

# **Multi-Modality and Multi-Sensor Image Registration for Satellite Images**

Aaron Abela



**Department of Communications and Computer Engineering**

**University of Malta**

**October 2020**

*Supervised by*

*Dr Ing. Reuben A Farrugia and Dr Ing. Gianluca Valentino*

*Submitted in partial fulfilment of the requirements for the degree of M.Sc. in  
Communications and Computer Engineering*



L-Università  
ta' Malta

## **University of Malta Library – Electronic Thesis & Dissertations (ETD) Repository**

The copyright of this thesis/dissertation belongs to the author. The author's rights in respect of this work are as defined by the Copyright Act (Chapter 415) of the Laws of Malta or as modified by any successive legislation.

Users may access this full-text thesis/dissertation and can make use of the information contained in accordance with the Copyright Act provided that the author must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the prior permission of the copyright holder.

### **Copyright Notice**

1. Copyright in text of this dissertation rests with the Author. Copies (by any process) either in full, or of extracts, may be made only in accordance with regulations held by the Library of the University of Malta. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) made in accordance with such instructions may only be made with permission (in writing) of the Author.
2. Ownership of the right over any original intellectual property, which may be contained in or derived from this dissertation, is vested in the University of Malta and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement

## Abstract

Satellite imagery provides information which is fundamental to remote sensing applications. Two of such applications are image registration and fusion of hyperspectral and multispectral imagery. Image registration is a fundamental pre-processing step to image fusion. Conjunctively, in remote sensing little previous work aimed at the registration of satellite imagery with significant scale differences and the registration of multi-modal satellite images. The aim of this work was to do a comprehensive analysis of registration techniques in remote sensing. The SIFT algorithm with different parameter sets was utilised to register thermal-thermal satellite imagery with significant scale differences. The work also examined and compared the use of other feature-based, area-based and optical flow-based techniques for the registration of multi-modal and multi-sensorial satellite imagery. The multi-modal data includes optical satellite imagery from Sentinel-2 and Landsat-8. SAR images from Sentinel-1 and thermal images from Landsat-8 and Sentinel-3. The findings of this study show that the most common type of modality utilised in the image registration of remote sensing data is Optical-Optical and synthetic aperture radar (SAR)-Optical. For the registration of thermal Landsat-8 to Landsat-8 and thermal Landsat-8 to Sentinel-3, the general pattern was that as one upscaled the sensed image, the misregistration and RMSE increased due to a higher scale difference. For the registration of SAR-Optical satellite imagery the overall best performing was SIFT-Flow. For the registration of single modality data, the overall best was SIFT followed by the Enhanced Correlation Coefficient (ECC). For the registration of multi-modal satellite imagery, the overall best was SIFT.

## **Acknowledgements**

I would like to express my sincere gratitude to my supervisors Dr Ing. Reuben A Farrugia and Dr Ing. Gianluca Valentino for their immense support, expert guidance, encouragement and understanding throughout this dissertation.

My sincere thanks also goes to Dr. David Lloyd who supported me along this entire project. I am truly grateful for his help and support whenever required whose assistance aided me immensely in completing this dissertation.

Special thanks also goes to my SAT-FIRE (Satellite data Fusion and Imaging Resolution Enhancement for coastal areas) colleagues, Dr. Anthony Galea and Ms. Ioanna Thoma. Last but not least, I would like to express my gratitude to my family and my friend, Ivan Tonna for their constant encouragement.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem Definition and Motivation . . . . .	1
1.2	Objectives . . . . .	5
1.3	Dissertation Outline . . . . .	5
<b>2</b>	<b>Background</b>	<b>6</b>
2.1	Remote Sensing Fundamentals . . . . .	6
2.1.1	Optical Images . . . . .	8
2.1.2	Thermal Infrared Images . . . . .	8
2.1.3	Synthetic Aperture Radar . . . . .	9
2.2	Processing Levels and Bands . . . . .	10
2.2.1	Sentinel-2 . . . . .	10
2.2.2	Landsat-8 . . . . .	10
2.2.3	Sentinel-3 . . . . .	12
2.3	Registration through Geocoding . . . . .	13
2.4	General Process of Full Image Registration . . . . .	14
2.5	SIFT Algorithm . . . . .	17
2.5.1	Scale Space Extrema Detection . . . . .	17
2.5.2	Local Extrema Detection . . . . .	19
2.5.3	Accurate Keypoint Localisation . . . . .	19
2.5.3.1	Local Extrema Detection . . . . .	19
2.5.3.2	Further Outlier Rejection by Eliminating Edge Re- sponses . . . . .	20
2.5.4	Orientation Assignment . . . . .	21
2.5.5	Local Descriptor Creation . . . . .	22
2.6	Matching Keypoints . . . . .	24
2.7	Elimination of Outliers: RANSAC . . . . .	24

<b>3</b>	<b>Literature Review</b>	<b>27</b>
3.1	Multi-Modal Image Registration for Satellite Imagery . . . . .	27
3.1.1	Area-based Methods . . . . .	28
3.1.2	Feature-based Methods . . . . .	29
3.1.3	Optical Flow-based Methods for Remote Sensing . . . . .	34
3.1.4	Evaluation of Registration Accuracy between Single and Multi- Modal Satellite Imagery . . . . .	35
3.1.4.1	Qualitative Comparisons . . . . .	36
3.1.4.2	Quantitative Comparisons . . . . .	38
<b>4</b>	<b>Investigation</b>	<b>40</b>
4.1	Assessing Multispectral Misregistration for Sentinel-2 . . . . .	40
4.2	Assessing Multispectral Misregistration for Landsat-8 . . . . .	44
4.3	Assessing Multispectral Misregistration between Sentinel-2 and Landsat- 8 . . . . .	47
4.4	Assessing Multispectral Misregistration between Sentinel-3 SLSTR and Landsat-8 TIRS . . . . .	48
<b>5</b>	<b>Methodology</b>	<b>51</b>
5.1	Implementation . . . . .	51
5.2	Preliminary Comparison between Area-based, Feature-based and Optical Flow-based Methods . . . . .	53
5.2.1	Test Case 1: Optical vs Optical . . . . .	53
5.2.2	Test Case 2: Optical vs SAR . . . . .	56
5.2.3	Test Case 3: Day vs Night . . . . .	59
5.3	Proposed Methodology . . . . .	62
5.3.1	Pre-processing: Extraction of Reference and Sensed Images . . . . .	62
5.3.2	Generation of Joined Cloud Mask . . . . .	66
5.3.3	Resampling and Min-Max Normalisation . . . . .	69
5.3.4	Selecting Best Matching Keypoints . . . . .	73
5.3.5	Homography Estimation . . . . .	75
5.3.5.1	Affine Transformation with Rotation, Scale and Trans- lation . . . . .	76
5.3.5.1.1	In-Depth Algebraic Example . . . . .	76
5.3.5.2	Affine Transformation with Translation Only . . . . .	78
5.3.6	Registration . . . . .	79

5.4	Improving the Parameter Search . . . . .	80
5.4.1	Salient Parameters . . . . .	80
5.4.1.1	RANSAC Threshold . . . . .	80
5.4.1.2	Number of Octaves . . . . .	80
5.4.1.3	Sigma . . . . .	81
5.4.1.4	Contrast Threshold . . . . .	83
5.4.1.5	Edge Threshold . . . . .	83
5.4.2	Default Parameters . . . . .	84
5.4.3	Monte Carlo . . . . .	85
5.4.4	Grid Search Method . . . . .	86
<b>6</b>	<b>Results and Evaluation</b>	<b>89</b>
6.1	Registration of Landsat-8 vs Landsat-8 . . . . .	89
6.1.1	Blurring the Sensed Image . . . . .	89
6.1.2	Translating the Sensed Image . . . . .	96
6.1.2.1	Location 1: Mallorca . . . . .	96
6.1.2.2	Location 2: Corfu . . . . .	98
6.1.3	Effect of Cloud Cover . . . . .	101
6.2	Registration of Landsat-8 vs Sentinel-3 . . . . .	106
6.2.1	Downscaling the Landsat-8 Image . . . . .	106
6.2.2	Upscaling the Sentinel-3 Image . . . . .	111
6.3	Registration of Multi-Modal Remote Sensing Data . . . . .	118
6.3.1	RIFT Dataset . . . . .	119
6.3.2	Optical to SAR . . . . .	128
6.3.3	Optical to Thermal . . . . .	131
6.4	Registration of Single Modality Remote Sensing Data . . . . .	133
6.4.1	Optical to Optical . . . . .	133
6.4.2	Thermal to Thermal . . . . .	136
<b>7</b>	<b>Conclusion</b>	<b>139</b>
7.1	Findings . . . . .	139
7.2	Limitations and Future Work . . . . .	143
<b>A</b>	<b>Numeric Example</b>	<b>154</b>



# Glossary

This glossary describes some common terminology used in various chapters of the dissertation concerning the remote sensing domain.

**Attitude** the orientation angle of a satellite with respect to a geographical reference system.

**Band** also known as *channel*. Consists of a segment of the electromagnetic spectrum. Satellite data typically contain multiple bands that collect radiation from various parts of the electromagnetic spectrum.

**Control Point (CP) or Ground Control Point (GCP)** an accurate known location on Earth which is typically used to georeference remote sensing data.

**Nadir** Point on Earth which is directly below a satellite.

**Image Pixel Size** size of the image data in pixels

**Spatial Resolution** the value allocated to each pixel in terms of area on the ground. For instance the thermal bands of Sentinel-3 have a spatial resolution of 1 km [1]. Thus, 1 pixel corresponds to a spatial resolution of 1000 m.

**Spectral Resolution** the utilised bandwidths in the electromagnetic spectrum.

**Temporal Resolution** refers to the observation frequency of the satellite. This is typically defined by the orbit of the satellite.

**Single-modality** originating from the same type of sensor.

**Multi-modality** originating from different types of sensors such as Optical, Thermal, Infrared and radar.

**Multi-sensor** having different spectral and spatial resolutions.

**Multi-Spectral Image** related to spectral resolution. Refers to a remote sensing image that contains data from more than one band.

**Hyperspectral Image** an image created from hundreds or thousands spectral images.

**Swath** the area that the sensor of the satellite can observe. A wider swath allows the satellite a faster revisit time.

**Top of Atmosphere (TOA)** refers to the altitude above the Earth's surface. This is typically around 100 km.

**Bottom of Atmosphere (BOA)** refers to the conversion of top of atmosphere reflectance to surface reflectance.

**Panchromatic** a panchromatic band is formed when an image sensor is sensitive to wavelengths of most of the *visible* spectrum.

**Normalized Difference Vegetation Index (NDVI)** is a graphical indicator that is utilised to analyse remote sensing measurements to assess the state of plant health.

**Synthetic-Aperture Radar (SAR)** is a type of radar that is utilised to from three-dimensional or two-dimensional images of objects such as landscapes.

**Reprojection** the act of transforming a georeferenced satellite image to another projected coordinate system.

**Orthorectification** the process of removing relief/terrain effects and tilt effects. The aim of orthorectification is to geometrically correct an satellite or aerial image such that the scale of the image is uniform.

**Geodesy** the science of understanding and measuring Earth's geometric shape, gravitational field and orientation in space.

# List of Figures

1.1	Remote sensing misregistration caused by orbital error. $I_2$ is the sensed image to be registered to the reference image $I_1$ . $u(x)$ denotes the displacement between the two images [2]. . . . .	3
1.2	Thermal satellite images of different spatial resolutions . . . . .	4
2.1	Electromagnetic spectrum (EMS) [3] . . . . .	6
2.2	Satellite images of the region around Malta International Airport. a) RGB image synthesised from red, green and blue bands from Sentinel-2 with 10 m spatial resolution. b) NDVI image synthesised from the red and NIR bands from Sentinel-2 with 10 m spatial resolution. c) SAR image from Sentinel-1 [4] with 10 m resolution. The normalised reflected amplitude in the vertical polarisation is shown. d) Thermal image from Landsat 8 showing brightness temperature in degrees Celsius with a spatial resolution of 100 m. The horizontal bars in each of the four images correspond to a distance of 1 km. Images shown in a), b) and d) were acquired on 04/07/2018, whereas c) was acquired on 06/07/2018. . . . .	7
2.3	Sentinel-3 SLSTR Levels of Processing [5] . . . . .	12
2.4	Ground Control Point [6] . . . . .	14
2.5	Difference of Gaussian (DoG) [7] . . . . .	18
2.6	Extrema Detection [7] . . . . .	19
2.7	Histogram of Gradient based on Orientation [8] . . . . .	22
2.8	$16 \times 16$ window broken down to into sixteen $4 \times 4$ windows [9] . . . . .	22
2.9	Allocating orientations into an 8-bin histogram [9] . . . . .	23
2.10	The weighted Gaussian filter is multiplied to the magnitude of orientations which results in the distance mentioned. [9] . . . . .	23
2.11	Feature descriptor for one keypoint. [9] . . . . .	23
2.12	Matching SIFT keypoints (including outliers). [10] . . . . .	24

2.13	Step 1 of the RANSAC algorithm. [11]	25
2.14	Step 2 of the RANSAC algorithm. [11]	25
2.15	Step 3 of the RANSAC algorithm. [11]	26
2.16	Regression model result using the Linear (Least Squares) and RANSAC regressors. [11]	26
3.1	Multi-modal sample data utilised in RIFT [12].	33
3.2	Multi-modal matching results [12]. The yellow and red lines refer to the correct matches and outliers, respectively.	34
3.3	Example of an overlay plot. In this example radar and lidar images are registered using the GeFolki algorithm [2]. The image on the left is the result before registration and the image on the right is the result after registration. The radar image is the reference image.	36
3.4	Example of a checkerboard plot. This example is taken from the proposed algorithm, OS-flow [13]. The image shows the misregistration between optical and SAR images.	37
4.1	Plot of concatenated subsets in grayscale to assess qualitatively the misregistration of geocoding in Sentinel-2B with level L1C processing taken on 2018/04/04. Location is the runway of the Malta International airport. Band 2 (Top Left), Band 3 (Top Right), Band 4 (Bottom Left) and Band 8 (Bottom Right). Spatial resolution of all bands is 10 m. Cloud cover is at 0%.	41
4.2	Multi-Spectral co-registration performance for S2B.	42
4.3	Quadrant diagram covering the runway of the Malta International airport. Bands are extracted from a Sentinel-2A L1C dataset on 2017/11/04. Top half of quadrant consists of band 2. Bottom left is extracted from B4. Bottom right is extracted from B8. Spatial resolution of all bands is 10 m. Cloud cover is at 100%. Quadrant size is of $500 \times 500$ pixels.	43
4.4	Multi-Spectral co-registration performance for S2A.	43
4.5	Landsat-8 B2 subset	45
4.6	Landsat-8 B10 subset	45
4.7	Sentinel-2 B2 subset	47
4.8	Landsat-8 B2 subset	47
4.9	Native Landsat-8 B10 subset of size $500 \times 500$ pixels	49

4.10	Native Sentinel-3 B8 subset of size $15 \times 15$ pixels. . . . .	49
4.11	Downscaled Landsat-8 B10 subset to a size of $15 \times 15$ pixels . . . . .	49
4.12	Native Sentinel-3 B8 subset of size $15 \times 15$ pixels. . . . .	49
4.13	Native Landsat-8 B10 subset of size $500 \times 500$ pixels . . . . .	50
4.14	Sentinel-3 B8 subset upscaled to a size of $500 \times 500$ pixels. . . . .	50
5.1	Kompsat-2 Panchromatic Band with 1 m spatial resolution of size $1211 \times 692$ pixels. Location: East-North of Malta . . . . .	54
5.2	Kompsat-2 Band 2 with 4 m spatial resolution of size $1211 \times 692$ pixels. . . . .	54
5.3	Registration results before and after registration for the first test case (optical vs optical). Magenta denotes the reference image and green denotes the sensed or registered image. The registration methods utilised were SIFT RANSAC, SIFT Flow and ECC. The chosen dataset is taken from the Kompsat-2 satellite. . . . .	55
5.4	Sentinel-2 Band 3 subset of size $1215 \times 692$ pixels. Location: Malta and Gozo. . . . .	56
5.5	Sentinel-1 Amplitude_VH_S1 Band subset of size $1215 \times 692$ pixels. . . . .	56
5.6	Registration results before and after registration for the second test case (optical vs SAR). Magenta denotes the reference image and green denotes the sensed or registered image. The registration methods utilised were SIFT RANSAC, SIFT Flow and ECC. The chosen dataset is taken from the Sentinel-2 and Sentinel-1 satellites. . . . .	58
5.7	Reference image: day shot of size $500 \times 500$ pixels taken from the RIFT multimodal dataset [12]. . . . .	59
5.8	Sensed image: night shot of size $500 \times 500$ pixels taken from the RIFT multimodal dataset [12]. . . . .	59
5.9	Registration results before and after registration for the third test case (Day vs Night). Magenta denotes the reference image and green denotes the sensed or registered image. The registration methods utilised were SIFT RANSAC, SIFT Flow and ECC. The chosen dataset is taken from the RIFT dataset. . . . .	60
5.10	High-level schematic of proposed algorithm. . . . .	64
5.11	Pre-processing steps to extract the Reference and Sensed Images. . . . .	65
5.12	Required steps to generate the joined cloud mask for the reference and sensed cloud masks. . . . .	68

5.13	Establishing the minimum and maximum temperatures from the reference and sensed images for min-max normalisation. . . . .	71
5.14	<b>Note:</b> The numbering of this figure, from 3a to 6 continue from the numbers shown in Fig. 5.13. Min-max normalising the reference and sensed images to convert the two images from brightness temperature to Uint8. . . . .	72
5.15	Extraction of best matching keypoints between the Reference and Sensed images. . . . .	74
5.16	Affine Transformations [14] . . . . .	75
5.17	Registered image of size $100 \times 100$ pixels for the example shown in Fig. 5.15. . . . .	79
5.18	Plot of Mean Number of Matches vs Sigma. . . . .	82
5.19	Graph showing the percent of keypoint locations that are repeatedly detected in a transformed image as a function of the $\sigma$ smoothing for the first level of each octave taken from David Lowe's work [7]. . . . .	82
5.20	Plot of Mean Number of Matches vs the Contrast Threshold $ D(\hat{x}) $ . . . . .	83
5.21	Plot of Mean Number of Matches vs the Edge Threshold $r$ . . . . .	84
6.1	Workflow of test that registers a Landsat-8 B10 image with its blurred version. . . . .	92
6.2	Reference images of the four chosen locations to be utilised for the test of registering a Landsat-8 B10 image with its blurred version. All chosen subset locations cover an area patch of $100 \text{ km} \times 100 \text{ km}$ which is equivalent to an image size of $3333 \times 3333$ pixels. All shown locations are in brightness temperature (Kelvin) and are extracted from their respective Landsat-8 L1C B10 data. . . . .	93
6.3	Plot of $\mu$ matches pre and post RANSAC vs scaling factor for all four locations as shown in Fig. 6.2 . . . . .	94
6.4	Plot of $\mu t_x$ vs scaling factor for all four locations as shown in Fig. 6.2	94
6.5	Plot of $\mu t_y$ vs scaling factor for all four locations. . . . .	95
6.6	Plot of $\mu$ RMSE before and after registration vs scaling factor for all four locations. . . . .	95
6.7	Reference and sensed images of the Mallorca location. Both images are located in the South-West of Mallorca shown in brightness temperature with a size of $1000 \times 1000$ pixels. . . . .	96

6.8	Absolute difference before and after registration for the reference and sensed images of the Mallorca location. . . . .	97
6.9	Overlay image before and after registration for the reference and sensed images of the Mallorca location. . . . .	98
6.10	Reference and sensed images of the Corfu location. Both images are located in the South of Corfu shown in brightness temperature with a size of $666 \times 666$ pixels. . . . .	99
6.11	Overlay image before and after registration for the reference and sensed images of the Corfu location. . . . .	100
6.12	Images of the Maltese archipelago taken in different days with different percentages of cloud cover. Fig. 6.12(a) is chosen as the reference image. Fig. 6.12(b) to Fig. 6.12(h) are sensed images. . . . .	102
6.13	Plot of $\mu$ matches vs the cloud cover percentage for each registered sensed image found in Fig. 6.12. . . . .	103
6.14	Plot of $\mu$ RMSE vs the cloud cover percentage for each registered sensed image found in Fig. 6.12. <b>Note:</b> for each image, the cloud cover % was rounded off to the nearest integer. . . . .	103
6.15	Matched features between the reference image Fig. 6.12(a) which has 0% cloud cover and the sensed image Fig. 6.12(h) which has 26.7%. . . . .	105
6.16	Registered image. . . . .	105
6.17	Workflow for the preprocessing steps done on the reference and sensed images for the registration of the downscaled L8 B10 subset with the S3 B8 subset. Covered geolocation for both images before shifting the sense image in step 2 is the same. Chosen locations for this example is Sardegna. . . . .	106
6.18	Reference images of the four chosen locations to be utilised for the test of registering a downscaled Landsat-8 B10 image with its respective Sentinel-3 B8 image. All chosen subset locations cover an area patch of $100 \text{ km} \times 100 \text{ km}$ . All shown locations are in brightness temperature and are extracted from their respective Landsat-8 L1C B10 data and Senintel-3 SLSTR data products respectively. . . . .	107
6.19	Plot of $\mu$ RMSE vs Shift using $\mathbf{H}_T$ . . . . .	109

6.20	Plot of $\mu t_x$ vs Shift using $\mathbf{H}_T$ . The green represents $y = x$ . The aim of this line is to illustrate how far off are the attained points from the green line. This represents the misregistration. The highest attained misregistration was at a zero shift using the default parameters of 0.1999 pixels. . . . .	109
6.21	Plot of $\mu t_y$ vs Shift using $\mathbf{H}_T$ . . . . .	110
6.22	Workflow for the preprocessing steps done on the reference and sensed images for the registration of the downscaled L8 B10 subset with the shifted and upscaled S3 B8 subset. Covered geolocation for both images before shifting the sense image is the same. Chosen location is Samothrace. . . . .	111
6.23	Plot of $\mu t_x$ vs Shift using $\mathbf{H}_T$ for $200 \times 200$ . . . . .	112
6.24	Plot of $\mu t_x$ vs Shift using $\mathbf{H}_T$ for $400 \times 400$ . . . . .	113
6.25	Plot of $\mu t_x$ vs Shift using $\mathbf{H}_T$ for $600 \times 600$ . . . . .	114
6.26	Plot of $\mu t_x$ vs Shift using $\mathbf{H}_T$ for $800 \times 800$ . . . . .	114
6.27	Plot of $\mu t_x$ vs Shift using $\mathbf{H}_T$ for $1000 \times 1000$ . . . . .	115
6.28	Plot of $\mu t_y$ for all the parameter methods vs Shift using $\mathbf{H}_T$ for the different upscaled sensed images . . . . .	116
6.29	Plot of $\mu$ RMSE vs Shift using $\mathbf{H}_T$ . . . . .	117
6.30	RIFT image dataset. Images on the left column are the sensed images whilst images on the right are the reference images. . . . .	120
6.31	Failed cases of registration when using SIFT Flow. . . . .	122
6.32	Overlay of the night-day image pair after registration. Magenta denotes the reference image and the green denotes the registered image. The overlay before registration is shown in Fig. 5.9(a). . . . .	123
6.33	Overlay of the Depth-Optical image pair before and after registration. Magenta denotes the reference image and the green denotes the registered image. Before registration green refers to the sensed image. . . . .	123
6.34	Failed matching for the Depth-Optical image pair using SIFT. . . . .	124
6.35	Failed matches when registering the infrared-optical image pair using SIFT. Left image is the optical. Right is the infrared. . . . .	124
6.36	Obtained matches after complementing the optical image of the infrared-optical pair using SIFT with the default parameters. Left image is the <i>complemented</i> optical. Right image is the infrared. . . . .	125



6.37	Overlay image before and after registration for the infrared and complemented optical image pair using SIFT with the default parameters.	125
6.38	Overlay image before and after registration for the infrared and complemented optical image pair using SIFT Flow with the default parameters. . . . .	126
6.39	Obtained failed matches using SAR-SIFT for the RIFT dataset Optical-SAR image pair. . . . .	127
6.40	Overlay image before and after registration for the SAR-Optical image pair using ECC. Magenta denotes the reference. Green denotes the sensed and registered image in the left and right image respectively. . . . .	127
6.41	Sensed image for the first registration test of Optical-SAR. Size of the image is $596 \times 698$ pixels. Covered Location is Sicily. . . . .	128
6.42	Reference image Landsat-8 L1TP B3 (green) for the first registration test of SAR-optical. Size of the image is $596 \times 698$ pixels. . . . .	128
6.43	Overlay image before and after registration for the first Optical-SAR pair using SIFT Flow with the default parameters. In both images, magenta is the reference. Green is the sensed in Fig. 6.43(a) and the reference in Fig. 6.43(b) respectively. . . . .	129
6.44	Obtained matches using SAR-SIFT for the first Optical-SAR test. .	130
6.45	Sensed and reference images for the second registration test of Optical-SAR. Size of the image is $596 \times 756$ pixels. Location: Cyprus. . . .	130
6.46	Registration before and after of the second Optical-SAR image pair using SIFT Flow. In both images magenta is the reference. Green is the sensed and registered in the left and right images respectively.	131
6.47	Reference and sensed images for the registration of Thermal-Optical images. Size of the images is $692 \times 767$ pixels. . . . .	132
6.48	Registration before and after of the image pair found in Fig. 6.47 using SIFT with the Monte Carlo parameters. . . . .	132
6.49	Reference and sensed images for the registration of Optical-Optical images. Size of the images is $692 \times 767$ pixels. Location: Cyprus . .	133
6.50	Registration before and after of the optical-optical image pair using SIFT with the GS parameters. . . . .	135
6.51	Reference and sensed images for the registration of Thermal-Thermal images. Size of the images is $1000 \times 1000$ pixels. Location: Samothrace	136

6.52	Reference and sensed images for the registration of second thermal-thermal image pair. Size of the images is $1000 \times 1000$ pixels. Location: Athens . . . . .	137
7.1	Visualisation of a deep learning intensity-based registration network [15]. . . . .	143

# List of Tables

2.1	(Non-exhaustive) Summary of currently active satellite instruments including the availability of the data. . . . .	9
2.2	Sentinel-2 product types [16] . . . . .	10
2.3	Spectral bands and spatial resolution of Sentinel-2 MSI [17] . . . . .	11
2.4	Landsat-8 product types [18,19] . . . . .	11
2.5	Spectral bands and spatial resolution of Landsat-8 OLI and TIRS [19]	11
2.6	Spectral bands and spatial resolution of Sentinel-3 SLSTR [20] . . . . .	13
3.1	Table denoting the multispectral misregistration in Sentinel-2A data in pixels [21]. . . . .	28
4.1	Phase correlation between the Sentinel-2B bands with a spatial resolution of 10 m. . . . .	41
4.2	Phase correlation for the same location of Fig. 4.3 between bands of Sentinel-2A with different spatial resolutions. The reference (B2, 10 m) remained with the same spatial resolution whereas the sensed image was upsampled to the same spatial resolution as the reference. The interpolation method used is bicubic interpolation. . . . .	44
4.3	Landsat-8 TIRS Geometric Performance Requirements [22]. . . . .	44
4.4	Landsat-8 OLI Geometric Performance Requirements [23]. . . . .	44
4.5	Phase correlation for the same location of Figures 4.5 and 4.6 between bands of Landsat-8 with different spatial resolutions. The chosen reference image is B2 with a spatial resolution of 30 m. Only B8 needed to be interpolated. It was downsampled by a factor of 2 to match the same spatial resolution as B2. The interpolation method used is bicubic interpolation. Bands 10 and 11 were not upsampled because they have the same image pixel size as B2 (but different spatial resolution). . . . .	46

4.6	Phase correlation between Band 2 of Sentinel 2 and Landsat 8 Bands 1-11. Same location and gap in time was chosen as Figures 4.7 and 4.8 between the reference and sensed images. <u>Note</u> : Bands 10 and 11 were upsampled with a scale factor of 3 and not 10 because the band is being represented in a 30 m grid despite a spatial resolution of 100 m. . . . .	48
5.1	RMSE results before and after registration for the first test case using SIFT Flow, ECC and SIFT RANSAC. . . . .	54
5.2	RMSE results before and after registration for the second test case using SIFT Flow, ECC and SIFT RANSAC. . . . .	57
5.3	RMSE results before and after registration for the second test case using SIFT Flow, ECC and SIFT RANSAC. <b>Note</b> : a large RMSE is expected as the reference and sensed image have a significant difference in intensity. . . . .	61
5.4	Theoretical number of octaves for an octave patch of size $64 \times 64$ . . . . .	81
6.1	Table illustrating the obtained $\mu$ RMSE before and after registration for the Mallorca location and the obtained $\mu$ misregistration values. Standard error for all obtained values was zero. . . . .	97
6.2	Table illustrating the obtained $\mu$ RMSE before and after registration for the Corfu location and the obtained $\mu$ misregistration values. Standard error for all obtained values was zero. . . . .	99
6.3	Obtained $\mu$ misregistration in the $x$ and $y$ direction for all registered sensed images of Fig. 6.12(h). The obtained standard error for all four values was zero. . . . .	104
6.4	Obtained $\mu$ misregistration in the $x$ direction for all registered shifted sensed images for the plot of Fig. 6.20 . . . . .	110
6.5	Obtained highest misregistration in the $x$ direction for each upscaled sensed image in meters . . . . .	115
6.6	Obtained highest misregistration in the $y$ direction for each upscaled sensed image in meters . . . . .	117
6.7	Table of results for each image pair found in the RIFT dataset. MC and GS refer to the proposed Monte Carlo (MC) and Grid Search (GS) parameter methods respectively. These methods were discussed in Sections 5.4.3 and 5.4.4. . . . .	121

6.8	Attained RMSE for the registration of the images found in Fig. 6.45.	130
6.9	Table of results for the registration of the optical-thermal pair of Fig. 6.47. . . . .	132
6.10	Table of the RMSE results for the registration of the optical-thermal pair of Fig. 6.47. . . . .	134
6.11	Table of the translation results for the registration of the optical-thermal pair of Fig. 6.47. . . . .	134
6.12	Table of the runtime results for the registration of the optical-thermal pair of Fig. 6.47. . . . .	134
6.13	Table of the RMSE results for the registration of the thermal-thermal pair of Fig. 6.51. . . . .	136
6.14	Table of the runtime results for the registration of the thermal-thermal pair of Fig. 6.51. . . . .	137
6.15	Table of the RMSE results for the registration of the second thermal-thermal pair of Fig. 6.52. . . . .	137
7.1	Table of the RMSE results for the registration of multi-modal and single modality remote sensing imagery. A number in red, blue and green denotes the first, second and third smallest RMSE respectively. The SAR-SIFT algorithm was only utilised for the registration of SAR-optical images. First six rows are the attained results for the images of the RIFT dataset. Remaining rows are the attained results for the chosen satellite imagery. . . . .	140

# List of Abbreviations

<b>BRIEF</b>	Binary Robust Independent Elementary Features
<b>BT</b>	Brightness Temperature
<b>CNN</b>	Convolutional Neural Network
<b>CRS</b>	Coordinate Reference System
<b>DEM</b>	Digital Elevation Model
<b>DoG</b>	Difference of Gaussian
<b>ECC</b>	Enhanced Correlation Coefficient
<b>EO</b>	Earth Observation
<b>EMS</b>	Electromagnetic spectrum
<b>ESA</b>	European Space Agency
<b>ETM</b>	Enhanced Thematic Mapper
<b>FANN</b>	Fast Approximate Nearest Neighbor
<b>FAST</b>	Features from Accelerated Segment Test
<b>FSC</b>	Fast Sample Consensus
<b>GCP</b>	Ground Control Point
<b>GeFolki</b>	Geoscience Extended Flow Lucas-Kanade Iterative
<b>GLOH</b>	Gradient Orientation and Location Histogram
<b>HS</b>	Horn–Schunck

<b>LK</b>	Lucas–Kanade
<b>LLS</b>	Linear least squares
<b>LSM</b>	Least squares matching
<b>LST</b>	Land Surface Temperature
<b>LoG</b>	Laplacian of Gaussian
<b>MAE</b>	Mean Absolute Error
<b>MIM</b>	Maximum Index Map
<b>MSI</b>	Multispectral Instrument
<b>MSSIM</b>	Mean Structural Similarity (MSSIM)
<b>MWR</b>	Microwave Radiometer
<b>NBCS</b>	Normalised barycentric coordinate system
<b>NDVI</b>	Normalized Difference Vegetation Index
<b>NIR</b>	Near Infra-Red
<b>NRD</b>	Nonlinear Radiation Distortions
<b>OLCI</b>	Ocean and Land Colour Instrument
<b>OLI</b>	Operational Land Imager
<b>ORB</b>	Oriented FAST and Rotated BRIEF
<b>OS</b>	Optical-to-SAR
<b>PC</b>	Phase Congruency
<b>POD</b>	Precise Orbit Determination
<b>RANSAC</b>	Random Sample Consensus
<b>RIFT</b>	Radiation-Variation Insensitive Feature Transform
<b>RMSE</b>	Root Mean Square Error

<b>SAR</b>	Synthetic-Aperture Radar
<b>SAT-FIRE</b>	Satellite data Fusion and Imaging Resolution Enhancement for coastal areas
<b>SIFT</b>	Scale Invariant Feature Transform
<b>SLSTR</b>	Sea and Land Surface Temperature Radiometer
<b>SNAP</b>	Sentinel Application Platform
<b>SRAL</b>	SAR Radar ALtimeter
<b>SST</b>	Sea Surface Temperature
<b>SURF</b>	Speeded Up Robust Features
<b>SWIR</b>	Short-wave infrared
<b>TIRS</b>	Thermal Infrared Sensor
<b>TOA</b>	Top of Atmosphere
<b>USGS</b>	United States Geological Survey



# Chapter 1

## Introduction

### 1.1 Problem Definition and Motivation

The process of spatially aligning two images of a scene in order to have corresponding points that have the same coordinate system is known as *Image Registration*. In image registration, for each point that is found in the first image, the corresponding point in the second image should also be found. In this process, the first image is denoted as the *reference* image whereas the second image to be registered is denoted as the *sensed* image. Throughout the entire steps of the image registration, the reference is kept the same whilst the sensed image is changed to map the spatial and geometry coordinates of the master image. Image registration is required as a preprocessing step in various image processing applications such as fusion based hyperspectral super-resolution and video enhancement.

Due to the drastic increase in data volumes and variety of sensors originating from a multitude of satellites, image registration is becoming one of the most crucial issues in remote sensing. An important consideration for this given problem is the *source of misregistration*. This misregistration can be caused by a change in sensor position, sensor viewpoint, object deformation, object movement or even from a difference in swath width from images originating from different satellites which could lead to geometric distortion. The width at which satellite sensors can observe is known as the *swath*.

In the remote sensing domain, satellite images have various characteristics that make image registration harder than when compared to image registration for nat-

ural images in applications such as for video, face recognition and medical images.

- *Multi-Modality.* The image registration method may be performed on images that are *taken from different sensors*. Such as *thermal-to-optical* sensors. Different satellites with distinct satellites imply that images are captured by different camera specifications which will make the registration harder.
- *The size of the data.* For instance, on average the size of a Landsat scene without any subset extractions is on average of size  $7000 \times 7000$  pixels. This contains 7 bands whose wavelength cover the visible to the thermal infrared range. Handling such huge amounts of data requires heavy computation. Thus, such requirements must be taken into account.
- *Lack of a known Ground Truth.* The lack of well-distributed points known as *Ground Control Points (GCP)* that aid to align images result in difficulties to validate image registration methods for remote sensing images. Although, prominent landmarks such as the Eiffel Tower can be utilised as fiducial points, such landmarks are very scarce, and are not evenly distributed around the world.

The most reliable way would be to actually record the GPS locations of millions of such sites, but such approach is too tedious and prohibitively expensive. Another approach is to compare the digital image with various other sources of ground reference data such as air photos or relevant reference maps. Nonetheless, such an approach can only be utilised if the features of interest are detectable using these sources.

- *Orbital Error.* Various types of errors may occur in the navigation-based correction, resulting in registration errors. Such errors may occur during spacecraft manoeuvres or due to ageing of the spacecraft and its resulting sensors which may cause them to perform differently from the way they were initially modelled. Fig. 1.1 illustrates misregistration caused by such an error in remote sensing.
- *Cloud and Atmospheric interactions.* Before radiation detected by remote sensing reaches the sensor it has to travel through considerable distance of the Earth's atmosphere. The incoming radiation and light can be affected by particles and gases found in the atmosphere. Atmospheric effects are

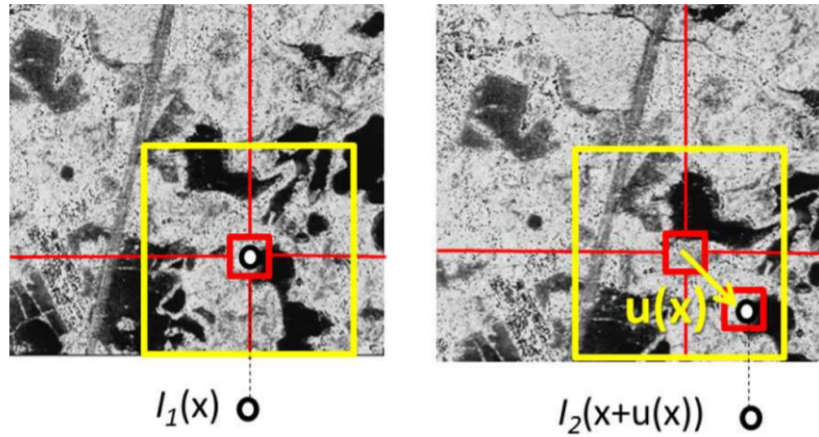


Figure 1.1: Remote sensing misregistration caused by orbital error.  $I_2$  is the sensed image to be registered to the reference image  $I_1$ .  $u(x)$  denotes the displacement between the two images [2].

caused by *scattering* and *absorption*. Examples of such effects are atmospheric humidity and volcanic aerosols.

Another issue in the image registration of remote sensing data is *cloud interaction*. Clouds might occlude a particular important feature of an image. Such occluded features are important for the image registration as these occluded features in a scene can improve the accuracy of the registration.

- *Multitemporal effects*. Multitemporal effects can be induced by *natural effects* or by *human-induced* effects that happen over time. Multitemporal effects caused by natural effects cause a change of the scene through natural occurrences such as a change in the Sun angle during the year. Season changes are also another example of natural multitemporal effects. The same scene in different seasons might look differently. To all those natural induced effects, the human-induced multitemporal effects must also be included. Human-induced effects are related to activities such as deforestation, urban development and agricultural practices.
- *Relief/terrain effects*. Another source that affects image registration of remote sensing imagery is the *terrain* or the *topography*. Depending on the time of the day, the slope of the geographic terrain, the characteristics of the sensor and the satellite orbit, terrain effects might appear exceptionally different in the images to be registered. Such topographic differences are generally corrected using *terrain models* but minor local effects will still be present.
- *Multisensor*. Multisensor refers to the difference in *spectral* and *spatial reso-*

*lutions*. When dealing with several sensors, with different spatial and spectral resolutions, the following image registration issues must be addressed:

1. Appropriate selection of geometric transformations that cater for the different spatial resolutions as **significant differences in spatial resolutions can cause spatial distortion**.
2. Extraction of image features which are invariant to radiometric differences due to multitemporal and multispectral resolutions.
3. Choice of similar bands that lie in a similar region in the electromagnetic spectrum when performing band-to-band registration.

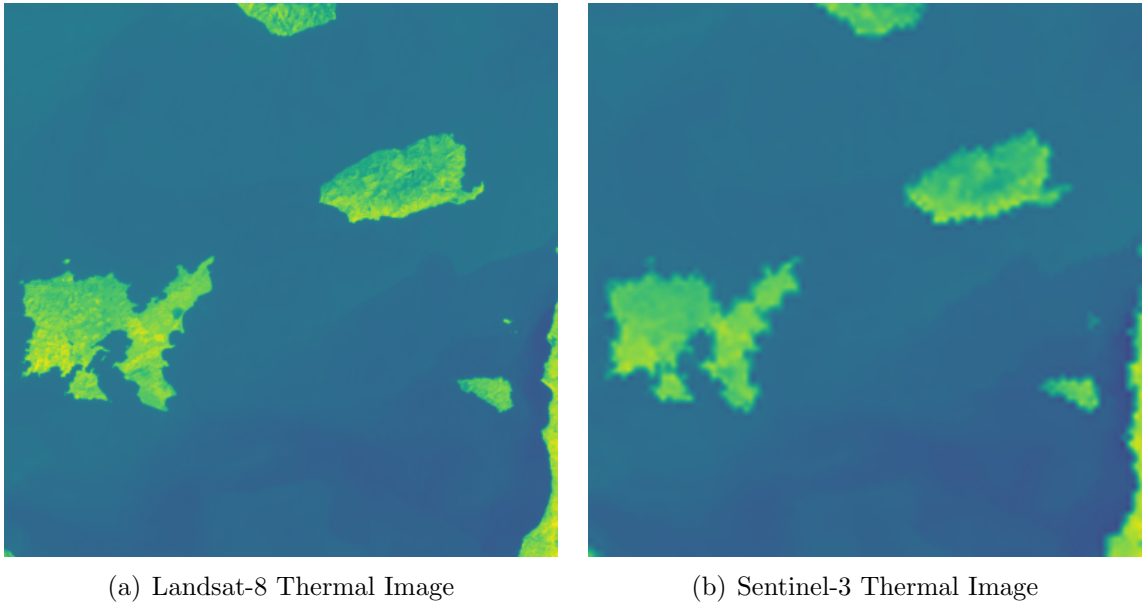


Figure 1.2: Thermal satellite images of different spatial resolutions

Fig. 1.2(a) shows the extracted Landsat-8 thermal image that has a spatial resolution of 100 m per pixel of size  $1000 \times 1000$  pixels. Fig.1.2(b) shows the extracted Sentinel-3 image which has an original spatial resolution of 1 km per pixel. Fig. 1.2(b) was upscaled to a size of  $1000 \times 1000$  pixels using bicubic interpolation to match the spatial resolution of the Fig. 1.2(a). Fig. 1.2 depicts the problem posed by the significant scale difference in spatial resolutions. The upscaling required for the S3 image caused significant distortion in Fig. 1.2(b). Induced artificial pixels can be observed in the coastal regions of the islands. Such an aspect can cause problems in registration as the false induced pixels can be detected as features to be

utilised for registration. In applications such as data fusion and super-resolution, such misregistration problems can be a hindrance to the resulting outcome.

Y. Zhou et al. proposed a fusion algorithm to register and fuse hyperspectral and multispectral images [24]. In their work they highlight the fact that despite image registration is a major application for remote sensing, there is little work aimed at the image registration of images with significant scale differences [24]. They also highlight that for such an application in the remote domain for real datasets there is no ground truth [24].

## 1.2 Objectives

1. Investigate the gap in literature concerning the image registration of remote sensing images for multi-modal and multi-sensorial satellite images.
2. Utilisation of a feature-based technique for multi-modal and multi-sensorial satellite image registration.
3. Evaluate and analyse the amount of misregistration between thermal-to-thermal satellite imagery with significant scale differences.
4. Compare area-based, feature-based and optical flow-based methods for the registration of multi-modal and multi-sensorial satellite imagery.

## 1.3 Dissertation Outline

In Chapters 2 and 3, the background and literature review are given. In Chapter 4 an investigation for the registration between multispectral bands of various satellites is analysed. In Chapter 5, the framework description of the proposed methodology is presented. In Chapter 6, Results and their respective analysis are shown. Finally, Chapter 7 draws some conclusions, outlines the limitations and mentions some directions for future work.

