{r_1} & -x_{s_1} \\
  0 & 0 & 0 & x_{r_1} & y_{r_1} & 1 & -y_{s_1} x_{s_1} & -y_{s_1} y_{r_1} & -y_{s_1} \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  x_{r_n} & y_{r_n} & 1 & 0 & 0 & 0 & -x_{s_n} x_{r_n} & -x_{s_n} y_{r_n} & -x_{s_n} \\
  0 & 0 & 0 & x_{r_n} & y_{r_n} & 1 & -y_{s_n} x_{s_n} & -y_{s_n} y_{r_n} & -y_{s_n}
\end{bmatrix}
\begin{bmatrix}
  h_{00} \\
  h_{01} \\
  h_{02} \\
  h_{10} \\
  h_{11} \\
  h_{12} \\
  h_{20} \\
  h_{21} \\
  h_{22}
\end{bmatrix}
= \begin{bmatrix}
  0 \\
  0 \\
  \vdots \\
  0
\end{bmatrix}
\]

(22)

D. Geolocation Presentation

Reference imagery, being obtained from databases of large Earth-imaging programs, has already been processed and includes a large amount of metadata; specifically, the geolocation of each pixel expressed in the form of an affine and coordinate reference system (CRS).

The most common coordinate system used for global coordinates is the Universal Transverse Mercator projection. This projection divides the Earth in 60 regions, each a six degree longitude-wide wedge with its own central meridian, as visualized in Figure 13. The use of zones aids minimize distortions caused by the mathematical procedure required to express a curved surface onto a two-dimensional surface. These zones are not of exclusive use, but the inaccuracy of the projected distance between points and angle between non-parallel lines increases as the zone’s bounds are exceeded. [27]

A CRS is expressed as an EPSG code, 4-6 digit number that identifies the ellipsoid and zone used within the Universal Transverse Mercator projection. Each CRS presents a series of deformations that allow to represent project a 3-dimensional surface (a scene) in two dimensions (a map). These codes are standardized and regularly updated.

Fig. 13. Section of the Universal Transverse Mercator Zone 12  
(credit: GISGeography)

Affine transformations are transformation matrices that, in Euclidean geometry, can be used to translate, stretch, and rotate a point. They are similar to homographies, but relate pixels to coordinates rather than pixels to pixels; in GIS, the affine transform is used to express a satellite image’s location within a CRS. More accurately, an affine transformation \( A \) expresses the location of the image’s upper
left pixel from the origin of the CRS and the image scale as

$$A = \begin{bmatrix}
S_x \cos(\theta) & -S_x \sin(\theta) & \Delta x \\
S_y \sin(\theta) & S_y \cos(\theta) & \Delta y \\
0 & 0 & 1
\end{bmatrix}, \quad (23)$$

where $S_x$ and $S_y$ are the spatial resolution, and $\Delta x$ and $\Delta y$ the translation from the origin, along the $x$- and $y$-axis, respectively; $\theta$ is the rotation angle. This way, a reference pixel’s position in the CRS was expressed as

$$\begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = A \begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix}, \quad (24)$$

where $x$ and $y$ are the pixel’s Cartesian coordinates in the reference image’s CRS.

Lastly, the reference scene’s affine $A$ was combined with the homography $H$ to create an affine for the sensed scene. The sensed scene’s affine inherited its datum from $A$’s CRS, thus having the same EPSG code. Finally, this affine and corresponding EPSG code were added to the sense scene’s file, completing the geolocation process laid out in this thesis.

---

**E. Image Enhancement for Feature Detection**

The ORB algorithm, used to locate corners in a scene, has customizable parameters to alter the thresholds and description complexity of keypoints, some of which are internally correlated; these aid in tuning the detector and matcher to optimize the processing of imagery. In raw imagery, intensity values often only populate a portion of the available range of DN values (e.g. the raw image in Figure 14 has DN values in the range $[0–60]$ from an available range $[0–255]$). Since each image may present a different range of DN values, the quantity of keypoints obtained with fixed ORB parameters differed greatly. Rather than developing an iterative or analytical methodology that adjusts these thresholds to an optimal value, it was considered to equalize the range of DN values in the imagery to reduce the need for the specification of new parameters for different scenes, as seen in previous work [16]. Precedent work has also shown effective results in the use of automated adaptive thresholds for feature detection [45], and contrast stretching is among image processing techniques used in other methodologies [52, 54]. The difficulty of combining OpenCV-native expressions of keypoint extraction with external ones resulted in the use of image enhancement techniques being favored and, therefore, implemented.