# Chapter 2

## Background

### 2.1 Remote Sensing Fundamentals

Remote sensing can be denoted as the process by which information about a phenomenon or object is captured from a remote location such as a satellite or an aircraft. More precisely, satellite or sensing images refer to images created from the use of sensors located on space platforms to capture electromagnetic energy which is emitted or reflected from planetary sources such as the Earth or the Sun. Sensors can be either *passive* or *active*. In passive satellite sensors the energy detected originates from the planetary surface (e.g. as a result of reflectance of solar energy), whilst active sensors, utilise their own source of energy to capture images such as radar systems.

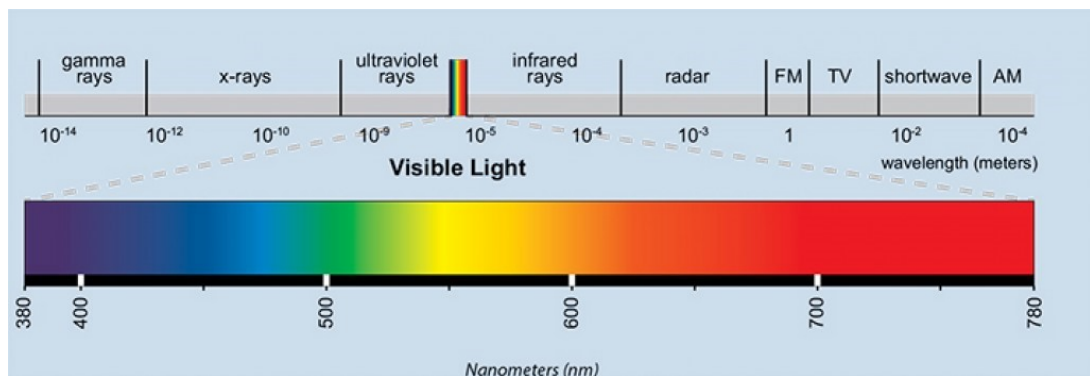


Figure 2.1: Electromagnetic spectrum (EMS) [3]

All objects emit radiation at different wavelengths, the emitted energy varies with

the temperature of the object and with the wavelength. The ideal body is a *black-body*, a body which allows the absorption of all incident radiation, without reflecting any. According to Wien's and Stefan-Boltzman's displacement laws, a *dominant wavelength*, can be determined for all blackbodies. A dominant wavelength is the wavelength at which the total amount of light, emitted by an area of surface of a radiating body is at maximum.

Assuming that the Sun and the Earth act like blackbodies, their respective dominant wavelengths are 500 nm (in the green visible part of the EMS) and 9700 nm (in the Infrared (IR) region of the EMS). This infers that the emitted energy by the Earth is best observed by sensors which operate in the microwave and thermal infrared regions of the EMS whilst reflected Sun energy by the Earth is best observed in the visible, near-infrared and mid-infrared regions of the EMS.

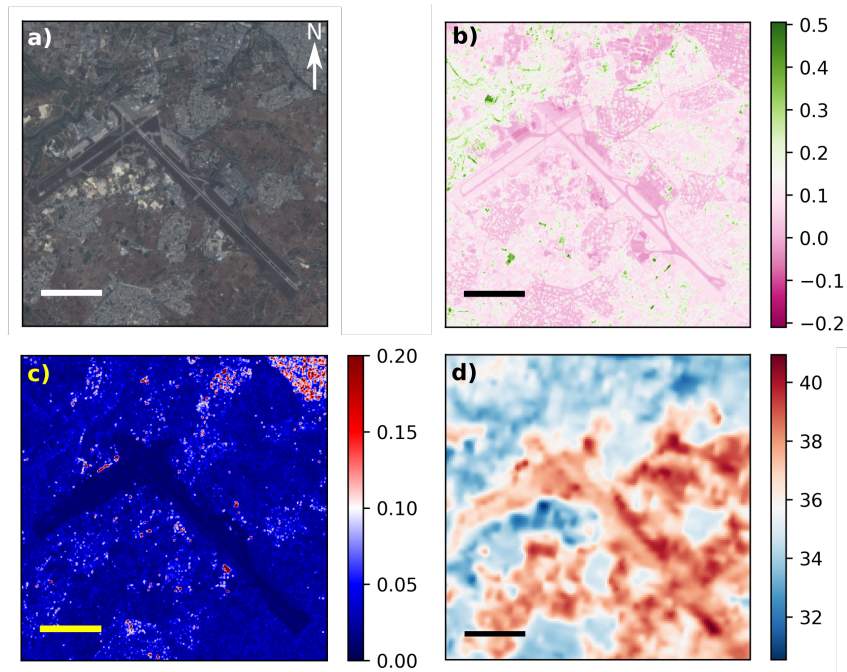


Figure 2.2: Satellite images of the region around Malta International Airport. a) RGB image synthesised from red, green and blue bands from Sentinel-2 with 10m spatial resolution. b) NDVI image synthesised from the red and NIR bands from Sentinel-2 with 10 m spatial resolution. c) SAR image from Sentinel-1 [4] with 10 m resolution. The normalised reflected amplitude in the vertical polarisation is shown. d) Thermal image from Landsat 8 showing brightness temperature in degrees Celsius with a spatial resolution of 100m. The horizontal bars in each of the four images correspond to a distance of 1 km. Images shown in a), b) and d) were acquired on 04/07/2018, whereas c) was acquired on 06/07/2018.

### 2.1.1 Optical Images

Optical images refer to recordings of the surface reflected solar radiation in the visible<sup>1</sup> and near-infrared<sup>2</sup> spectral regions. A typical instrument may record images simultaneously in a number of discrete spectral bands, allowing a 3-dimensional (two space and one spectrum) image of the surface to be derived from the data. With respect to the satellites listed in Table 2.1, the highest resolution, freely available optical images are provided by Sentinel-2 [25] with a spatial resolution of 10 m per pixel. In remote sensing, *spatial resolution* refers to the value allocated to each pixel in terms of area on the ground in meters.

Multi-spectral instruments, contain discrete, separated bands, offering partial coverage of the electromagnetic spectrum. An example is Operational Land Imager on board Landsat 8 which records in 9 discrete bands. The central wavelength and bandwidth of each of the bands are chosen with specific applications in mind. For example, band 9 of the Sentinel-2 satellite covers wavelengths in the range  $945 \pm 10$  nm. This region coincides with strong absorption by water vapour. Hence the band 9 image can be used to estimate the density of water vapour in the atmospheric column between the surface and the satellite. A composite RGB image formed from three bands of the Sentinel-2 satellite is shown in Fig. 2.2 a).

Less common are hyperspectral instruments, such as the Hyperion imaging spectrometer on board the now-decommissioned Earth Observation 1 satellite [26]. For these instruments, the spectral coverage is more complete: in the case of Hyperion there were 220 contiguous bands covering the spectral region between 0.4 and  $2.5 \mu\text{m}$ . The trade-off is a very narrow swath (7.5 km) for the spatial resolution (30 m). Nevertheless, spectroscopic applications, including element and mineral detection [27] are enhanced by the higher spectral sensitivity of hyperspectral instruments.

### 2.1.2 Thermal Infrared Images

Thermal infrared images refer to recordings of light radiated by the Earth's surface as a result of its temperature. In the spectral region spanning  $8 - 25 \mu\text{m}$  in wavelength, the satellite-recorded signal is dominated by thermal light, with reflected solar radiation negligible in comparison. According to Planck's law, for a fixed

---

<sup>1</sup>Wavelengths between 400 – 700 nm.

<sup>2</sup>Wavelengths between 700 – 1400 nm.

<sup>3</sup>At a mid-latitude location (i.e. the Mediterranean Sea), correct as of May 2019.



Satellite Instrument Name & Type	Spatial Resolution (m)	Revisit Time (days) <sup>3</sup>	Open Data
Sentinel-1: C-SAR (SAR)	5, 20 & 25	2	Yes
Sentinel-2: MSI (Optical)	10, 20 & 60	2-3	Yes
Sentinel-3: OLCI (Optical) [28]	300	< 1.4	Yes
Sentinel-3: SLSTR (Thermal)	1000	< 1	Yes
Landsat 7: ETM+ (Optical) [29]	30	16	Yes
Landsat 7: ETM+ (Thermal)	60	16	Yes
Landsat 8: OLI (Optical) [30]	30	16	Yes
Landsat 8: TIRS (Thermal)	100	16	Yes
SPOT5: HRG (Optical) [31]	5, 10 & 20	3	No
RapidEye: MS (Optical) [32]	6.5	1	No
Pleiades: HR (Optical) [33]	0.5	26	No
TerraSAR-X: HR (SAR) [34]	1	2.5	No
RADARSAT-2: HR (SAR) [35]	3	3	No

Table 2.1: (Non-exhaustive) Summary of currently active satellite instruments including the availability of the data.

wavelength, the brightness of the measured signal is related to the temperature of the emitting body: the brighter the emission the higher the temperature. Hence the *radiance*, as measured in the thermal infrared spectral region by a satellite, can be related to a *brightness temperature*: the temperature of an equivalent black body. Accounting for atmospheric absorption and the emissivity of the ground location allows the brightness temperature to be corrected to better correspond with the actual surface temperature. A brightness temperature image acquired by the Landsat 8 satellite is shown in Fig. 2.2 d). With respect to the satellites listed in Table 2.1, the highest spatial resolution, freely available thermal images are provided by Landsat 7 ETM+ instrument, with a spatial resolution of 60 m. This high spatial resolution comes at the cost of a large revisit time (16 days).

### 2.1.3 Synthetic Aperture Radar

As opposed to optical and thermal satellite imagery, Synthetic Aperture Radar (SAR) images are produced by active remote sensing, in which microwaves are beamed from the satellite towards Earth, and the reflected waves are detected by sensors on board the satellite. The advantage of SAR remote sensing is that images can be acquired in any type weather conditions, day or night, including in the presence of cloud cover. A disadvantage is that SAR imagery suffers from speckle and topography effects. Typically, side looking radar is used to correctly

decipher the amplitudes and phases of the received signals. An example of a SAR image acquired by Sentinel-1 is shown in Fig. 2.2 c). With respect to the satellites listed in Table 2.1, the highest spatial resolution, freely available SAR images are provided by Sentinel-1 in “stripmap” mode, with a resolution of just 5 m.

## 2.2 Processing Levels and Bands

This section defines the level of data processing and the bands for the most mentioned and utilised satellites in this dissertation.

### 2.2.1 Sentinel-2

The Sentinel-2 MSI (optical) samples 13 bands whose wavelengths range from the Blue spectral region to the Short-wave infrared (SWIR) region in the EMS. Sentinel-2 MSI data products contain two levels of processing: Level-1LC and Level-2A. Levels 0 to 1B are not available to the users, only level-1LC and level-2A products are available.

Name	Description	Size of Tile
Level-1C	TOA reflectances in cartographic geometry	100×100 km <sup>2</sup>
Level-2A	BOA reflectances in cartographic geometry	100×100 km <sup>2</sup>

Table 2.2: Sentinel-2 product types [16]

### 2.2.2 Landsat-8

Landsat-8 is the most recent satellite launched by Landsat. It carries two instruments: the optical instrument known as the *Operational Land Imager (OLI)* and the *Thermal Infrared Sensor (TIRS)* instrument. Landsat-8 samples 11 band whose wavelengths range from the Blue spectral region to the TIRS region. Landsat-8 data products contain three levels of processing: *L1TP*, *L1GT* and *L1GS* as denoted in Table 2.4.

The Landsat-8 data product utilised in this dissertation and in most reviewed literature is the Landsat-8 L1TP product. It is the highest quality level-1 product from all three. It is orthorectified and radiometrically calibrated using digital elevation model (DEM) data and ground control points (GCPs) to correct for

Band	Spectral Region	Spatial Resolution ( $m$ )	Central Wavelength ( $\mu m$ )
B1	Ultra Blue (Coastal and Aerosol)	60	0.443
B2	Blue	10	0.490
B3	Green	10	0.560
B4	Red	10	0.665
B5	Vegetation Red edge	20	0.705
B6	Vegetation Red edge	20	0.740
B7	Vegetation Red edge	20	0.783
B8	NIR	10	0.842
B8a	NIR narrow	20	0.865
B9	Water Vapour	60	0.940
B10	SWIR (Cirrus)	60	1.375
B11	SWIR	20	1.610
B12	SWIR	20	2.190

Table 2.3: Spectral bands and spatial resolution of Sentinel-2 MSI [17]

relief displacement. GCPS utilised for the L1TP correction are extracted from the Global Land Survey 2000 (GLS2000) data set [18].

Name	Description	Size of Tile
L1TP	Terrain Precision Correction	$185 \times 180 \text{ km}^2$
L1GT	Systematic Terrain Correction	$185 \times 180 \text{ km}^2$
L1GS	Geometric Systematic Correction	$185 \times 180 \text{ km}^2$

Table 2.4: Landsat-8 product types [18, 19]

Band	Spectral Region	Spatial Resolution ( $m$ )	Central Wavelength ( $\mu m$ )
1	Coastal aerosol	0.43-0.45	30
2	Blue	0.45-0.51	30
3	Green	0.53-0.59	30
4	Red	0.64-0.67	30
5	NIR	0.85-0.88	30
6	SWIR 1	1.57-1.65	30
7	SWIR 2	2.11-2.29	30
8	Panchromatic	0.50-0.68	15
9	Cirrus	1.36-1.38	30
10	TIRS 1	10.60-11.19	100
11	TIRS 2	11.5-12.51	100

Table 2.5: Spectral bands and spatial resolution of Landsat-8 OLI and TIRS [19]

A salient point to be noted is that the thermal TIRS bands of Landsat-8 have a

spatial resolution of 100 m but are *resampled at a pixel size of 30 m to match the multispectral bands* [36].

### 2.2.3 Sentinel-3

The objective of the Sentinel-3 mission is to measure sea and land-surface temperature, sea-surface topography and ocean-and land-surface colour. Sentinel-3 samples 11 spectral bands and carries multiple instruments. *Ocean and Land Colour Instrument (OLCI), Sea and Land Surface Temperature Radiometer (SLSTR), SAR Radar ALtimeter (SRAL), Microwave Radiometer (MWR) and Precise Orbit Determination (POD).*

Sentinel-3 SLSTR thermal data contains three levels of processing. Level-0, Level-1 and Level-2. Only Levels-1 and Levels-2 are available to the public. This dissertation utilises the Level-1 *SL\_1\_RBT* Sentinel-3 SLSTR data product that contains radiance and brightness temperature measurements for both oblique and nadir views.

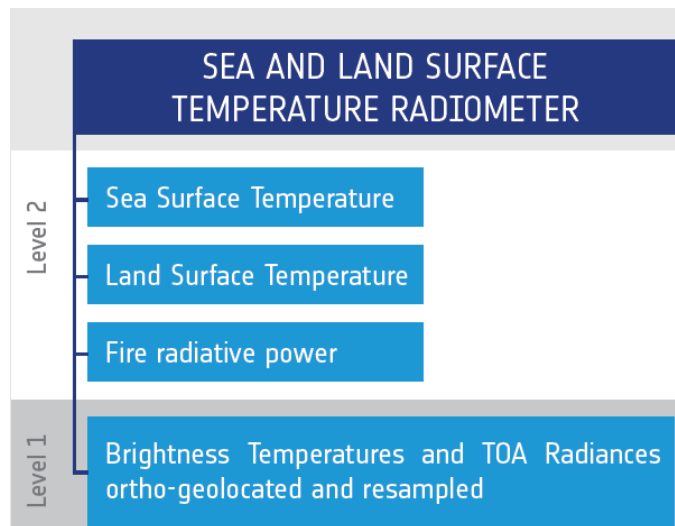


Figure 2.3: Sentinel-3 SLSTR Levels of Processing [5]

Band	Spectral Region	Spatial Resolution (m)	Central Wavelength	Function
S1	VNIR	500	0.55427	Vegetation monitoring, cloud screening, aerosol
S2	VNIR	500	0.65947	Vegetation monitoring, NDVI, aerosol
S3	VNIR	500	0.868	NDVI, cloud flagging
S4	SWIR	500	1.3748	Cirrus detection
S5	SWIR	500	1.6134	Ice, snow, cloud clearing, vegetation monitoring
S6	SWIR	500	2.2557	Cloud clearing, vegetation state
S7	Thermal IR	1000	3.742	SST, LST, Active fire
S8	Thermal IR	1000	10.854	SST, LST, Active fire
S9	Thermal IR	1000	12.0225	SST, LST
F1	Thermal IR	1000	3.742	Active Fire
F2	Thermal IR	1000	10.854	Active Fire

Table 2.6: Spectral bands and spatial resolution of Sentinel-3 SLSTR [20]

## 2.3 Registration through Geocoding

Multi spectral and single modality satellite imagery are *geocoded* by assigned coordinates to the image through the usage of *Ground Control Point (GCP) collection*. The units that represent these assigned coordinates in a geographic coordinate system are known as the *latitude* and *longitude*.

A GCP is an accurate known location on Earth which is typically used to georeference remote sensing data. An example of a GCP is the Eiffel Tower. As illustrated in Fig. 2.4, a GCP determines the relationship between the raw image and the ground by associating the pixel (P) and line (L) image coordinates to the x, y, and z coordinates on the ground. These ground coordinate can come from various sources such as ground surveys, topographic maps and GPS.

Both the geometrically corrected Sentinel-2 L1C TOA data and the equivalent Landsat-8 L1T TOA utilise onboard GPS sensors, an attitude determination sys-

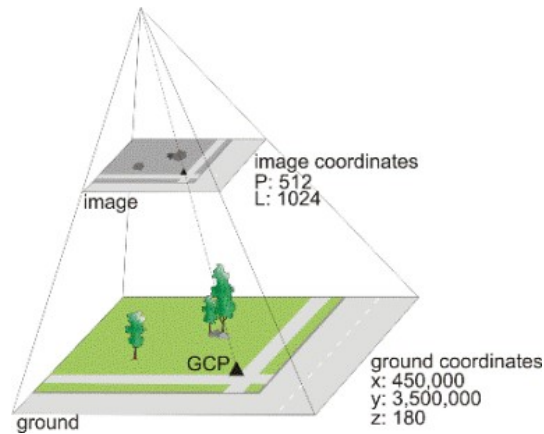


Figure 2.4: Ground Control Point [6]

tem and digital elevation models (DEM) to geolocate the sensed field of view [37]. Landsat-8 L1T products are their highest level one products. Orthorectified and radiometrically calibrated Landsat-8 L1T products utilise digital elevation model data and ground control points to correct for relief displacement. GCPs utilised for the L1TP correction are extracted from the Global Land Survey 2000 (GLS2000) data set [18].

Sentinel-3 data products are geolocated by calculating the geolocation parameters (*altitude*, *altitude rate*, *latitude* and *longitude*) which are calculated using the orbit file which is generated by the *Precision Orbit Determination (POD)* Sentinel-3 instrument [38].

## 2.4 General Process of Full Image Registration

In addition to geocoding, an image matching algorithm is required for full image registration to achieve subpixel accuracy in the georeferenced satellite data.

A full image *feature-based* registration process typically consists of four main steps:

1. *Feature extraction*. In this step features such as contours, edges and regions are extracted and utilised as tie-points. These extracted tie-points are then utilised to match the images to be registered. Getting the correct tie-points is essential for accurate registration.

Feature-based methods rely on the extraction of features from the images to

be registered. Examples of such extracted features are lines, edges, intersections and contours. The choice of feature-based method utilised is application dependent. Examples of feature-based algorithms include: *Scale Invariant Feature Transform (SIFT)* [7], *Speedup robust features (SURF)* [39], *KAZE* [40], *Accelerated-KAZE (AKAZE)* [41], *Oriented FAST and Rotated BRIEF (ORB)* [42] and *Binary robust invariant scalable keypoints (BRISK)* [43]. From all the mentioned feature-based algorithms, for natural images, SIFT was found to be the most accurate feature-detector descriptor for rotation, affine and scale variations (overall) [44].

2. *Feature matching*. In this step, the correspondence between the features of the two images to be registered is found. In feature-based methods the correspondence is generally found by using *distance metrics* such as *Hamming distance* or *Euclidean distance*.
3. *Selection of Best Matching Keypoints*. In this step the best matching keypoints are selected by rejecting outliers. A outlier detection method is typically utilised to reject outliers such as the *Random Sample Consensus (RANSAC)* Algorithm [45].
4. *Transformation estimation* is the mapping function utilised to estimate transformation parameters based on the previously established correspondence between the extracted features. In remote sensing applications, the most commonly used geometric transformation is *affine transform*.
5. *Image Warping* is the process that maps the sensed image to the reference image based on the obtained transformation parameters from the affine matrix. In the case that the images are of different sizes *image resampling* is also required. Typical interpolation methods utilised for image resampling are *nearest neighbour*, *bilinear interpolation* or *bicubic interpolation*.

In addition to feature-based methods, three other methods can also be utilised for full image registration. These are *area-based* methods, a *hybrid approach* where both feature-based and area-based methods are combined and *optical-flow based* methods.

Area-based methods rely on the *brightness consistency*. The gray-levels of the images are utilised directly to match them. This matching is typically done through statistical comparison of pixel values in a window area. The matching in area-based methods is typically done using *phase-correlation* or *cross-correlation*. *Area-based methods are best suited for images that are captured from a highly similar sensor or the same sensor*. An example of area-based method is *Enhanced cross-correlation*. [46].

*In contrast to area-based methods, feature-based methods are robust to differences in scale and rotation between the two images by detecting reliable features in both reference and sensed images. Thus, feature-based methods can perform well for the registration of images coming from heterogeneous sensors*. Hybrid-approaches utilise a combination of area-based and feature-based methods. These methods are also useful when registering data coming from heterogeneous sensors.

Contrary to feature-based and area-based methods which typically require an estimation of a transformation homography optical-flow based methods try to find the pixel wise correspondence. *Optical-flow algorithms* rely on motion vectors. An *optical-flow* algorithm takes two frames and for each pixel in the first frame a vector is derived. This vector is then utilised to display each pixel from the first to the second frame. *Optical-flow methods are best suited for video-analysis and to register nonrigid images*. Examples of optical-flow methods include *Lucas–Kanade* [47] method and *SIFT-Flow* [48]. However, optical-flow methods are also considered to register remote sensing images.



## 2.5 SIFT Algorithm

SIFT is a patented feature detection algorithm that was published by David Lowe [7] in 2004 to match or stitch images coming from different sources. It is used to *detect and describe local features*. SIFT descriptors are *invariant to illumination, scale and rotation*.

The SIFT algorithm consists of *four* main steps:

- **Scale Space Extrema Detection**
- **Accurate Keypoint Localisation**
- **Orientation Assignment**
- **Local Descriptor Creation**

### 2.5.1 Scale Space Extrema Detection

The scale space of an image is defined as a function  $L(x, y, \sigma)$ , that is produced from the convolution of a variable-scale Gaussian  $G(x, y, \sigma)$  with an input image  $I(x, y)$ .

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2.1)$$

where  $*$  is the convolution operator in  $x$  and  $y$ ,  $(x, y)$  are the spatial coordinates and  $\sigma$  is the spread of the Gaussian also known as the “scale” parameter.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2.2)$$

To efficiently detect stable key point locations proposed the difference-of-Gaussian function convolved with the image,  $D(x, y, \sigma)$ , which can be computed from the difference of two nearby scales separated by a constant multiplicative factor  $k$ .

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma)) - G(x, y, \sigma) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (2.3)$$

where typically  $k = \sqrt{2}$  and  $\sigma = 1.6$  [7].

As shown on the left of Fig. 2.5 an *octave* is a set of images where the blur of the last image is double the blur of the first image. Within an octave the adjacent images differ by a blur of  $k \times \sigma$ . For each octave, the initial image is repeatedly convolved with the Gaussian operator to form the blurred images known as the scale space images shown on the left of Fig. 2.5.

The adjacent blurred images in each octave are subtracted to form the DoG images shown on the right of Fig. 2.5. After each octave, the blurred image is downsampled by half, and this process is iterated for the chosen number of octaves. The number of octaves proposed by Lowe is 4 [7].

The *DoG* is an approximation for the *Laplacian of Gaussian (LoG)*.

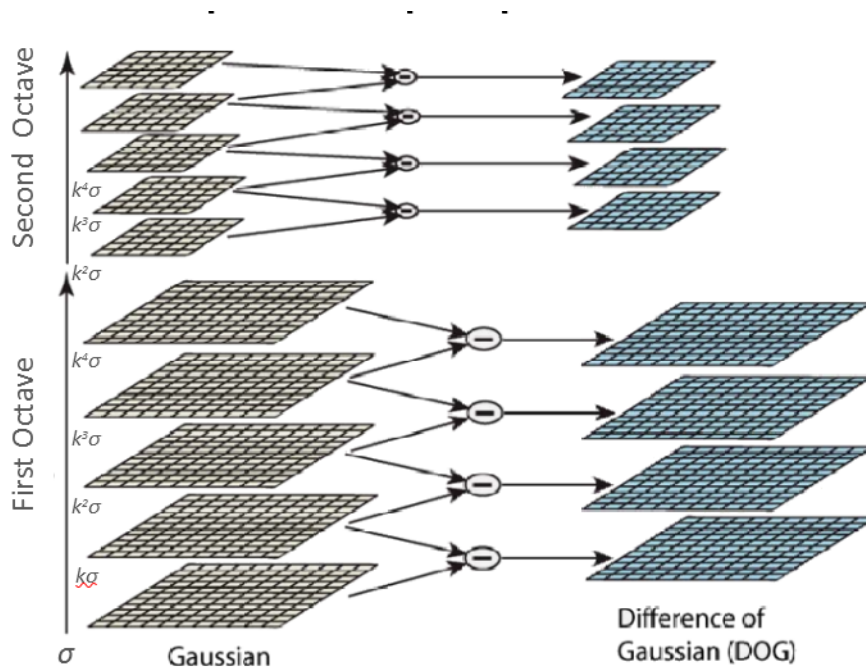


Figure 2.5: Difference of Gaussian (DoG) [7]

## 2.5.2 Local Extrema Detection

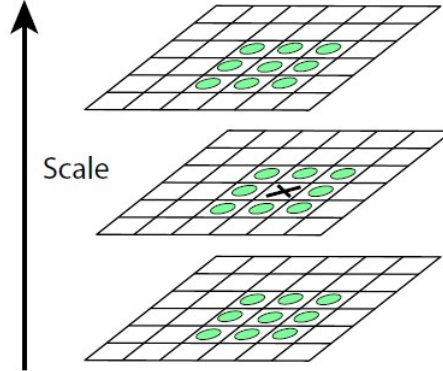


Figure 2.6: Extrema Detection [7]

A check is done through each pixel in the DoG image of Fig. 2.6. In order to detect the *local maxima* and *local minima* of  $D(x, y, \sigma)$ , the point marked in  $X$  is compared to its eight neighbours in the current image and nine neighbours in the scale above and below as shown in Fig. 2.6. The point  $X$  is only selected if it is larger than all of these neighbours or smaller than all of them. This process is used to denote a list of candidate keypoints.

## 2.5.3 Accurate Keypoint Localisation

### 2.5.3.1 Local Extrema Detection

The next step is to reject low contrast and poorly localized edges. Therefore, this is computed by approximating the difference of Gaussians using a Second Order Taylor series denoted by:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} \quad (2.4)$$

where  $\mathbf{x} = (x, y, \sigma)^T$

The author of SIFT showed that the location of the extrema,  $\hat{\mathbf{x}}$ , is found by taking the derivative of Eq. 2.4 and initialising it to zero, resulting in

$$\hat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D}{\partial \mathbf{x}} \quad (2.5)$$

Thus,  $D(\hat{\mathbf{x}})$  is equivalent to

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \mathbf{x}} \hat{\mathbf{x}} \quad (2.6)$$

$|D(\hat{\mathbf{x}})|$  is known as the *contrast threshold* and its value is utilised to eliminate low contrast keypoints. All extrema values of  $|D(\hat{\mathbf{x}})|$  less than a particular floating value are rejected. Image pixel values in the algorithm are assumed to be within the range  $[0, 1]$  and the contrast threshold utilised by Lowe is 0.03 [7].

### 2.5.3.2 Further Outlier Rejection by Eliminating Edge Responses

It is not enough to reject keypoints with low contrast, keypoints which have a strong response along edges must also be rejected. Mathematically, this is accomplished using the *Hessian Matrix*. The eigenvalue decomposition of the Hessian  $\mathbf{H}$  is computed to derive the eigenvalues  $\alpha$  and  $\beta$  as similarly done for Harris Corner detection.

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (2.7)$$

Taking the differences of the neighbouring sample points estimates the derivatives. The eigenvalues of  $\mathbf{H}$  are proportional to the second order derivatives of  $D$ . They are denoted by  $D_{xx}$ ,  $D_{xy}$  and  $D_{yy}$ .

$\alpha$  is denoted as the eigenvalue with the largest magnitude and  $\beta$  is denoted as the eigenvalue with the smallest magnitude. The sum of the eigenvalues is computed from the *trace* of  $\mathbf{H}$  and their product from the determinant:

$$Tr(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta \quad (2.8)$$

$$Det(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \quad (2.9)$$

Let  $r$  be the ratio between  $\alpha$  and  $\beta$ , such that  $\alpha = r\beta$ . Then,

$$\frac{Tr(\mathbf{H})^2}{Det(\mathbf{H})} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r} \quad (2.10)$$

$r$  is denoted as the *edge threshold*. To detect whether there is a *corner* or an *edge* around the keypoint of an image, the SIFT algorithm checks

$$\frac{Tr(\mathbf{H})^2}{Det(\mathbf{H})} < \frac{(r+1)^2}{r} \quad (2.11)$$

Based on the image around the keypoint *three* possibilities exist:

- A *flat region* - both gradients will be small and  $|r|$  is small
- An *edge* - one gradient will be big the other will be small,  $r < 0$
- A *corner* - both gradients will be big,  $r > 0$

*Corners are the most discriminative features. The aim is to only keep corners.* This method only selects keypoints where  $r > 0$ . Lowe recommends that the edge threshold  $r = 10$  [7].

## 2.5.4 Orientation Assignment

In this step an orientation is assigned to prominent keypoints. *This orientation is required to achieve rotation invariance.* This is done by computing the gradient magnitude  $m$  and orientation  $\theta$  at a scale of keypoint  $(x, y)$ . Therefore, *for every detected keypoint both  $m(x, y)$  and  $\theta(x, y)$  are computed using:*

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (2.12)$$

$$\theta(x, y) = \arctan\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right) \quad (2.13)$$

The orientation is computed by computing the histogram in a neighbourhood of each keypoint identified by the blue circle in Fig. 2.7 (36 bins where one bin corresponds to  $10^\circ$ ).

The amount added to each bin is based on the *magnitude of gradient* at that point and the *Spatial gaussian filter* (windows size) with  $\sigma = 1.5$ . The spatial gaussian filter gives higher weights to vectors closer to the keypoint. *The orientation of the interest point is then assigned to the bin with the highest peak.* In case of multiple peaks (within 80% of highest peak), the histogram is assigned different directions.

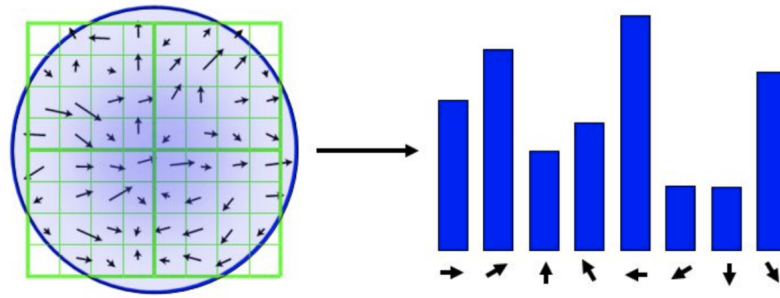


Figure 2.7: Histogram of Gradient based on Orientation [8]

### 2.5.5 Local Descriptor Creation

The previous operations have assigned an *image location*, *scale* and *orientation* to each *keypoint*. Thus, these operations provide invariance to such parameters. The next step *computes a feature vector for each keypoint*.

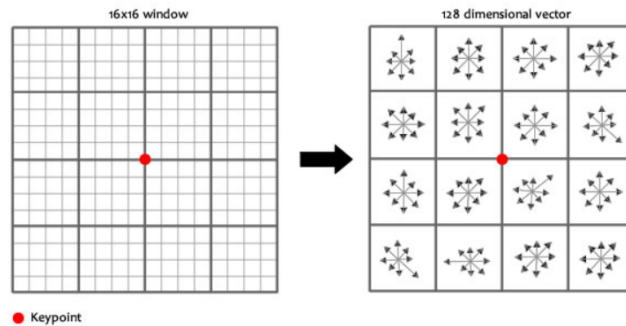


Figure 2.8:  $16 \times 16$  window broken down to into sixteen  $4 \times 4$  windows [9]

This is accomplished by creating a  $16 \times 16$  window around the keypoint. This  $16 \times 16$  is broken into sixteen  $4 \times 4$  windows as shown in Fig. 2.8. Within each  $4 \times 4$  window, the gradient magnitudes and orientations are computed. These orientations are allocated into an 8-bin histogram as shown in Fig. 2.9.

An orientation in the range  $0^\circ - 44^\circ$  is added to the first bin.  $45^\circ - 89^\circ$  is added to the second bin and so on. The amount added to each bin depends on the magnitude of the gradient and the distance from the keypoint. The distance is computed using the *weighted Gaussian filter* which generates a gradient as shown in Fig. 2.10.

The previous steps are repeated for all sixteen  $4 \times 4$  regions which results in  $4 \times$

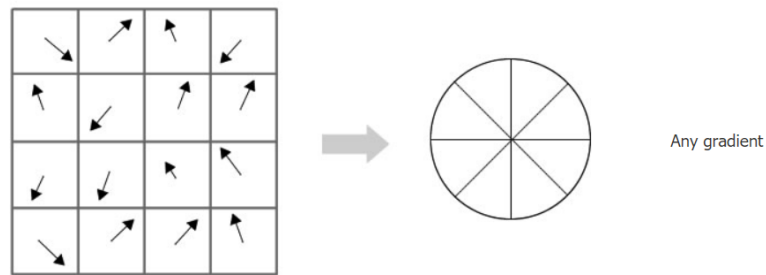


Figure 2.9: Allocating orientations into an 8-bin histogram [9]

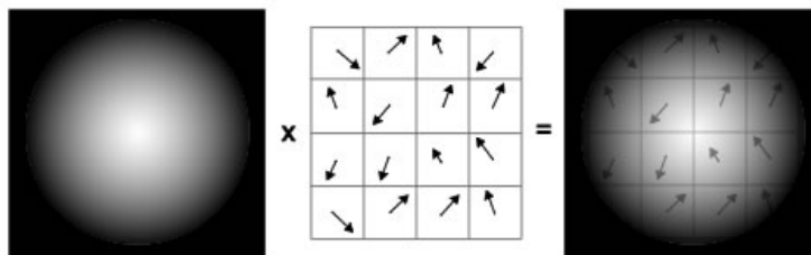


Figure 2.10: The weighted Gaussian filter is multiplied to the magnitude of orientations which results in the distance mentioned. [9]

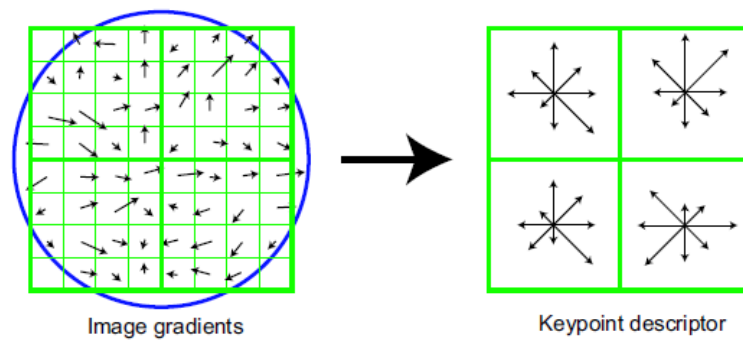


Figure 2.11: Feature descriptor for one keypoint. [9]

$4 \times 8 = 128$  numbers. These 128 numbers form the feature vector/descriptor. Each keypoint in the image is identified by its own feature descriptor as shown in Fig. 2.11.

## 2.6 Matching Keypoints

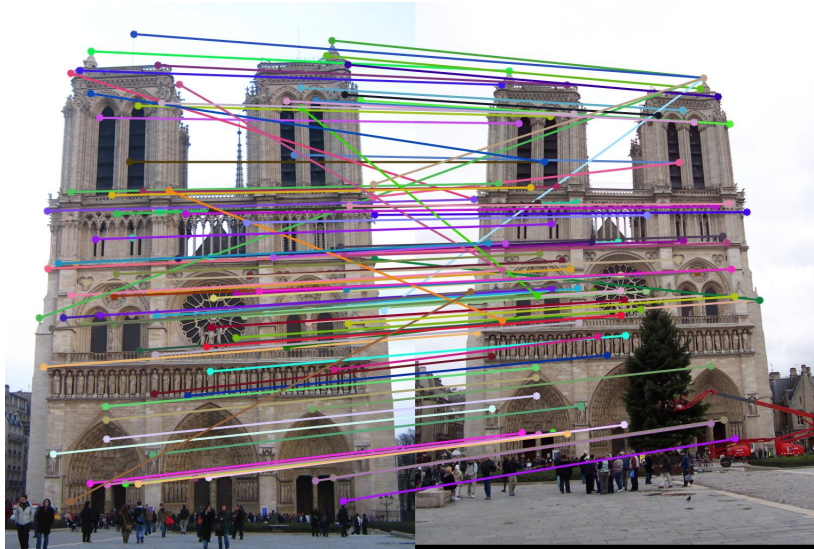


Figure 2.12: Matching SIFT keypoints (including outliers). [10]

Once the SIFT keypoints in the reference and sensed images are extracted, the next step is to *match* each keypoint descriptor of the reference image against the keypoint descriptor of another image.

This matching is accomplished by *finding the nearest neighbour* i.e. a keypoint with *minimum Euclidean distance*. The ratio between the smallest distance and second smallest distance is computed. Lowe rejects all matches which have a distance ratio greater than 0.8 [7]. In Fig. 2.12, correct matched keypoints (inliers) are depicted as *parallel lines* whilst incorrect matches (outliers) are depicted as *non-parallel lines*. Thus, the rejection of additional outliers is still required.

## 2.7 Elimination of Outliers: RANSAC

To eliminate outliers in keypoint matching one utilises an outlier detection method such as the *Random Sample Consensus (RANSAC) algorithm*. RANSAC estimates parameters of a model by random sampling of observed data. Given a set of data keypoints there are a number of inliers (parallel lines) and outliers, RANSAC uses a form of voting scheme to find the optimal set of inliers. The RANSAC model is an iterative model which is generally repeated for a *maximum number of iterations*.



The RANSAC algorithm starts by selecting  $n$  points at random. In this case two (marked in blue) as shown in Fig. 2.13.

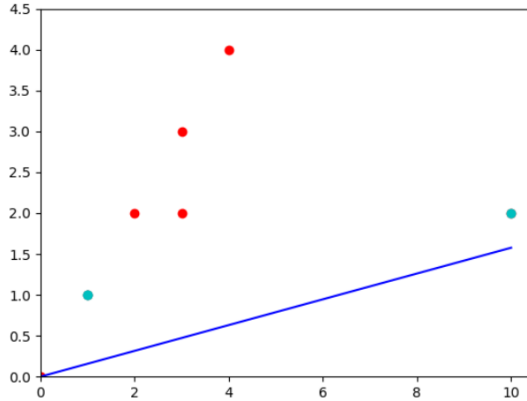


Figure 2.13: Step 1 of the RANSAC algorithm. [11]

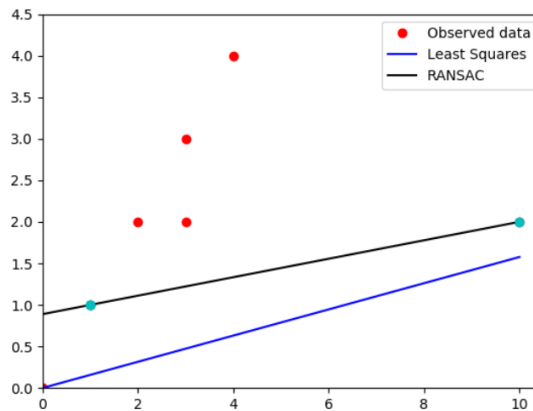


Figure 2.14: Step 2 of the RANSAC algorithm. [11]

In Fig. 2.14 assuming that both randomly chosen data points marked in blue are inliers, the least squares method is utilised to estimate the best straight line between these  $n$  points. In this case  $n = 2$ .

Considering a tolerance margin (*RANSAC threshold*) as shown in Fig. 2.15 - all the points lying within this margin are considered as inliers (for this first iteration!). From the shown testing data (not utilised to model the (black) straight line), one can observe that two blue (data) points are within the tolerance margin. *Thus, this model will give two inliers (in blue) and two outliers (in red). The above steps are then repeated for a maximum number of iterations.*

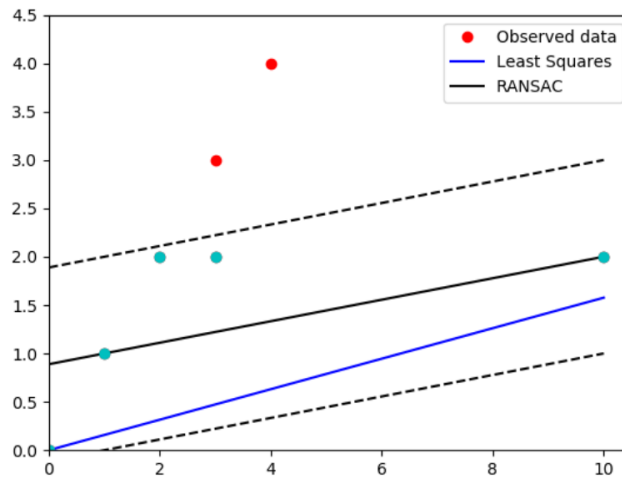


Figure 2.15: Step 3 of the RANSAC algorithm. [11]

Consider an experimental example where the chosen number of random data points is  $n = 3$ . The set maximum number of iterations for the RANSAC algorithm is 100 iterations and the set RANSAC threshold/tolerance margin for this example is 0.1.

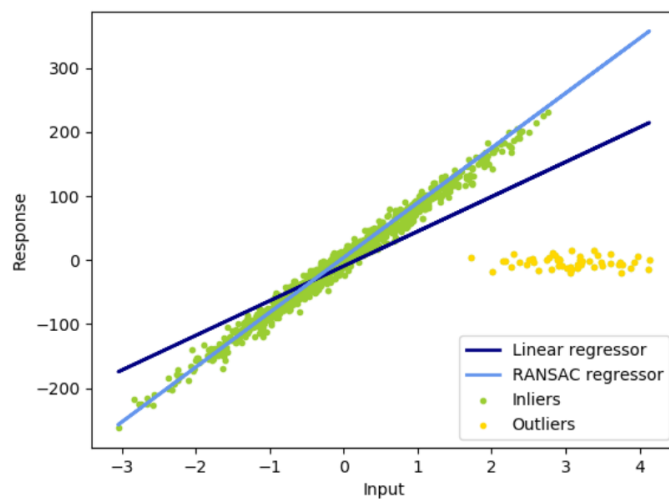


Figure 2.16: Regression model result using the Linear (Least Squares) and RANSAC regressors. [11]

In the experimental result shown in Fig. 2.16, the RANSAC algorithm performs better than linear (least squares) regressor. Whilst the RANSAC algorithm manages to ignore the outliers (marked in yellow) in the regression model, the linear regression is more susceptible to outliers.