Thus, a linear contrast stretch was performed to both sensed and reference imagery prior to feeding them to CAGE. A linear contrast stretch enhances the contrast of the image proportionally to each DN value, so that the pixels in an image are still equal relative to each other. Figure 14 illustrates the difference in visual perception yielded by this; the larger difference in intensity values allows ORB to extract more keypoints without having to tailor its parameters to each scene.

---

**F. Non-optimal Conditions: Clouds**

Space-based observatories represent one of the most modern ways to image large areas at regular intervals. It presents many benefits, hence its popularity, but it does not come without challenges. In satellite imagery, the presence of clouds strongly decreases the performance of image processing algorithms, as most visible and infrared wavelengths cannot pass through clouds. For this, cloud detection and masking constitutes an essential phase of the pre-processing of satellite imagery. [6, 58]

Clouds, while diverse in types and nature, are generally categorized in remote sensing as either thin or thick. Thick clouds reflect most of the incoming solar radiation (hence being perceived as opaque in visible spectral bands), while thin clouds allow incoming radiation through. This causes for thick clouds to be much brighter than the ground in the visible spectrum. Then, as clouds are at sub-zero temperatures, they absorb infrared radiation, which makes them much darker than the ground in TIR bands. This phenomena is illustrated in Figure 15. [6, 18]

The presence of thick clouds can affect remote sensing in different ways; focusing on computer vision methods, there
are three main effects rooted in the presence of clouds that affect CAGE:

- The opacity of clouds impedes the observation of the ground underneath.
- The notably higher brightness of thick clouds in the visible spectrum creates large differences in intensity values in neighboring pixels, this way skewing the detection of features in a scene.
- A cloud’s shadow produces sharp changes in intensity among neighboring pixels. Most keypoints detected are placed over the cloud’s edges, further skewing the detection of features in a scene.

The largest Earth Observation programs (i.e. the Landsat and Sentinel programs) have sensors on-board exclusively dedicated to cloud detection and categorization. The lower atmosphere absorbs all radiation in certain frequencies (around 1.4 µm), allowing for only clouds to be observed [15, 69]. Even with the usage of these cloud-targeting bands, the cloud detection algorithm used by ESA in Sentinel-2, Sen2Cor, yields significant enough errors for the scientific community to suggest caution when automating cloud detection in operative chains [70], having an overall accuracy of 85%. Moreover, the USGS’s Landsat cloud-detection algorithm, Fmask, yields an estimated 90% accuracy. [25]

The complexity of cloud detection algorithms, their lack of commonality for diverse remote sensing platforms, and the need for additional image processing discouraged the development of a cloud detection algorithm within CAGE, as it can be considered a different project in its own. Nonetheless, the foundations of such an algorithm were considered and presented hereafter.

Guardian’s VIS and TIR bands are useful in the development of a cloud detection and masking algorithm. As shown in Figure 15, thick clouds are brighter than the ground in the visible bands and darker in the thermal infrared bands. Therefore, a simple cloud masking algorithm can be used for general cases of thick clouds, where a pixel can be deemed to (probably) be cloudy if its DN in the visible band is above a given threshold and its DN in the thermal band is below another one. This thresholding can be automated through Otsu’s method, a global thresholding technique, which segments an image’s histogram into two clusters of intensity ranges, hence creating a binary mask [54, 55]. The threshold is determined through the minimization of the weighted variance of these two clusters, which is denoted by

\[
\sigma^2_i(t) = w_1(t) \sigma^2_1(t) + w_2(t) \sigma^2_2(t),
\]

where \( w_1 \) and \( w_2 \) are the probabilities of each cluster and \( t \) is the threshold. The probability of each cluster was calculated through

\[
w_1(t) = \frac{\sum_{i=1}^{t} n_i}{N}
\]

and

\[
w_2(t) = \frac{\sum_{i=t+1}^{N} n_i}{N},
\]

where \( i \) is each DN value in the \([0–I]\) range (i.e. [0–255] for 8-bit imagery), \( n_i \) is the number of pixels with a DN value \( i \), and \( N \) is the total number of pixels.

This first approximation of a cloud mask was complemented by means of a buffer; for every cloudy pixel, all those pixels within a certain radius are also masked as an overestimation of the presence of clouds to reduce the occurrence of false negatives. Buffering is common in other cloud masking algorithms; eight of ten algorithms compared in [25] have a predetermined buffer of up to 300 m. Figure 16 shows the functionality of this basic cloud mask in covering thick clouds; the black pixels are masked, and hence ignored when the feature extraction module processes the scene.