# Chapter 3

## Literature Review

The first section of this chapter provides an overview of the literature concerning multi-multimodal and multi-sensor registration for satellite and remote-sensing imagery. The aim is to investigate the gap in literature with regards to the use of multi-modal sensors utilised.

### 3.1 Multi-Modal Image Registration for Satellite Imagery

Several approaches have been put forth to solve the problem of image registration in remote-sensing. Tondewad and Dale presented a review of different registration methodologies [49] with their involved challenges. In their findings they stated that in literature concerning image registration in the remote sensing domain, the most commonly used methodology are feature-based algorithms [49]. SIFT was found to be the most utilised feature detection algorithms [49]. Based on their work it was found that feature-based algorithms are more suitable for this task compared to area-based methods as feature-based methods focus more on features rather than intensity thus making area-based methods more sensitive to changes in illumination. Furthermore, their review also stated that utilising SIFT alone is not enough since SIFT based techniques tend to suffer from insufficient feature points and a high outlier ratio under severe appearance change. A suggested improved approach is to combine traditional based methods such as SIFT with neural networks such as a *Convolutional Neural Network (CNN)* [49].

### 3.1.1 Area-based Methods

S. Skakun et al. [50] performed registration of *optical-to-optical* satellite imagery. Multi-temporal Landsat-8 OLI images and Sentinel-2 MSI images at a spatial resolution of 30 m were utilised for registration. *Phase-Correlation* was utilised to generate control points between the two images. The RANSAC model was then utilised to establish the linear transformation model that detects inliers and outliers. S. Skakun et al. also applied a phase correlation method to detect pixel shifts between multi-spectral bands of Sentinel-2A. Thus, the phase correlation assessed the registration accuracy between intra multi-spectral bands [21]. Fig. 3.1 shows the misregistration in Sentinel-2A intra multi-spectral bands in pixels.

AVERAGE SHIFTS (IN PIXELS AT 10 m) WITH STANDARD DEVIATION  
BETWEEN MULTISPECTRAL SENTINEL-2A IMAGES

Scene	Number of points	Band combination	Shift, pixels
Dubai, 13 June 2017	154449	B3-B2	0.037±0.028
		B4-B3	0.033±0.025
		B4-B2	0.043±0.089
US, 15 June 2016	104329	B3-B2	0.049±0.057
		B4-B3	0.051±0.053
		B4-B2	0.041±0.063

Table 3.1: Table denoting the multispectral misregistration in Sentinel-2A data in pixels [21].

For the S2 L1C image, the red band (B4, an optical) was used as the reference image whereas the *panchromatic image* was utilised for L8. This panchromatic image was generated from three Sentinel-2 bands found in the visible spectra. Thus, the work proposed in [51] co-registered *optical-to-optical* satellite imagery. Combined binary cloud mask was first generated using the *FMask algorithm* [52]. The aim was to exclude clouds, snow or water before commencing registration. The images were then matched using an *image-correlation algorithm* [53] which yielded correlation coefficients. These coefficients were then mapped using a systematic affine transformation. The resulting homologous points were then improved further by using the RANSAC algorithm.

R. Behling et al. [54] co-registered *optical-to-optical* images coming from different satellites such as Landsat, RapidEye, SPOT-1, SPOT-5 and Aster. The focus was on the *multi-temporal* aspect of optical satellite imagery. Common tie-points between the reference and sensed images were generated using an *area-based cross-*

*correlation* method to warp the reference and the sensed. A 30 m spatial resolution was considered. The validation of tie-points which excluded potential outliers was established using an *affine transform* that considered translation, rotation and scale.

D. Scheffler et al. [55] co-registered satellite imagery coming from different satellites such as L8, S2A, and RapidEye at different spatial resolutions (such as 10, 15, 20 and 30 m). The work in [55] registered *optical-to-optical* images and *optical-to-radar*. A *phase-correlation* method in conjunction with the *Mean Structural Similarity Index (MSSIM)* similarity metric was utilised.

D. Evangelidis et al. [46] proposed a modified version of the correlation coefficient as the performance criterion for image registration. The proposed registration method is called the *Enhanced Correlation Coefficient (ECC)*. The proposed modification aimed at maximizing the Enhanced Correlation Coefficient function to be robust against geometric and photometric distortions. This method was compared to the Forward Additive Lukas-Kanade method. Under photometric and noisy conditions this proposed outperformed the Lukas-Kanade method by converging faster and achieving better image registration.

### 3.1.2 Feature-based Methods

L. Yan et al. [56] coregistered L8 OLI L1T and S2A MSI L1C *optical-to-optical* imagery in the NIR region. The proposed work utilises a hybrid-based method which consists of feature and area-based methods. The proposed method has four major steps: *Image Gaussian pyramids*: which is a hierarchy of low-pass images. The pyramid utilised contains low-pass images, from the lowest spatial resolution to the highest native spatial resolution. In the work proposed, four levels (120 m, 60 m, 30 m, 10 m) were utilised. *Features were detected in one image* (the S2A image which has higher native resolution). The *locally adaptive Forstner operator* was utilised at each pyramid layer to detect the features. An *Area-based matching* between the feature points was applied. This was done sequentially from the lowest to the highest spatial resolution pyramid level. The area-based matching was computed by comparing the NIR reflectance values over a small square image patch ( $n \times n$  pixels) surrounding each feature point obtained in the S2 image with the corresponding sized patch found in the L8 image. Rather than utilising the conven-

tional cross-correlation method, the *least square matching (LSM)* was utilised. The LSM results in a set of corresponding locations (tie-points) between the S2 and L8 images. Use of least-squares regression and various transformation functions such as translation, affine, and second order polynomial followed to compute the needed homography for image registration.

R. Ambati et al. [57] performed co-registration between *optical-to-optical* L8 L1T and S2 L1C bands using a *combined feature-based* method. The proposed algorithm has five major steps: *Extraction of feature points* using *SURF* [39] and *SIFT* [7] features. Matching of keypoints between the L8 and S2 images using *Fast Approximate Nearest Neighbor (FANN)* [58]. *Selection of best-keypoints* based on a threshold. The threshold value (20%) that corresponds to the percentage of the Euclidean distance from a perfect match. Chosen threshold value was found through empirical data that minimised the RMSE. *Estimation of homography* was computed using *conformal*, *affine* and *projective* transformations. *Image resampling* was then performed using *Nearest Neighbour*, *Bilinear* and *Bicubic* interpolation.

Eugenio et al. [59] proposed a feature-based algorithm for the registration of multitemporal and multisensorial oceanographic satellite imagery. *Optical and SST images* from different dates and sensors were utilised for registration. Examples of satellites utilised are MODIS and SeaWiFS. The proposed algorithm first utilised the MODIS cloud mask to exclude clouds. *Feature extraction* was then performed through contour extraction of the coastline region using the *Sobel* operator to estimate the edges. Furthermore, the cloud overlay was also utilised to obtain reliable non-occluded coastline areas (edges). *Feature matching* was performed through the use of an *algorithm that minimised local energy* [59]. The local energy refers to the matrix of gradient energies. Once points of the overlapping reliable edges were obtained, an *affine* transformation matrix which catered for translation, scaling and rotation was utilised.

Wahed et al. [60] performed registration between optical, panchromatic and near-infrared images from the Landsat and QuickBird satellites. Images utilised for registration all lie close the visible spectra. Both the reference and sensed image were first image enhanced through the use of a median filter. Then both images were also decomposed in a *Steerable Pyramid Transform* [61]. A steerable pyramid

transform is a linear multi-scale, multi-orientation image decomposition that consists of a class of filters. It was developed to overcome the limitations imposed by wavelet decompositions. For both images, features at every decomposition pyramid level were extracted using SIFT. Features in both images were matched by calculating the ratio of the Euclidean distance of the feature descriptor vector from the closest neighbour to the distance of the second closest. Outliers were then filtered out using RANSAC. The homography was then estimated using an affine transform. According to the obtained transform parameters image resampling on the sensed image was then performed.

Ding et al. [62] proposed a multi-sensor registration algorithm between *optical-to-infrared* images. Satellite images utilised were taken from the satellite, Seasat. Both images have a size of  $225 \times 225$  pixels. The proposed algorithm has four major steps: *Feature extraction* is done by extracting corners using the *Harris Corner Detector* algorithm. Once the corner points in the reference and sensed images were found, establishing the correspondences between the points in both images followed. The *corner point matching* was performed in two steps. Firstly, cross-correlation was utilised to find the cross correlation coefficients of every corner in both images. Secondly, the corresponding corners that have a cross correlation coefficient greater than a chosen threshold were selected. These were then sorted according to their value and the top twenty were chosen as the matched features. The removal of mis-matching corners was performed using a *mis-matching corner removal algorithm*. The algorithm utilised the *Mahalanobis distance* to remove mis-matches. After a minimum of at least three corner points are established, the affine matrix was utilised to compute the needed homography for image registration. The affine matrix was estimated using the Least Square Method. In the computed affine matrix scale, translation and rotation were considered. In the final step, based on the obtained affine transformation parameters, the sensed image was resampled on to the reference using bi-linear interpolation.

F. Dellinger et al. proposed an algorithm to register SAR images called *SAR-SIFT* [63]. In this work they proposed a modified version of the SIFT algorithm for SAR images. The major step changed in SAR-SIFT is the keypoint detection step. Rather than utilise the DoG approach they utilised the *LoG-Harris detector* [63] to detect keypoints to cater for the speckle-noise found in SAR images.

The SAR data utilised for this work originated from TerraSAR-X which has a 1 m and 2 m spatial resolution.

Xiang Y. et al. proposed an algorithm known as *Optical-to-SAR (OS) SIFT (OS-SIFT)* [64] to register *optical-to-SAR* satellite images. In the traditional SIFT algorithm, a Gaussian image is constructed by convolving the image with Gaussian filters at various scales then a series of DoG images were computed. This is an approximation of LoG. However, the LoG method cannot detect reliable key points for SAR images. The cause of detecting unreliable key points is due to speckle noise. Instead of constructing the DoG scale space, OS-SIFT constructs two Harris scale spaces. In the two Harris scale spaces, key points are detected. The local maxima in the Harris scale space was then established to find candidate key points at each level. This was followed by suppressing non-maxima and thresholding. The conventional SIFT method localises key points by using the Hessian matrix, which is not suitable for SAR data. OS-SIFT proposed a *localisation refinement* method to replace the step of the Hessian matrix.

Contrary to SIFT, OS-SIFT restricted the orientation between an interval  $[0, 180^\circ]$  due to the fact that in multisensor images it is very common that the gradients of corresponding parts of the images change their direction by  $180^\circ$ . The next step involved finding the feature descriptor. In the conventional SIFT algorithm a  $4 \times 4$  square neighbourhood is utilised. In OS-SIFT a GLOH-like circular neighbourhood with a radius of  $12\alpha$  and 17 location bins were utilised to form the feature descriptor. The next step is keypoint matching. The conventional SIFT algorithm matches key points by using a matching strategy that finds the distance between descriptors. In OS-SIFT, the NNDR method was utilised. The *Nearest Neighbour Distance Ratio (NNDR)* method [7] consisted of finding the nearest Euclidean distance between the descriptors. Then a threshold was applied on the ratio of the closest distance to the second closest distance to filter out unreliable matches. The outliers were then removed by using the *Fast Sample Consensus (FSC)* algorithm over the RANSAC. The FSC was found to establish more correct matches over RANSAC in less number of iterations [65].

J. Li et al. proposed a feature-based method for *multi-modal image matching* [12]. The proposed algorithm is called *Radiation-Variation Insensitive Feature Trans-*



form (*RIFT*). Traditional feature-based methods such as SIFT, utilise intensity or gradient information to detect feature points; nonetheless, such methods that utilise intensity or gradient information are easily affected by *nonlinear radiation distortions (NRD)*.

Instead of image intensity, RIFT used *phase congruency (PC)* for feature detection. PC is based on the convolution of a 2D image with even-symmetric and odd-symmetric wavelets. These wavelets resulted from the *2D Log-Gabor Filter*. On the PC map for feature detection, RIFT considers both *corners and edges*. To construct the feature vector, contrary to SIFT which utilises the histogram distribution, RIFT used a measure known as *Maximum Index Map (MIM)*. MIM is established from the log-Gabor convolution sequence. Through the utilisation of MIM maps, RIFT is also *rotation invariant*. Outliers after matching were removed by using the *Normalised barycentric coordinate system (NBCS)* method [66].

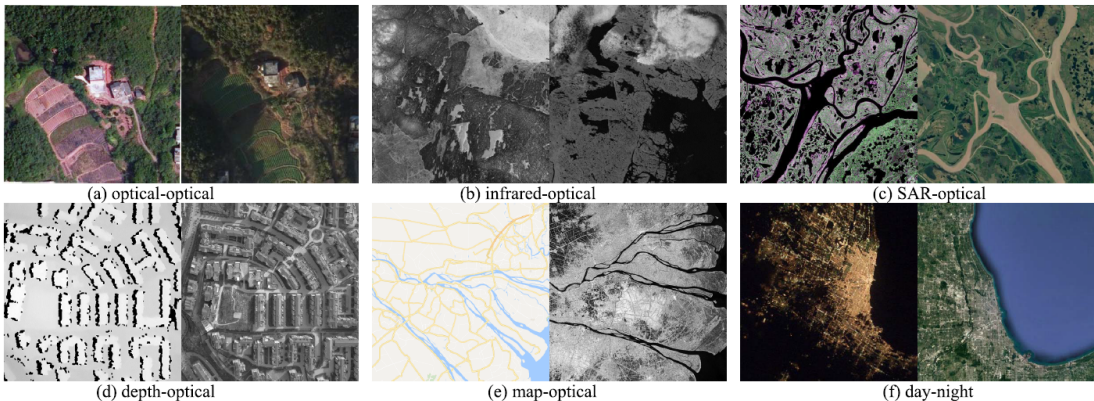


Figure 3.1: Multi-modal sample data utilised in RIFT [12].

As shown in Fig. 3.1, *six types* of multi-modal image datasets are utilised in RIFT. Among these, (a) has translation, small rotation and small-scale changes; (b) contains a translational change and a  $90^\circ$  rotation change; (c), (e) and (b) include both translational and rotation changes; (d) includes a translational change. Since these images are multi-modal pairs, they contain significant NRD.

Based on the results shown in Fig. 3.2, RIFT outperforms SIFT and SAR-SIFT [63]. This is due to the fact that gradient maps utilised in SIFT are very susceptible to NRD [12]. It is very important to note that in this work, the SIFT parameters

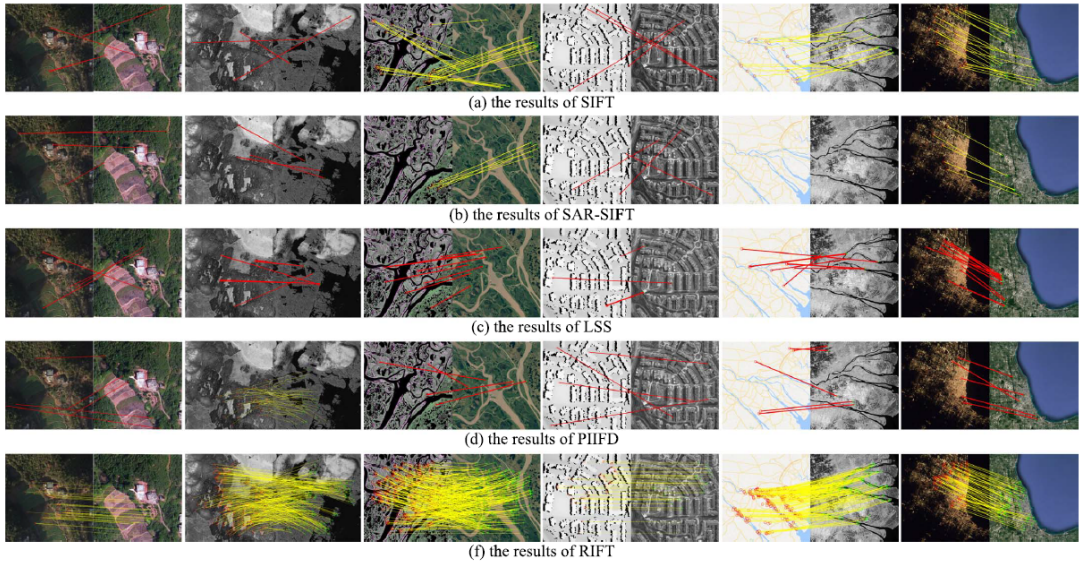


Figure 3.2: Multi-modal matching results [12]. The yellow and red lines refer to the correct matches and outliers, respectively.

utilised were the (default) Lowe SIFT parameters and that the RIFT algorithm is not invariant to significant scale differences due to a lack of scale space during feature detection. The authors of RIFT pointed this aspect in the limitations section [12].

### 3.1.3 Optical Flow-based Methods for Remote Sensing

The current state of the art Optical-flow algorithms for the registration of remote-sensing imagery are: *GeFolki* [2], *SIFT Flow* [48] and *OS-Flow* [13].

*GeFolki* [2] stands for *Geoscience Extended Flow Optical Lucas–Kanade Iterative*. GeFolki is an optical-flow method that registered SAR-Optical images and SAR-LIDAR images. The GeFolki algorithm is based on the classical *Lucas-Kanade* optical flow algorithm [67]. In GeFolki, a multi-scale implementation of the Lucas-Kanade method was utilised with various filters such as rank filtering, rolling guidance filtering and low contrast inversion.

SIFT-Flow [48] is an optical-flow algorithm that is based on SIFT features. The SIFT-Flow algorithm consists of matching densely sampled, pixel-wise SIFT features amongst two images. In the process spatial discontinuities are also preserve. The utilised SIFT features allow robust matching across different scenes. Addi-

tionally, the use of the proposed discontinuity preserving spatial model allows the matching of objects that are found at different parts of a scene [48]. The purpose of SIFT-Flow was to match non-rigid scenes and different instance of the same scene. Chen. J et al. [68] utilised SIFT-Flow to register remote sensing images. The chosen images were panchromatic aerial photographs provided by the Chinese Academy of Surveying and Mapping. These chosen images had structural deformation. To eliminate incorrect matches after matching the SAR-SIFT feature descriptors in the reference and sensed images, a modified version of RANSAC called *MAC-RANSAC* [69] was utilised.

The proposed *Optical-to-SAR Flow (OS-Flow)* [13] tries to find pixel wise correspondence between optical and SAR images. This OS-Flow involves three main steps: Establish the *Dense Feature Descriptors* for both images. For SAR images, the dense feature descriptors were extracted using the *SAR-SIFT* algorithm [70]. For the optical images, the *SOBEL edge detector* was utilised to compute the gradients in the optical images. To compute the descriptor similar to OS-SIFT, the *GLOH structure* with 17 location bins is utilised. The second step of OS-Flow involves *Flow-Estimation*. In OS-Flow *two optical flow methods were utilised*. The first method is a global method that used an objective function known as  $E_{OS-HS}$  which is based on the global *Horn and Schunck* method. Whereas the local method used an objective function  $E_{OS-LK}$  based on the *Lucas-Kanade* method. The third step involved the optimisation for *Flow-Field Computation* for both global and local methods. A coarse to fine pyramid structure approach was utilised. In the descriptor pyramid for OS-HS the *Belief Propagation* method was utilised for flow-field estimation whereas for OS-LK the *Gauss Newton* method was utilised. In the majority of the results both qualitatively and quantitatively, OS-Flow shows higher registration accuracy [13] as compared to GeFolki [2], SIFT-Flow [48] and OS-SIFT [64].

### 3.1.4 Evaluation of Registration Accuracy between Single and Multi-Modal Satellite Imagery

The biggest limitation of evaluating the registration of remote sensing images is **the lack of known ground truth**. If one took an image from the same satellite and registered that image with a shifted version of itself, the RMSE after registra-

tion should be zero as the ground truth is known. **But, in images coming from multi-modal sensors, the RMSE before registration and the amount of pixel shifts before registration are not known.**

### 3.1.4.1 Qualitative Comparisons

*Qualitatively*, one can utilise various approaches such as: *Error plots*, *Overlay plots*, *checker board and quadrant plots* and the utilisation of known control controls where one knows the parameters of an affine transform. Thus, the sensed image is mapped to the geocoded known location by using the affine transform.

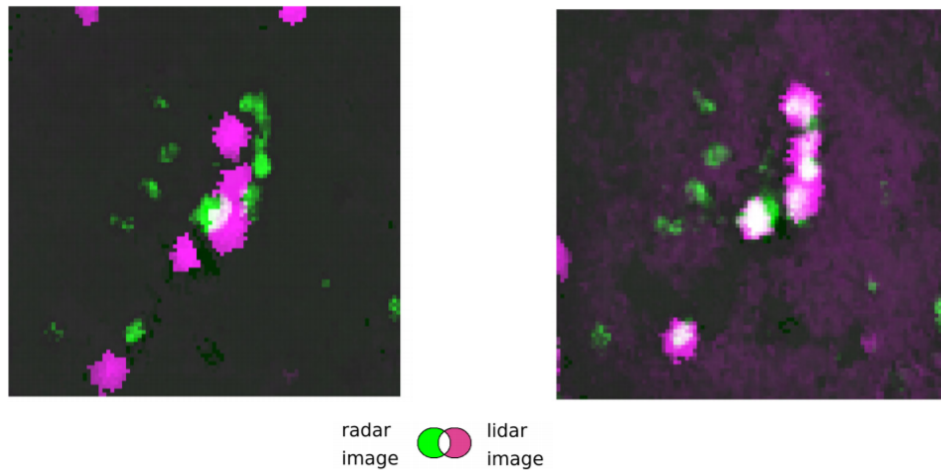


Figure 3.3: Example of an overlay plot. In this example radar and lidar images are registered using the GeFolki algorithm [2]. The image on the left is the result before registration and the image on the right is the result after registration. The radar image is the reference image.

In *overlay plots*, the reference, sensed and registered are assigned colours. For instance the reference image is assigned a magenta colour whilst the sensed and registered images are assigned a green colour. In this plot, the reference is overlaid/blended with the sensed image before registration. If the images are not well registered one should see an offset in the colours. Whereas when the reference image is overlaid with the registered image, assuming very good registration, the magenta and green colours should overlap in the well registered regions. An example of such an overlay plot is shown in Fig. 3.3.

In the case of *error plots*, if one considers the registration of a coastal region. Depending on the amount of misregistration between the two images to be registered, the error field generated before registration in the coastal regions should result in thicker coastal edges whilst after registration one expects the lines at the coastal edges to be thinner due to improved subpixel registration accuracy.

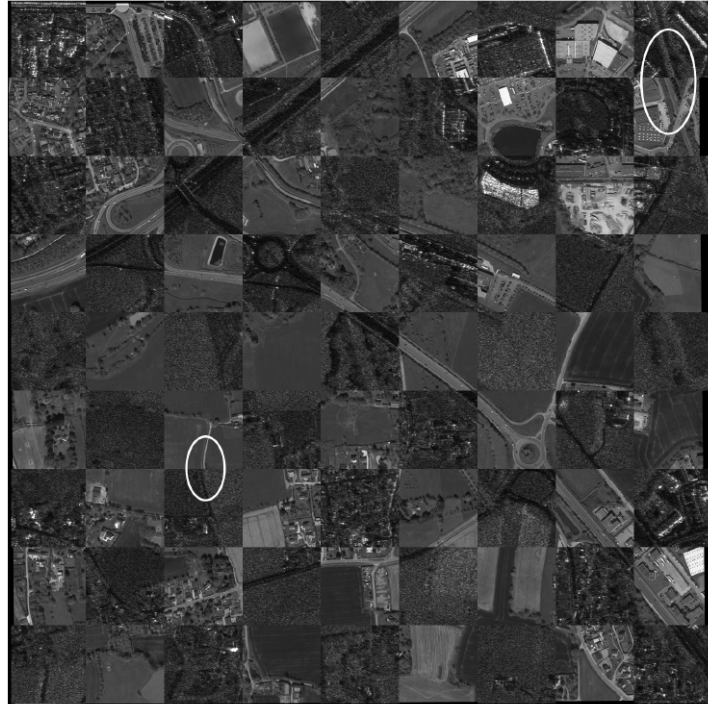


Figure 3.4: Example of a checkerboard plot. This example is taken from the proposed algorithm, OS-flow [13]. The image shows the misregistration between optical and SAR images.

*Checker board* and *quadrant* plots consist of combining images of the same scene in the form of a checker board to form a single image. The idea is to use an image that is already well registered and combine this with the registered image. If the image is well registered the formed checker board image should lead to the same structure as the well registered images but with an induced change in texture similar to a checker board. When using such an approach, typical structures such as roads and roundabouts are utilised to assess qualitatively the misregistration in pixel shifts. An example of a checkerboard plot is shown in Fig. 3.4.

### 3.1.4.2 Quantitative Comparisons

In the presented review of different registration methodologies in the previous subsections of Section 3.1, the authors found that the mostly commonly used quantitative metric for the registration of remote sensing imagery using feature-based and area-based methods was the RMSE [49]. Another quantitative metric utilised in remote sensing is the *Mean Absolute Error (MAE)*. The equations of the RMSE and MAE are denoted as follows:

1. Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i,j} [I_2(x_i, y_j) - I_1(x_i, y_j)]^2} \quad (3.1)$$

2. Mean Absolute Error (MAE)

$$MAE = \frac{1}{N} \sum_{i,j} |I_2(x_i, y_j) - I_1(x_i, y_j)| \quad (3.2)$$

where  $N$  is the total number of pixels in the image,  $(x, y)$  are the spatial coordinates, image  $I_1$  is the ground truth image which refers to the reference image and image  $I_2$  is the sensed or registered image.

The smaller the RMSE or MAE the more identical the two images are, with similar images having a RMSE or MAE of 0.

With the use of the RMSE metric, the biggest limitation is its *sensitivity to outliers* [71]. For the application of image registration this implies that if an image has any zeros and the corresponding image does not have any zeros in those pixel locations the RMSE will increase due to this bias. Thus, in practice, it might be justifiable to discard these outliers when computing the RMSE [71]. For the work of this dissertation any NaNs or zeros are excluded from RMSE calculation. This step is required because if zeros are not excluded, it would affect the min-max normalisation to convert the reference and sensed images to uint8. The joined cloud masks are also utilised to exclude cloud pixels from Landsat-8 and Sentinel-3 images.

One distinct advantage of the RMSE metric over the MAE is that RMSE avoids the use of the absolute value [71]. Thus, contrary to the MSE, the

RMSE does account for negative or positive values. This aspect is very important for mathematical calculations. For instance, it might be hard to calculate the gradient or sensitivity of MAEs. The RMSE is also better in terms of reflecting performance when dealing with large error values whilst the MAE is less biased for larger values and might not adequately reflect the actual performance when dealing with larger error values.

# Chapter 4

## Investigation

In this chapter, the multispectral intra-band misregistration of both Sentinel-2 and Landsat-8 will be assessed separately. Secondly, the misregistration of equivalent bands between Landsat-8 and Sentinel-2 is also assessed. Lastly, the misregistration of equivalent thermal bands between Landsat-8 and Sentinel-3 is examined.

Phase-correlation is an area-based method that is utilised to estimate the relative translative offset between two images. Contrary to cross-correlation which works in the spatial domain, phase correlation works in the Fourier domain. In the remote-sensing domain, the obtained translative offset is equivalent to the misregistration.

The utilised Phase-Correlation method for this entire chapter is based on the work proposed by H. Foroosh et al. [72] where they proposed a phase correlation method for the subpixel registration of images of different nature and across different spectral bands.

### 4.1 Assessing Multispectral Misregistration for Sentinel-2

Fig. 4.1 illustrates the registration through geocoding of bands 2, 3, 4 and 8 for multispectral bands of Sentinel-2B. The size of the quadrant diagram is  $500 \times 500$  pixels. The joined circle which is formed as a result of the four geocoded bands shows that the geocoding registration amongst the multispectral bands of a spatial resolution of 10 m is of excellent quality.



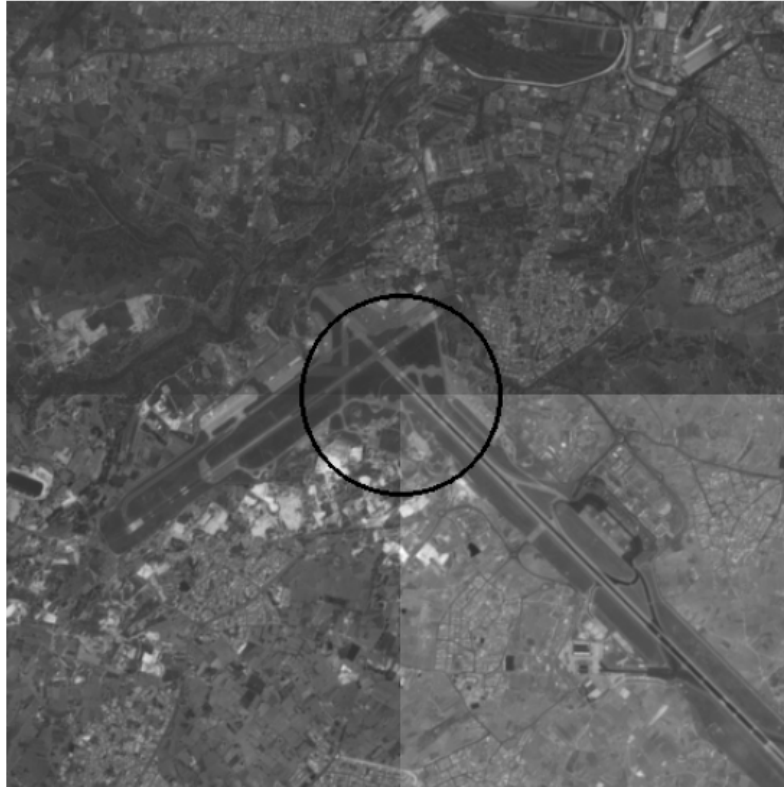


Figure 4.1: Plot of concatenated subsets in grayscale to assess qualitatively the misregistration of geocoding in Sentinel-2B with level L1C processing taken on 2018/04/04. Location is the runway of the Malta International airport. Band 2 (Top Left), Band 3 (Top Right), Band 4 (Bottom Left) and Band 8 (Bottom Right). Spatial resolution of all bands is 10 m. Cloud cover is at 0%.

Reference Band	Sensed Band	Misregistration (pixels)		
		$\partial x$	$\partial y$	$\partial x^2 + \partial y^2$
B2	Quadrant	0.0674	0.0752	0.1010
B2	B3	0.1543	0.0801	0.1738
B2	B4	0.0930	0.1139	0.1471
B2	B8	0.0973	0.1586	0.1860

Table 4.1: Phase correlation between the Sentinel-2B bands with a spatial resolution of 10 m.

The quadrant image utilised in Table 4.1 is the same as used in Fig. 4.1 but without the circle. The chosen reference image is Band 2. Since all chosen bands have the same spatial resolution, no interpolation was required for the sensed image. Size of both the reference and sensed images in Table 4.1 are  $500 \times 500$  pixels. All chosen band subsets originate from the same date as utilised in Fig. 4.1.

Table 4.1 shows the obtained misregistration values between the reference and respective sensed images. The misregistration was calculated using the phase corre-

S2B				
Bsec/Bref	B01	B04	B05	B11
B02		0,124		
B03		0,092		
B06			0,069	
B07			0,077	
B08		0,123		
B8A			0,092	
B09	0,238			
B11			0,172	
B12			0,167	0,104

Figure 4.2: Multi-Spectral co-registration performance for S2B.

lation method proposed by by H. Foroosh et al. [72]. From this table, it was shown that ESA indeed reached the required target of 0.3 pixel for the co-registration of multispectral bands for Sentinel-2B. Fig. 4.2 highlights the latest co-registration performance for the multispectral bands of Sentinel-2B products. In meters, the highest misregistration for Table 4.1 for a spatial resolution of 10 m was 1.86 m. Key difference between S2A and S2B is that S2B has better control of on-board vibrations.

The added red circles depicted in Fig. 4.3 outline subtle details of the accurate qualitative registration obtained through geocoding for Sentinel-2B bands 2, 4 and 8. Despite a cloud cover percentage of 100% qualitative registration is very good.

The bands utilised in Table 4.2 have the same location, date, size and cloud cover percentage as utilised in Fig. 4.3. Bands 3, 4 and 8 did not require any upsampling as they have the same spatial resolution as band 2 (10 m).

Fig. 4.4 highlights the latest co-registration performance for the multispectral bands of Sentinel-2A products which was taken from the Sentinel-2 product data quality report issue 55 of September 2020 [73]. This report states that the *ESA target is of 0.3 pixels* [73]. With the exception of the cirrus band (B10) and the SWIR band (B11), from Table 4.2 it was shown that ESA also reached the required target of 0.3 pixels for the co-registration of multispectral bands for Sentinel-2A. In all cases in Table 4.2, the misregistration was below 0.65 pixels. For a spatial resolution of 60 m, the achieved misregistration in meters was not greater than 20 m.

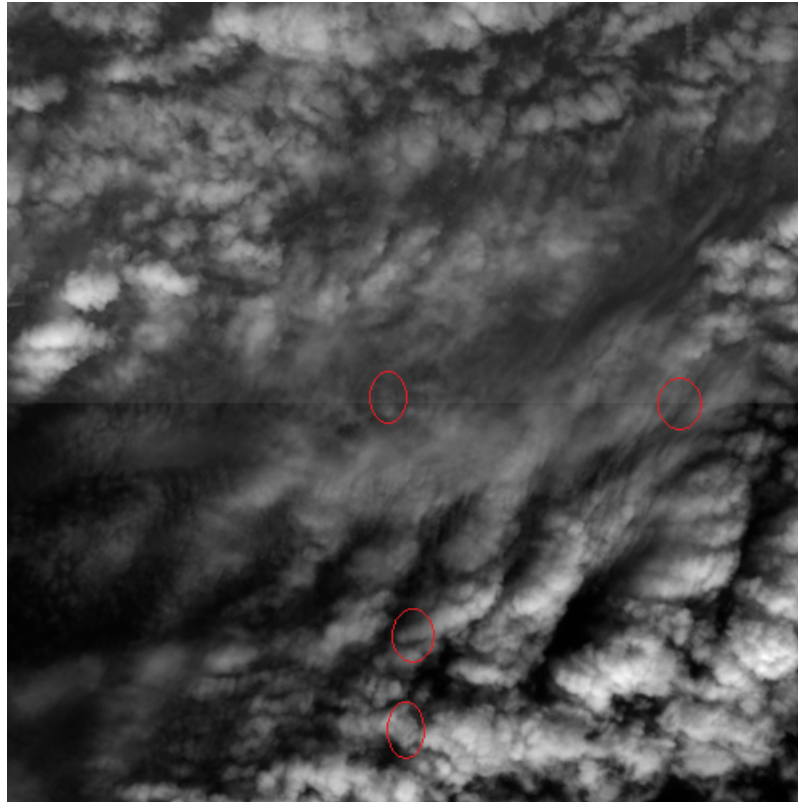


Figure 4.3: Quadrant diagram covering the runway of the Malta Interational airport. Bands are extracted from a Sentinel-2A L1C dataset on 2017/11/04. Top half of quadrant consists of band 2. Bottom left is extracted from B4. Bottom right is extracted from B8. Spatial resolution of all bands is 10 m. Cloud cover is at 100%. Quadrant size is of  $500 \times 500$  pixels.

S2A				
Bsec/Bref	B01	B04	B05	B11
B02		0,188		
B03		0,236		
B06			0,146	
B07			0,199	
B08		0,199		
B8A			0,240	
B09	0,184			
B11			0,260	
B12			0,270	0,215

Figure 4.4: Multi-Spectral co-registration performance for S2A.

Reference		Sensed		Misregistration (pixels)		
Band	U.F. <sup>a</sup>	Band	S.R.(m) <sup>b</sup>	$\partial x$	$\partial y$	$\partial x^2 + \partial y^2$
B2	/	B3	10	0.1206	0.0722	0.1405
B2	/	B4	10	0.1228	0.3185	0.3413
B2	/	B8	10	0.1573	0.0222	0.1589
B2	2	B5	20	0.2867	0.0182	0.2873
B2	2	B6	20	0.2757	0.2481	0.2808
B2	2	B11	20	0.2777	0.0502	0.2823
B2	2	B12	20	0.1899	0.6076	0.6366
B2	6	B9	60	0.2213	0.2109	0.3057
B2	6	B10	60	0.2314	0.3281	0.4015
B2	6	B1	60	0.2272	0.2184	0.3152

Table 4.2: Phase correlation for the same location of Fig. 4.3 between bands of Sentinel-2A with different spatial resolutions. The reference (B2, 10 m) remained with the same spatial resolution whereas the sensed image was upsampled to the same spatial resolution as the reference. The interpolation method used is bicubic interpolation.

<sup>a</sup>Upsampling Factor

<sup>b</sup>Spatial Resolution

## 4.2 Assessing Multispectral Misregistration for Landsat-8

Requirement	Specification
TIRS-to-OLI Registration Accuracy	30 m
TIRS Band-to-Band Registration Accuracy	18 m
TIRS Absolute Geodetic Accuracy	76 m
TIRS Geometric (L1T) Accuracy	42 m
TIRS Image Registration Accuracy	45 m

Table 4.3: Landsat-8 TIRS Geometric Performance Requirements [22].

Requirement	OLI Specification
OLI Absolute Geodetic Accuracy	65
OLI Geometric (Terrain Corrected) Accuracy	12
OLI Band-to-Band Registration Accuracy	4.5
OLI Image Registration Accuracy	12

Table 4.4: Landsat-8 OLI Geometric Performance Requirements [23].

Absolute geodetic accuracy refers to the geolocation accuracy of geometrically corrected products prior to the application of ground control points. The geolocation

accuracy achieved is mainly dependent on the spacecraft performance such as the satellite trajectory and its velocity. The geolocation accuracy also depends on knowledge of the OLI instrument. [23]

The OLI band-to-band registration accuracy specification defines the accuracy with which corresponding Level 1T product pixels from different OLI spectral bands must be co-aligned. This requirement only applies to products after geometric correction and resampling, i.e L1T products [23]. For TIRS Band-to-Band registration accuracy it is the same notion as OLI Band-to-Band, but applied for the TIRS sensor.

The term utilised in both Tables 4.3 and 4.4 of *image registration accuracy* in this case refers to the quantification of the accuracy with which corresponding L1T product pixels from images of the same area acquired at separate times (multitemporal), must be co-aligned. Thus, for this subsection the two salient aspects are the *TIRS-to-OLI Registration Accuracy* and *OLI Band-to-Band Registration Accuracy* which are 30 m and 4.5 m respectively. Thus, for an Landsat-8 OLI to OLI the tolerated misregistration for 1 pixel is 4.5 m.



Figure 4.5: Landsat-8 B2 subset

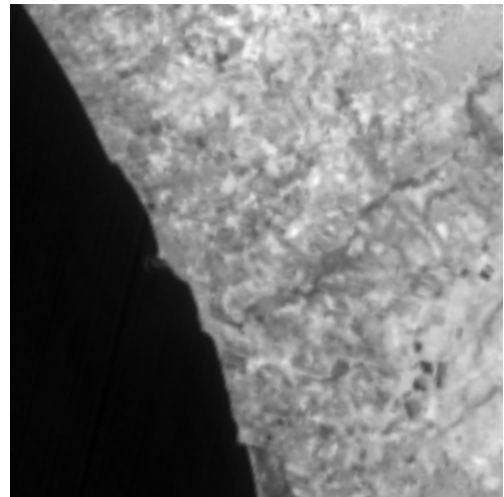


Figure 4.6: Landsat-8 B10 subset

Both Figures 4.5 and 4.6 have a size of  $500 \times 500$  pixels and they were both extracted from the same Landsat-8 L1T product on 2018/07/17. The reference image is Fig. 4.5 and the sensed image is Fig. 4.6. No interpolation was required for the sensed image as B10 has the same pixel size as B2 (B10 is resampled at 30 m grid size). At-

tained phase correlation amongst Figures 4.5 and 4.6:  $\sqrt{\partial x^2 + \partial y^2} = 0.6222$  pixels.

Reference	Sensed		Misregistration (pixels)		
Band	Band	S.R. (m)	$\partial x$	$\partial y$	$\sqrt{\partial x^2 + \partial y^2}$
B2	B3	30	0.0319	0.0685	0.0755
B2	B1	30	0.0059	0.0262	0.0268
B2	B4	30	0.0231	0.0861	0.0891
B2	B5	30	0.0422	0.0423	0.0953
B2	B6	30	0.0370	0.0576	0.0685
B2	B7	30	0.0082	0.0712	0.0717
B2	B9	30	0.2854	0.4782	0.5569
B2	B10	100	0.5066	0.3613	0.6222
B2	B11	100	0.5434	0.3357	0.6388
B2	B8	15	0.1959	0.2561	0.3225

Table 4.5: Phase correlation for the same location of Figures 4.5 and 4.6 between bands of Landsat-8 with different spatial resolutions. The chosen reference image is B2 with a spatial resolution of 30 m. Only B8 needed to be interpolated. It was downsampled by a factor of 2 to match the same spatial resolution as B2. The interpolation method used is bicubic interpolation. Bands 10 and 11 were not upsampled because they have the same image pixel size as B2 (but different spatial resolution).

In Table 4.5, for the optical bands 1, 3, 4 and 5 the largest obtained misregistration was of  $0.0953 \times 30 = 2.859$  m which is less than the tolerated misregistration of 4.5 m between OLI Band-to-Band Registration accuracy as quoted from Table 4.4. For the panchromatic band (in Table 4.5), B8, the obtained misregistration was of  $0.3225 \times 30 = 9.675$  m. For both the SWIR bands, B6 and B7, the largest obtained misregistration was of  $0.0717 \times 30 = 2.151$  m. For the cirrus band, B9, the obtained misregistration was of  $0.5569 \times 30 = 16.707$  m. For the two thermal bands the largest obtained misregistration was  $0.6388 \times 30 = 19.164$  m misregistration of 30 m between TIRS-to-OLI Registration accuracy as quoted from Table 4.4.

### 4.3 Assessing Multispectral Misregistration between Sentinel-2 and Landsat-8

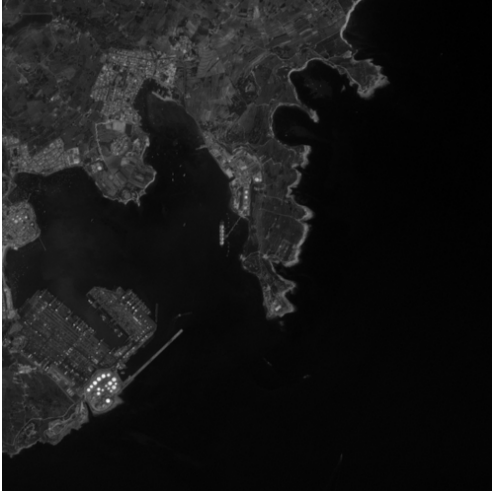


Figure 4.7: Sentinel-2 B2 subset



Figure 4.8: Landsat-8 B2 subset

Both Figures 4.7 and 4.8 have a size of  $500 \times 500$  pixels. The Landsat-8 subset was extracted from an L1T product on 2018/07/04 whereas the Sentinel-2 subset was extracted from an L1C product on 2018/07/04. The gap in time between both products is 17 minutes. Location: South Eastern region of Malta, next to Wied il-Ghajj. The sensed imaged which has a spatial resolution of 30 m needed to be resampled to the same spatial resolution as band 2 of S2 (10 m). Interpolation method utilised is bicubic interpolation. Attained phase correlation between the subsets:  $\sqrt{\partial x^2 + \partial y^2} = 0.2927$  pixels.

In Table 4.6, band 2 was assumed as the reference image whereas the Landsat-8 bands were assumed as the sensed images. The sensed subsets were upsampled using bicubic interpolation to obtain the same spatial resolution as the reference image. The highest obtained misregistration was between B2 of Sentinel-2 and B9 (cirrus) of Landsat-8, which represents a misregistration of  $0.4309 \times 10 = 4.309$  m which is less than the tolerated misregistration of 4.5 m between OLI Band-to-Band Registration accuracy as quoted from Table 4.4. Thus, from Table 4.6 it was also shown that the registration between multispectral bands of Sentinel-2 and Landast-8 is also very good. The optical-to-optical bands between Sentinel-2 and Landsat-8 bands achieved the overall lowest misregistration.

Reference	Sensed			Misregistration (pixels)		
	Band	U.F.	Band	S.R. (m)	dx	dy
B2	3	B1	30	0.2278	0.1831	0.2922
B2	3	B2	30	0.2362	0.1729	0.2927
B2	3	B3	30	0.2774	0.1166	0.3009
B2	3	B4	30	0.2777	0.0913	0.2923
B2	3	B5	30	0.2036	0.0482	0.2092
B2	3	B6	30	0.2378	0.0062	0.2379
B2	3	B7	30	0.2525	0.0466	0.2567
B2	1.5	B8	15	0.2903	0.2548	0.3862
B2	3	B9	30	0.3542	0.2455	0.4309
B2	3	B10	100	0.1609	0.1326	0.2085
B2	3	B11	100	0.1687	0.1369	0.2173

Table 4.6: Phase correlation between Band 2 of Sentinel 2 and Landsat 8 Bands 1-11. Same location and gap in time was chosen as Figures 4.7 and 4.8 between the reference and sensed images. *Note:* Bands 10 and 11 were upsampled with a scale factor of 3 and not 10 because the band is being represented in a 30 m grid despite a spatial resolution of 100 m.

## 4.4 Assessing Multispectral Misregistration between Sentinel-3 SLSTR and Landsat-8 TIRS

The utilised phase correlation in sections of this chapter proposed by H. Foroosh et al. [72] was *meant for the sub-pixel registration for similar images*. Even if the images were similar, but one of them had a significant translational difference, the obtained results for this phase correlation method would be inaccurate. *The H. Foroosh phase correlation method should only be utilised for identical images with subpixel misregistration [72].*

The phase correlation method is known to fail in the following situations:

- Error caused from nonoverlapping regions amongst the two images.
- Errors caused from periodicity due to not having a single peak.
- Errors caused due to wide band random noise.
- Errors caused due to aliasing. The phase correlation method is known to fail for large scale and rotation differences due to aliasing in low frequencies, which result in false peaks, resulting in an inaccurate alignment [74].

Both Figures 4.9 and 4.10 were extracted on 04/07/2018 covering an area patch of 15 km  $\times$  15 km. Chosen location is Gozo. The time gap between L8 B10 and S3 B8 is



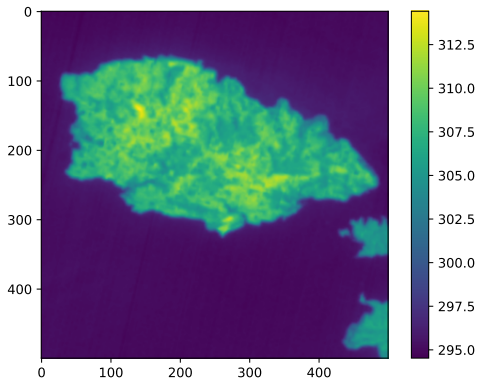


Figure 4.9: Native Landsat-8 B10 subset of size  $500 \times 500$  pixels

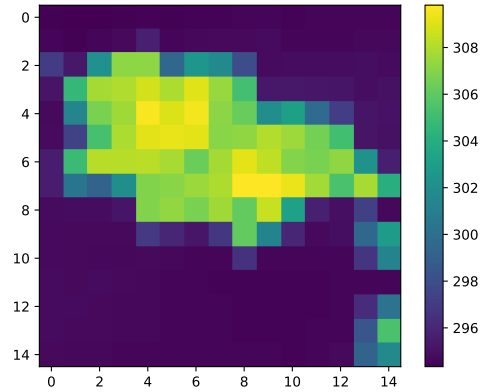


Figure 4.10: Native Sentinel-3 B8 subset of size  $15 \times 15$  pixels.

8 minutes apart. Since the resampled spatial resolution of L8 B10 and S3 B8 is 30 m and 1 km per pixel the native size is  $500 \times 500$  pixels and  $15 \times 15$  pixels respectively.

*It is important to note that the Landsat-8 thermal B10 is initially in radiance and this had to be converted to brightness temperature as the Sentinel-3 thermal B8 is in brightness temperature. This conversion will be explained in the next chapter.*

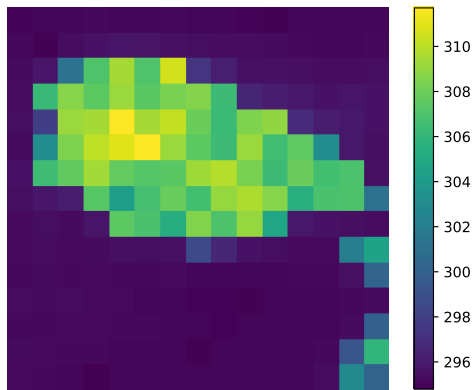


Figure 4.11: Downscaled Landsat-8 B10 subset to a size of  $15 \times 15$  pixels

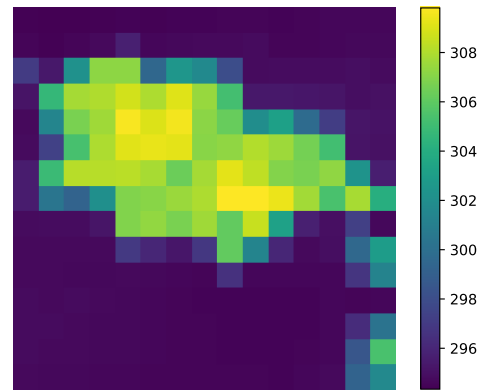


Figure 4.12: Native Sentinel-3 B8 subset of size  $15 \times 15$  pixels.

In Fig. 4.11, Landsat-8 B10 was downscaled to the same spatial resolution as Sentinel-3 B8. The obtained misregistration (in pixels) between Figures 4.11 and 4.10 in the  $x$  and  $y$  direction is  $\partial x = 0.2082$  pixels and  $\partial y = 0.2295$  pixels respectively. In metres, the obtained misregistration in the  $x$  and  $y$  direction for a spatial resolution of 1 km is 208.2 m and 229.5 m respectively. For a spatial resolution of 30 m the obtained misregistration in the  $x$  and  $y$  direction is 6.246 m and 6.885 m

respectively.

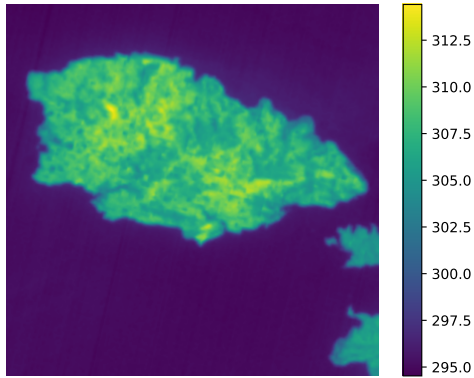


Figure 4.13: Native Landsat-8 B10 subset of size  $500 \times 500$  pixels

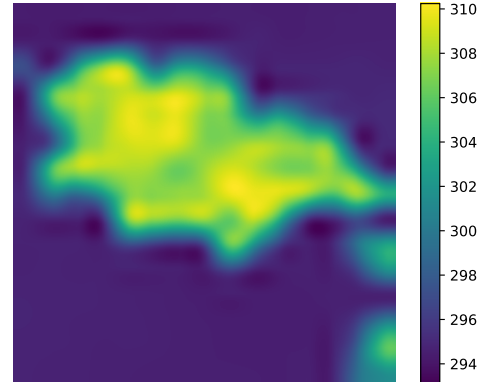


Figure 4.14: Sentinel-3 B8 subset upscaled to a size of  $500 \times 500$  pixels.

In Fig. 4.14, Sentinel-3 B8 was upscaled to the same spatial resolution as Landsat-8 B10. The obtained misregistration (in pixels) between Figures 4.13 and 4.14 in the  $x$  and  $y$  direction is  $\partial x = 0.1034$  pixels and  $\partial y = 0.1661$  pixels respectively. In metres, the obtained misregistration in the  $x$  and  $y$  direction for a spatial resolution of 1 km is 103.4 m and 166.1 m respectively. For a spatial resolution of 30 m the obtained misregistration in the  $x$  and  $y$  direction is 3.102 m and 4.983 m respectively.

# Chapter 5

## Methodology

The aims of this chapter are the following:

- Outline what tools were utilised to implement the main methodology/algorithm and other subsidiary algorithms utilised for comparisons.
- Give a detailed explanation of how the problem was tackled through the implementation of the main algorithm. This detailed explanation also involves justifications for all decisions taken in the course of the solution.

### 5.1 Implementation

The main algorithm of this dissertation was implemented in Python version 3.7.6. Salient details of the programming tools utilised:

- Sentinel-3 data was read using the Python library *Satpy* [75] version 0.20.0.
- The Landsat-8 data was read using the Python library *rasterio* [76] version 1.1.1.
- The Sentinel-S3 cloud masks were extracted from the *cloud\_in* band found in the Sentinel-3 SLSTR *L1 SL\_1\_RBT product*.
- The Landsat-8 cloud masks were extracted from the *BQA* band found in the Landsat-8 *L1T* data product.
- For the SIFT algorithm, the SIFT implementation from the *opencv-contrib-python* Python library [77] was utilised. Version 4.4.0.44.

- Implementation of the *RANSAC* algorithm is taken from the book [78] written by H. Singh.
- The proposed SIFT Python implementation for the registration of multi-modal and single modality data in Section 6.3 can be found in the following GitHub repository [79].

In addition to the main algorithm, subsidiary algorithms are also utilised for comparisons. Salient details of the programming tools utilised for the implementation of these subsidiary algorithms include:

- Unless stated otherwise, the implementation of these subsidiary algorithms was done in *MATLAB* version R2020a.
- The implementation of the Area-based method known as *Enhanced Correlation Coefficient (ECC)* is taken from the work proposed by Georgios Evangelidis [46, 80].
- The implementation of the Optical Flow based method known as *SIFT Flow* is implemented by the authors of the paper [48].
- For the implementation of the SIFT image registration in *MATLAB* with default parameters, the *VLF EAT* open source library was utilised [81]. This SIFT *MATLAB* implementation was only utilised for the next subsection for the preliminary comparisons between area-based, feature-based and optical flow-based methods. For the remainder of the results, the main Python SIFT algorithm was utilised.
- The utilised Phase Correlation method in Chapter 4 is based on the work by H. Foroosh et al. [72] implemented in Python.
- The implementation of the RIFT algorithm [12] is based on the work of the authors . It is implemented in C++ and *MATLAB* [82].
- The implementation of the SAR-SIFT is based on the work of F.Dellinger et al. [70] and also implemented by the authors in *MATLAB* and C++ [83].

## 5.2 Preliminary Comparison between Area-based, Feature-based and Optical Flow-based Methods

In this subsection the ECC (area-based) method, the SIFT RANSAC (feature-based) method and SIFT Flow (optical flow-based) method will be compared as a preliminary analysis to justify why a feature-based method was chosen over an area-based method. Generally area-based methods should fail to register images that have different intensities. As mentioned in Section 5.1 all these subsidiary algorithms were implemented in MATLAB. *In all the upcoming tests cases, the parameters used were the default parameters.*

### 5.2.1 Test Case 1: Optical vs Optical

The chosen satellite for the first test case is Kompsat-2 [84]. The data is acquired on 2009/08/17 at 09:07:55. Both chosen bands are extracted from the same Kompsat-2 product. The chosen reference image is Fig. 5.1 and the chosen sensed image is Fig. 5.2. Prior to registration one can observe that the sensed image has vertical misregistration.

In Fig. 5.3(a) magenta denotes the reference image and green denotes the sensed image. From this figure, the vertical misregistration is quite apparent as the green segment is not perfectly overlaid on the magenta segment.

In Figures 5.3(b), 5.3(c) and 5.3(d), the magenta represents the reference image and the green represents the registered image. Compared to Fig. 5.3(a), SIFT Flow did improve the registration (Fig. 5.3(b)), but the ECC and SIFT RANSAC obtained better qualitative registration as the green part overlaid exactly on the magenta part as shown in Figures 5.3(c) and 5.3(d) respectively.

Quantitatively, the ECC algorithm obtained the smallest RMSE followed by the SIFT RANSAC and SIFT Flow. It is to be noted that the ECC improvement in RMSE was marginal. The ECC algorithm was expected to work well as the intensities between the reference and sensed image in the first test case are similar.



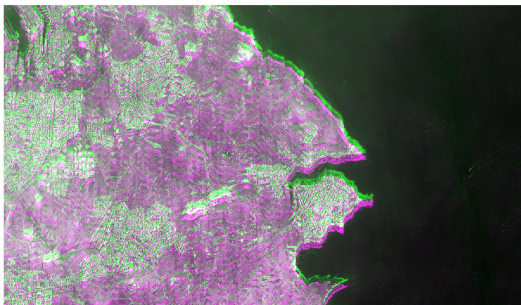
Figure 5.1: Kompsat-2 Panchromatic Band with 1 m spatial resolution of size  $1211 \times 692$  pixels. Location: East-North of Malta



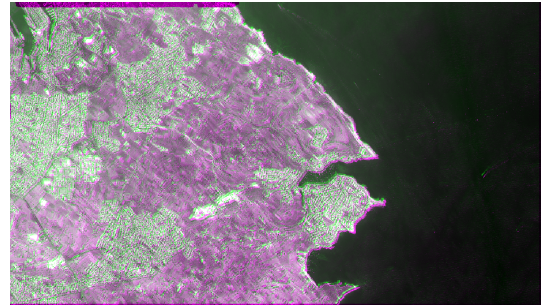
Figure 5.2: Kompsat-2 Band 2 with 4 m spatial resolution of size  $1211 \times 692$  pixels.

<b>RMSE Before Registration</b>		
9.5486		
<b>RMSE After Registration</b>		
<b>SIFT Flow</b>	<b>ECC</b>	<b>SIFT RANSAC</b>
8.6060	8.5590	8.5628

Table 5.1: RMSE results before and after registration for the first test case using SIFT Flow, ECC and SIFT RANSAC.



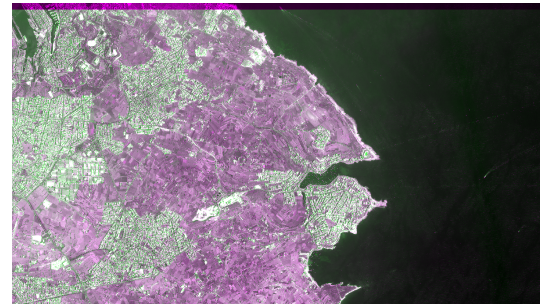
(a) Overlay of the reference and sensed image before registration



(b) Overlay of the reference and registered image after registration using SIFT Flow



(c) Overlay of the reference and registered image after registration using ECC



(d) Overlay of the reference and registered image after registration using SIFT RANSAC

Figure 5.3: Registration results before and after registration for the first test case (optical vs optical). Magenta denotes the reference image and green denotes the sensed or registered image. The registration methods utilised were SIFT RANSAC, SIFT Flow and ECC. The chosen dataset is taken from the Kompsat-2 satellite.

### 5.2.2 Test Case 2: Optical vs SAR

The chosen satellites for the second test case are Sentinel-1 and Sentinel-2. The chosen reference image is Band 3 extracted from a Sentinel-2A L1C data product acquired on 2020/07/31. The selected sensed image is the *Amplitude\_VH\_S1* SAR band extracted from a *S1B-IW-GRDH* data product acquired on 2019/11/11. The aims of this test case are to show registration between different modalities and to show multitemporal registration.

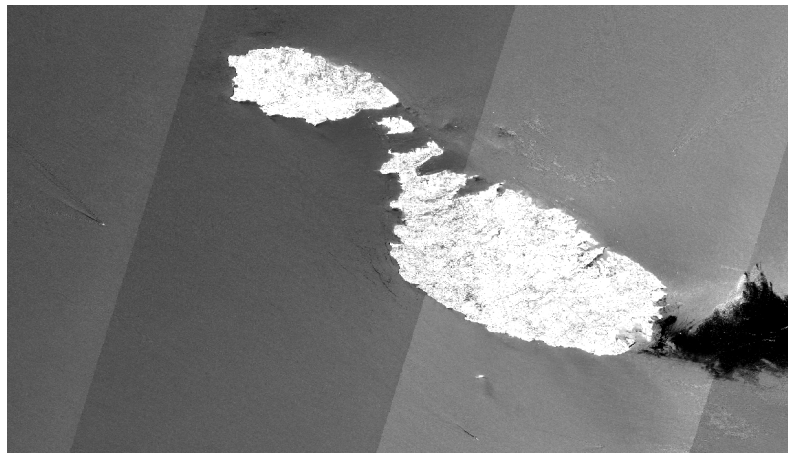


Figure 5.4: Sentinel-2 Band 3 subset of size  $1215 \times 692$  pixels. Location: Malta and Gozo.



Figure 5.5: Sentinel-1 Amplitude\_VH\_S1 Band subset of size  $1215 \times 692$  pixels.

As the registration between Fig. 5.4 and Fig. 5.5 was already very good, the sensed image in Fig. 5.5 (originally without translation) was translated in the x and y-direction by 10 pixels respectively.



Both Figures 5.4 and 5.5 were resampled and reprojected using the *Sentinel Application Platform (SNAP)* [85] toolbox. It is to be noted that the Sentinel-1 image was also terrain corrected using SNAP to remove any geometric distortions typically found in SAR imagery.

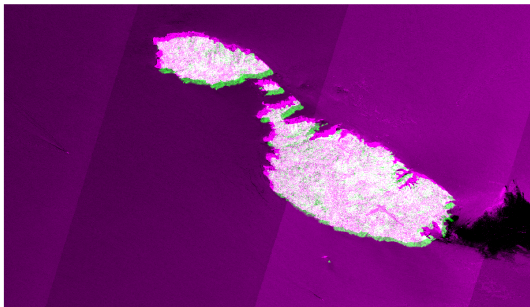
Fig. 5.6(a) illustrates the registration before in which the sensed image in green is translated in the x and y direction by 10 pixels respectively.

In all three algorithms, compared to Fig. 5.6(a) where the sensed image in green was translated by 10 pixels both in the x and y direction, the registration improved as shown in Figures 5.6(b), 5.6(c) and 5.6(d). In all three figures the green segment is overlapping on the magenta segment. The improvement in registration is evident in the coastal regions of Malta and Gozo.

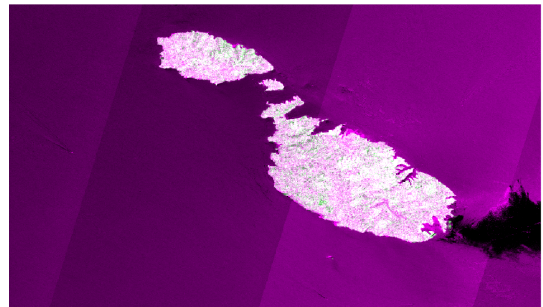
<b>RMSE Before Registration with Translated Sensed Image</b>		
2.4143		
<b>RMSE After Registration of the Translated Sensed Image</b>		
<b>SIFT Flow</b>	<b>ECC</b>	<b>SIFT RANSAC</b>
1.8048	1.8024	1.8864

Table 5.2: RMSE results before and after registration for the second test case using SIFT Flow, ECC and SIFT RANSAC.

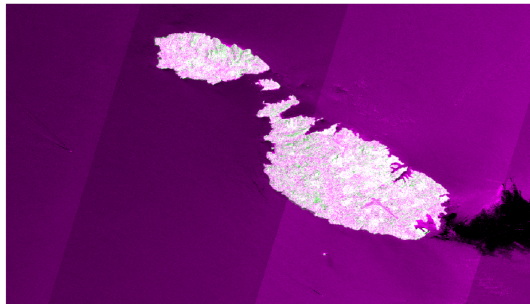
As shown in Table 5.2, SIFT Flow and the ECC algorithms resulted in similar RMSE followed by the SIFT RANSAC algorithm. Nonetheless, all three algorithms had a reduction in the RMSE for the registration of Fig. 5.4 and the translated image of Fig. 5.5.



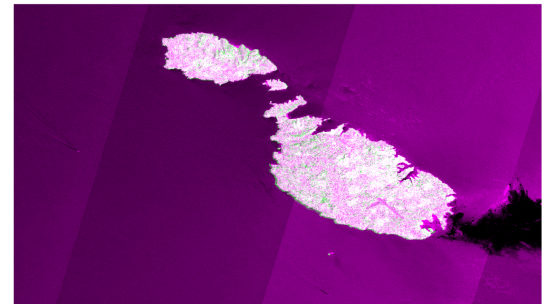
(a) Overlay of the reference and translated sensed image before registration



(b) Overlay of the reference and registered image after registration using SIFT Flow



(c) Overlay of the reference and registered image after registration using ECC



(d) Overlay of the reference and registered image after registration using SIFT RANSAC

Figure 5.6: Registration results before and after registration for the second test case (optical vs SAR). Magenta denotes the reference image and green denotes the sensed or registered image. The registration methods utilised were SIFT RANSAC, SIFT Flow and ECC. The chosen dataset is taken from the Sentinel-2 and Sentinel-1 satellites.

### 5.2.3 Test Case 3: Day vs Night

The chosen images for this subsection are taken from the multimodal dataset [82] of the RIFT algorithm [12]. The reference is the day view whilst the sensed image is the night view.

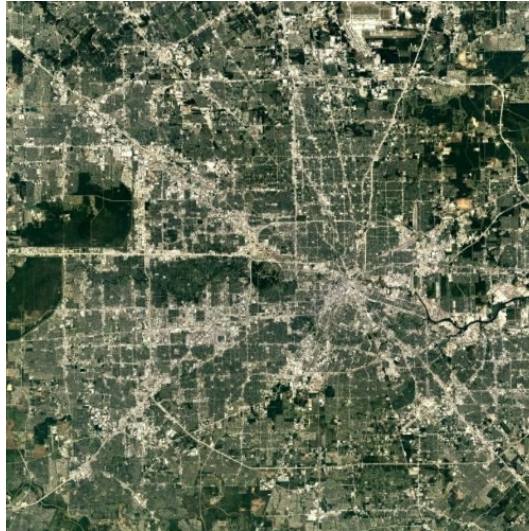


Figure 5.7: Reference image: day shot of size  $500 \times 500$  pixels taken from the RIFT multimodal dataset [12].

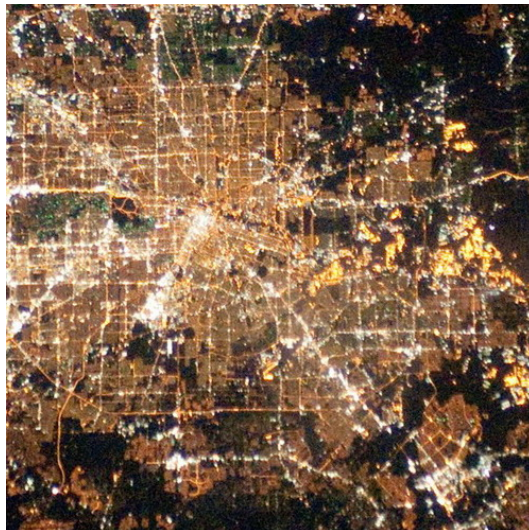
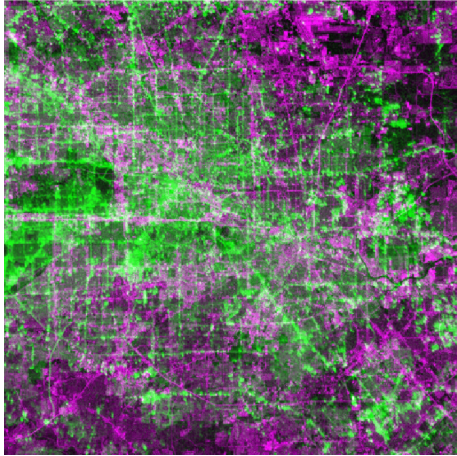


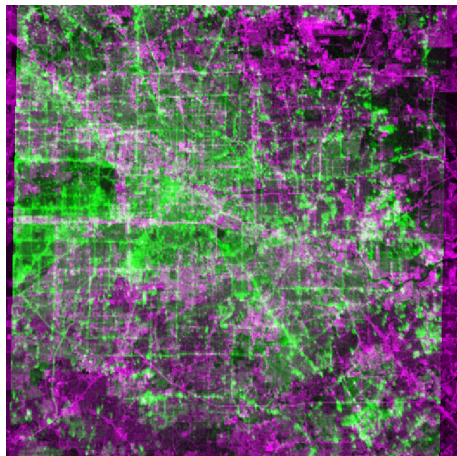
Figure 5.8: Sensed image: night shot of size  $500 \times 500$  pixels taken from the RIFT multimodal dataset [12].



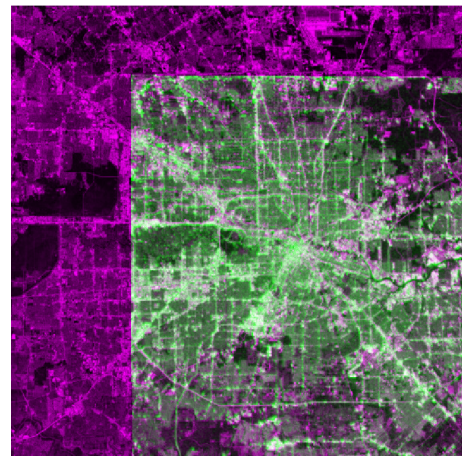
(a) Overlay of the reference and sensed image before registration



(b) Overlay of the reference and registered image after registration using SIFT Flow



(c) Overlay of the reference and registered image after registration using ECC



(d) Overlay of the reference and registered image after registration using SIFT RANSAC

Figure 5.9: Registration results before and after registration for the third test case (Day vs Night). Magenta denotes the reference image and green denotes the sensed or registered image. The registration methods utilised were SIFT RANSAC, SIFT Flow and ECC. The chosen dataset is taken from the RIFT dataset.

<b>RMSE Before Registration</b>		
10.9705		
<b>RMSE After Registration</b>		
<b>SIFT Flow</b>	<b>ECC</b>	<b>SIFT RANSAC</b>
10.9590	11.8	8.8586

Table 5.3: RMSE results before and after registration for the second test case using SIFT Flow, ECC and SIFT RANSAC. **Note:** a large RMSE is expected as the reference and sensed image have a significant difference in intensity.

The best registration from all three algorithms resulted from the SIFT RANSAC algorithm. This is both apparent qualitatively and quantitatively, as shown in Fig. 5.9(d) and Table 5.3 respectively. Additionally, contrary to a feature-based method (SIFT RANSAC), this test case showed that if the two images have different intensities the area-based method (ECC) will fail to register both images. Furthermore, SIFT even outperforms SIFT flow for the registration of the multi-modal aspect.

## 5.3 Proposed Methodology

In this section, the proposed methodology will be explained thoroughly. In the first part of this section, an illustration is given which depicts a high level overview of the proposed methodology. Subsequently, each step of the proposed methodology will be outlined in detail.

### 5.3.1 Pre-processing: Extraction of Reference and Sensed Images

In Fig. 5.10, the chosen reference and sensed images are Landsat-8 and Sentinel-3 Thermal brightness temperature (BT) bands respectively. Originally, the Landsat-8 thermal band is in *radiance*. This needs to be converted to brightness temperature. Furthermore, both Landsat-8 and Sentinel-3 subset images need to be extracted from their respective data products. The steps required to extract the two images as shown in Fig. 5.11 are as follows:

1. The Landsat-8 B10 tile and the Sentinel-3 B8 tile were extracted from Landsat-8 L1C and S3A\_SL\_1\_RBT data products respectively. The size of the Landsat-8 tile and Sentinel-3 tile is  $7831 \times 7701$  and  $1200 \times 1500$  respectively. Both products were acquired on 2020/08/06.
2. As depicted in Fig. 5.11 the size of the area-patch and the geo-coordinates must be initialised for both tiles. The choice depends on the required size of the subset to be extracted and the required location. In Fig. 5.11 the chosen area-size is  $100 \text{ km} \times 100 \text{ km}$  and the chosen location is Annaba (Algiers).
3. Steps for the extraction of the reference image include:
  - Extract the L8 B10 radiance subset from the L8 tile depending on the chosen area-patch and location. In this case the subset has a size of  $3333 \times 3333$  pixels due to the 30 m spatial resolution.
  - Check whether the subset contains any NaNs. If so an new location has to be chosen for *both* the reference and sensed images otherwise one goes to the next step.
  - The chosen L8 B10 subset needs to be converted from *radiance to brightness temperature*. This is done by using the following equation which is

derived from the Landsat-8 Data Users Handbook [86]:

$$T = \frac{K_2}{\ln \frac{K_1}{L_\alpha} + 1} \quad (5.1)$$

where

- $T$  is the TOA brightness temperature in Kelvin
- $L_\alpha$  is the spectral radiance in  $\frac{W_a}{m^2 \times sr \times \mu m}$ .  $L_\alpha$  refers to the extracted radiance subset multiplied by (Multiplicative Factor (ML) + Additive Factor (AL)) chosen for the respective band. For B10,  $L_\alpha = extracted \ radiance \ subset \times (ML + AL)$  where  $ML = 3.3420E - 04$  and  $AL = 0.10000$
- $K_1$  is Band-specific thermal conversion constant from the metadata of the L1C data product.  $K_{B10} = 774.8853$
- $K_2$  is Band-specific thermal conversion constant from the metadata.  $K_{B10} = 1321.0789$
- The converted Landsat-8 B10 brightness temperature subset results in the Reference Image to be utilised in the proposed algorithm.

4. Steps for the extraction of the sensed image include:

- Extract the S3 B8 brightness temperature subset from the S3 tile depending on the chosen area-patch and location. In this case the subset has a size of  $100 \times 100$  pixels due to the 1000 m spatial resolution.
- Check whether the subset contains any NaNs. If so a new location has to be chosen for *both* the reference and sensed images otherwise one goes to the next step.
- The extracted S3 B8 brightness temperature subset is the Sensed Image to be utilised in the proposed algorithm.

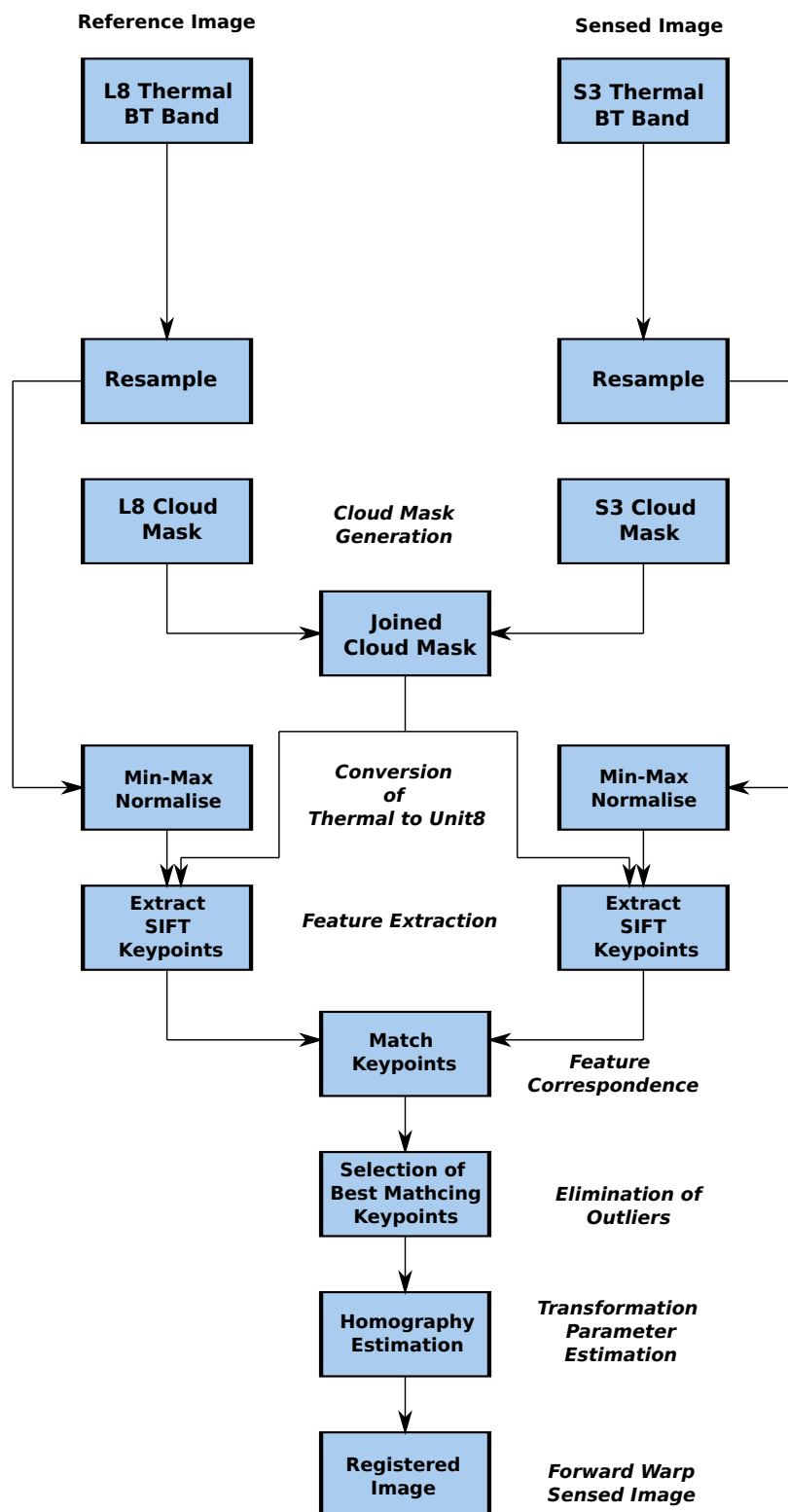


Figure 5.10: High-level schematic of proposed algorithm.



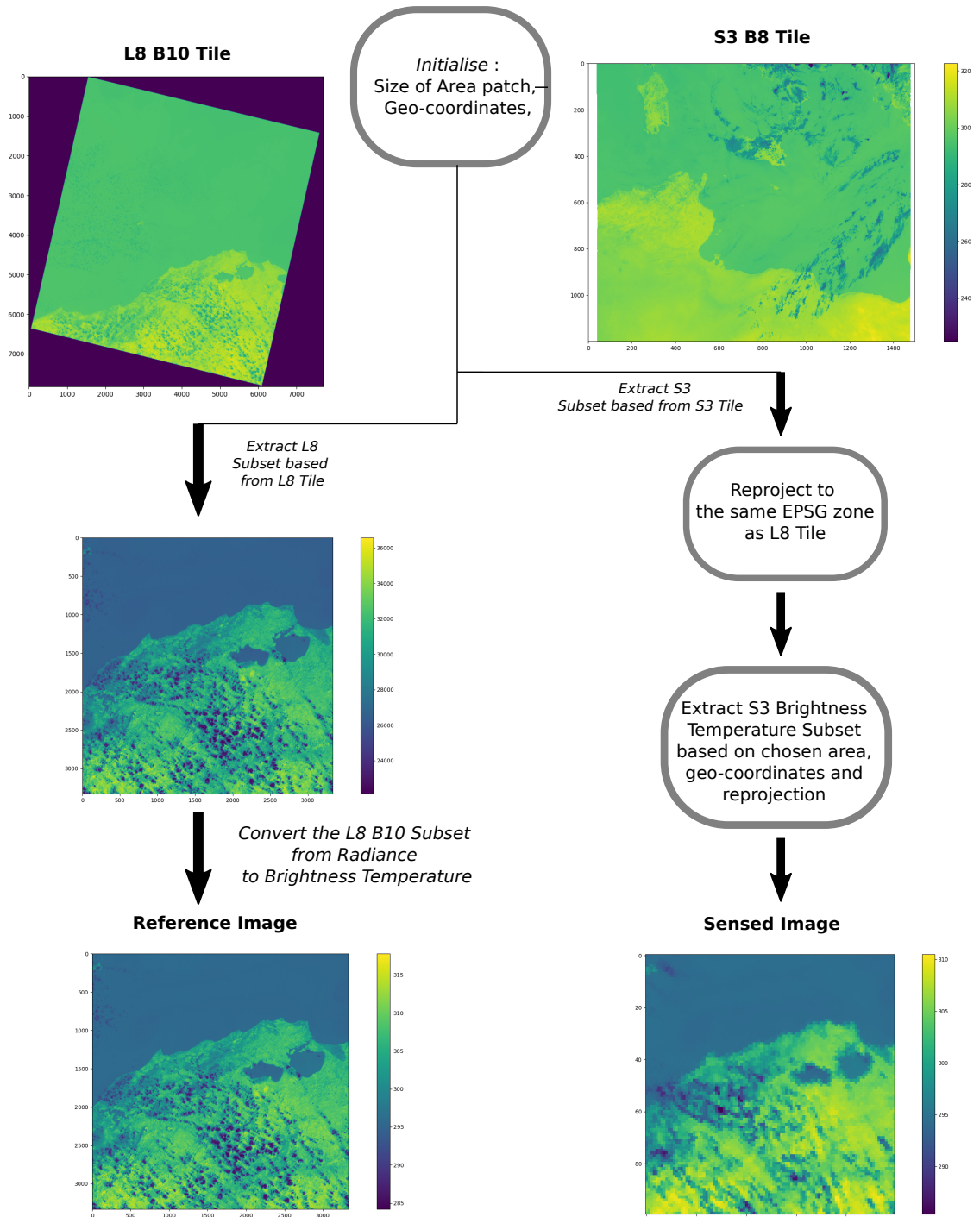


Figure 5.11: Pre-processing steps to extract the Reference and Sensed Images.

### 5.3.2 Generation of Joined Cloud Mask

The basic technique of cloud detection consists in the use of a threshold. This technique consists in thresholding the image of a spectral band in the short wavelength range. Typically, the blue band is preferred. The pixels which have a reflectance value higher than the threshold are defined as clouds. This method generally does not detect thin clouds and is very susceptible to false positives.

For both reference and sensed cloud masks the same area patch and geo-coordinates are utilised as mentioned in Fig. 5.11. The steps required to generate the joined cloud mask for the reference and sensed image are as follows:

- 1a. In Fig. 5.12, the L8 cloud mask is first extracted from the BQA L8 band found in the L8 L1C data product.
- 1b. The L8 cloud mask in **1a.** in Fig. 5.12 is resampled to the same spatial resolution as the S3 sensed image i.e.  $100 \times 100$  pixels. The L8 cloud mask is resampled using *nearest neighbour* interpolation. This resampled cloud mask is also mapped to a binary mask. Where 3 represents clouds, 2 might be clouds and 1 is no clouds. These are mapped as shown in **1b.** in Fig 5.12.
- 1c. The clouds in the L8 cloud mask in **1b.** are first eroded using an erosion kernel of 1 pixel followed by dilation using a dilation kernel of 2 pixels. Lastly, an NOT operation is performed to represent the clouds as a 0 and the non-clouded parts as a 1.
- 2a. The original S3 cloud mask is extracted from the *cloud\_in* band found in the *S3\_L1\_RBT* product.
- 2b. The S3 cloud mask in **2a.** is reprojected to the same *coordinate reference system (CRS)* as the reference cloud mask.
- 2c. The cloud mask in **2b.** is first eroded using an erosion kernel of 1 pixel followed by dilation using a dilation kernel of 2 pixels.
- 2d. The binary values in the S3 cloud mask in **2c.** are inverted using an NOT operation to represent the clouds as a 0 and the non-clouded parts as a 1 as done for the reference L8 cloud mask.

3. As illustrated in Fig. 5.12, to form the required joined cloud mask for both the reference and sensed images, a *bitwise AND* operation between the reference and sensed cloud mask is utilised. This results in the joined cloud mask which will be used for both the reference and sensed images to ignore any features that are detected as cloud pixels when using the SIFT algorithm.

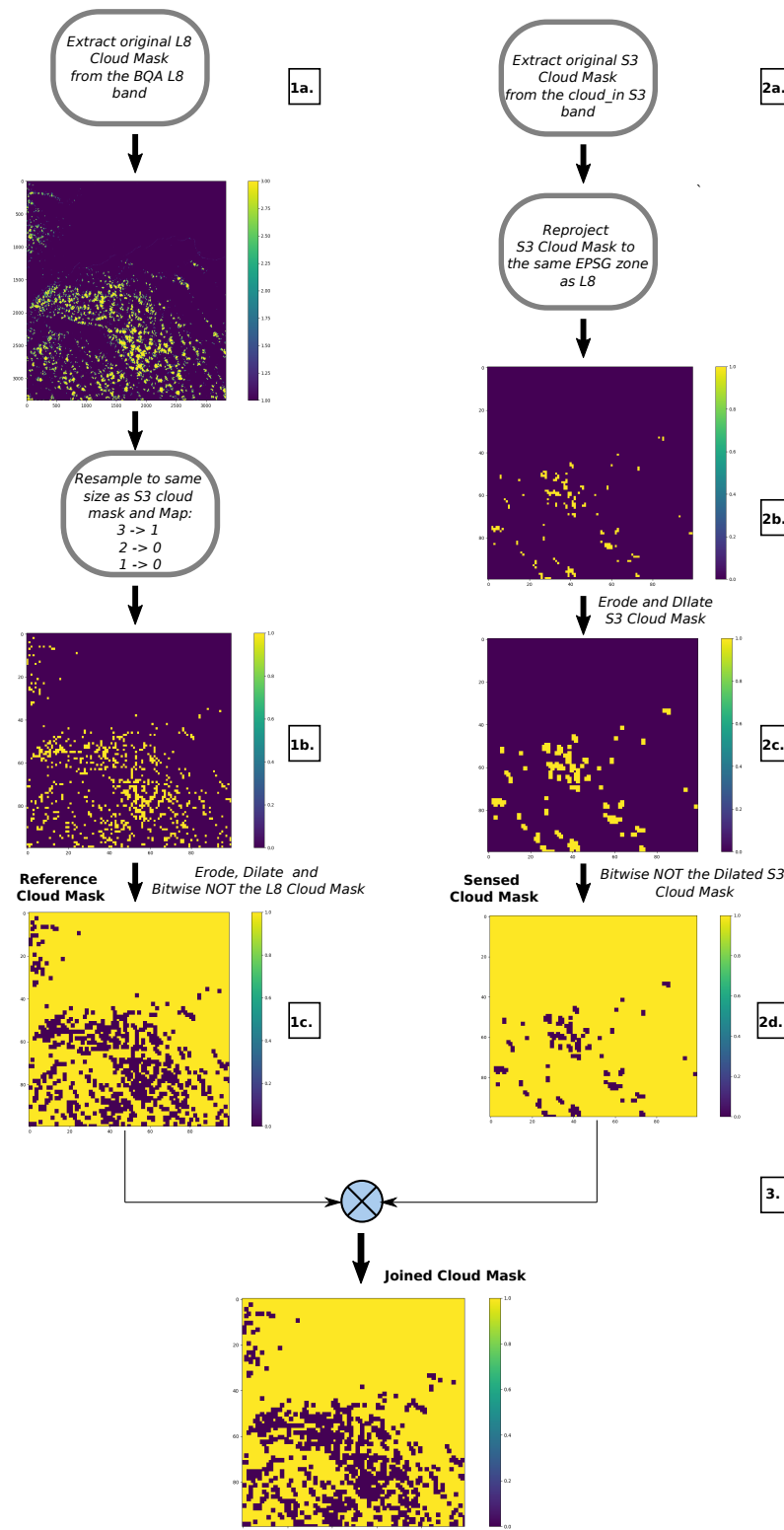


Figure 5.12: Required steps to generate the joined cloud mask for the reference and sensed cloud masks.