Nonetheless, there are several additional layers of complexity to be assessed prior to this cloud detection algorithm being considered suitable for use in satellite imagery analysis: the overestimation of clouds needs to be minimized, snow differentiated from clouds, and cirrus (thin) clouds detected.

The NIR band could prove useful for snow-cloud differentiation. Ice crystals present in the clouds are smaller than snow grains, they absorb notably less radiation in wavelengths in the NIR-SWIR range [6]. Cirrus detection poses the biggest challenge, as cirrus clouds absorbs thermal
radiation but only a minimal amount of visible electromagnetic radiation.

Furthermore, to refine the cloud detection algorithm, processing of the DN values is required. Most of the algorithms presented for other remote sensing platforms make use of ToA reflectance values, Brightness Temperature or other quantities derived from the DN values perceived by the telescope [49, 50, 52, 56, 57]. To obtain a ToA reflectance value, the DN value must be corrected to express radiance, \( L_\lambda \), such that

\[
L_\lambda = g \, \text{DN} + b,
\]

where \( g \) and \( b \) are radiometric coefficients for the gain and bias of the corresponding band, respectively. Then, radiance is converted to ToA reflectance, \( \rho_\lambda \), such that

\[
\rho_\lambda = \frac{\pi L_\lambda d^2}{E_\lambda \cos(\theta_s)},
\]

where \( d \) is the Earth-Sun distance in Astronomical Units, \( E_\lambda \) is the mean solar irradiance in the spectral band, and \( \theta_s \) is the solar zenith angle. These radiometric and geometric coefficients are needed to process Level-0 satellite imagery, and are either obtained through look-up tables (i.e. \( d \) and \( E_\lambda \)) or from the calibration of the telescope (i.e. \( g \) and \( b \)). As Guardian has not yet been calibrated in orbit, the development of a cloud mask is further proved to be outside the purview of this thesis. [69]

V. UNCERTAINTY ASSESSMENT

CAGE was designed as a multi-stage approach, designed so that each phase improves the margin of the one before, in addition to allowing for further processing to be seamlessly included in the geoprocessor.

In this section, the determination of absolute uncertainty in the physical model and relative uncertainty between the sensed and reference scenes in the optical model is presented.

A. Analytical Uncertainty of the Physical Model

The disturbances within the physical model originate in the accuracy and precision of the GPS location, ADCS subsystem, Guardian’s watchdog, and the assumption that Earth is an ellipsoid.

The global average GPS range error is, per its standards, within the order of 2 m for healthy signals [71–73]. As time is provided and expected to be updated continuously for accuracy, the limitation is set by Guardian’s software; this timekeeping error is estimated by the manufacturer, OrbAstro, to be in the order of 2 ms. Moreover, the positional error within 24 hours of the ephemerides for the TLE of a spacecraft in an orbit similar of Guardian is of 1 km [63]. The ADCS error was, per OrbAstro’s specifications, within 0.01°; when projected onto Earth’s surface from Guardian’s orbit, this is equivalent to a linear error of, approximately, 90 m. Lastly, the error intrinsic to the assumption that Earth is an ellipsoid (that of the WGS84 model, more precisely) is found within the algorithm itself and was quantified to be in the order of 1 m [29, 30]. These uncertainties are summarized in Table IV.

<table>
<thead>
<tr>
<th>Uncertainty Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS location</td>
<td>2 m</td>
</tr>
<tr>
<td>ADCS pointing knowledge</td>
<td>0.01°</td>
</tr>
<tr>
<td>Watchdog</td>
<td>2 ms</td>
</tr>
<tr>
<td>Earth Ellipsoid</td>
<td>1 m</td>
</tr>
</tbody>
</table>

A preliminary analysis of worst-case scenario assumptions, where all errors add linearly, yields an uncertainty of a horizontal 106 m radius from the scene’s expected center on Earth’s surface. To better characterize the effect of these uncorrelated uncertainty sources in the precision of the physical model, a Monte Carlo simulation was performed. Monte Carlo simulations are statistical models to solve mathematical problems through random sampling; knowing the range of input values and their distributions, an expected outcome can be determined.

In this scenario, as none of the uncertainty distributions are defined in their respective sources, these were conservatively approximated as uniform. A total of six random numbers are required to describe the set of random parameters, \( X \), for each sample taken:

- three were used for the GPS uncertainty, as it can be any point of a 3-dimensional sphere of 2 m radius. One value indicated the error vector’s magnitude and the other two determined the direction in spherical coordinates. The azimuthal and polar angles were arbitrarily taken according to the most common notation; that is, with respect to the \( xz \)-plane and \( z \)-axis, respectively. The assumption of a uniform distribution allows this.

- two more were used for the quaternion uncertainty, which could be any direction within a cone centered at the MST’s optical vector and of 0.01° half-angle. The first indicated the magnitude of the error (up to 0.01°), and the second the rotation around the optical vector.

- the last random number served to simulate the watchdog uncertainty. It was used to determine a value within the ±2 ms range specified.

Upon the determination of a simulated geolocation through the physical model, the spherical distance \( D_{SH} \) between the estimated and absolute coordinates was calculated through the Haversine formula. Equation (32) shows the Haversine formula, where \( \theta_d \) and \( \phi_d \) are the estimated latitude and longitude, and \( \theta_{real} \) and \( \phi_{real} \) the absolute latitude and longitude, respectively.

The results of the simulations performed to study the random set of parameters were used to calculate an expected value \( m_X \) and variance \( s_X^2 \) (whose square root is the standard deviation, \( s_X \)) to characterize the uncertainty of the physical model. These are defined as

\[
m_X = \frac{1}{S} \sum_{i=1}^{S} x_i
\]
B. Experimental Uncertainty of the Optical Model

The precision with which the optical model determined the geolocation of a sensed scene was calculated through the RMSE and MCE between checkpoints after the geolocation homography is performed to the target imagery. The RMSE was calculated by first separating the valid matches output by RANSAC, and separating it into two groups: 80% of the matches are used to determine a homography, while the remaining 20% are used to check the accuracy of the homography. Then, the homography was determined as explained in Section IV-C, and the resulting affine matrix used to convert the sensed points of the matching matches into the predicted reference points. This way, the residual between these predicted reference points, \( x_r, y_r \), and the real reference points \( x_{\text{check}}, y_{\text{check}} \), was used to calculate the RMSE per

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{q} (x_{\text{check}_i} - x_{r_i})^2 + (y_{\text{check}_i} - y_{r_i})^2}{q - 1}},
\]

where \( q \) is the total number of checking points [6]. A diagram of the RMSE calculation process is presented in Figure 17.

In some scenarios, the RMSE can tend to minimal values albeit an erroneous match. To characterize the origin of these cases, an additional, less-conventional parameter was used in the classification of geolocation results during the testing phase: the MCE. The MCE was used to threshold the allowable distance between a sensed scene’s corners calculated geolocation to their real value. Taking the absolute difference between each corner \( x_{rc}, y_{rc} \) and their real location \( x_{\text{real}}, y_{\text{real}} \), and averaging them as

\[
\text{MCE} = \frac{1}{4} \sum_{i=1}^{4} \sqrt{(x_{\text{real}_i} - x_{rc_i})^2 + (y_{\text{real}_i} - y_{rc_i})^2}.
\]

As established in the project requirements, as well as in literature [6], a sensed scene’s RMSE shall be lower than 1 for the geolocation to be considered valid. Moreover, the MCE limit was arbitrarily established to be 500 m. This was considered to be an appropriate limit within which an operator could correct an image in a processing stage subsequent to the ones presented in this thesis; this is in line with the requirements established for this project. It is noted that the deformed geolocation would still be within the reference scene, per the reference scene’s size margin.

VI. RESULTS

The functionality of CAGE was assessed through a set of test cases. The focus of this was to prove the functionality of the algorithm presented and identify any weaknesses it may present to be assessed through future work. To reduce randomness in the comparison of methods, the same seed was used to initialize the random number generator.

A. Test Data

Sentinel-2 imagery was used as sensed imagery, as a placeholder of Guardian imagery. From all of Guardian’s VIS bands, the red band (band 3) was used for testing, as it has been suggested that it presents more detail than the green and blue visible bands by both literature [6] and by Aistech Space’s Geospatial Application Specialist. From the Sentinel-2 bands common to Guardian, band 4 was selected as equivalent, as it overlaps in its entirety with part of Guardian’s VIS-Red band. Sensed imagery was cropped prior to processing, such that its new dimensions would accommodate for the size of a sensed scene (5 × 7 km²). The cropped size of reference scenes was determined by the uncertainty of the physical model. This way, the placeholder sensed imagery would have the same resolution and dimensions as Guardian imagery, and a similar radiometric
Figure 18. Test sensed (left) and reference (right) scenes over Lubbock.