### 5.3.3 Resampling and Min-Max Normalisation

In this section the step of Min-Max normalisation illustrated in Fig. 5.10 is explained. The SIFT algorithm was designed to be utilised for natural images not thermal images. Thus, this step converts the reference and sensed images from brightness temperature to Uint8 images.

- 1a. In Fig. 5.13, the reference image is downsampled to the same spatial resolution as the sensed i.e.  $100 \times 100$  pixels. It is important to note that the reference image can remain the same and instead the sensed image is upsampled. Most important aspect is that both the reference and sensed image have the same image size.
- 1b. For the sensed image since the reference image is downsampled, the sensed image remains in its own native resolution.
- 2a. A histogram for all the respective brightness temperatures (in Kelvin) for each pixel in the reference image (L8 BT image) is plotted.
- 2b. Similarly, a histogram for all the respective brightness temperatures for each pixel in the sensed image (S3 BT image) is plotted.
3. As illustrated in Fig. 5.13, both histograms in **2a.** and **2b.** are joined together to find the minimum and maximum temperature for both images. It is to be noted that the minimum temperature chosen correlates with the temperature of the land. Cloud and sea temperatures are excluded. In this example the  $\min = 292.9K$  and the  $\max = 316.3K$ .
- 3a. In Fig. 5.14, in the downsampled (BT) reference image, any temperature that is below the min temperature is assigned this minimum. In this example, since the max temperature was extracted from the blue histogram (in Fig. 5.13) in this aspect the reference image remains unchanged.
- 3b. Once the min and max temperatures are assigned to the downsampled reference image, the next step required is min-max normalisation:

$$I_N = (I - \min) \frac{\max_{new} - \min_{new}}{\max - \min} + \min_{new} \quad (5.2)$$

where

- max and min are as explained in step **3**.
  - $max_{new}$  is 255
  - $min_{new}$  is 0
  - $I$  is the downscaled (BT) reference with the assigned min and max temperatures as explained in step **3a**.
  - $I_N$  is the Uint8 reference image.
4. In Fig. 5.14, after min-max normalisation, the downscaled (BT) reference image is converted to Uint8.
- 5a. In Fig. 5.14, in the (BT) sensed image, any temperature that is below the min (of step **3**.) temperature is assigned this minimum. Similarly, if any temperature in the sensed image is greater or equal to the max (of step **3**.) temperature it is assigned the max temperature. Since the new maximum is taken from the downscaled (BT) reference image, in this aspect the sensed image remains unchanged.
- 5b Once the min and max temperatures are assigned to the downscaled sensed image, the next step required is min-max normalisation. This is computed by using eq. (5.2). In this case  $I$  is the sensed (BT) image with the assigned min and max temperatures as explained in step **5a**. and  $I_N$  is the Uint8 sensed image.
6. In Fig. 5.14, after min-max normalisation, the sensed (BT) reference image is converted to Uint8.

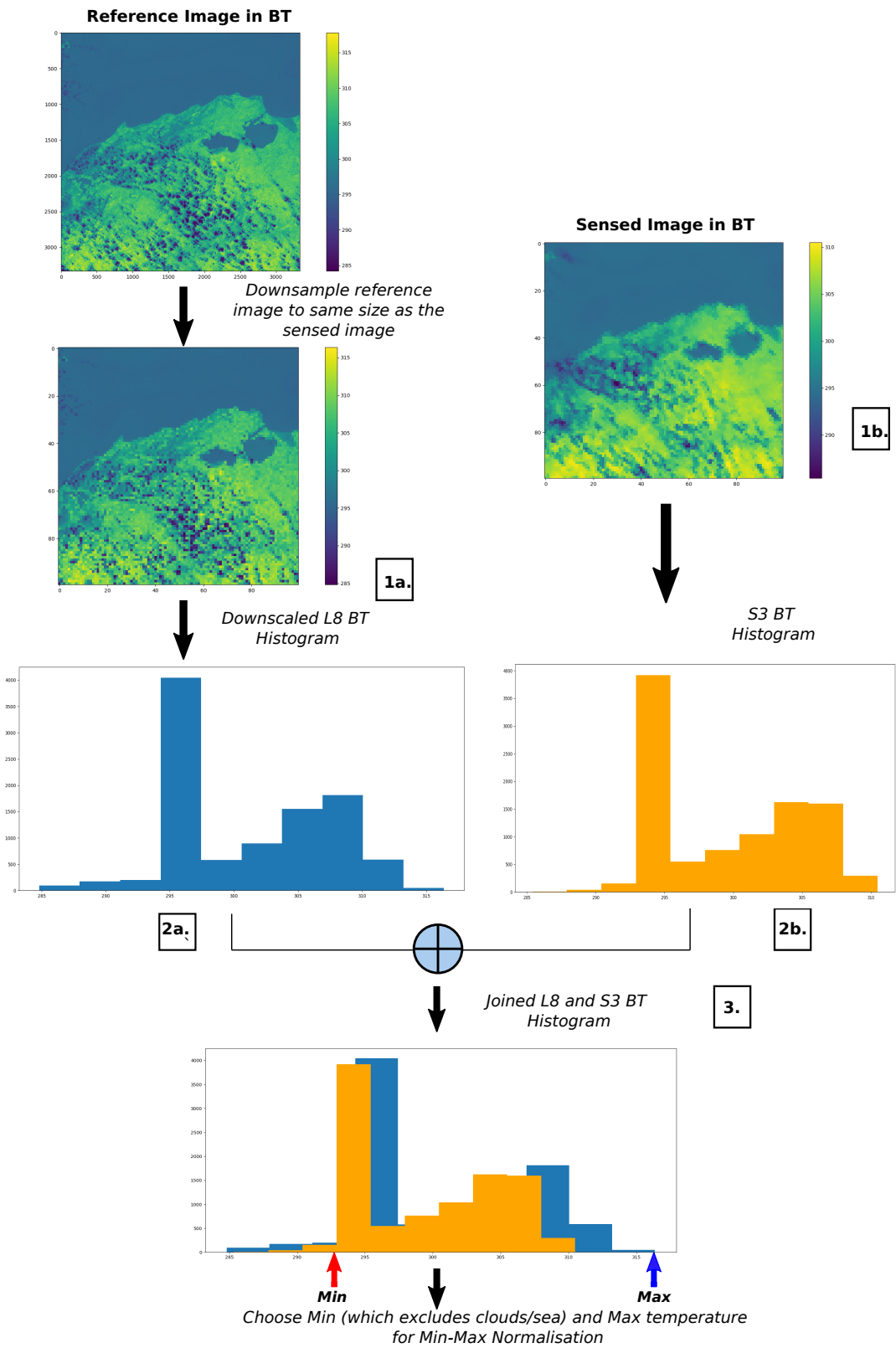


Figure 5.13: Establishing the minimum and maximum temperatures from the reference and sensed images for min-max normalisation.

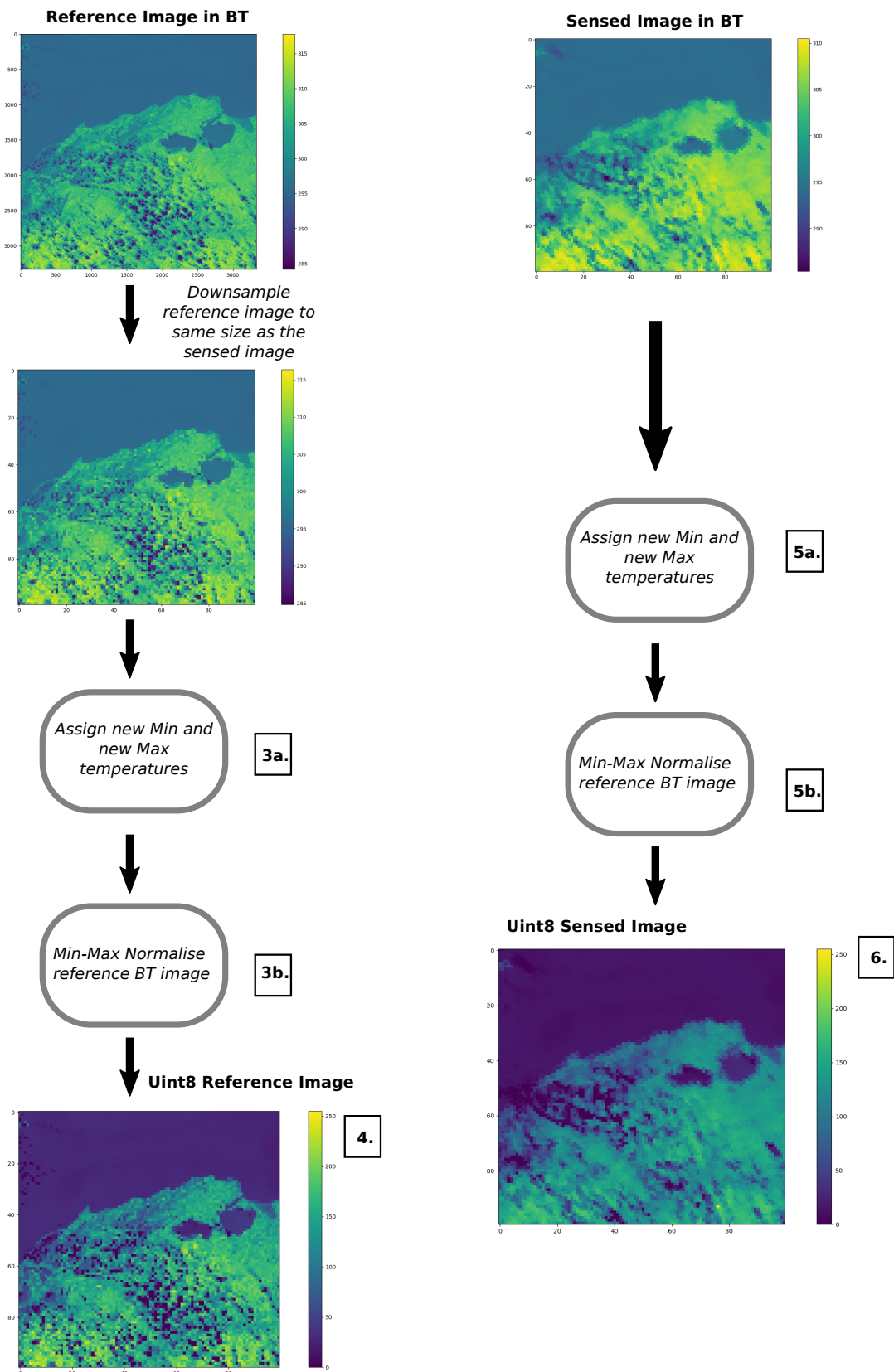


Figure 5.14: **Note:** The numbering of this figure, from 3a to 6 continue from the numbers shown in Fig. 5.13. Min-max normalising the reference and sensed images to convert the two images from brightness temperature to Uint8.



### 5.3.4 Selecting Best Matching Keypoints

1. In Fig. 5.15, the SIFT keypoints and their respective descriptors are extracted for both the reference and sensed images. The SIFT parameters utilised are the default parameters for both images. It is important to note that the unit8 images for both the reference and sensed images are only utilised for the SIFT algorithm. The rest of the methodology utilises the brightness temperature images.
2. The SIFT feature descriptors of both images are matched by finding the nearest neighbour. The *Euclidean distance* between the SIFT keypoints of the reference and sensed image is calculated. The ratio between the smallest distance and second smallest distance is computed. All matches which have a ratio greater than 0.9 are rejected.
3. Additionally, an outlier detection method is utilised to eliminate incorrect matches found in step **2.** of Fig. 5.15. The outlier detection method utilised is the RANSAC algorithm as explained in Section 2.7. For the example in Fig. 5.14, the utilised RANSAC threshold is 3. Thus, the RANSAC algorithm only retains the best matches between the sensed and reference images.

*Assuming there is no rotation and scale differences between the two images, a correct match is generally depicted as a parallel line.* The remaining lines that are not parallel are incorrect matches and they are discarded by RANSAC. This aspect is clearly observed in Fig. 5.15. This assumptions comes from the fact that the two images to be registered originate from the same modality.

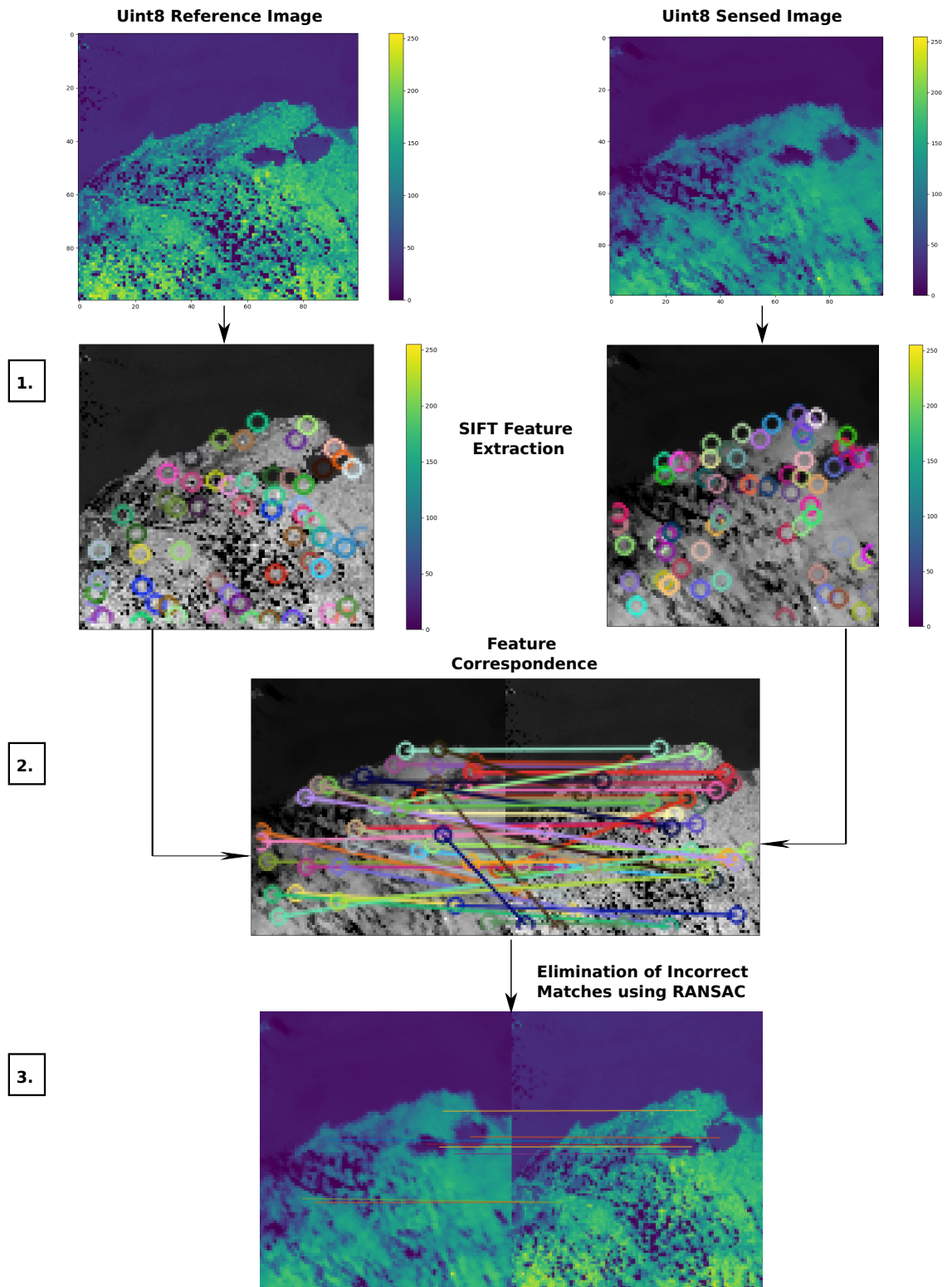


Figure 5.15: Extraction of best matching keypoints between the Reference and Sensed images.

### 5.3.5 Homography Estimation

Once the inliers/best matches (from RANSAC) between both images were found, the next step is to register the two images. A geometric transformation known as an *Affine transform* is utilised to compute the homography. In this proposed methodology, two type of affine transformations are utilised.

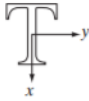





Transformation Name	Affine Matrix, T	Coordinate Equations	Example
Identity	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v$ $y = w$	
Scaling	$\begin{bmatrix} c_x & 0 & 0 \\ 0 & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = c_x v$ $y = c_y w$	
Rotation	$\begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v \cos \theta - w \sin \theta$ $y = v \sin \theta + w \cos \theta$	
Translation	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix}$	$x = v + t_x$ $y = w + t_y$	
Shear (vertical)	$\begin{bmatrix} 1 & 0 & 0 \\ s_v & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v + s_v w$ $y = w$	
Shear (horizontal)	$\begin{bmatrix} 1 & s_h & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v$ $y = s_h v + w$	

Figure 5.16: Affine Transformations [14]

As depicted in Fig. 5.16, there are five fundamental types of affine transformations. These include the *identity*, *scaling*, *rotation*, *translation* and *shear*. Based on these five affine matrices, one can combine one or more of the mentioned affine transformations to form a single affine transformation. In fact, the first affine transformation utilised in this proposed methodology considers rotation, scale and translation. Whilst the second utilised affine transformation considers only translation.

### 5.3.5.1 Affine Transformation with Rotation, Scale and Translation

Assume that a matrix  $\mathbf{A}$  contains the  $x$  and  $y$  coordinates of the inliers (after RANSAC) in the sensed image. Additionally, assume that the row vector  $\mathbf{b}$  contains the  $x$  and  $y$  coordinates of the inliers in the reference image. The aim is to solve:

$$\mathbf{Ax} = \mathbf{b} \quad (5.3)$$

Aim is to solve eq. (5.3) using the *Linear Least Squares* method.

$$\mathbf{x} = \|\mathbf{Ax} - \mathbf{b}\|^2 = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \quad (5.4)$$

Matrix  $\mathbf{X}$  is the reshaped vector  $\mathbf{x}$ . The proof of equations (5.3) and (5.4) can be found in [87].

$$\mathbf{H} = \mathbf{X}^T \quad (5.5)$$

where  $\mathbf{H}$  is of the following form:

$$\mathbf{H} = \begin{pmatrix} a_{11} & a_{12} & a_{tx} \\ a_{13} & a_{14} & a_{ty} \\ 0 & 0 & 1 \end{pmatrix} \quad (5.6)$$

where  $a_{11} = c_x \cos \theta$ ,  $a_{12} = \sin \theta$ ,  $a_{13} = -\sin \theta$ ,  $a_{14} = c_y \cos \theta$ ,  $a_{tx}$  is the horizontal translation  $t_x$ ,  $a_{ty}$  is the vertical translation  $t_y$ ,  $c_x$  and  $c_y$  cater for scale and  $\theta$  represents the rotation.

#### 5.3.5.1.1 In-Depth Algebraic Example

1. Assume one has 2 arrays of size  $(2, n)$  called  $\mathbf{s}$  and  $\mathbf{r}$ 
  - $n$  denotes the number of inliers i.e. the number of sensed and matched reference keypoints
  - Array  $\mathbf{s}$  and array  $\mathbf{r}$  contain the pixel locations of the matched keypoints resulting after using SIFT and RANSAC
  - First row in both arrays contains the x-coordinate
  - Second row in both arrays contains the y-coordinate

2. Aim is to solve the equation  $\mathbf{Ax} = \mathbf{b}$  using the Linear Least Squares method

- $\mathbf{x} = \|\mathbf{Ax} - \mathbf{b}\|^2 = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$

3.

$$\mathbf{A} = \begin{bmatrix} s[0, i] & s[1, i] & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & s[0, i] & s[1, i] & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s[0, n] & s[1, n] & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & s[0, n] & s[1, n] & 1 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} r[0, i] \\ r[1, i] \\ \vdots \\ r[0, n] \\ r[1, n] \end{bmatrix}$$

from  $i$  to  $n$ , where  $\mathbf{A}$  is of size  $(2n, 6)$  and  $\mathbf{b}$  is of size  $(2n, 1)$

4. •  $\mathbf{x} = \text{pinv}(\mathbf{A}) \cdot \mathbf{b}$  in python is accomplished by: `np.dot(np.linalg.pinv(A), b)` or by `np.linalg.lstsq(A, b, rcond=None)` to compute eq. (5.4)

- Reshape  $\mathbf{x}$  of size  $(6, 1)$  to matrix  $\mathbf{X}$  of size  $(3, 3)$

- Initialise right column of matrix  $\mathbf{X}$  to  $[1, 0, 0]^T$

- $\mathbf{H} = \mathbf{X}^T$  where  $\mathbf{H}$  is of the following form as shown in eq. (5.6)

Refer to Appendix A for a numeric example.

### 5.3.5.2 Affine Transformation with Translation Only

From the chosen reference and sensed images as shown in Fig. 5.14, after reprojecting the Sentinel-3 B8 (sensed) image and bringing both images to the same spatial resolution (i.e downscale the reference to the same size as the sensed or upscale the sensed to the same size as the reference), there is minimal difference in terms of rotation and scale between the reference and sensed images. Thus, the affine transformation with translation alone was also considered.

The procedure to compute the  $\mathbf{H}$  homography for translation alone is as follows:

1. Plot  $x = v + t_x$ 
  - Where  $x$  is the x-coordinates of the inliers (after RANSAC) of the reference image which is to be plotted in the y-axis.
  - $v$  is the x-coordinates of the inliers of the sensed image which is to be plotted in the x-axis.
  - A line of best-fit using the *Non-linear least squares* method [88] is taken for these points and the obtained y-intercept of this line is equivalent to  $t_x$ .
2. Similarly, plot  $y = w + t_y$ 
  - Where  $y$  is the y-coordinates of the reference image which is to be plotted in the y-axis.
  - $w$  is the y-coordinates of the inliers of the sensed image which is to be plotted in the x-axis.
  - A line of best-fit using the *Non-linear least squares* method [88] is taken for these points and the obtained y-intercept of this line is equivalent to  $t_y$ .
3. In both plots the gradient is fixed to 1.
4. Once  $t_x$  and  $t_y$  are found, the homography with translation alone is:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \quad (5.7)$$

### 5.3.6 Registration

As illustrated in Fig. 5.10 which depicts a high-level schematic of the proposed algorithm, once the homography is found, the next step is to find the registered image. This is accomplished by using the established homography and *forward warp* the sensed image to the reference image.

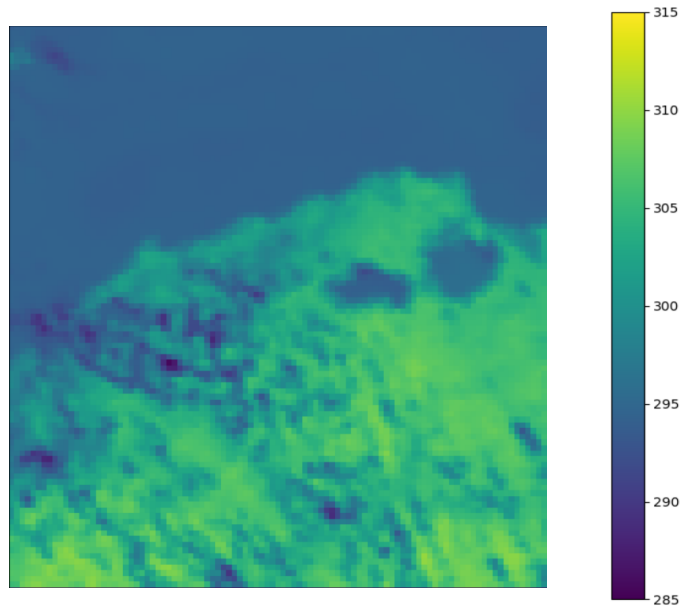


Figure 5.17: Registered image of size  $100 \times 100$  pixels for the example shown in Fig. 5.15.

The vertical and horizontal paddings shown in the registered image of Fig. 5.17 illustrate the vertical and horizontal displacement of the sensed image after registration. The obtained homography for Fig. 5.17 which includes translation only is as follows:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & -0.4073 \\ 0 & 1 & -0.2613 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.8)$$

Thus, from eq. (5.8), the obtained displacement in the x-direction  $t_x = -0.4073$  pixels and the obtained displacement in the y-direction  $t_y = -0.2613$ .

## 5.4 Improving the Parameter Search

In Section 5.3, the default SIFT parameters were utilised. In this proposed methodology, a major focus was to find parameters that are best suited to register the two chosen remote sensing images such as a subset of Landsat-8 B10 and a subset of Sentinel-3 B8 as shown in the previous sections of chapter 5.

### 5.4.1 Salient Parameters

The salient parameters that were chosen to be optimised to improve registration are the following: *RANSAC threshold*, *Number of Octaves*, *Sigma Blur*, *Edge Threshold* and the *Contrast Threshold*.

#### 5.4.1.1 RANSAC Threshold

As explained in Section 2.7, the RANSAC threshold is a tolerance margin where all the points lying within this margin are considered as inliers. Concerning the choice of this parameter, the aim is to choose a value that results in retaining the best keypoint matches possible with the fewest possible number of false positives. The value of the RANSAC threshold can range from 1 to 10. 10 being the most lenient which results in many false positive matches whilst 1 being the most strict which might result in no matches if few features were detected between the reference and sensed images.

#### 5.4.1.2 Number of Octaves

The number of octaves utilised is dependent on the size of the reference image. This is determined through the following equation:

$$n_{oct} = \text{floor}(\log_2(\min \text{width}, \text{height}) - 3) \quad (5.9)$$

Subtracting 3 in eq. (5.9) ensures that there is a minimum size patch that the SIFT feature detection requires to obtain a good output. Thus, subtracting 3 ensures that this minimum patch size is met once one gets to the lowest octave of the DoG pyramid in the *Scale Space Extrema* detection step as outlined in Section 2.5.1.

Taking a numerical example. Assume one has an octave patch of size  $64 \times 64$  pixels. As outlined in Section 2.5.1, one knows that at each octave, the size of each



dimension is divided by 2. Thus, taking the  $\log_2$  of the smallest of the width and height will theoretically result in the total number of octaves.

Octave	Size
1	$64 \times 64$
2	$32 \times 32$
3	$16 \times 16$
4	$8 \times 8$
5	$4 \times 4$
6	$2 \times 2$
7	$1 \times 1$

Table 5.4: Theoretical number of octaves for an octave patch of size  $64 \times 64$ .

Nonetheless, from Table 5.4 one can observe that at octaves 5, 6 and 7 will result in no useful information and can be ignored. Thus by subtracting 3 from the total number of octaves, one will stop analysing the information at octave 4. Hence, the smallest patch to analyse is  $8 \times 8$  pixels.

#### 5.4.1.3 Sigma

As outlined in Section 2.5.1, sigma refers to the spread of the Gaussian operator applied to the input image at the first octave. It determines the amount of blur added to the image. A larger  $\sigma$  increases the added blur. Typically this value ranges from 0.5 to 2.

In Fig. 5.20, the reference image chosen is a L8 B10 patch found in parts of Croatia taken on 2019/06/03 with a size of  $1000 \times 1000$  pixels covering an area patch of 30 km. The sensed image to be registered is the same L8 B10 patch but down-scaled and up-scaled by a scale factor of 10 and it is also shifted to the left by 1 pixel.

In this test the RANSAC threshold was set to 3 and the number of iterations for RANSAC was set to 2000 iterations, all the SIFT parameters are the default parameters and the only parameter that was modified is  $\sigma$ . From Fig. 5.20 it can be observed that beyond a  $\sigma$  value of 1.6 the number of matches saturate. Interestingly enough this is a similar observation noted by the author of SIFT in his work. In the repeatability plot shown in Fig. 5.19, the plot flattens beyond a  $\sigma$  value of 1.6.

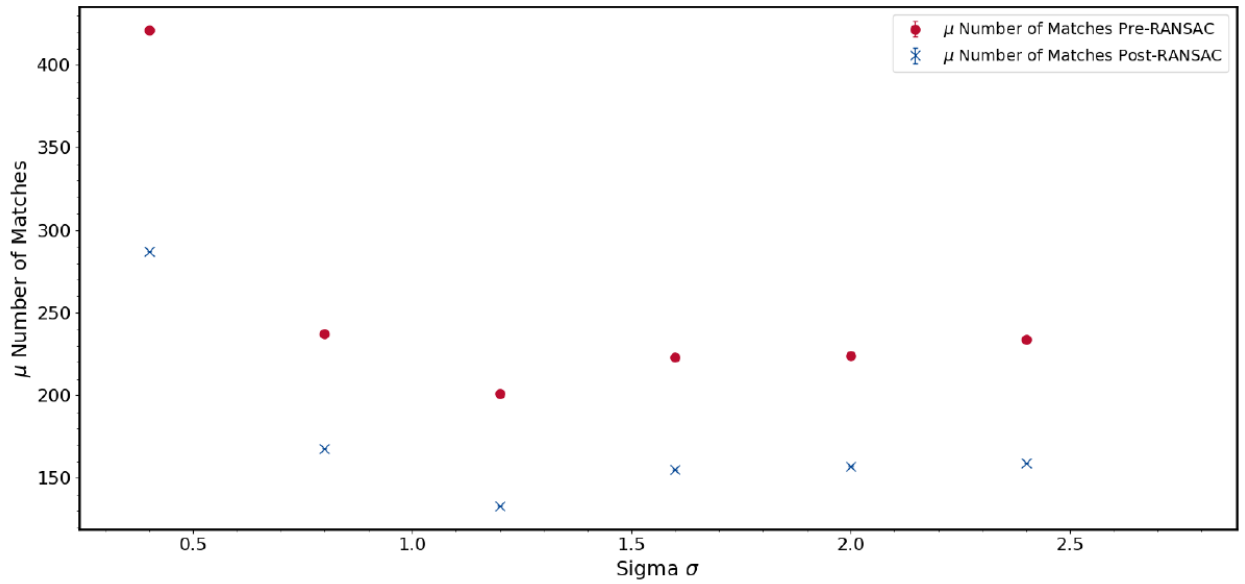


Figure 5.18: Plot of Mean Number of Matches vs Sigma.

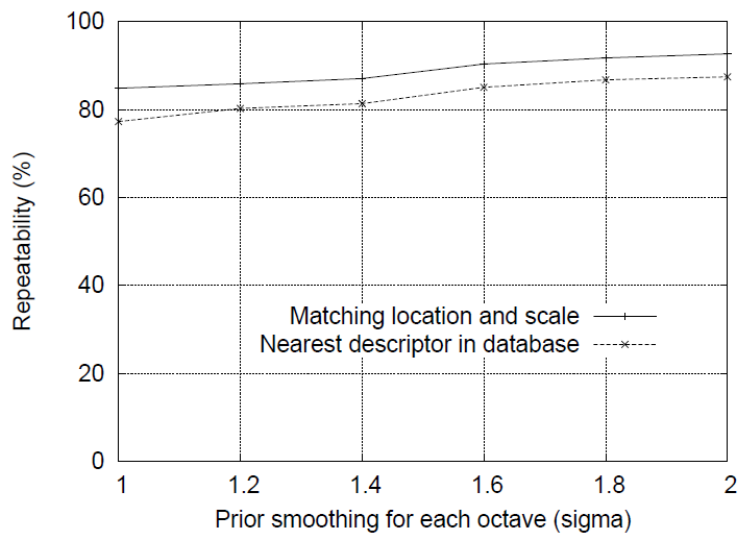


Figure 5.19: Graph showing the percent of keypoint locations that are repeatedly detected in a transformed image as a function of the  $\sigma$  smoothing for the first level of each octave taken from David Lowe's work [7].

#### 5.4.1.4 Contrast Threshold

As outlined in Section 2.5.3.2, the contrast threshold  $|D(\hat{x})|$  is the threshold utilised to discard low contrast keypoints. *The larger the contrast threshold, the less features are produced by the detector.* Similarly to the  $\sigma$  plot, the same test was conducted for the contrast threshold. Thus, all parameters remained fixed other than contrast threshold. From Fig. 5.20 it can be observed that as the contrast threshold increases, less features are produced by the detector.

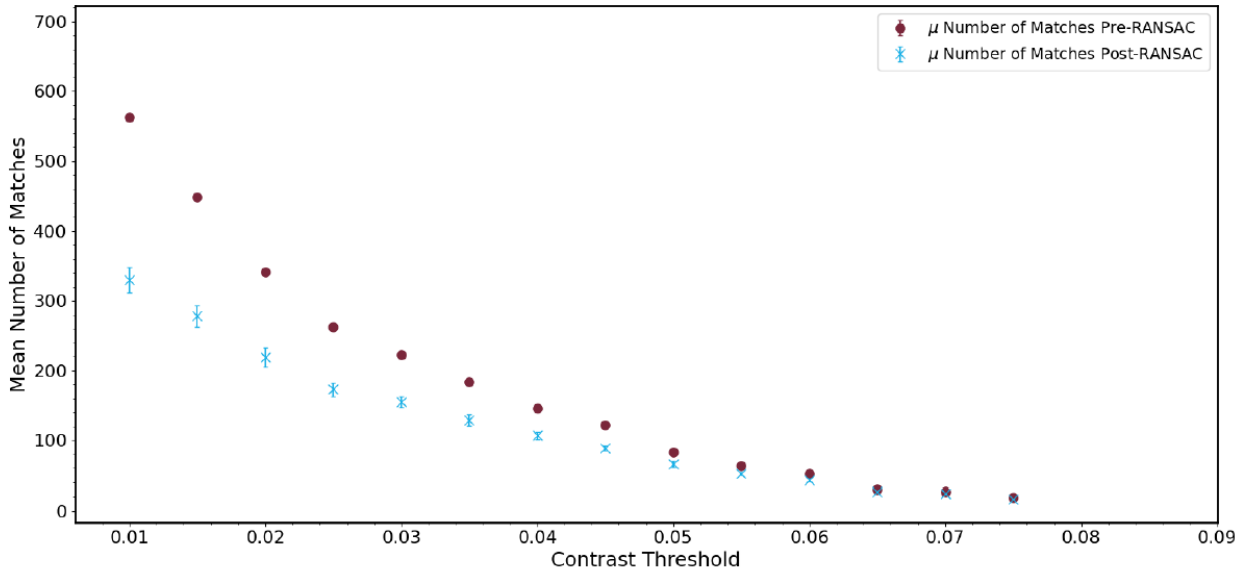
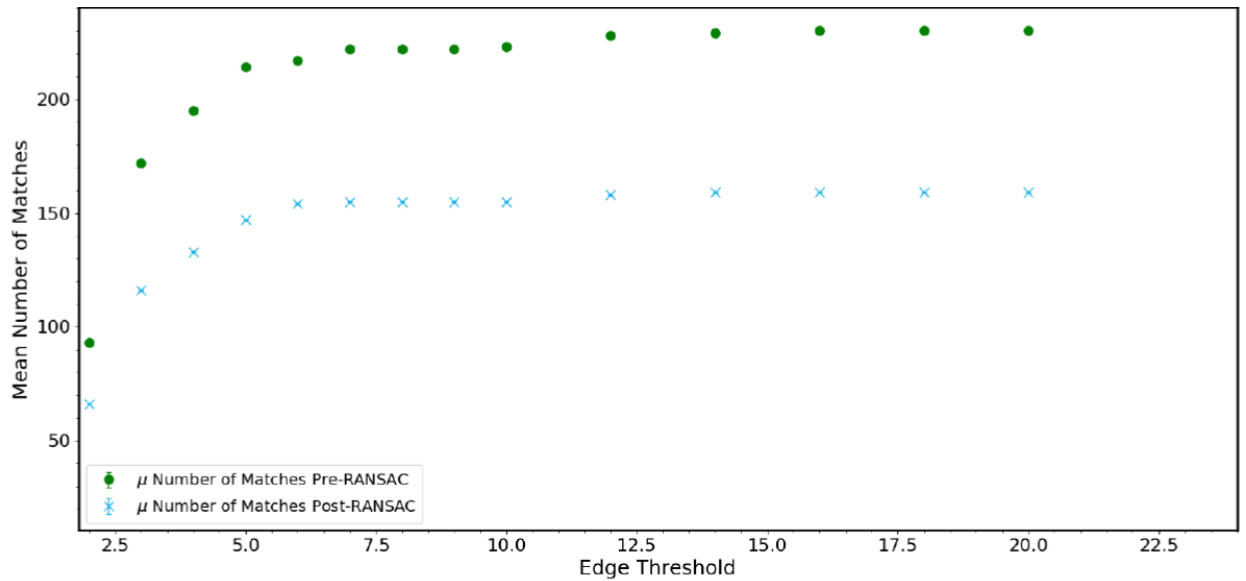


Figure 5.20: Plot of Mean Number of Matches vs the Contrast Threshold  $|D(\hat{x})|$ .

#### 5.4.1.5 Edge Threshold

As outlined in Section 2.5.3.2, the edge threshold  $r$  is the threshold utilised to eliminate keypoints which have a strong response along the edges. The aim of this threshold is to *detect keypoints found at corners*. Contrary to the contrast threshold, *the larger the edge threshold the less features are filtered out.*

Similarly to the previous plots, the same test was conducted for the edge threshold. Thus, all parameters remained fixed other than edge threshold. From Fig. 5.21, one can observe that contrary to the contrast threshold, the larger the edge threshold the less features were filtered out.

Figure 5.21: Plot of Mean Number of Matches vs the Edge Threshold  $r$ .

### 5.4.2 Default Parameters

In the results of this dissertation, one of the major focus was to find values of the mentioned parameters that are best suited to register two chosen remote sensing images. The initial parameters utilised are the default parameters used by the author of the SIFT algorithm [7]:

- $\sigma = 1.6$
- Contrast Threshold  $|D(\hat{x})| = 0.03$
- Edge Threshold  $r = 10$

With regards to the RANSAC threshold for the first chosen parameter set this was set to 3.

### 5.4.3 Monte Carlo

The second approach utilised to find values for the mentioned parameters to register two chosen remote sensing images is the *Monte Carlo* approach. The Monte Carlo search method is a problem solving technique used to approximate the probability of certain outcomes by running multiple trial runs using random variables. This requires a good source of random numbers.

---

#### Algorithm 1 Monte Carlo Approach

---

```

1:  $RMSE_{Old} = RMSE_{Before\ Reg}$ 
2: for  $iterations = 1, 2, \dots, N$  do
3:    $RANSAC_{Threshold\ Initial} = random.uniform(a_{rans}, b_{rans})$ 
4:    $Sigma_{Initial} = random.uniform(a_{sig}, b_{sig})$ 
5:    $Contrast_{Threshold\ Initial} = random.uniform(a_{cont}, b_{cont})$ 
6:    $Edge_{Threshold\ Initial} = random.uniform(a_{edg}, b_{edge})$ 
7:
8:   Run registration algorithm using above parameters
9:
10:   $RMSE_{New} \leftarrow RMSE_{After\ Reg}$ 
11:
12:  if  $RMSE_{New} < RMSE_{Old}$  then
13:     $RMSE_{New} \leftarrow RMSE_{Old}$ 
14:     $RANSAC_{Threshold\ Final} \leftarrow RANSAC_{Threshold\ Initial}$ 
15:     $Sigma_{Final} \leftarrow Sigma_{Initial}$ 
16:     $Contrast_{Threshold\ Final} \leftarrow Contrast_{Threshold\ Initial}$ 
17:     $Edge_{Threshold\ Final} \leftarrow Edge_{Threshold\ Initial}$ 
18:  end if
19: end for

```

---

In Algorithm 1, *the utilised registration algorithm is the proposed algorithm outlined in Section 5.3*,  $N$  refers to the number of Monte Carlo iterations.  $N$  is set to 1000 iterations. The set number of iterations for the RANSAC algorithm is 2000 iterations.  $a$  and  $b$  for the four salient parameters refer to limits of the respective randomly generated parameter. The SIFT parameter limits are chosen based on the illustrated plots shown before. The assumed limits are as follows:

- $a_{rans} = 3$  and  $b_{rans} = 5$
- $a_{sig} = 1.2$  and  $b_{sig} = 1.8$
- $a_{cont} = 0.025$  and  $b_{cont} = 0.055$

- $a_{edg} = 5$  and  $b_{edg} = 13$

As shown in the Monte Carlo algorithm, the objective function to be minimised is the RMSE. When registering thermal images such as Landsat-8 B10 and Sentinel-3 B8, the RMSE excludes any cloud pixels and any NaNs that might result from rotation or translation differences before and after registration. Additionally, before and after registration, the RMSE is always computed on brightness temperature subsets in Kelvin. The RMSE excludes the cloud pixels through the utilisation of the joined cloud mask as shown in Fig. 5.12. **The defined way of how the RMSE is computed in the Monte Carlo algorithm also applies for all the results of this dissertation.**

#### 5.4.4 Grid Search Method

The last approach utilised to find values for the mentioned parameters to register two chosen remote sensing images is the *Grid Search* approach. The aim of this method is to find the *minimum* value that satisfies a *cost function*.

One chooses a point in the cost function and starts moving in steps towards the minimum value. The step size to converge to this minimum value is known as the *Learning Rate*  $\alpha$ . With a higher learning rate one can cover more area. Nonetheless, this can lead to overshooting the minima. Contrarily, choosing a very small step size can lead to no convergence.

**The cost function to be minimised for the problem of registering two remote sensing images is the RMSE and the input parameters of that cost function are the SIFT parameters and the RANSAC threshold. Utilising the RMSE as the cost function to be minimised leads to multiple minima. Thus, there are multiple parameter sets that can lead to that same minimum value.**

In Algorithm 2, the utilised registration algorithm is the proposed algorithm outlined in Section 5.3.  $N$  refers to the number of Grid Search iterations which is set to 100 iterations. The set number of iterations for the RANSAC algorithm is 2000 iterations.  $a$  and  $b$  for the four salient parameters refer to limits of the respective parameters as outlined in Section 5.4.3.