Sample sensed and reference scenes are presented in Figure 18. For absolute geolocation accuracy assessment, the reference imagery was sourced in data from Landsat 8 and 9, which have been historically deemed accurately geolocated [36]. Band 4 of Landsat reference imagery was used, as it overlaps Guardian’s VIS-Red band similarly to Sentinel-2 band 4. Reference imagery was obtained from the EROS database, which procures Landsat imagery. Each reference image was then cropped and centered at the geolocation approximated by the physical model, accounting for the uncertainty of the physical model, a factor of safety, and any rotation of the sensed scene relative to the reference scene. This way, the size of the reference scene is minimized to reduce computation time and increase the likelihood of similar keypoints being extracted, although it must be large enough that the sensed scene should still be located entirely within the reference scene. This is illustrated in Figure 19.

Reference imagery was cropped to $10 \times 10$ km$^2$ scenes for simplicity and given that smaller values, closer to the estimated error, yielded no higher performance while decreasing the available margin. Lastly, as the detection and masking of clouds and clouds shadow were deemed out of the current scope of the thesis, all test imagery used was cloudless.

The test dataset consisted of 10 pairs of Sentinel-2 and Landsat scenes, from which sensed and reference test scenes were extracted, respectively. This way, multiple test cases were generated from each image, totalling 48 test cases. The geographical distribution of all original imagery is illustrated in Figure 20, and summarized and classified in Table V and Table VI. Test scenes covered agricultural, coastal, desert, forest, mountainous, and urban regions, as well as areas that included several of these land types.

### B. Physical Model Results

A Monte Carlo simulation of 5,000 simulations of five samples each was performed. The Monte Carlo simulation provided an expected value $m_X$ of 44.0 m and a standard deviation $s_X$ of 11.2 m; the distribution is shown in Figure 21. The 99.9% tolerance interval of this was $44.0 \pm 36.8$ m; the upper limit of this interval was 80.8 m.

Further accounting for any rotation of the sensed scene relative to the reference image, as visualized in Figure 19,

<table>
<thead>
<tr>
<th>Location</th>
<th>Number of Test Scenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lubbock, Texas, United States</td>
<td>5</td>
</tr>
<tr>
<td>Karachi, Pakistan</td>
<td>5</td>
</tr>
<tr>
<td>Mumbai, India</td>
<td>4</td>
</tr>
<tr>
<td>Gouro, Chad</td>
<td>5</td>
</tr>
<tr>
<td>Melbourne, Australia</td>
<td>7</td>
</tr>
<tr>
<td>La Palma, Spain</td>
<td>2</td>
</tr>
<tr>
<td>Port Elizabeth, South Africa</td>
<td>7</td>
</tr>
<tr>
<td>San Luis Río Colorado, Mexico</td>
<td>4</td>
</tr>
<tr>
<td>Hudson Bay, Canada</td>
<td>5</td>
</tr>
<tr>
<td>Campeche, Mexico</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>48</strong></td>
</tr>
</tbody>
</table>
TABLE VI
TEST SCENE PAIRS PER LAND TYPE

<table>
<thead>
<tr>
<th>Land Type</th>
<th>Number of Test Scenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>6</td>
</tr>
<tr>
<td>Coast</td>
<td>2</td>
</tr>
<tr>
<td>Desert</td>
<td>5</td>
</tr>
<tr>
<td>Forest</td>
<td>2</td>
</tr>
<tr>
<td>Mountain</td>
<td>7</td>
</tr>
<tr>
<td>Urban</td>
<td>5</td>
</tr>
<tr>
<td>Agriculture &amp; mountain</td>
<td>5</td>
</tr>
<tr>
<td>Coast &amp; forest</td>
<td>5</td>
</tr>
<tr>
<td>Coast &amp; mountain</td>
<td>3</td>
</tr>
<tr>
<td>Coast &amp; urban</td>
<td>4</td>
</tr>
<tr>
<td>Desert &amp; mountain</td>
<td>2</td>
</tr>
<tr>
<td>Urban &amp; mountain</td>
<td>2</td>
</tr>
</tbody>
</table>

would make it have an apparent size of up to 8.60 km. Hence, a reference image should be cropped to sides of dimensions no smaller than 8.77 km. The use of variance reduction methods was disregarded, as the presented variance’s magnitude relative to that of a scene’s size is almost negligible. An effort to decrease the variance of the physical model’s uncertainty analysis would not have a noteworthy effect on the rest of the algorithm.

C. Optical Model Results

The effect of cropping the original Landsat reference image to a minimal size such that its size would be as close to the sensed scene’s as possible was tested through a simple iterative process, where only the dimensions of the reference scene were varied. The results of this series of simulations can be found in Table VII.