---

**Algorithm 2** Grid Search Approach

---

```
1:  $RMSE_{Old} = RMSE_{Before\ Reg}$ 
2:  $Contrast_{Threshold} = b_{cont}$ 
3:  $Edge_{Threshold} = b_{edg}$ 
4:  $Sigma = b_{sig}$ 
5:  $RANSAC_{Threshold} = b_{rans}$ 
6:
7:  $\alpha_0 = (b_{cont} - a_{cont})/N$ 
8:  $\alpha_1 = (b_{edg} - a_{edg})/N$ 
9:  $\alpha_2 = (b_{sig} - a_{sig})/N$ 
10:  $\alpha_3 = (b_{rans} - a_{rans})/N$ 
11:
12: for  $iterations = 1, 2, \dots, N$  do
13:
14:   Run registration algorithm using the initialised threshold and  $\alpha$  parameters
15:
16:    $RMSE_{New} \leftarrow$  Output of registration algorithm
17:
18:   if  $RMSE_{New} < RMSE_{Old}$  then
19:      $RMSE_{New} \leftarrow RMSE_{Old}$ 
20:      $RANSAC_{Threshold\ Final} \leftarrow RANSAC_{Threshold}$ 
21:      $Sigma_{Final} \leftarrow Sigma$ 
22:      $Contrast_{Threshold\ Final} \leftarrow Contrast_{Threshold}$ 
23:      $Edge_{Threshold\ Final} \leftarrow Edge_{Threshold}$ 
24:   end if
25:
26:    $Contrast_{Threshold} = Contrast_{Threshold} - \alpha_0$ 
27:    $Edge_{Threshold} = Edge_{Threshold} - \alpha_1$ 
28:    $Sigma = Sigma - \alpha_2$ 
29:    $RANSAC_{Threshold} = RANSAC_{Threshold} - \alpha_3$ 
30:
31: end for
```

---

The initial starting values of the SIFT and RANSAC parameters for the Grid Search method are the upper limits  $b$  for each parameter as defined in Section 5.4.3. The step size  $\alpha$  for each of those parameters is based on the difference between their lower and limits divided by the total number of Grid Search iterations.

**In both the Monte Carlo and Grid Search methods, the final SIFT and RANSAC values established are then utilised as the input parameters for the registration of the two chosen remote sensing images such as Landsat-8 B10 and Sentinel-3 B8.**



# Chapter 6

## Results and Evaluation

The process of registering two remote sensing images has no ground truth. Inherently, even if the registration was a success the RMSE can never be zero. Thus, the best way to initially show that the proposed registration algorithm indeed works is through the use of single modality satellite imagery. Thermal images from Landsat-8 were used in the following Section 6.1.

In the first section a Landsat-8 B10 is registered with the same Landsat-8 B10 image but with some induced bias in the sensed image such as blurring and translation shifts. Since the same Landsat-8 is utilised, one knows that the RMSE after registration should be close to zero. In the second section, the Landsat-8 B10 is registered with a Sentinel-3 B8. In the third section, the proposed registration method is compared to phase correlation, RIFT, ECC and SIFT Flow algorithms for the registration of single multi-modal and single modality remote sensing data.

### 6.1 Registration of Landsat-8 vs Landsat-8

#### 6.1.1 Blurring the Sensed Image

The first aim of this subsection is to create a registration test that mimics the scale difference between Landsat-8 B10 and Sentinel-3 B8, but using only Landsat-8 B10 images. The reference BT image in Fig. 6.1 was extracted from a Landsat-8 B10 band. The location is Samothrace, a Greek island taken on 2020/07/31. There is no cloud cover in this chosen patch. The covered area patch is of  $100 \text{ km} \times 100 \text{ km}$ . For a spatial resolution of 30 m this represents a size of  $3333 \times 3333$  pixels. As shown

in Fig. 6.1 this reference image was downsampled from  $3333 \times 3333$  to  $1000 \times 1000$  pixels using bicubic interpolation. The sensed image initially is the reference BT image at  $3333 \times 3333$  pixels and then was downsampled and upsampled by a scale factor  $n$  to mimic the scale difference between Landsat-8 B10 and Sentinel-3 B8. In Fig. 6.1, the chosen scale factor was 10.

The workflow shown in Fig. 6.1, is repeated for different scale factors, from  $n = 1$  to  $n = 10$ . The reference and sensed images of size  $1000 \times 1000$  are then registered using the proposed method of Chapter 5. The RANSAC algorithm was set to 2000 iterations and the SIFT parameters utilised are the default parameters as outlined in Section 5.4.2. The number of Octaves used are 6 calculated using eq. (5.9). The two homographies utilised to register the images is as described in Section 5.3.5. The registration result for each scale factor was repeated for 100 times taking the mean for each result.

The above steps explained for the location of Samothrace were also repeated for three other locations. All the specified parameters and processing are as explained for the first location. The aim is to get the mean result for all locations and show the mean result in various plots. All the chosen locations are shown in Fig. 6.2.

Fig. 6.3 shows the effect of introducing more blur to the sensed image as a result of downscaling and upscaling by the same scale factor. As this scaling factor increases the number of matches decrease due to this added blur. It should also be noted that at a high scaling factor such as 10, false information is introduced in the sensed images. This can be observed in the coastal region areas in the Sensed BT image after upscaling in Fig. 6.1. This introduced false information can cause misregistration.

In Fig. 6.4, as the scaling factor increases the misregistration in the horizontal direction  $t_x$  for the homography which includes rotation, scale and translation increases. For the homography that includes translation only misregistration in the horizontal direction is almost zero. The achieved horizontal misregistration for the homography with translation only is not greater than 0.2 pixels. For a spatial resolution of 30 m this represents a misregistration of 6 m.

In Fig. 6.5, a similar pattern for the misregistration in the vertical direction  $t_y$  was observed. The homography with translation only achieved the lowest misregistration. The achieved vertical misregistration for the homography with translation only is not greater than 0.24 pixels. For a spatial resolution of 30 m this represents a misregistration of 7.2 m.

From Fig. 6.6 one can observe that the both the RMSE before and after registration increase as the scaling factor increases. This is to be expected as increasing the scaling factor causes the sensed image to be more dissimilar than the reference image. Thus, the RMSE increases. For all scaling factors the RMSE after registration for the homography with translation only is equal or smaller than the RMSE before registration.

*The general trend observed in the mean plots shown in Figures 6.3, 6.4, 6.5, and 6.6 was observed in all the plots for each individual location.*

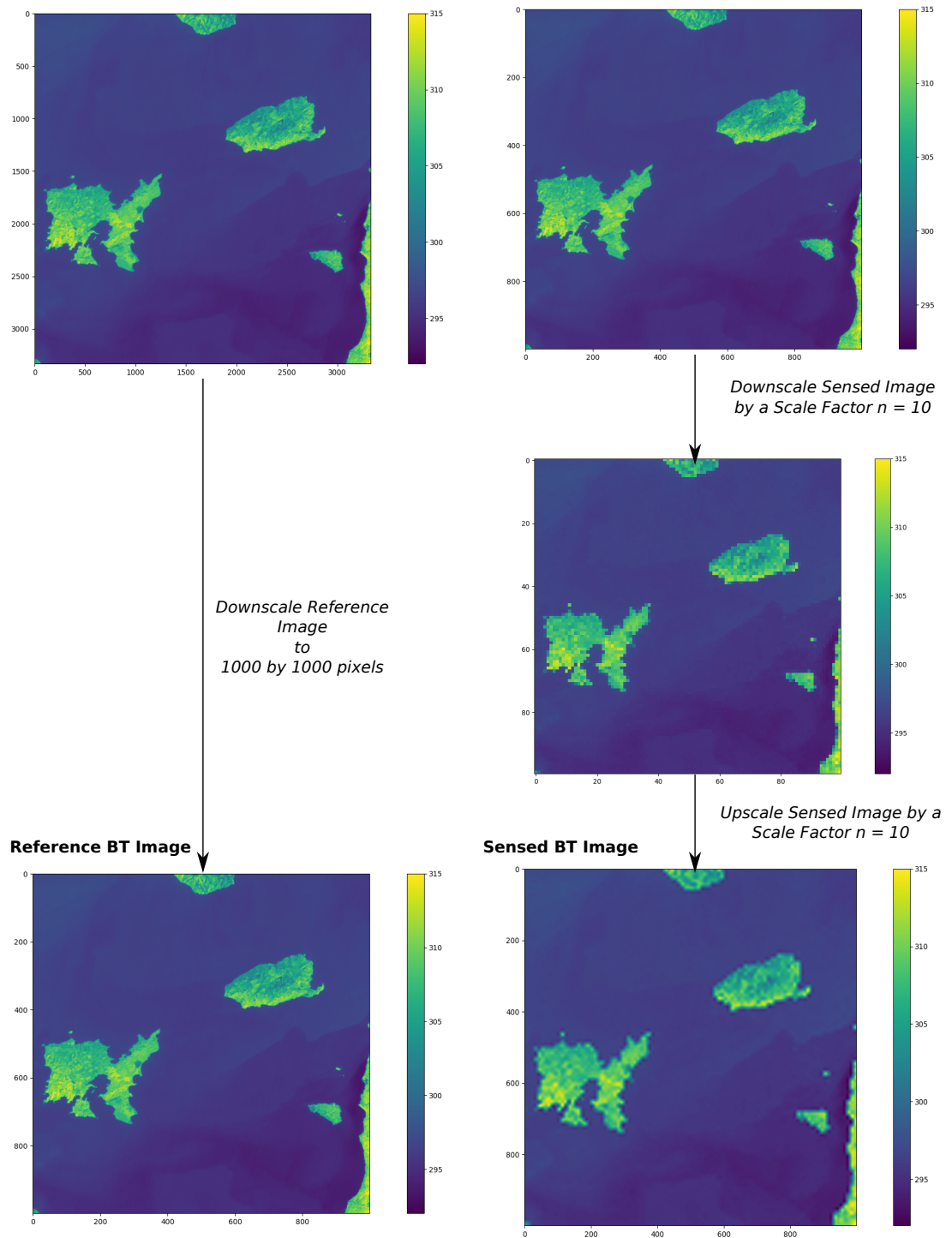
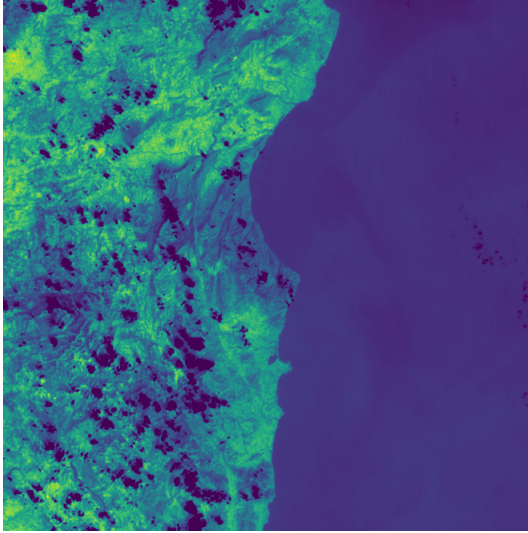
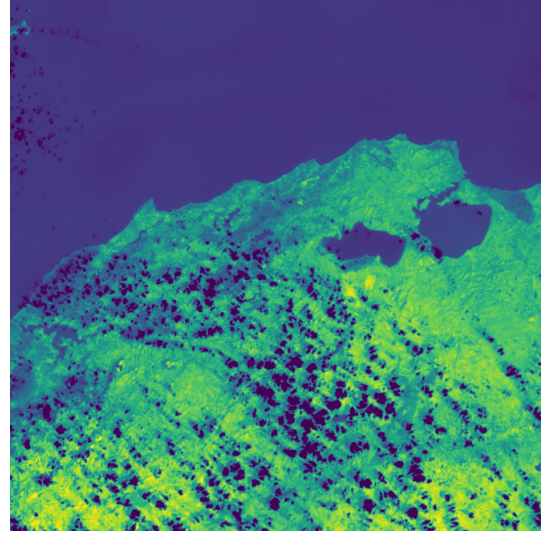


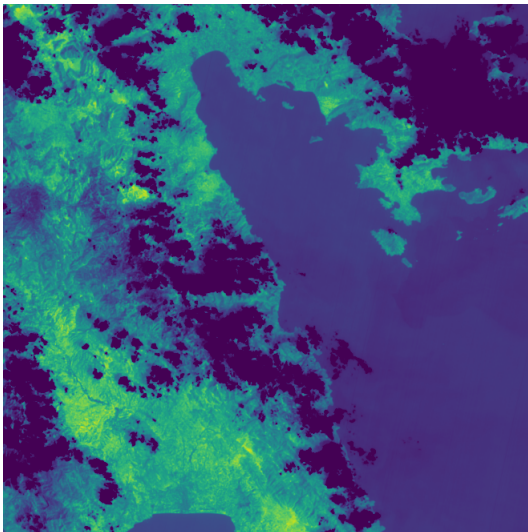
Figure 6.1: Workflow of test that registers a Landsat-8 B10 image with its blurred version.



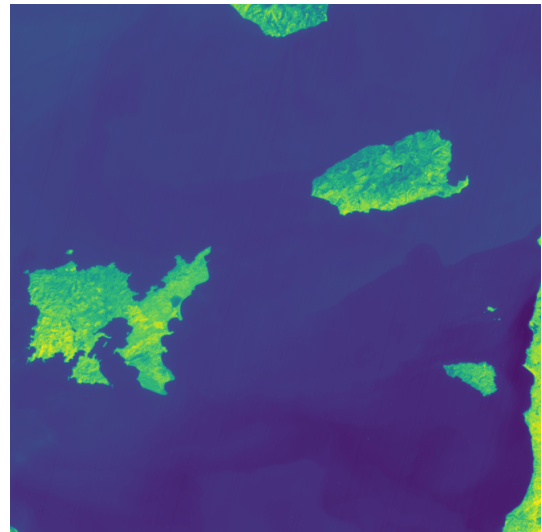
(a) Reference Image of Location 1 (Sardegna) taken on 2020/08/06.



(b) Reference Image of Location 2 (Annaba in Algiers) taken on 2020/08/06.



(c) Reference Image of Location 3 (Athens) taken on 2020/08/07.



(d) Reference Image of Location 4 (Samothrace, Greek Island) taken on 2020/07/31.

Figure 6.2: Reference images of the four chosen locations to be utilised for the test of registering a Landsat-8 B10 image with its blurred version. All chosen subset locations cover an area patch of  $100 \text{ km} \times 100 \text{ km}$  which is equivalent to an image size of  $3333 \times 3333$  pixels. All shown locations are in brightness temperature (Kelvin) and are extracted from their respective Landsat-8 L1C B10 data.

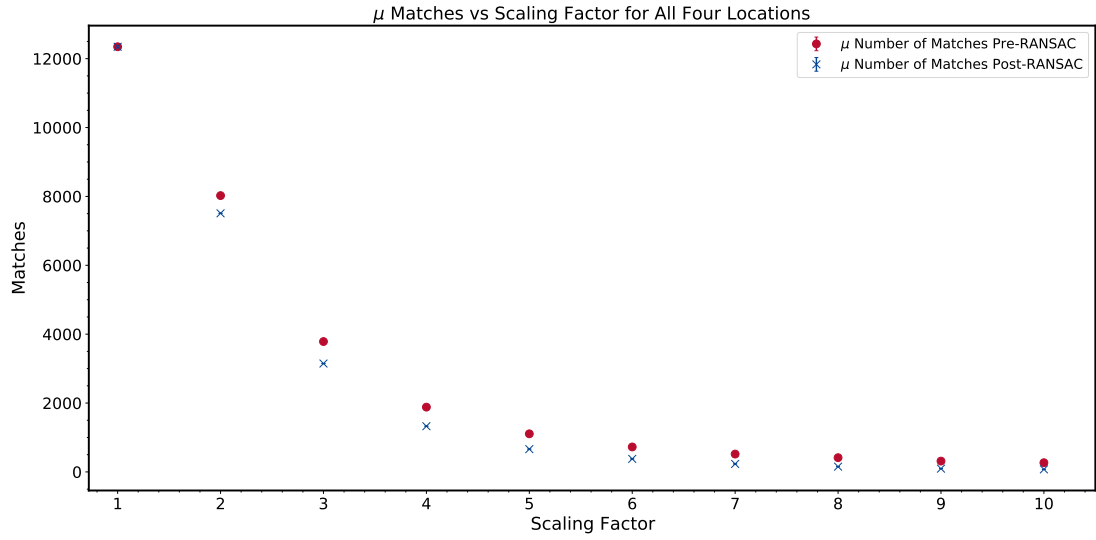


Figure 6.3: Plot of  $\mu$  matches pre and post RANSAC vs scaling factor for all four locations as shown in Fig. 6.2

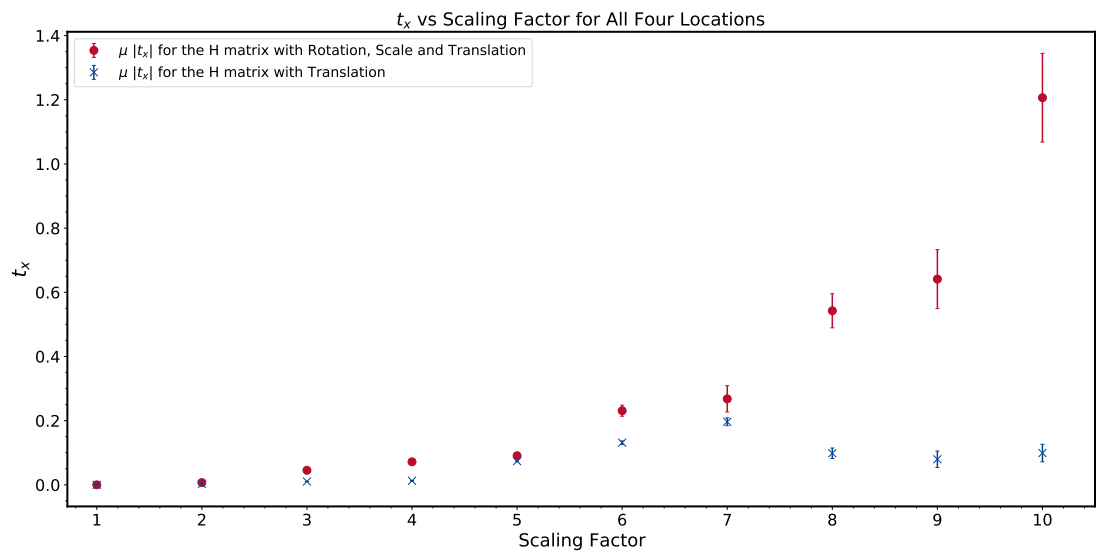


Figure 6.4: Plot of  $\mu t_x$  vs scaling factor for all four locations as shown in Fig. 6.2

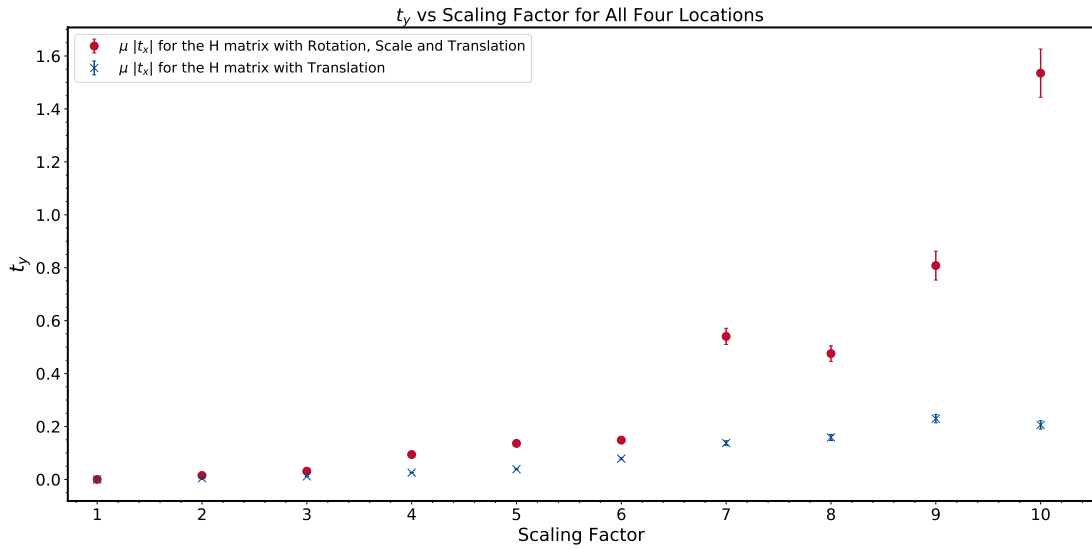


Figure 6.5: Plot of  $\mu t_y$  vs scaling factor for all four locations.

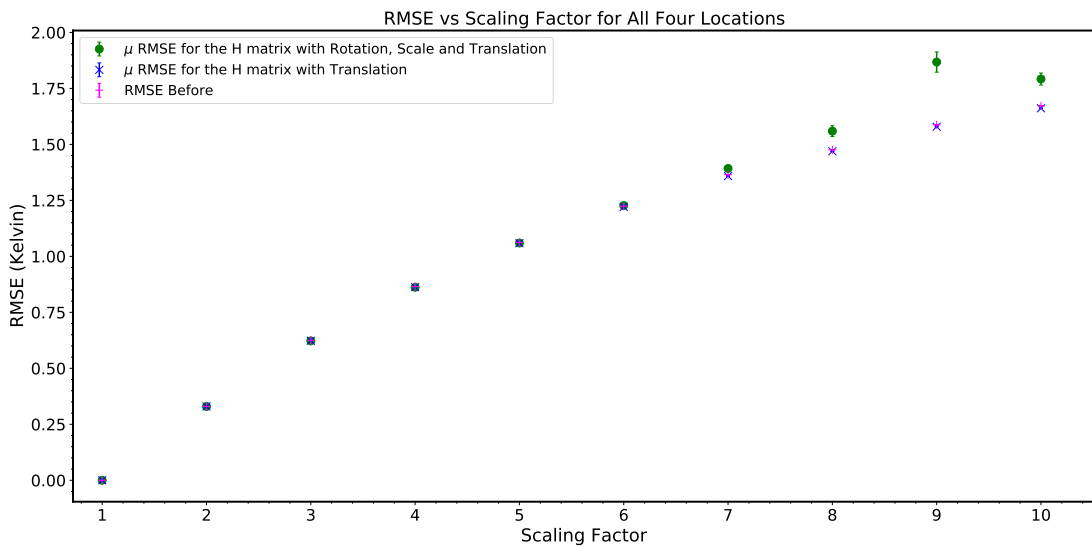


Figure 6.6: Plot of  $\mu$  RMSE before and after registration vs scaling factor for all four locations.

## 6.1.2 Translating the Sensed Image

In this section, the sensed image is the reference BT image with added translation. The induced translation is done through a change in the geolocation coordinates which equates to pixels shifts.

### 6.1.2.1 Location 1: Mallorca

The first chosen location is in Mallorca. The reference and sensed images are extracted on 2020/08/09 from their respective Landsat-8 L1C B10 data.

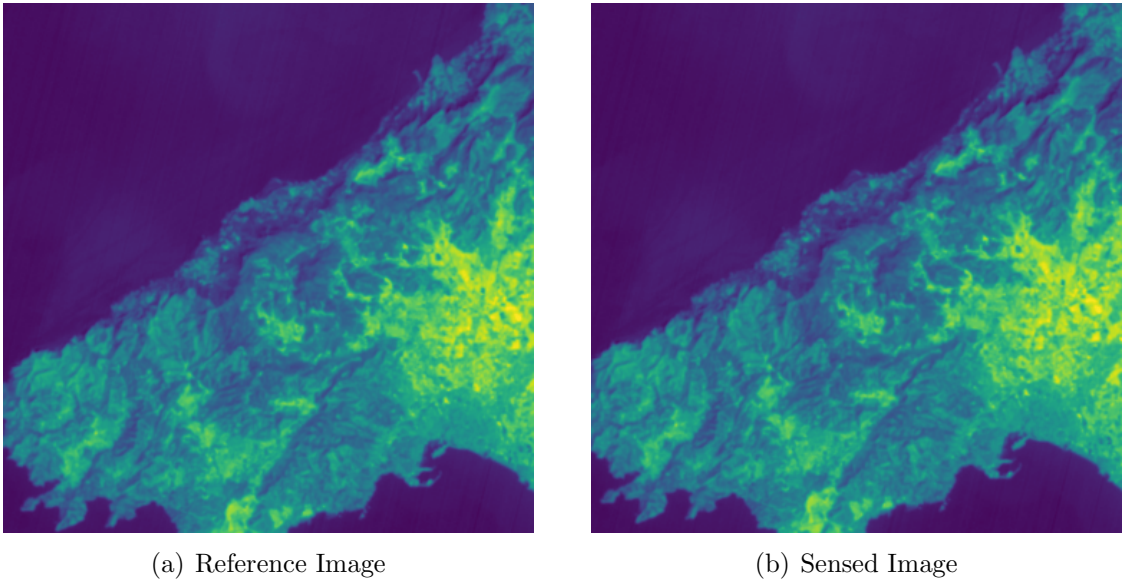


Figure 6.7: Reference and sensed images of the Mallorca location. Both images are located in the South-West of Mallorca shown in brightness temperature with a size of  $1000 \times 1000$  pixels.

Both reference and sensed images of Fig. 6.7 cover an area patch of  $30 \text{ km} \times 30 \text{ km}$ . For a spatial resolution of  $30 \text{ m}$  this is equivalent to a size of  $1000 \times 1000$  pixels. The difference between the reference and sensed images is that in the sensed image, Fig. 6.7(b), the geolocation coordinates are translated in the both the  $t_x$  and  $t_y$  direction by  $450 \text{ m}$  which is equivalent to a pixel shift of 15 pixels respectively.

The two images are registered using the proposed method of Chapter 5. The number of RANSAC iterations is 2000, the RANSAC threshold is set to 3 and the SIFT parameters are the default parameters. The two homographies utilised are



as discussed in Section 5.3.5. The result was simulated for 100 iterations and the mean values are taken.

RMSE Before (Kelvin)	RMSE After (Kelvin)		$t_x$		$t_y$	
	$H_{R,T,S}$	$H_T$	$H_{R,T,S}$	$H_T$	$H_{R,T,S}$	$H_T$
2.1030	0.0012	0	15.0186	15.0047	15.0072	14.9984

Table 6.1: Table illustrating the obtained  $\mu$  RMSE before and after registration for the Mallorca location and the obtained  $\mu$  misregistration values. Standard error for all obtained values was zero.

As shown in Table 6.1, for both homographies the RMSE after registration is approximately 0. Similarly, for the obtained misregistration values in the horizontal  $t_x$  and vertical  $t_y$  direction is approximately 15 pixels respectively.

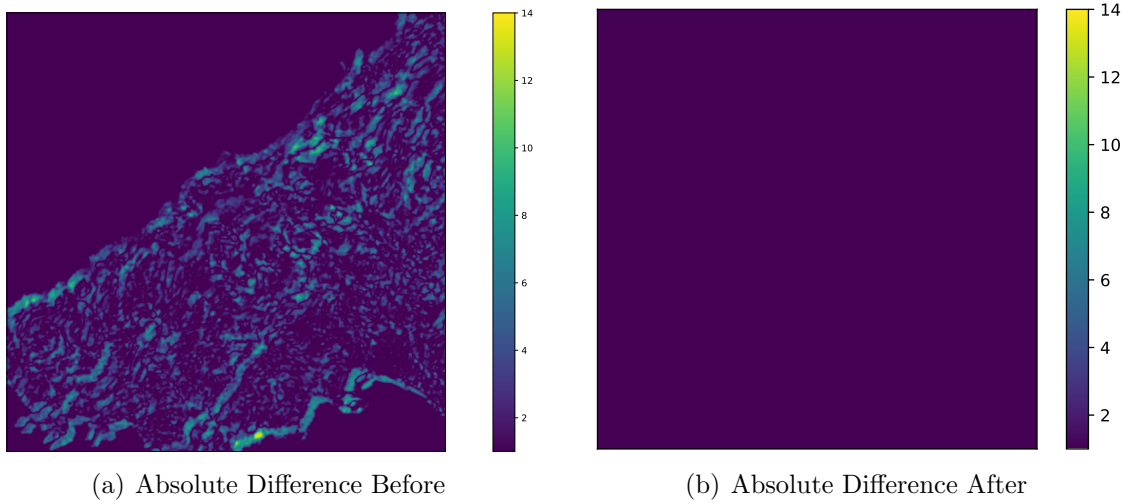


Figure 6.8: Absolute difference before and after registration for the reference and sensed images of the Mallorca location.

In Fig. 6.8(a), the yellow and light blue indicate an error between the reference and sensed images due to the added translations caused by the difference in the geolocation coordinates. In Fig. 6.8(b), the absolute difference is solely a purple colour as after registration the observed patch between the reference and the registered image is the same. Thus, there is no difference.

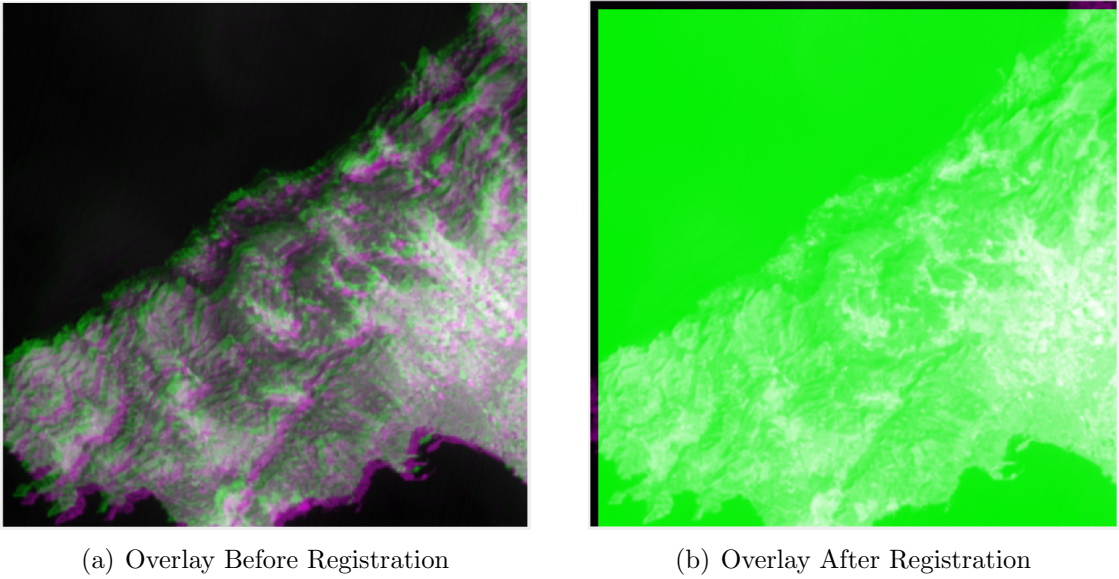


Figure 6.9: Overlay image before and after registration for the reference and sensed images of the Mallorca location.

In Fig. 6.9(a), the green colour represents the sensed image whilst the magenta colour represents the reference image before registration. Similarly, to the absolute difference before registration, the two colours do not overlap due to the translational difference between the two.

In Fig. 6.9(b), the green colour represents the registered image whilst the magenta colour represents the reference image after registration. The green square represents the overlapping region between the two images. The padded parts in magenta represent the translational difference between the reference and registered image.

### 6.1.2.2 Location 2: Corfu

The exact same test as shown in the first location (Mallorca) is done for a second location. The second location chosen covers an area found in Corfu (Greece). The chosen area patch is of  $20 \text{ km} \times 20 \text{ km}$ . For a spatial resolution of 30 m this represents an image of size  $666 \times 666$  pixels.

For this location, compared to the reference image shown in Fig. 6.10(a), the sensed image geolocation coordinates have an induced translation of 30 m in the horizontal direction  $t_x$  and 60 m in the vertical direction  $t_y$  which represents 1 and

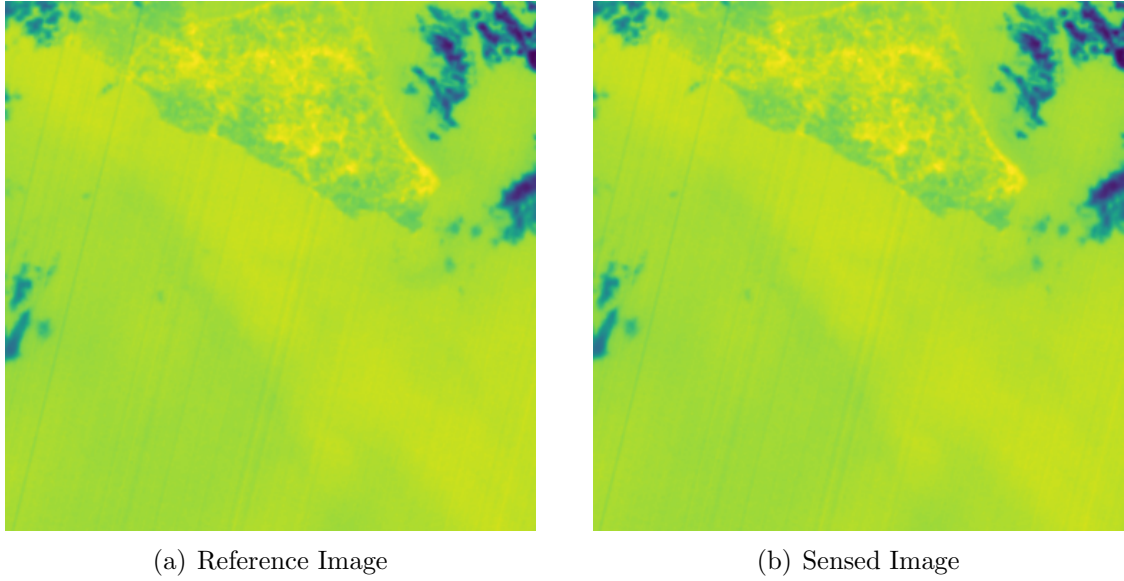


Figure 6.10: Reference and sensed images of the Corfu location. Both images are located in the South of Corfu shown in brightness temperature with a size of  $666 \times 666$  pixels.

2 pixels respectively as shown in Fig. 6.10(b).

RMSE Before (Kelvin)	RMSE After (Kelvin)		$t_x$		$t_y$	
	$H_{R,T,S}$	$H_T$	$H_{R,T,S}$	$H_T$	$H_{R,T,S}$	$H_T$
0.17436	0.01324	0.0030	1.0437	0.9758	1.9571	2.0038

Table 6.2: Table illustrating the obtained  $\mu$  RMSE before and after registration for the Corfu location and the obtained  $\mu$  misregistration values. Standard error for all obtained values was zero.

As shown in Table 6.2, for both homographies the RMSE after registration is approximately 0. Similarly, for the obtained misregistration values in the horizontal  $t_x$  and vertical  $t_y$  direction is approximately 1 and 2 pixels respectively.

In Fig. 6.11(a), the green colour represents the sensed image whilst the magenta colour represents the reference image before registration. Qualitatively the difference is minimal due to the small translational difference.

In Fig. 6.11(b), the green colour represents the registered image whilst the magenta colour represents the reference image after registration. The green square represents the overlapping region between the two images. The padded parts in magenta

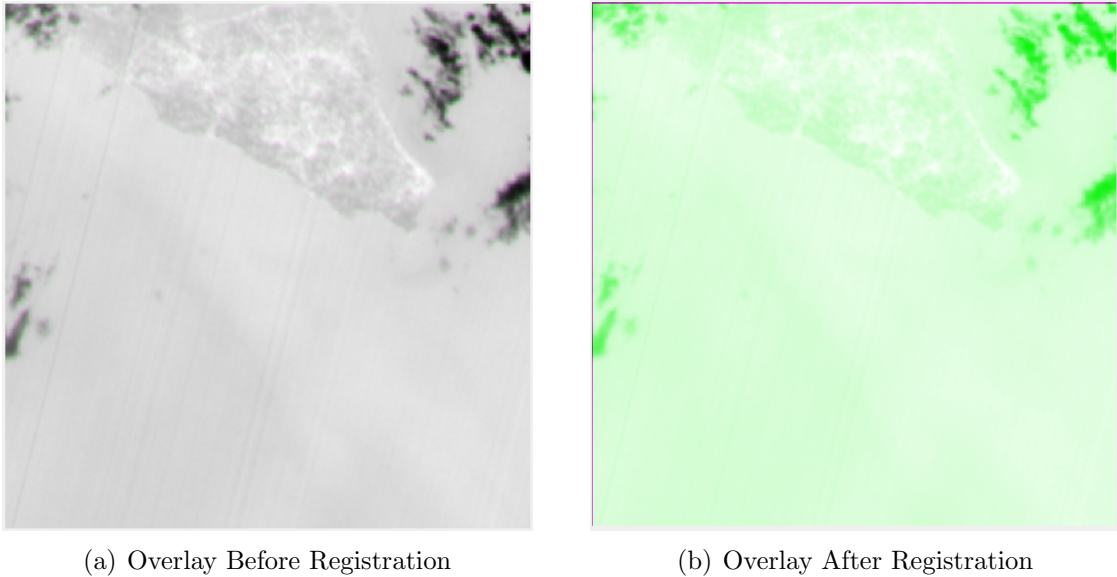


Figure 6.11: Overlay image before and after registration for the reference and sensed images of the Corfu location.

represent the translational difference between the reference and registered image. Since the translational difference is not greater than 2 pixels the non-overlapping area is less than compared to the Mallorca result. Nonetheless, one can observe that in the vertical direction the magenta area that is not overlapping is twice as much as compared to the horizontal direction.

### 6.1.3 Effect of Cloud Cover

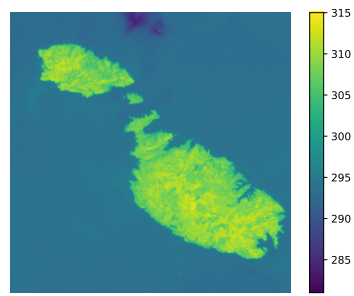
The aim of this section is to show the effect of cloud cover on the registration between two Landsat-8 B10 BT images. The chosen location is the Maltese archipelago. Fig. 6.12(a) is chosen as the reference image whilst Fig. 6.12(b) to Fig. 6.12(h) are chosen as the sensed images to be registered. The chosen reference image has 0% cloud cover (CC) whilst the sensed images have a cloud cover that varies from 11.58% to 61.25%.

*Both the reference and sensed images are extracted using the same geolocation coordinates from L8 B10 L1C products and the sensed images have no shifts or added blur.* The chosen area patch for all images is  $40 \text{ km} \times 40 \text{ km}$  which for a spatial resolution of 30 m represents an image size of  $1333 \times 1333$  pixels. The size of the dilation and erosion kernel is 3 and 2 pixels respectively. The cloud cover percentage was computed from the L8 cloud mask of each chosen image in Fig. 6.12 through the following equation:

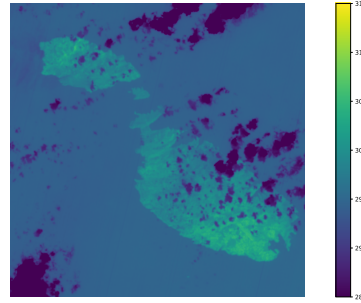
$$CC\% = \frac{100 \times n_c}{N} \quad (6.1)$$

where  $n_c$  is the number of cloud pixels found in the L8 cloud mask and  $N$  is the total number of pixels found in the image.

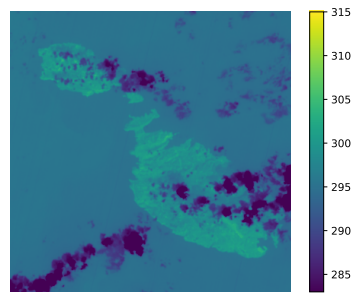
The registration algorithm utilised is the one proposed in Chapter 5. The SIFT parameters utilised are the default. The RANSAC threshold was set to 3 and the number of RANSAC iterations was set to 2000. The chosen reference image was registered respectively to all the sensed images found in Fig. 6.12. Each registration result was repeated for 100 times and the mean for that respective result is taken.



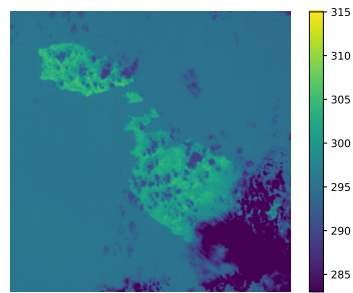
(a) 0.04% cloud cover taken on 2018/06/02.



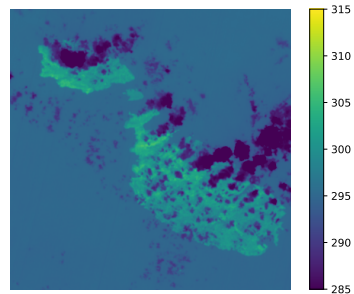
(b) 11.58% taken on 2018/09/22.



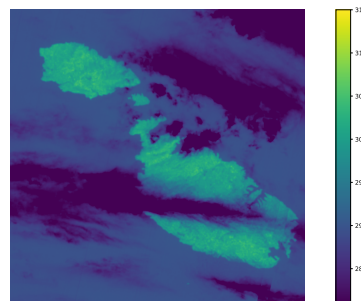
(c) 13.41% taken on 2018/10/08.



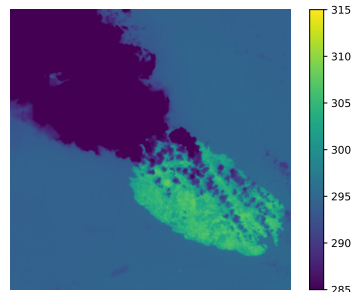
(d) 13.78% taken on 2020/09/11.



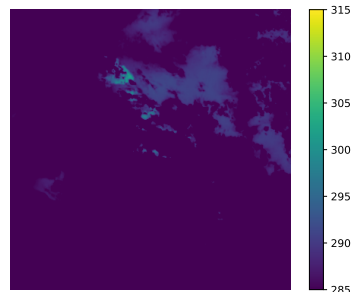
(e) 15.26% taken on 2018/08/21.



(f) 24.9% taken on 2018/05/01.



(g) 26.7% taken on 2018/08/05.



(h) 61.65% taken on 2016/06/12.

Figure 6.12: Images of the Maltese archipelago taken in different days with different percentages of cloud cover. Fig. 6.12(a) is chosen as the reference image. Fig. 6.12(b) to Fig. 6.12(h) are sensed images.

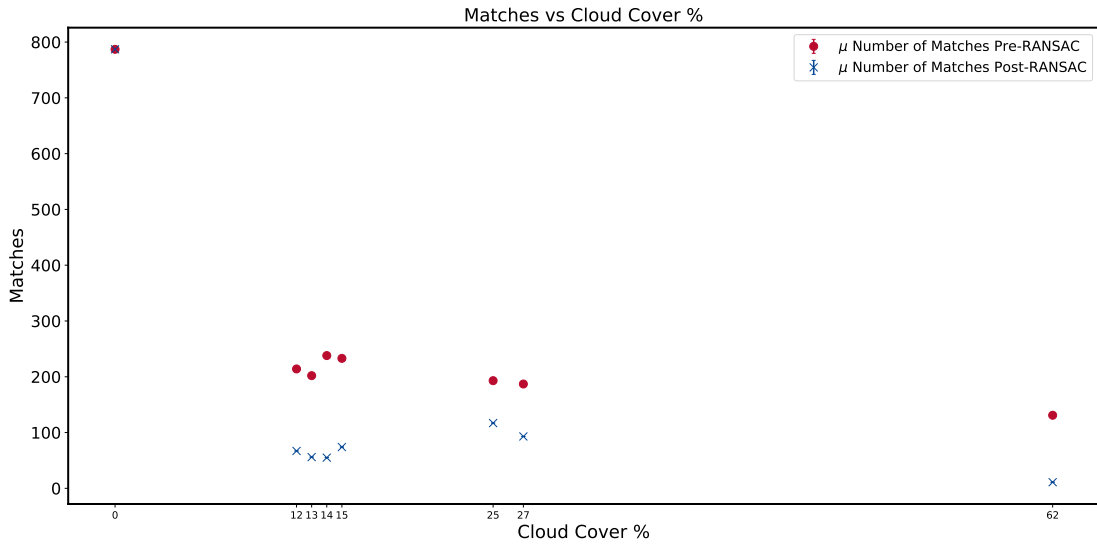


Figure 6.13: Plot of  $\mu$  matches vs the cloud cover percentage for each registered sensed image found in Fig. 6.12.