The optical model requires an absolute minimum of four matches to be able to determine a homography between two scenes, as mentioned in Section IV-C. Additionally, at least two more matches are required to be able to calculate the RMSE. All of the 48 test scenarios successfully completed the geolocation process through the physical and optical model. The performance of CAGE was measured in terms of the RMSE and MCE; Figures 22 and 23 present the relation between these two parameters, the amount of GCPs extracted, and the outcome of the test case, and Figure 24
Fig. 25. True positive result of CAGE. The image, a Lubbock test case, shows a sensed scene (left) matched into the reference scene (right). This test case has RMSE of 0.56 pixels, MCE of 26.2 m, and 189 valid matches.

<table>
<thead>
<tr>
<th>Reference Scene Size (km × km)</th>
<th>Number &amp; Percentage of Valid Matches (%)</th>
<th>Computation time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 × 100</td>
<td>2/48 (4%)</td>
<td>252</td>
</tr>
<tr>
<td>50 × 50</td>
<td>7/48 (15%)</td>
<td>112</td>
</tr>
<tr>
<td>30 × 30</td>
<td>14/48 (30%)</td>
<td>72</td>
</tr>
<tr>
<td>20 × 20</td>
<td>29/48 (60%)</td>
<td>55</td>
</tr>
<tr>
<td>10 × 10</td>
<td>37/48 (77%)</td>
<td>43</td>
</tr>
<tr>
<td>9 × 9</td>
<td>37/48 (77%)</td>
<td>42</td>
</tr>
</tbody>
</table>

Table VII: Effect of Reference Scene Size in the Amount of Valid Matches

illustrates the cumulative density function of true positives and false negatives.

VII. DISCUSSION

The results can be categorized in four different scenarios, according to whether the results were expected (true or false) and if CAGE successfully matched the scenes (positive or negative):

- true positives, where the RMSE and MCE are within the established limits and the scenes are visually a match, Figure 25 is an example of this classification.
- false negatives, where the RMSE or MCE are too large but the scenes are visually a match, Figure 26 shows several examples of this classification.
- true negatives, where the RMSE or MCE are too large and the scenes are visually not a match.
- false positives, where the RMSE and MCE are within the established limits but the scenes are visually not a match.

The occurrence of true positives and false negatives was assessed through the use of the test cases presented in Section VI-A. All of these scene pairs fed to CAGE were expected to match. The overall efficiency of CAGE was defined as how often it correctly matched a sensed and

Fig. 26. False negative results of CAGE. These are test cases from: top: Karachi, RMSE: 0.61, MCE: 1.07 km, 15 valid matches; mid: San Luis Colorado, RMSE: 83.14, MCE: 11.78 km, 13 valid matches; and bottom: Port Elizabeth, RMSE: 1.94, MCE: 6.22 km, 188 valid matches.
reference scene pair. This way, and as presented in Table VII, CAGE has an efficiency of, approximately, 77%.

The misclassification of test cases as true positive was reduced through the introduction of the MCE. There are three subcategories to false negative scenes, as they can fall under this class for different reasons:

- uneven keypoint distribution, which causes even small errors on keypoint location to significantly deform the resulting projection, leading to a MCE ≥ 500 m. These are considered to allow for operator input a posteriori to visually assess and rectify a sensed scene’s geolocation by manually assigning GCPs. This subcategory represents over a third (36%) of all false negatives, and an 8% of all test cases.

- incorrect matching, where an incorrect set of matches output by the filtering methodologies results in an incorrect projection of the sensed scene within the reference scene. The occurrence of these cases represents almost half (46%) of all false negative test cases.

- single-keypoint matching, where most sensed keypoints are related to a single reference keypoint. This last subcategory represents 18% of false negatives and 4% of all test cases and could be mitigated through a spatially-informed filtering algorithm; it has been proven in literature that spatial post-processing can increase accuracy in related image processing stages [16].

Examples of these three subcategories of false negative test cases are reflected in the three test cases in Figure 26, respectively. The distribution of these subcategories can be hinted in Figures 22 and 23, where false negative test cases group at low-RMSE and high-RMSE values in test scenes with a low amount of valid GCPs. Noticing from the observation of Figures 23 and 24 that the population of false negatives is composed by test cases with a relatively low amount of GCPs, it can be determined that the prime solution for reducing the recurrence of false negatives could be to ensure the detection of a higher value of valid matches; 95% of false negatives occur when less than 20 valid GCPs are extracted from a scene, while only 5% of all true positives are classified with less than 20 GCPs. Hence, it can be gathered that future efforts should focus mainly on an amelioration of ORB’s performance. Additionally, the performance of RANSAC was compared with other RANSAC-based methods [48], and minimal improvement of results followed (up to 3%). Though it was acknowledged that some of these alternatives could be implemented instead, a deeper understanding of their methodology was required to better tune ORB and boost the adeptness of this algorithm, and hence deemed part of future improvements to CAGE.