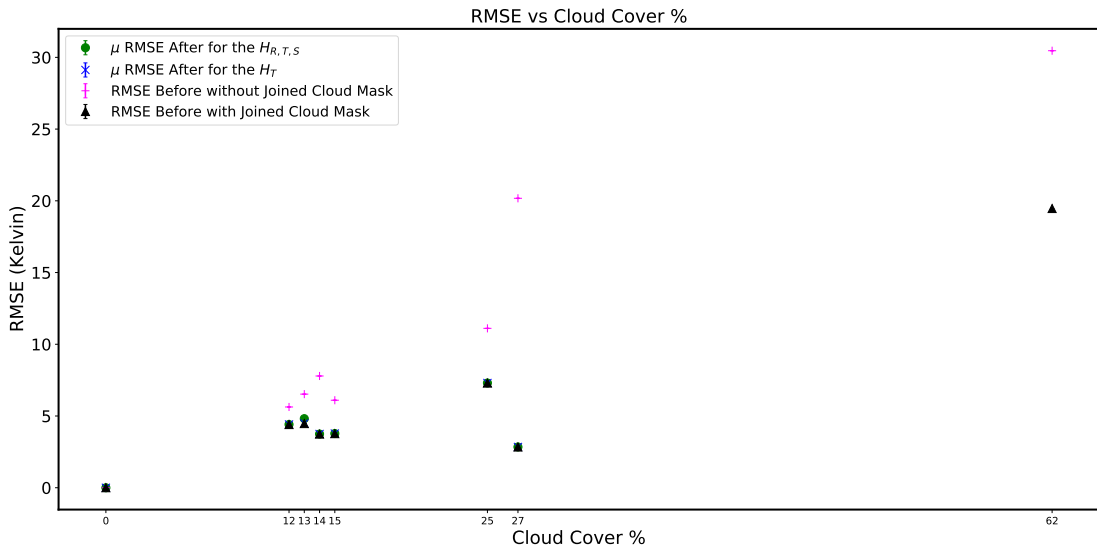


Figure 6.14: Plot of  $\mu$  RMSE vs the cloud cover percentage for each registered sensed image found in Fig. 6.12. **Note:** for each image, the cloud cover % was rounded off to the nearest integer.

From Fig. 6.13, one can observe that as the cloud cover percentage increases, the number of matches before and after RANSAC decrease. This is expected since more cloud cover will occlude features in the sensed image. Thus, SIFT will detect less features to utilise for registration. With regards to the sensed image of Fig. 6.12(h) which had the highest cloud cover (61.6%), the registration failed. Fig. 6.12(h) is too occluded to detect any meaningful matches.

In Fig. 6.14, one can observe that as the cloud occlusion increases the RMSE before registration without the joined cloud mask increases. This is expected as more cloud coverage implies a bigger temperature difference between the reference and sensed images since the reference does not have any clouds. Apart from the sensed image of Fig. 6.12(h) in which the registration failed, the RMSE after registration for both homographies was equal to the RMSE before registration with the joined cloud mask.

In Fig. 6.14, the registered sensed image with 27% cloud cover obtained the second smallest RMSE after registration. It is even smaller than the sensed image with 11% cloud cover. This is due to the fact that in Fig. 6.12(g), the clouds are mainly concentrated in one region, at sea.

The RMSE after registration along with the joined cloud mask, is not expected to decrease less than the RMSE before registration since the images before are already well registered as they originate from the same band of the same satellite. It is to be noted that the only way the RMSE before and after registration both using the joined cloud mask can decrease is by increasing the size of the dilation kernel. Nonetheless, increasing the dilation will lead to less features which will also effect the matching.

$t_x$		$t_y$	
$\mathbf{H}_{\mathbf{R},\mathbf{T},\mathbf{S}}$	$\mathbf{H}_{\mathbf{T}}$	$\mathbf{H}_{\mathbf{R},\mathbf{T},\mathbf{S}}$	$\mathbf{H}_{\mathbf{T}}$
0.2856	0.1153	0.2239	0.1183

Table 6.3: Obtained  $\mu$  misregistration in the  $x$  and  $y$  direction for all registered sensed images of Fig. 6.12(h). The obtained standard error for all four values was zero.

In Table 6.3, the highest obtained misregistration in the  $x$  and  $y$  direction was 0.2856 pixels and 0.2239 pixels respectively. In meters, for a spatial resolution



of 30 m, the obtained misregistration was 8.568 m and 6.717 m respectively. The highest registration in  $x$  and  $y$  both resulted from, the homography  $\mathbf{H}_{\mathbf{R},\mathbf{T},\mathbf{S}}$  which considered rotation, scale and translation.

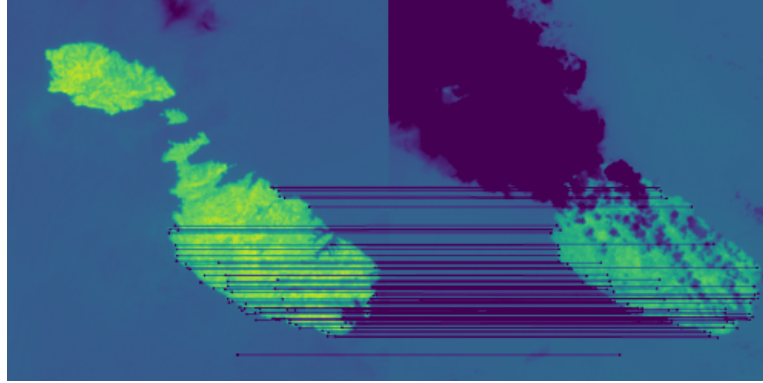


Figure 6.15: Matched features between the reference image Fig. 6.12(a) which has 0% cloud cover and the sensed image Fig. 6.12(h) which has 26.7%.

In Fig. 6.15, since the joined cloud mask is utilised to avoid detecting features from clouds, before even commencing the algorithm, SIFT avoids detecting any features coming from cloud pixels. The predominant matches in Fig. 6.15 occurred in the coastal regions, the areas with the highest change in gradients between land and sea and the areas with the most corners.

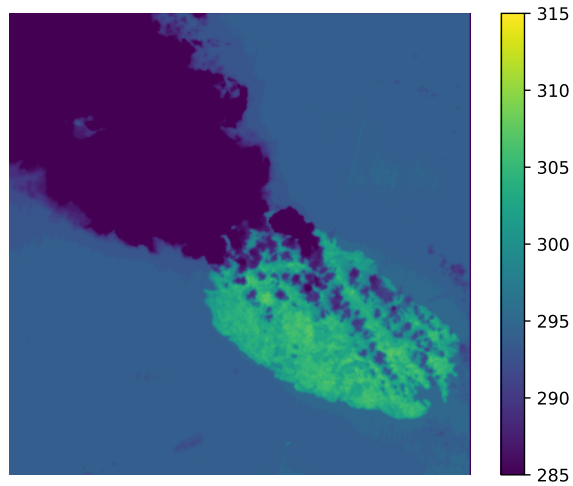


Figure 6.16: Registered image.

Fig. 6.16 shows the registered image for the images found in Fig. 6.15. Since the attained misregistration in Table 6.3 is subpixel in accuracy, no translational borders appear.

## 6.2 Registration of Landsat-8 vs Sentinel-3

### 6.2.1 Downscaling the Landsat-8 Image

In this subsection the Landsat-8 B10 subset is *downscaled to the same spatial resolution of Sentinel-3 B8 subset* which is 1 km. The chosen area patch for both images is  $100 \text{ km} \times 100 \text{ km}$ . For the Landsat-8 image at a spatial resolution of 30 m this equates to a size of  $3333 \times 3333$  pixels and for the Sentinel-3 image at 1 km spatial resolution this equates to a size of  $100 \times 100$  pixels.

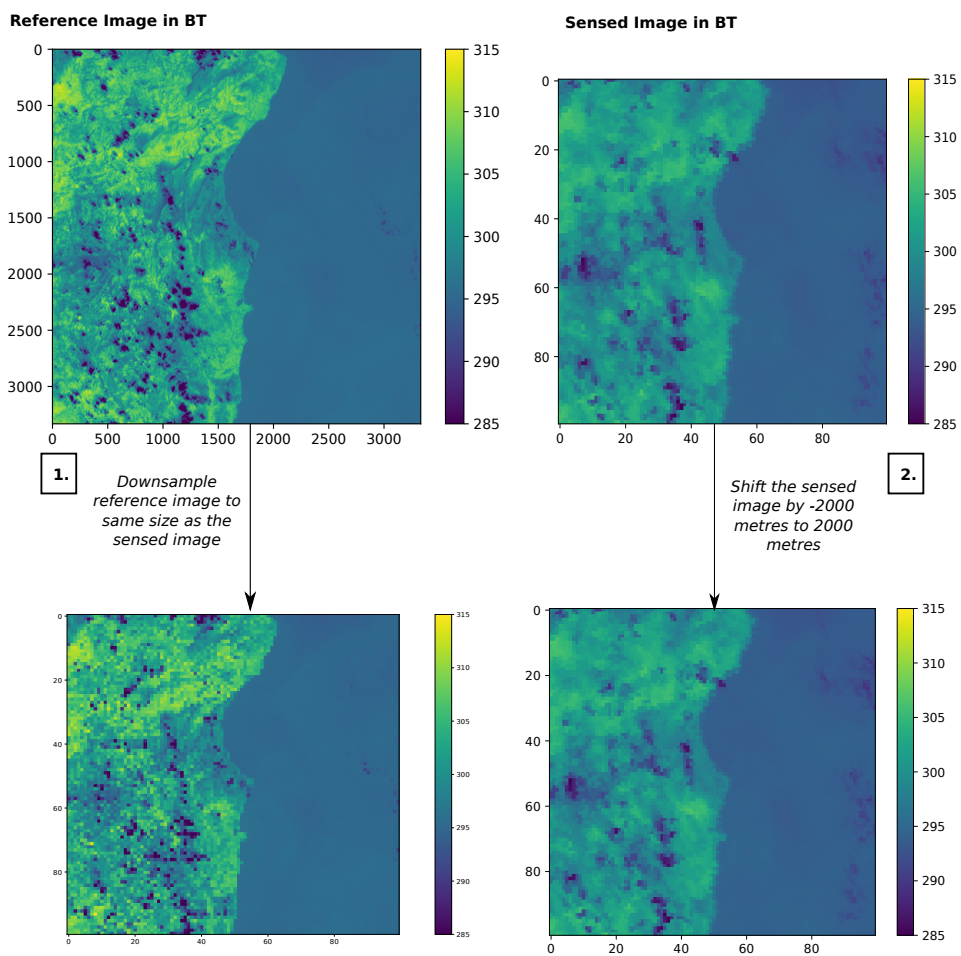
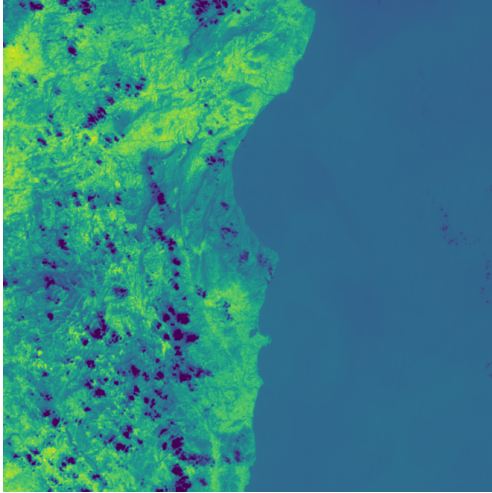
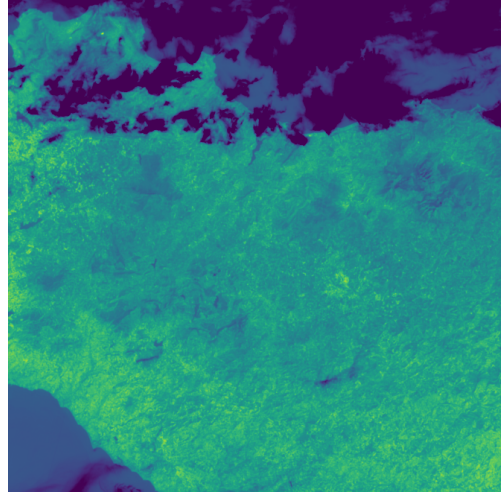


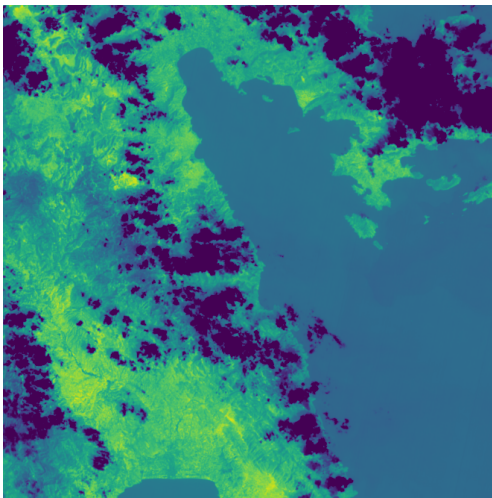
Figure 6.17: Workflow for the preprocessing steps done on the reference and sensed images for the registration of the downsampled L8 B10 subset with the S3 B8 subset. Covered geolocation for both images before shifting the sense image in step 2 is the same. Chosen locations for this example is Sardegna.



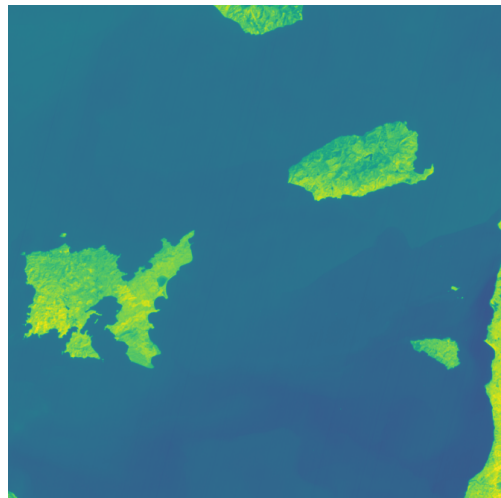
(a) Reference Image of Location 1 (Sardegna) taken on 2020/08/06.



(b) Reference Image of Location 2 (Sicily) taken on 2019/05/11.



(c) Reference Image of Location 3 (Athens) taken on 2020/08/07.



(d) Reference Image of Location 4 (Samothrace, Greek Island) taken on 2020/07/31.

Figure 6.18: Reference images of the four chosen locations to be utilised for the test of registering a downscaled Landsat-8 B10 image with its respective Sentinel-3 B8 image. All chosen subset locations cover an area patch of  $100 \text{ km} \times 100 \text{ km}$ . All shown locations are in brightness temperature and are extracted from their respective Landsat-8 L1C B10 data and Sentinel-3 SLSTR data products respectively.

1. In Fig. 6.17, the reference Landsat-8 B10 image of size  $3333 \times 3333$  pixels is *downscaled* to the same spatial resolution as the sensed Sentinel-3 B8 image of size  $100 \times 100$  pixels using *bicubic interpolation*.
2. The geolocation coordinates of the sensed image is shifted from  $-2000$  m to  $2000$  m. In pixels for a spatial resolution of  $1$  km this equates to  $-2$  to  $2$  pixel shifts.
3. The reference image is registered respectively to each of the shifted sensed images. For registration the default parameters with RANSAC threshold of  $3$ , the parameters generated using the Monte Carlo approach and the parameters generated using the Grid Search method were utilised. For all methods the Monte Carlo and Grid Search method were repeated for each shifted sensed image. RANSAC in all utilised approaches was set to  $2000$  iterations and the size of the dilation and erosion kernels were set to  $2$  and  $1$  pixel respectively. The Monte Carlo approach was repeated for  $1000$  iterations whereas the Grid Search approach was repeated for  $100$  iterations.
4. Steps 1 to 3 are repeated for three additional locations shown in Fig. 6.18. For all locations, both the Landsat-8 and Sentinel-3 images are taken on the same day approximately  $40$  minutes apart. In the results of Landsat-8 vs Landsat-8,  $\mathbf{H}_T$  outperformed  $\mathbf{H}_{R,T,S}$  (as shown in Fig. 6.2). Thus, for the remaining results of this section, only  $\mathbf{H}_T$  is considered.

Fig. 6.19, shows the plot of the  $\mu$  RMSE for all locations before and after registration for all the respective shifts using all three parameter methods. The RMSE before registration shows a parabolic effect. This is expected due to the shifting. As the shift in the sensed image increases, the sensed and reference images become more dissimilar. Thus the RMSE before registration should increase. The standard error for the  $100$  repetitions for each result of each method for each shift was approximately zero implying that the obtained result for all the  $100$  repetitions was identical. All three methods resulted in a smaller RMSE after registration than the RMSE before registration. The Grid Search method obtained the smallest RMSE after registration followed by the Monte Carlo and default parameters. It is to be noted that the improvement is marginal as all three methods obtained a similar result. The RMSE after registration can never be zero as the two images are coming from two different satellites.

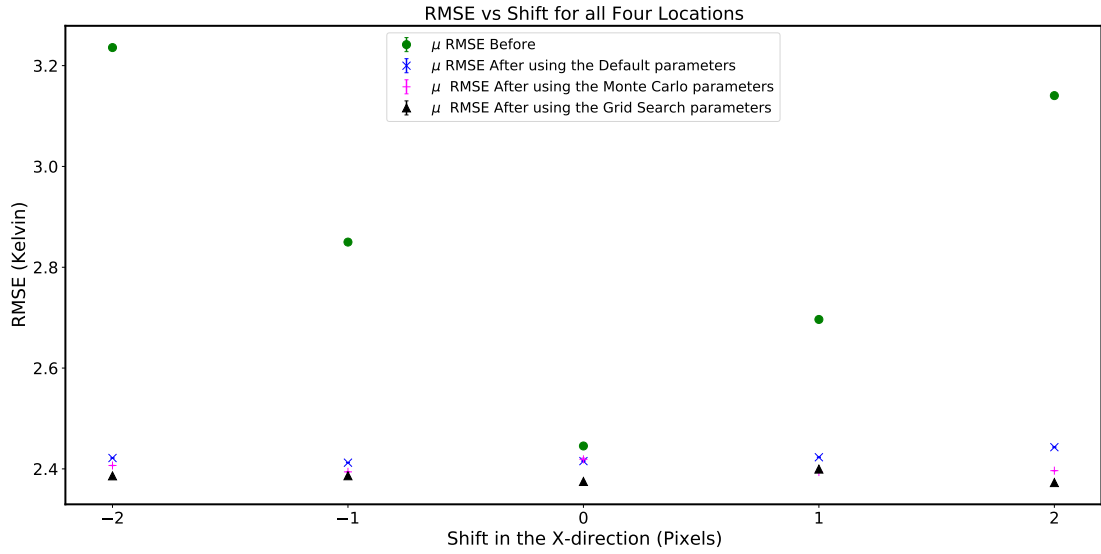


Figure 6.19: Plot of  $\mu$  RMSE vs Shift using  $\mathbf{H}_T$

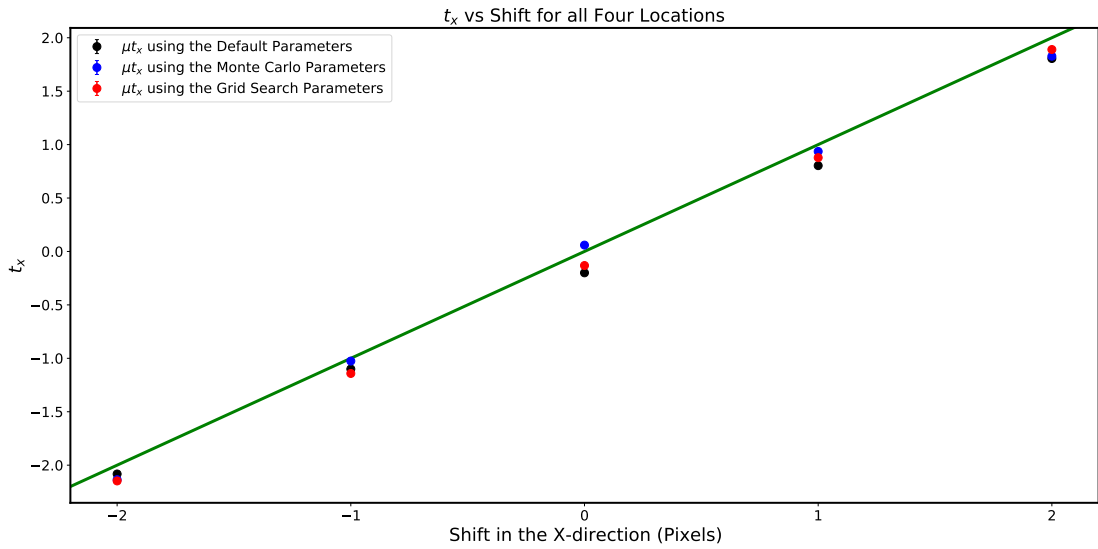


Figure 6.20: Plot of  $\mu t_x$  vs Shift using  $\mathbf{H}_T$ . The green represents  $y = x$ . The aim of this line is to illustrate how far off are the attained points from the green line. This represents the misregistration. The highest attained misregistration was at a zero shift using the default parameters of 0.1999 pixels.

Fig. 6.20 and Table 6.4 show the obtained misregistration in the  $x$  direction. The obtained misregistration is subpixel. The largest obtained misregistration was at a zero shift using the default parameters of 0.1999 pixels which represents a misregistration of 199.9 m for a spatial resolution of 1 km. This misregistration might imply that the downscaled Landsat-8 image and the Sentinel-3 image might be slightly

	$t_x$				
<b>Default parameters</b>	-2.0825	-1.1004	-0.1999	0.8032	1.8057
<b>Monte Carlo parameters</b>	-2.1354	-1.0256	0.0594	0.9360	1.8271
<b>Grid Search parameters</b>	-2.1476	-1.1418	-0.1316	0.8781	1.8899

Table 6.4: Obtained  $\mu$  misregistration in the  $x$  direction for all registered shifted sensed images for the plot of Fig. 6.20

misregistered before registration. It should be noted that since the two images are coming from two different satellites the correct misregistration value is not known due to a lack of ground truth. This notion applies for the explained misregistration analysis between all multi-modal satellite imagery.

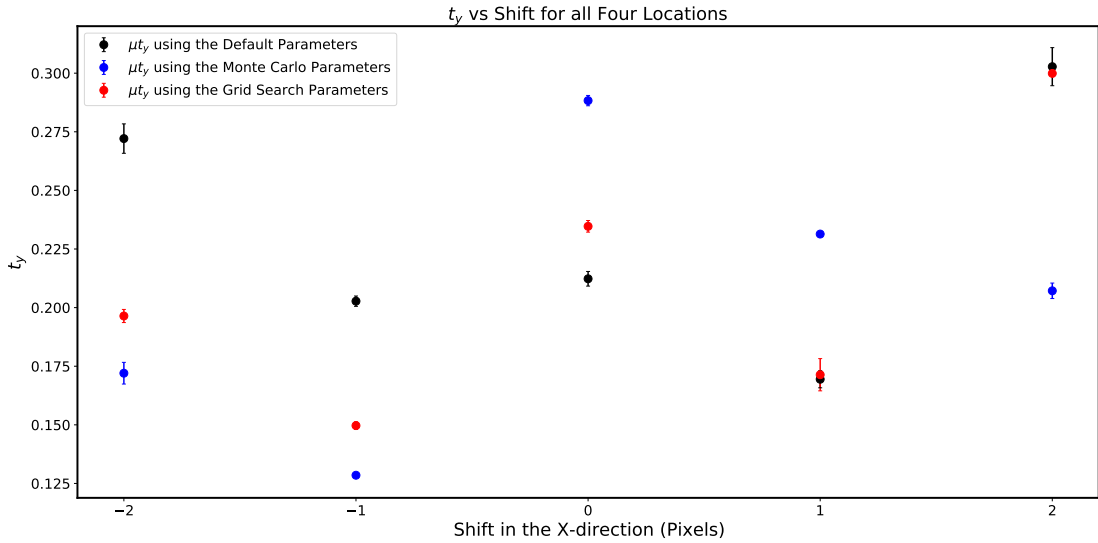


Figure 6.21: Plot of  $\mu t_y$  vs Shift using  $\mathbf{H}_T$

Fig. 6.21 shows the obtained misregistration in the  $y$  direction. The obtained misregistration for all three parameter methods for all shifts in the  $y$  direction is also subpixel, below 0.3 pixels which is equivalent to a misregistration of 300 m for a spatial resolution of 1 km.

### 6.2.2 Upscaling the Sentinel-3 Image

In this subsection, the Sentinel-3 B8 subset (sensed image) is shifted then *upscaled* at different upscaling factors whilst the Landsat-8 B10 subset (reference image) is downscaled to the same resolution as the upscaled sensed image. The chosen *initial* area patch for both images before downscaling or upscaling is  $100 \text{ km} \times 100 \text{ km}$  which equates to a size of  $3333 \times 3333$  pixels (for L8) and  $100 \times 100$  (for S3) for a spatial resolution of 30 m and 1 km respectively. The chosen location is Samothrace as shown in Fig. 6.18(d).

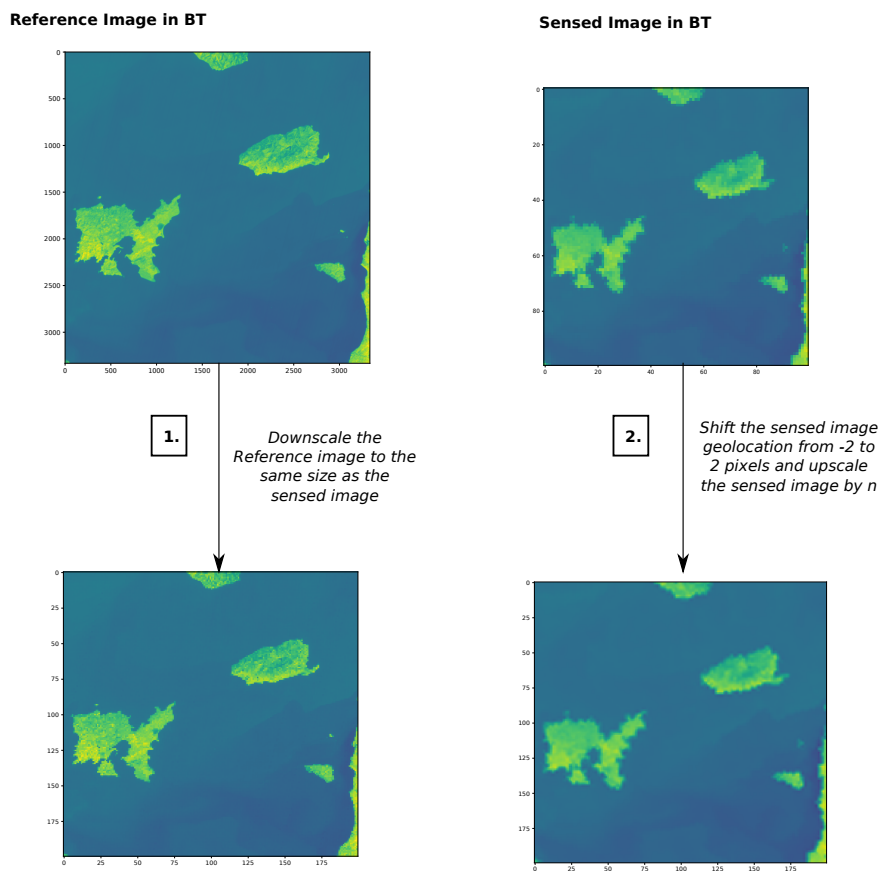


Figure 6.22: Workflow for the preprocessing steps done on the reference and sensed images for the registration of the downscaled L8 B10 subset with the shifted and upscaled S3 B8 subset. Covered geolocation for both images before shifting the sense image is the same. Chosen location is Samothrace.

1. In Fig. 6.22, the reference Landsat-8 B10 image of size  $3333 \times 3333$  pixels is *downscaled* to the same spatial resolution as the upscaled sensed Sentinel-3 B8. Both downscaling and upscaling were done using *bicubic interpolation*.

2. The geolocation coordinates of the sensed image is shifted from  $-2$  to  $2$  pixel shifts. The geolocation is shifted in m, but the amount of meters is dependent on the chosen upscaled size of the Sentinel-3 sensed image. If the sensed image is upscaled by  $n = 2$  and shifted by 2 pixels this is equivalent to a shift of 500 m. The upscaled sensed is shifted from  $-2$  to  $2$  pixels for each upscaled image from  $n = 2$  to  $n = 10$  as shown in Fig. 6.22.
3. The reference image is then registered respectively to each of the shifted and upscaled sensed images. For registration the set parameter conditions are as utilised in the case of downscaling the Landsat-8 image to the same native spatial resolution as Sentinel-3 as shown and explained in Section 6.2.1.

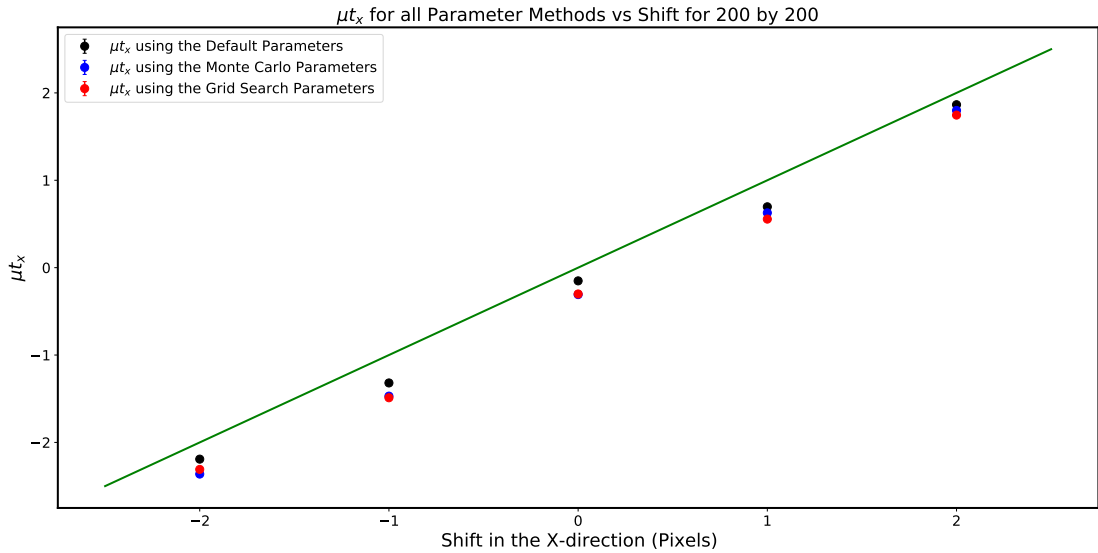


Figure 6.23: Plot of  $\mu t_x$  vs Shift using  $\mathbf{H}_T$  for  $200 \times 200$

Fig. 6.23 shows the plot of the attained mean  $\mu$  misregistration in the  $x$  direction for each shift using  $\mathbf{H}_T$  for an upscaled size of  $200 \times 200$ . *Attained standard error for all shifts was approximately zero for all upscaled sensed images.* In this plot the best parameter method was the method with default parameters since the values of that parameter set are the closest to the green line. The largest attained misregistration occurred at a shift of  $-1$  using the Grid Search method which attained  $-1.4887$  with a misregistration of  $0.4887$  pixels and  $244.35$  m for a spatial resolution of  $500$  m. It is worth noting that all three methods seem to have a common offset that forms a straight line. This might infer that there is a common misregistration amongst the Landsat-8 and Sentinel-3 images. The obtained  $|\mu|$



offset for the Default, Monte Carlo and Grid Search was 0.1930, 0.3428 and 0.3593 pixels respectively.

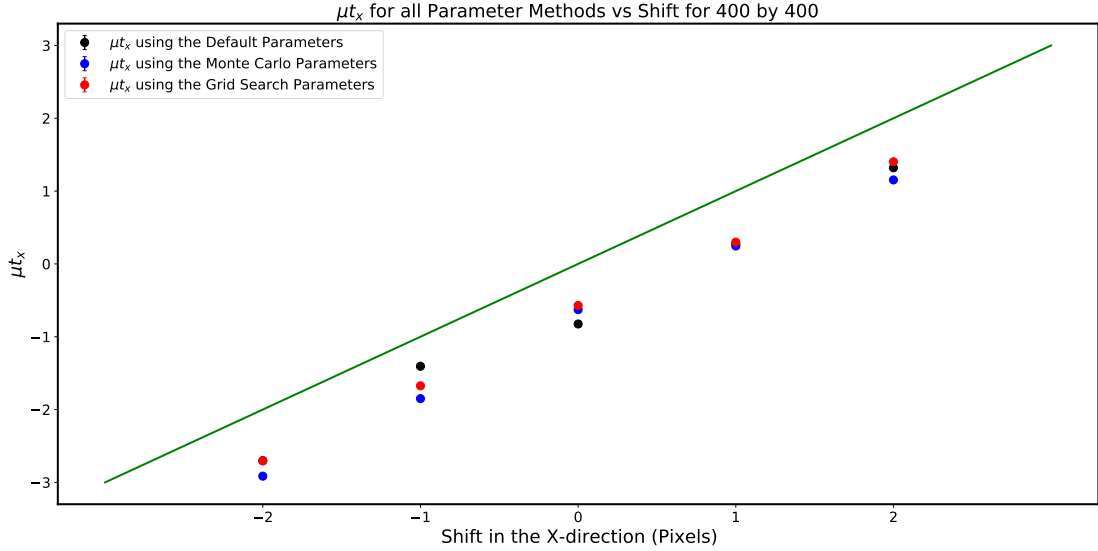


Figure 6.24: Plot of  $\mu t_x$  vs Shift using  $\mathbf{H}_T$  for  $400 \times 400$

In Fig. 6.24, the best parameter method was the Grid Search followed by the Default and Monte Carlo methods. The largest attained misregistration occurred at a shift of  $-2$  using the Monte Carlo method which attained  $-2.9148$  with a misregistration of  $0.9148$  pixels and  $228.7$  m for a spatial resolution of  $250$  m. The obtained  $|\mu|$  offset for the Default, Monte Carlo and Grid Search at  $400 \times 400$  was  $0.6674$ ,  $0.7986$  and  $0.5692$  pixels respectively. Compared to the  $200 \times 200$ , the pattern is an increase in misregistration when the sensed image gets upscaled by a larger upscaling factor. In fact, this can be observed both from the plots as the values are getting further away from the green line and from the  $|\mu|$  offset which is getting larger. The  $|\mu|$  offset approximately increased by a multiplication factor of  $2$ .

In Fig. 6.25, the obtained  $|\mu|$  offset for the Default, Monte Carlo and Grid Search at  $600 \times 600$  was  $0.9777$ ,  $1.0924$  and  $1.0962$  pixels respectively. The largest attained misregistration occurred at a shift of  $1$  using the Grid Search method which attained  $-0.8120$  with a misregistration of  $1.8120$  pixels and  $301.87$  m for a spatial resolution of  $166.6$  m.

In Fig. 6.26, the obtained  $|\mu|$  offset for the Default, Monte Carlo and Grid Search

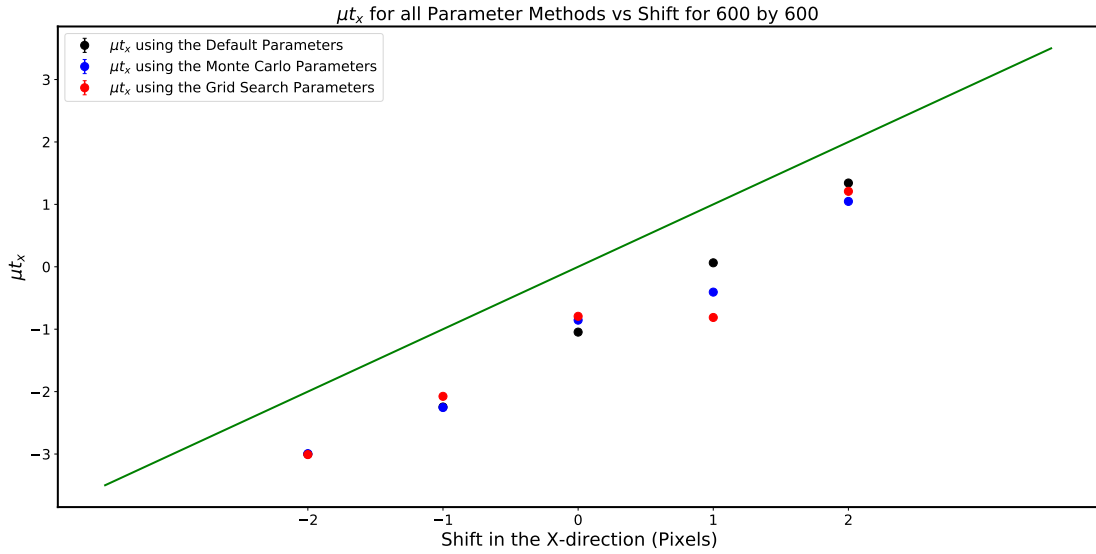


Figure 6.25: Plot of  $\mu t_x$  vs Shift using  $\mathbf{H}_T$  for  $600 \times 600$

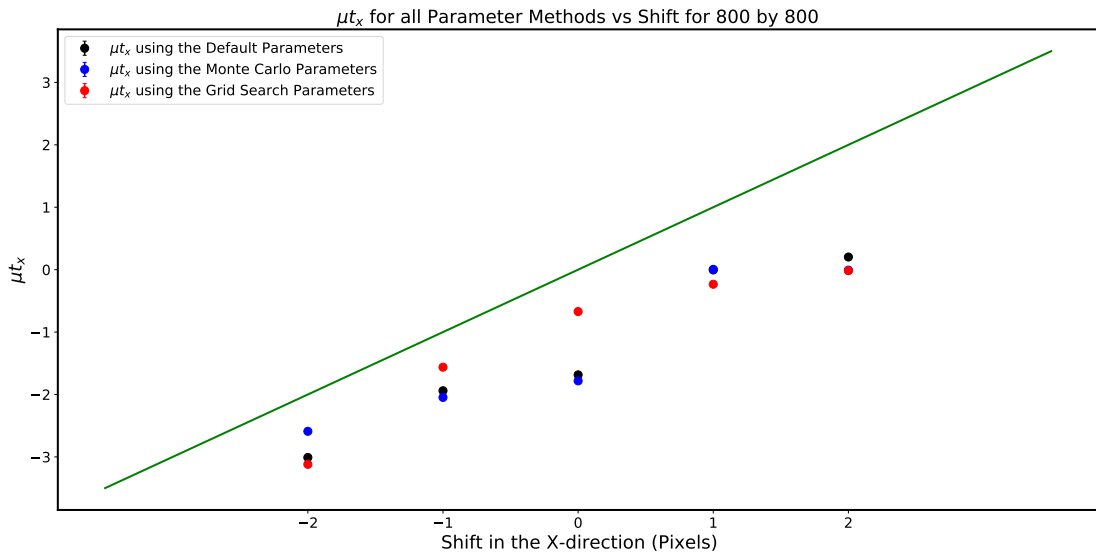
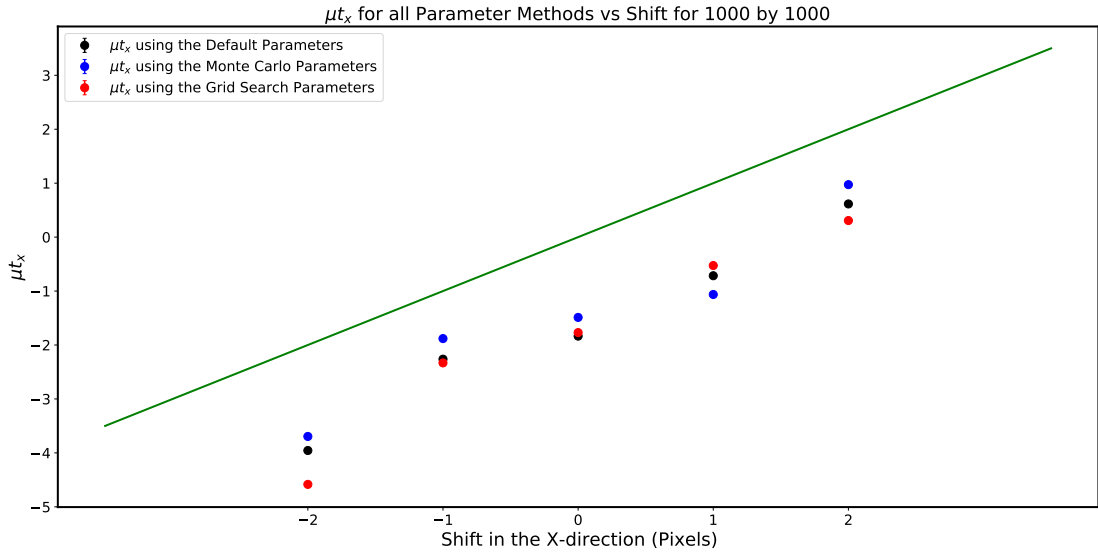


Figure 6.26: Plot of  $\mu t_x$  vs Shift using  $\mathbf{H}_T$  for  $800 \times 800$

at  $800 \times 800$  was 1.2874, 1.2843 and 1.1203 pixels respectively. The largest attained misregistration occurred at a shift of 2 using the Grid Search method which attained  $-0.0151$  with a misregistration of 2.0151 pixels and 251.89 m for a spatial resolution of 125 m.


 Figure 6.27: Plot of  $\mu t_x$  vs Shift using  $\mathbf{H}_T$  for  $1000 \times 1000$ 

In 6.27, the obtained  $|\mu|$  offset for the Default, Monte Carlo and Grid Search at  $1000 \times 1000$  was 1.6308, 1.4310 and 1.7809 pixels respectively. The largest attained misregistration occurred at a shift of  $-2$  using the Grid Search method which attained  $-2$  with a misregistration of 2.5840 pixels and 258.4 m for a spatial resolution of 100 m.

Highest Misregistration (pixels)	Highest Misregistration (m)	Upscaled Size of Sensed Image (pixels)	Spatial Resolution (m)
0.4887	244.35	$200 \times 200$	500
0.9148	228.7	$400 \times 400$	250
1.8119	301.87	$600 \times 600$	166.6
2.0144	251.8	$800 \times 800$	125
2.584	258.4	$1000 \times 1000$	100

 Table 6.5: Obtained highest misregistration in the  $x$  direction for each upscaled sensed image in meters

From Fig. 6.23 to Fig. 6.27, two general trends were observed:

1. As the upscaling of the sensed image increased, the misregistration in the  $x$  direction increased. This is evident from the points as they get further away from the green line and from the absolute  $|\mu|$  offset which increases. This increase in misregistration is caused from the effect of upscaling using bicubic

interpolation. Bicubic interpolation not only adds blur to the upscaled sensed image but also introduces false information in the image. This is clearly illustrated in Fig. 1.2 where the sensed image was upscaled from  $100 \times 100$  to  $1000 \times 1000$  pixels. In the coastal regions of the sensed image false parts of land are introduced.

2. Despite the increase in misregistration in pixels, the largest misregistration for the  $x$  direction in meters for all upscaled sensed images was roughly similar. This is shown in Table 6.5, for all sizes a  $\mu$  misregistration in the  $x$  direction of 257 m was attained. This occurred as a result of the upscaling which causes the spatial resolution to increase accordingly.