ORB’s inefficiency in keypoint extraction in some of the test cases appears to not be a systematic error. Looking at the distribution of results per land type, summarized in Table VIII and accounted cumulatively in Table IX, it was inferred that the current version of the algorithm performs best for mountainous and desert regions, and below average for urban and forested areas. Through the manual optimization of ORB’s 10+ parameters carried out, it was noted that certain combinations of parameters skewed the results in Table IX to heighten or lower those of different land types. This suggests that spatially-informed post-processing stages could benefit from being informed of the main land types in a scene to tune ORB and optimize the geolocation process.

Moreover, the incidence of true negatives and false positive was studied through a secondary set of test cases, where incorrect pairs of sensed and reference scenes were fed to CAGE. Here, 10 incorrect pairs were generated from the test data; half of them were composed of sensed and reference scenes from different original images and the other half were composed of different test scenes of the same original image (e.g. the sensed image of the first Hudson Bay test case was paired with the reference scene of the second Hudson Bay test case). The combination of RMSE (error assessment on a pixel-to-pixel basis) and the MCE (error assessment in the image-to-image basis) yielded a conservative set of products, resulting in this desired lack of false positives. This result means non-corresponding imagery being wrongly matched by CAGE can be considered an unlikely scenario in the presented experimental layout. Nonetheless, in real-time operations, processing Guardian imagery, the MCE will not be available; this likely increases the risk of false positives.

### A. Socioenvironmental Impact of Geoprocessors and Image Processing

Geoprocessors are central components to satellite imagery analysis; they make use of physical knowledge to determine the precise and accurate location of a satellite image, and
can make use of reference imagery to contrast and complement its results. As a software, geoprocessors have little direct ramifications to society. However, indirectly, their products and the methodology implemented within them have an effect on every aspect of our life, from the use of image processing software for other applications (e.g. Google Images, Google Reverse Image Search, autonomous driving, and many, many others) to the use of global imagery on apps that the general population uses daily (e.g. Maps, Instagram). However, one of the main uses of this technology is that of surveillance and Earth observation. The most visible example of geoprocessors and image processing technologies in today’s society has been shown throughout Russia’s invasion of Ukraine.

Over the last months, hundreds (if not thousands) of images and videos of the conflict have been distributed online every day. Maxar’s high-resolution satellite imagery was pivotal in the documentation of Russian advancements before the invasion began on February 24, 2022; and throughout the ongoing war, georeferenced imagery has played a key role in following the movement of Russia’s troops through different sources and the corroboration of their movements through comparative analysis of this imagery [74,75]. More notably, the comparison of space- and land-based imagery has supported the accusations of Russian war crimes in Bucha, and are expected to play a decisive role in any actions taken by the International Criminal Court [76,77]. In this era of misinformation, the ability to georeference and compare imagery from different sources can be crucial in the verification of controversial imagery.

Furthermore, geoprocessors can entail significant computational requirements for large-scale operations, which in turn can be translated into high energy consumption and the need for dedicated hardware. For this, geoprocessors can carry a significant environmental impact over time as their processes grow in complexity and the amount and size of imagery being fed to them grows with the addition of newer, higher-resolution remote sensing platforms.

Finally, geoprocessing and image processing techniques are, like many other advanced technologies, double-edged swords. This technology can be invasive if applied to someone’s personal data (e.g. any Instagram picture with a characteristic background can, in theory, be geolocated without the user’s knowledge or consent). Espionage and surveillance are just some of the many applications of these technologies, both ethically right and wrong.

VIII. CONCLUSION

CAGE has been proven in its ability to successfully geolocate imagery of characteristics similar to those of Guardian’s imagery with an overall success rate of 77%. Additionally, it was estimated that operator input would aid correct up to 8% of the cases that fail the geolocation process. CAGE has shown a well-balanced ability to geolocate imagery composed of different land types through the use of a physical model and publicly available satellite imagery. Its main weakness is the large occurrence of failed geolocations for imagery from which less than 20 GCPs are extracted; this is the foremost item aspect to be considered for the improvement of the algorithm. Overall, these results provide a positive outlook towards the development of an in-house, considerably cheaper solution to the need for a geoprocessing platform for Guardian imagery.

It must be considered that a large amount of external factors can affect the image geolocation process through the present methodology. The temporal change in terrain and availability of reference data, the optimization of ORB’s parameters for different terrains, the distribution of keypoints through a scene, etc. This algorithm presents the groundwork for targeted optimization of its different sections. Overall, it can be concluded that, while the results are not comparable to the extent of other commercially available geolocation algorithms, CAGE displays a promising and significant step towards an in-house solution for Aistech Space.

Ultimately, CAGE has proven to be a useful tool in satellite imagery analysis, and is expected to perform adequately with imagery generated by the multispectral telescope developed by Aistech Space. Further development of CAGE and testing with its native imagery will enable optimized and innovative additions to the work here presented.

IX. FUTURE WORK

CAGE’s main area of improvement is the feature extractor and matcher, which must be further developed to augment the amount of scenes with at least 20 valid GCPs to diminish the incidence of false negatives. Alternatively, to mitigate the occurrence of false negatives, the expansion of CAGE through a spatial post-processing stage and the inclusion of an operator input should be considered. While the use of RANSAC was compared with other derivations based on the same process [48] and minimal improvement of results was noted, it was considered that the parameters provided to ORB could represent a local optimum for RANSAC, but not the derived algorithms. Further work on the effects and correlations of ORB’s parameters may benefit the determination of local optima for RANSAC-derived algorithms, but may require considerable computational capacity. This could increase the volume of valid GCPs, hence increasing the frequency of the desired, true positive outcomes. Then, the addition of spatial information in later stages of scene processing has been shown to improve accuracy [16], so it should be considered that a geolocation process with additional stages (i.e. spatially-aware homography, manual GCP input) to the two presented in this thesis would further increase the precision and accuracy of the geolocation.

Moreover, the use of ORB as the predetermined keypoint extractor for Guardian should be reassessed once sufficient MST-produced imagery is retrieved, as it has been proven that, while ORB may currently be one of the best-performing keypoint extractors [39,42,43], different remote sensing platforms may perform better with different feature-based extractors [67]. The feature extraction process in CAGE
could be further improved through a shift from the use of image enhancement techniques to automated adaptative thresholding, as previous work has shown promising results [45]. The interfacing between the OpenCV-based keypoint matcher and an external keypoint extractor is expected to incur in considerable compatibility issues, but may yield a significant improvement in the quality of extracted keypoints.

Cloud detection is, to this day, an evolving field where a multitude of algorithms have been developed for years, but are still to be perfected. Hence, a global solution for cloud selection is yet to be developed. Specifically, software-based solutions for reduced data\(^1\) such as that provided by the VIS and TIR bands of the MST are uncommon. A breakthrough in this topic would lead to a drastic increase in the accessibility of small-scale projects to independent analysis of remotely sensed imagery.

Further expansion of CAGE and its functionality would be to contrast Guardian’s scenes with reference imagery from other sources such as the Terra and Aqua satellites, or Sentinel-2 itself; since the latter’s imagery can be obtained with an ESA-procured cloud mask, the lack of TIR band is not critical to the geolocation process. Furthermore, CAGE is ready in its current state to support operator input of GCPs, which could return a significant improvement in the frequency of true positive scene matching. Moreover, a more complex interpolation of GPS and TLE locations, as well as quaternions, rather than a linear interpolation between the two temporally-closest values. Lastly, the inclusion of a Digital Elevation Model in the geolocation process is expected increase the precision of the geolocation [1, 6]. All in all, there is a myriad of improvements that would present a valuable addition to CAGE and Aistech Space’s development of an in-house geoprocessor.

X. ACKNOWLEDGEMENTS

Aistech Space provided an incredible opportunity to spearhead the development of an in-house solution to the geolocation of their upcoming satellite imagery. The author would like to thank the company and its staff for the support he received there. Especially, the author would like to thank David Paredes, who supervised his progress through the project, was always available for discussion and feedback sessions, and would always come into the office whistling a different tune. As well, the author would like to also thank Pablo Iglesias, Aistech Space’s Geospatial Application Specialist, for he was always available to discuss remote sensing, whether it was related to CAGE or not, regularly dispensed constructive feedback, and was just as eager to learn about the geolocation algorithm as the author himself.

Additionally, the author would like to thank professors Michael Swartwout, Ph.D., and John George, Ph.D., for it was their passion for space that made the author pursue a graduate degree in the first place, and Anton Lager, a fellow student and friend, who helped the author with his poor Swedish skills. Finally, the author would like to thank those people who, through late night FaceTime calls and random messages, kept him sane and focused throughout this project, even when he wanted to run away to a remote tropical island instead of completing this thesis.

REFERENCES


\(^1\)reduced relative to the data available in the 10+ bands of larger Earth observation satellites