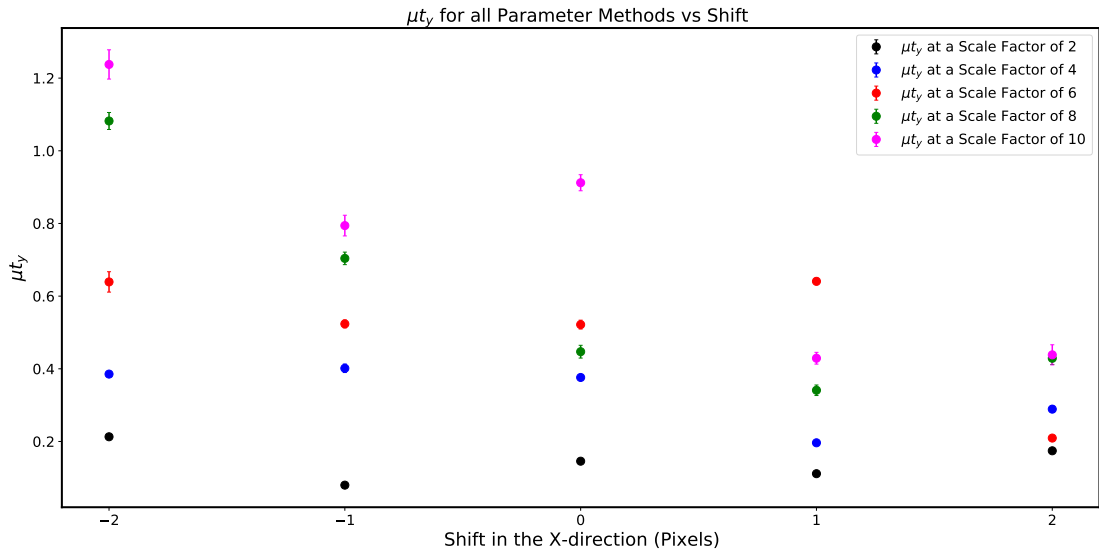


Figure 6.28: Plot of  $\mu t_y$  for all the parameter methods vs Shift using  $\mathbf{H}_T$  for the different upscaled sensed images

In Fig. 6.28, a similar misregistration in the  $y$  direction was also observed, as the upscaling increased, the misregistration in the  $y$  direction also increased. Since no shift was induced in the  $y$  direction of the sensed image, the obtained misregistration in the  $y$  direction was less than the obtained misregistration in the  $x$  direction.

Similarly to the  $x$  direction, despite the increase in misregistration in pixels, the largest misregistration for the  $y$  direction in meters for all upscaled sensed images was roughly similar. This is shown in Table 6.6, for all sizes a  $\mu$  misregistration in the  $y$  direction of 121.45 m was attained. This occurred as a result of the upscaling which causes the spatial resolution to increase accordingly.

Highest Misregistration (pixels)	Highest Misregistration (m)	Upscaled Size of Sensed Image (pixels)	Spatial Resolution (m)
0.2130	106.5	200 × 200	500
0.4014	100.35	400 × 400	250
0.6407	106.74	600 × 600	166.6
1.0820	135.25	800 × 800	125
1.2376	123.76	1000 × 1000	100

Table 6.6: Obtained highest misregistration in the  $y$  direction for each upscaled sensed image in meters

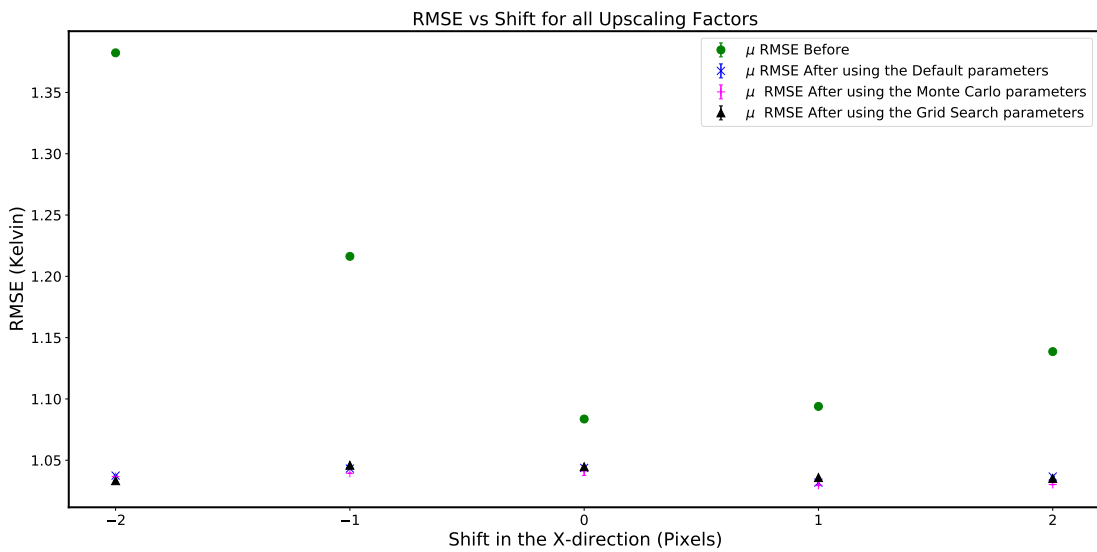


Figure 6.29: Plot of  $\mu$  RMSE vs Shift using  $\mathbf{H}_T$

Fig. 6.29 shows the plot of the  $\mu$  RMSE of all the upscaled sensed images for each parameter method at each respective shift. The RMSE before registration shows a parabolic effect. This is expected due to the shifting. As the shift in the sensed image increases, the sensed and reference images become more dissimilar. Thus the RMSE before registration should increase. Standard error for each result was zero. All the parameter methods obtained a smaller RMSE after registration with an attained identical RMSE performance. **Inherently, this shows that for the tackled problem of registering thermal-to-thermal remote sensing data different parameter sets will lead to an identical RMSE after registration.**

## 6.3 Registration of Multi-Modal Remote Sensing Data

In this section remote sensing single/multi-modal data is registered using the following registration algorithms:

- **Area-based Algorithms:**

1. Phase Correlation (POC) [89]
2. Enhanced Correlation Coefficient (ECC) [46]

- **Feature-based Algorithms:**

1. SIFT based algorithm outlined in Chapter 5
2. Radiation-Variation Insensitive Feature Transform (RIFT) [12]

- **Optical Flow-based Algorithm:**

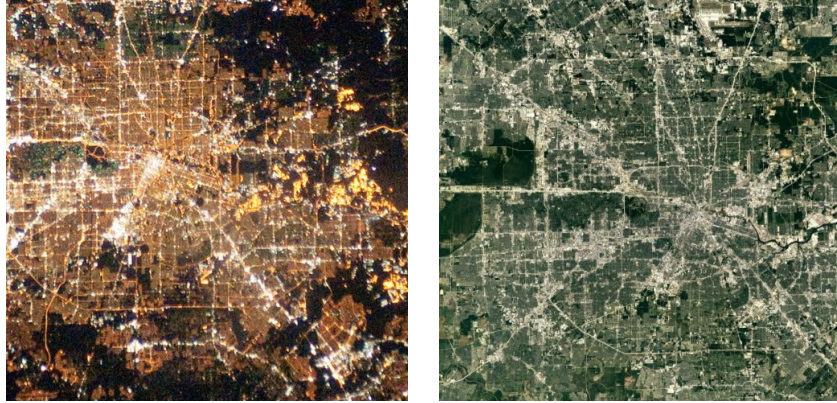
1. SIFT Flow [48]

The implementation utilised for the phase correlation method is done by Y. Ri and H. Fujimoto [89] where they utilised it to estimate drift-free motion estimation from video images. The utilised phase correlation method simply consists of using 2D FFT to convert images to the frequency domains, and then they calculate the IFFT of the correlation of those two images resulting in a peak which indicates the highest correlated area between the two images [89]. In the case no peak is established, the proposed method did not find any correlated parts between the two images. In addition Y. Ri et al. also utilise linear optimization to find any translational differences between the two images [89]. The proposed phase correlation method is implemented in Python or MATLAB [90].

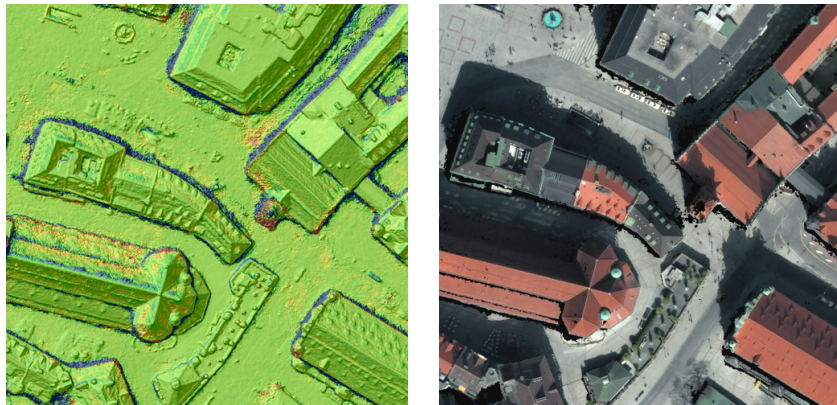
The implementation of the SIFT algorithm is in Python as explained in Chapter 5. For the registration of the RIFT dataset, the only difference in the implementation was that the min-max normalisation step was not required since the images are already in Uint8. In addition, all the implementation used for the remainder of the algorithms is as mentioned in Section 5.1. Unless stated otherwise, all methods used the default parameters. In Sections 6.3.1 and 6.3.2, the SAR-SIFT algorithm [70] is also utilised to register Optical to SAR image pairs.

### 6.3.1 RIFT Dataset

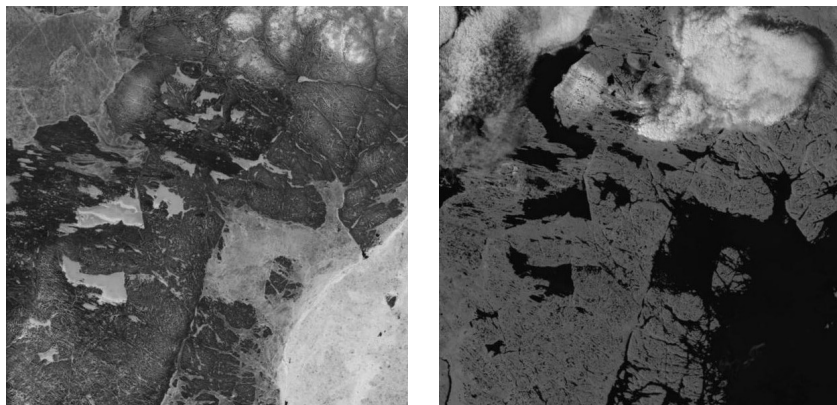
The utilised image dataset in this section involves six image pairs taken from the RIFT dataset [82].



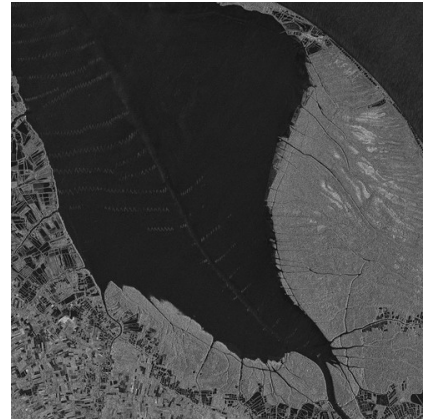
(a) Night-Day



(b) Depth-Optical



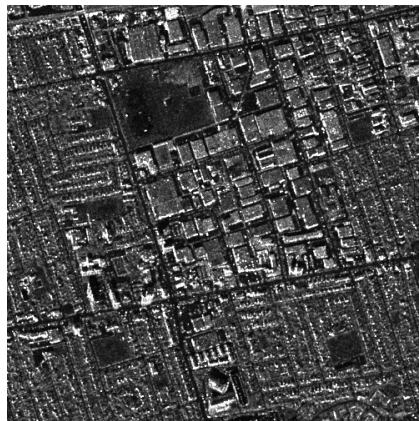
(c) Infrared-Optical



(d) Map-Optical



(e) Optical-Optical



(f) SAR-Optical

Figure 6.30: RIFT image dataset. Images on the left column are the sensed images whilst images on the right are the reference images.

Authors of RIFT state that these image pairs shown in Fig. 6.30 are very difficult to register as they contain severe radiation distortions [12]. Fig. 6.30(b) contains



translation change and a  $90^\circ$  rotation change [12]. With regards to the five remaining image pairs whether they have any translational, scale or rotation changes is not known as the remaining image pairs provided in the dataset are not the same images utilised in the results of the RIFT paper [12].

Modality of Image Pair	RMSE Before	POC	ECC	SIFT Flow	SIFT Default	SIFT MC	SIFT GS	RIFT
Night-Day	10.9705	<b>X</b>	<b>X</b>	<b>X</b>	8.9586	8.9199	8.8567	9.0077
Depth-Optical	6.2562	<b>X</b>	<b>X</b>	5.6783	<b>X</b>	<b>X</b>	<b>X</b>	6.1291
Infrared-Optical	12.7390	12.5521	10.9320	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	10.9018
Map-Optical	0.3660	<b>X</b>	0.1802	0.1744	0.1874	0.1478	0.2025	0.2374
Optical-Optical	9.4798	8.7781	8.2049	8.9239	8.2134	8.7065	8.2111	8.9228
SAR-Optical	14.0206	<b>X</b>	14.0502	14.1113	<b>X</b>	<b>X</b>	<b>X</b>	13.9702

Table 6.7: Table of results for each image pair found in the RIFT dataset. MC and GS refer to the proposed Monte Carlo (MC) and Grid Search (GS) parameter methods respectively. These methods were discussed in Sections 5.4.3 and 5.4.4.

Table 6.7 shows whether each registration result for each image pair using the respective algorithm was successful or not. Fields that have a number indicate the respective attained RMSE after registration whilst a **X** indicates that the registration failed. **The registration is deemed as a failure if the sensed image is not correctly warped qualitatively. Additionally, the smallest RMSE denotes the algorithm which best registered the sensed image both quantitatively and qualitatively.**

*For the rmse calculation any translational borders are excluded from calculation and for the infrared-to-optical when using the ECC algorithm, a translation transformation over a affine (default) was utilised. In all image pairs, the  $\mathbf{H}$  that considers rotation, translation and scale was utilised for the SIFT algorithm.*

In the Night-day image pair, the two feature-based algorithms, SIFT and RIFT were successful. POC, ECC and SIFT flow failed the registration. With regards to the two area-based methods, they were expected to fail as the two image pairs have a huge difference in intensity. Furthermore, the POC method may yield ambiguous results for images that have several peaks such as periodic images. It is to be noted that the Enhanced Correlation Coefficient function was modified to be robust to geometric and photometric distortions such as contrast and brightness but in this Night-day image pair the intensities are too different. In the attempt to register the two images, SIFT Flow gave a higher priority to the deformation found in the top-left of the reference in (day) image in Night-Day pair. Thus, causing failure in

registration. Fig. 6.32(a) shows the overlay of the failed registration. SIFT Flow was not designed for images that have little overlap or geometric distortions. Fig. 6.31 shows two failed examples of when using SIFT Flow. SIFT flow was designed for scene alignment with common overlap such as motion detection. Additionally, SIFT Flow can also be utilised for the registration of non-rigid deformations.

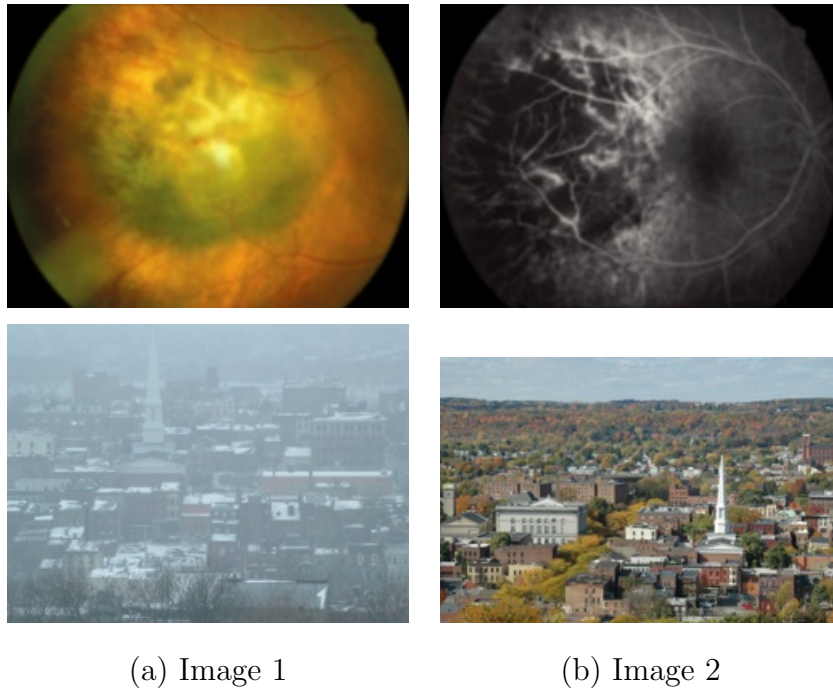


Figure 6.31: Failed cases of registration when using SIFT Flow.

As shown in Table 6.7, in the Depth-Optical image pair, the only two algorithms that had successful registration were SIFT Flow and RIFT. Fig. 6.33 shows the overlay after registration using SIFT Flow. The two area-based algorithms, POC and ECC were expected to fail as the two images have completely different intensity. Based on the RIFT literature, the SIFT algorithm was expected to fail due to the fact that the matching in SIFT depends heavily on the similarity of the gradient histograms of the image pair. They emphasise that the fundamental reason for its poor matching is down to its sensitivity to “nonlinear radiation distortion” (NRD). Fig. 6.34 shows the failed matching between the depth-optical using SIFT.

As given in Table 6.7, in the Infrared-Optical image pair, the algorithms that were successful in the registration were the two area-based methods (POC and ECC) and RIFT. Before registration, it was expected that the two area-based methods succeed in the registration of this pair as both images lie in a similar wavelength of the electromagnetic spectrum with similar intensity. Fig. 6.35 shows the failed

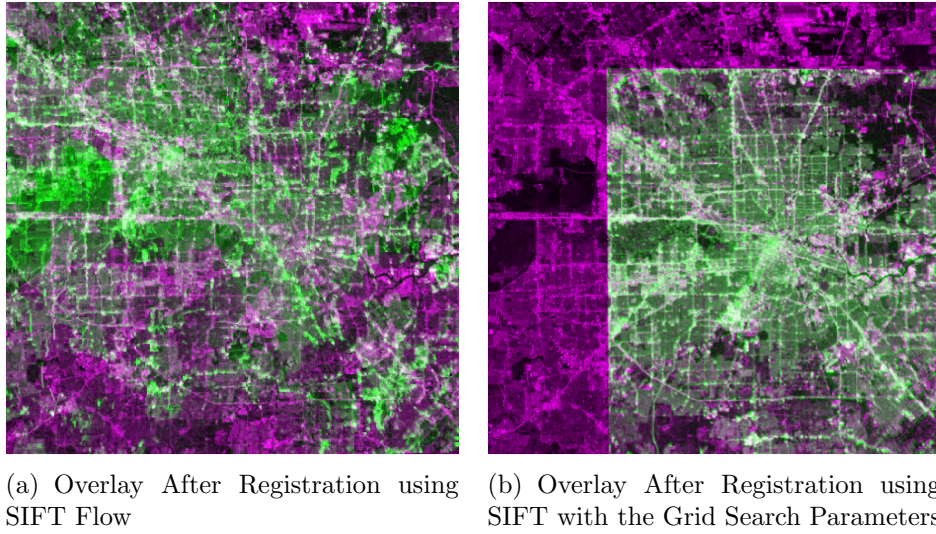


Figure 6.32: Overlay of the night-day image pair after registration. Magenta denotes the reference image and the green denotes the registered image. The overlay before registration is shown in Fig. 5.9(a).

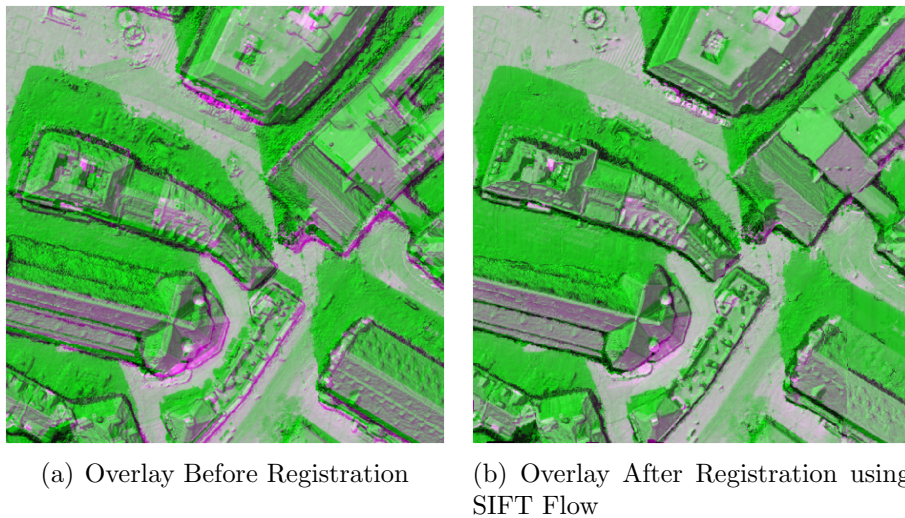


Figure 6.33: Overlay of the Depth-Optical image pair before and after registration. Magenta denotes the reference image and the green denotes the registered image. Before registration green refers to the sensed image.

matching when using SIFT. It was not expected that SIFT Flow and especially SIFT do not register this image pair. Since SIFT is the most widely feature-based algorithm to register optical images in the remote sensing domain as outlined in literature.

To challenge this failed registration case a test was made. In this test the optical image of Fig. 6.30(d) was complemented to reverse the black and white of the

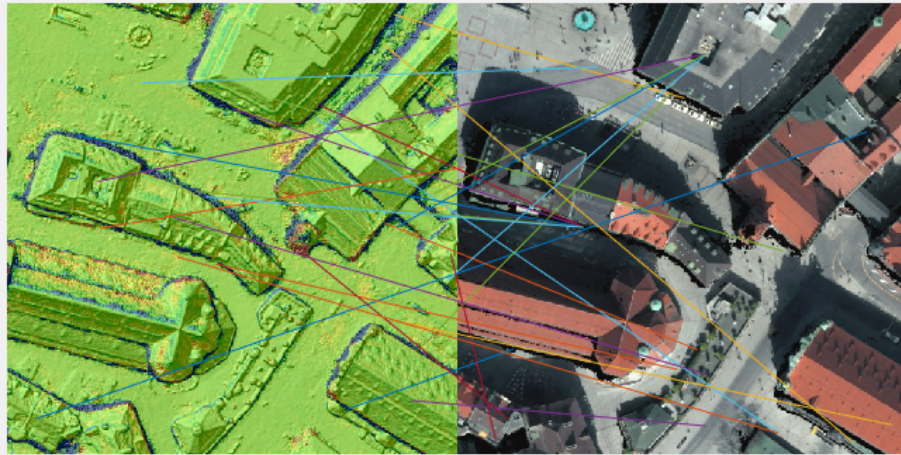


Figure 6.34: Failed matching for the Depth-Optical image pair using SIFT.

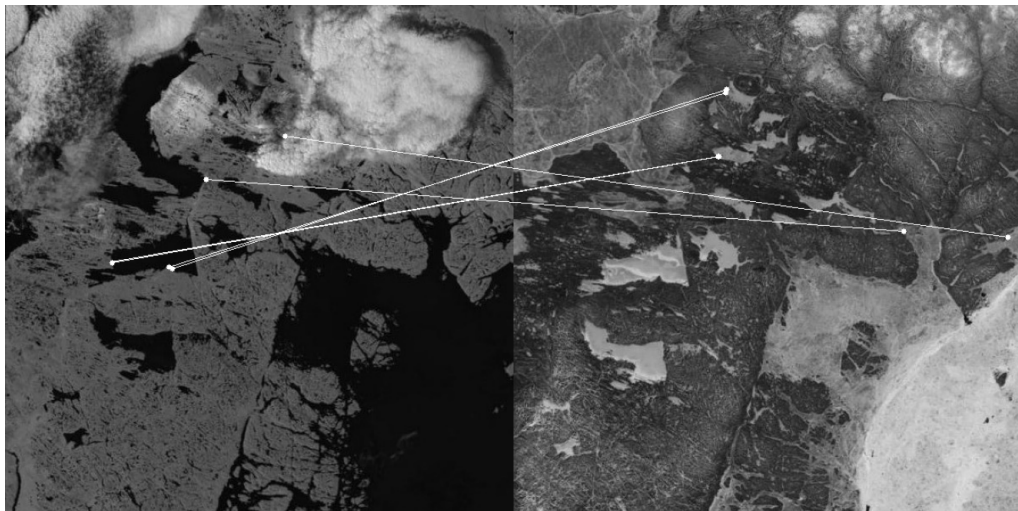


Figure 6.35: Failed matches when registering the infrared-optical image pair using SIFT. Left image is the optical. Right is the infrared.

optical image. The infrared image remained untouched.

After complementing the optical image for the infrared-optical pair SIFT was successful in matching and registering the images. Fig. 6.36 shows the matched features using SIFT with the default parameters.

Fig. 6.37 shows the successful registration of the infrared-optical pair after complementing the optical image using SIFT. In the left image the magenta colour represents the reference image which is the complemented optical image and the green image represents the infrared image. On the right, the magenta represents

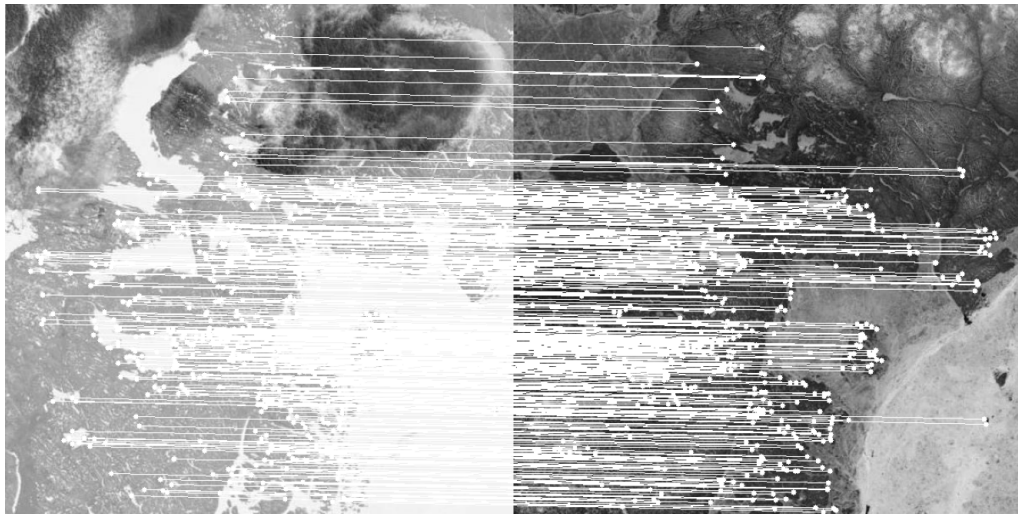


Figure 6.36: Obtained matches after complementing the optical image of the infrared-optical pair using SIFT with the default parameters. Left image is the *complemented* optical. Right image is the infrared.

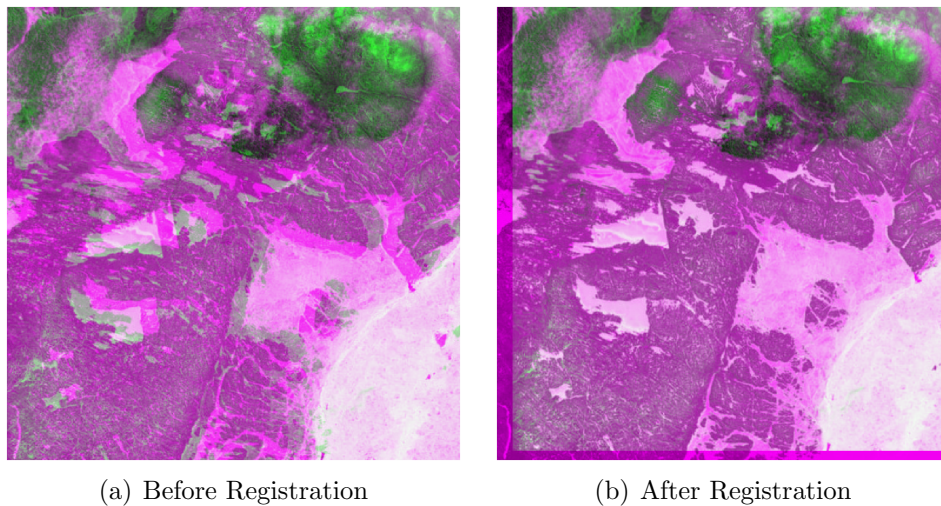


Figure 6.37: Overlay image before and after registration for the infrared and complemented optical image pair using SIFT with the default parameters.

the reference which is the complemented optical and the green represents the registered image.

Furthermore, similarly to the SIFT algorithm, after complementing the optical image the SIFT Flow algorithm was also successful in registering the infrared-optical image pair. Fig. 6.38 shows the overlay plot between the reference and registered image using SIFT Flow.

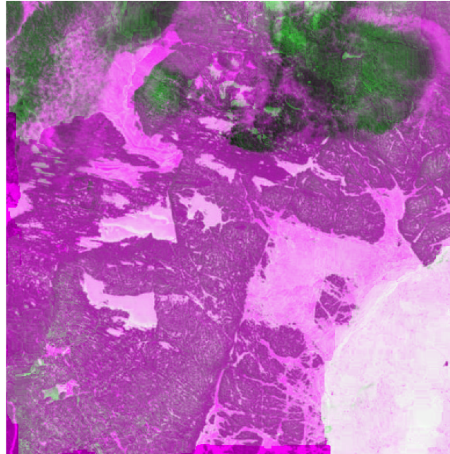


Figure 6.38: Overlay image before and after registration for the infrared and complemented optical image pair using SIFT Flow with the default parameters.

As shown in Table 6.7, the only algorithm that failed to register the map-optical image pair is the POC. Contrary to the POC, the ECC algorithm managed to register the map-optical pair. This is due to the fact that Evangelidis et al. modified the criterion of the correlation coefficient to be invariant to *geometric distortions and photometric distortions* in contrast and brightness [46].

For the Optical-Optical image pair all algorithms were successful in their registration. This was expected as the image pair has a single modality with similar intensity. All the SIFT parameter methods obtained a smaller RMSE than RIFT.

For the SAR-Optical image pair, POC and SIFT failed to register the image pair. The POC was expected to fail as no single peak was found and the two images are of different intensity. The ECC managed due to the modification done to the criterion of the correlation coefficient which makes it invariant to contrast and brightness changes. For the RMSE values, the RMSE for the ECC and POC is slightly larger as the image pair is already well registered before registration and the RMSE is always expected to differ when images are of different modality. RIFT obtained the best registration result, but decrease in RMSE was still minimal as the image pair is already well registered before registration.

For SIFT, it is worth noting the SAR-SIFT algorithm was created to tailor for such an image pair. The SIFT algorithm was not primarily designed to register SAR-

optical images and in some instances such as this case it might fail. For the optical-SAR image pair, the SAR-SIFT was also utilised. *Using the default parameters of SAR-SIFT, SAR-SIFT also failed to register the image pair. Fig. 6.39 shows these failed matches.* In the proposed work of RIFT, the authors showed that for their tested multi-modal datasets, in terms of matching RIFT outperformed both SIFT and SAR-SIFT [12].



Figure 6.39: Obtained failed matches using SAR-SIFT for the RIFT dataset Optical-SAR image pair.

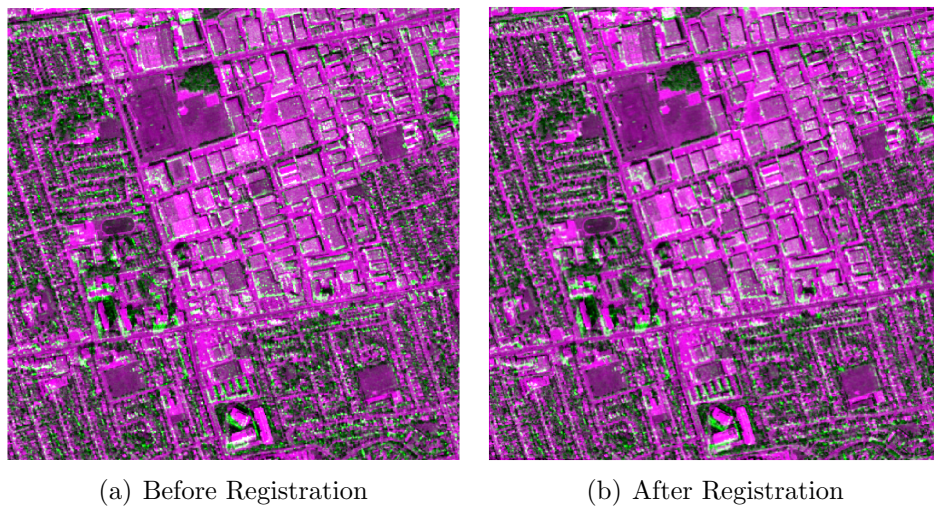
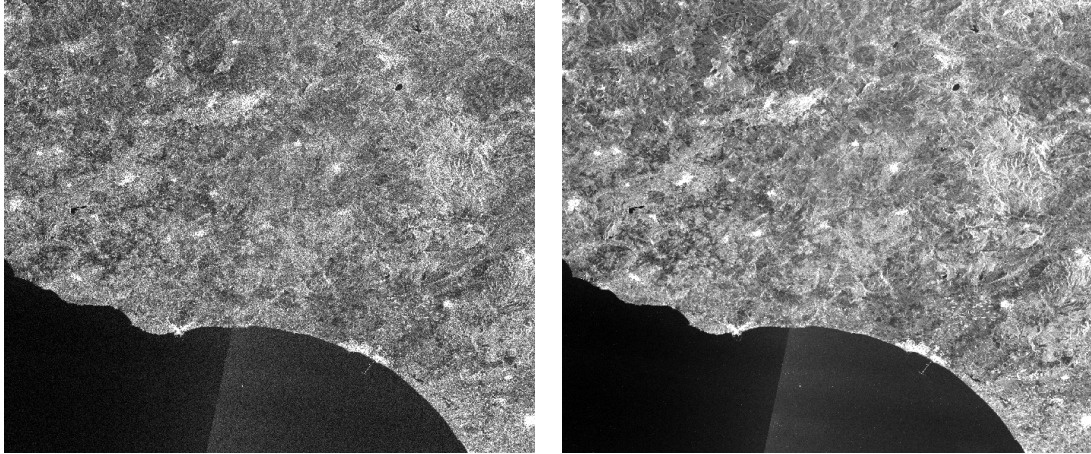


Figure 6.40: Overlay image before and after registration for the SAR-Optical image pair using ECC. Magenta denotes the reference. Green denotes the sensed and registered image in the left and right image respectively.

In the top-right and bottom left of Fig. 6.40(b) one can actually observe translational borders due a slight translational difference in the reference and sensed images.

### 6.3.2 Optical to SAR



(a) Sentinel-1 GRD Amplitude VH with speckle noise      (b) Sentinel-1 GRD Amplitude VH with reduced speckle noise

Figure 6.41: Sensed image for the first registration test of Optical-SAR. Size of the image is  $596 \times 698$  pixels. Covered Location is Sicily.



Figure 6.42: Reference image Landsat-8 L1TP B3 (green) for the first registration test of SAR-optical. Size of the image is  $596 \times 698$  pixels.

The shown sensed image of Fig. 6.41 was terrain corrected, reprojected and re-sampled to the same CRS and spatial resolution of the reference Landsat-8 image shown in Fig. 6.42. Terrain correction is required to remove any geometric distortions in the Sentinel-1. Without terrain correction, the SAR image is inverted. Reprojection is required to bring the SAR image to the same coordinate system as the reference Landsat-8 image. Terrain correction, reprojection and resampling



were done using the SNAP ESA toolbox [85]. The reference image was acquired on 2020/08/01 whilst the sensed image was acquired on 2020/10/18.

In addition, the sensed SAR image needs to be filtered to reduce the speckle noise found in SAR images. The speckle noise was reduced using the Lee sigma filter proposed by Lee et al. [91] for the speckle filtering of SAR imagery. A  $\sigma$  of 0.9 and a window size of  $9 \times 9$  were utilised with this filter. Fig. 6.41(a) shows the sensed image without speckle noise reduction and Fig. 6.41(b) shows the sensed image with speckle noise reduction.

Amongst all algorithms, the only algorithm that registered Fig. 6.41(b) and Fig. 6.42 was SIFT Flow. Fig. 6.43 shows the attained registration before and after. The biggest culprit of the failed registration was the speckle noise. Despite reducing the speckle noise, the POC algorithm failed to find a single peak. The SIFT algorithm picked up mostly features from the speckles and RIFT picked up incorrect matches which failed the registration. *The SAR-SIFT was also utilised to try and register this image pair using the default parameters and it also failed to match the images as not enough matches were found. Fig. 6.44 shows the established matches using SAR-SIFT.*

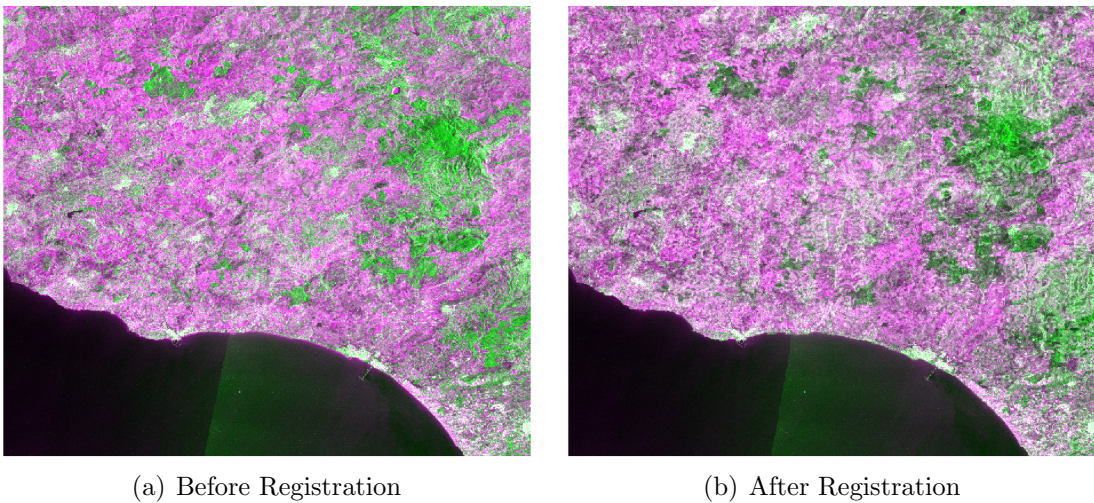


Figure 6.43: Overlay image before and after registration for the first Optical-SAR pair using SIFT Flow with the default parameters. In both images, magenta is the reference. Green is the sensed in Fig. 6.43(a) and the reference in Fig. 6.43(b) respectively.

As claimed in their work the RIFT algorithm did manage to find more matches than both SAR-SIFT and SIFT, but the obtained matches by RIFT were still not good enough to register the first Optical-SAR pair.

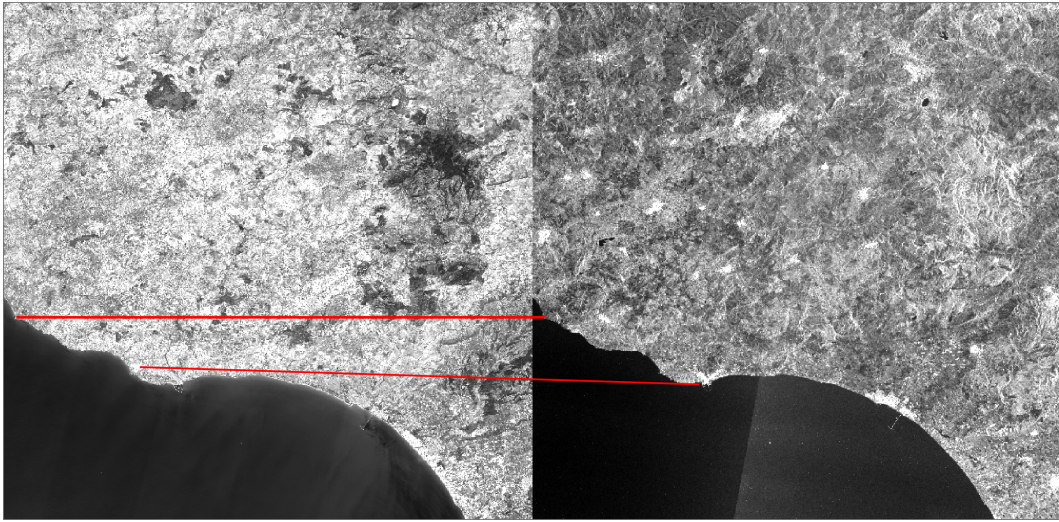
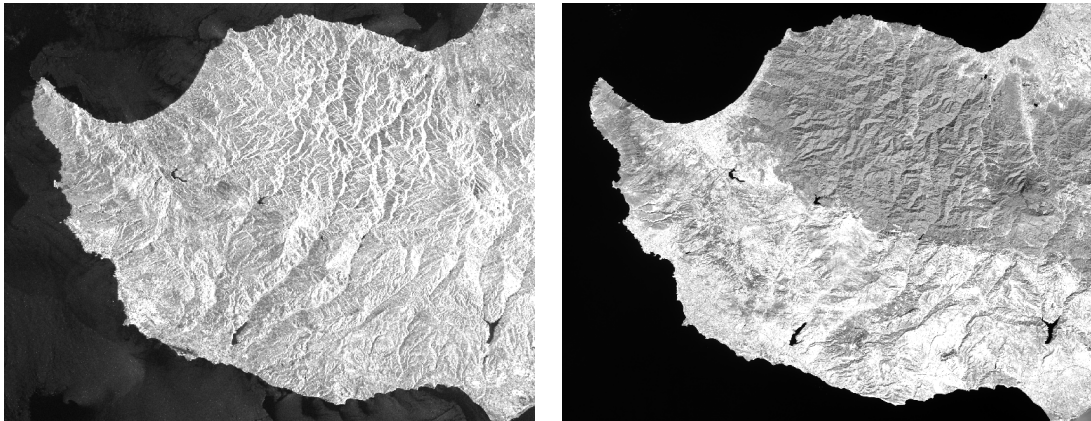


Figure 6.44: Obtained matches using SAR-SIFT for the first Optical-SAR test.



(a) Sentinel-1 GRD Amplitude VV with reduced speckle noise

(b) Landsat-8 L1TP B5 (NIR)

Figure 6.45: Sensed and reference images for the second registration test of Optical-SAR. Size of the image is  $596 \times 756$  pixels. Location: Cyprus.

The sensed image 6.45(a) was again preprocessed as done in the first Optical-SAR test. The chosen optical band for this test is the NIR shown in Fig. 6.45(b).

Image Pair	RMSE Before	POC	ECC	SIFT Flow	SIFT Default	SIFT MC	SIFT GS	SAR-SIFT	RIFT
L8 B5 - S1 Amplitude VV	8.2695	✗	8.2011	8.0615	8.1752	8.1689	8.1622	8.1590	✗

Table 6.8: Attained RMSE for the registration of the images found in Fig. 6.45.

Table 6.8 shows the attained RMSE for the registration of the images in Fig. 6.45(b). Fig. 6.46 shows the attained qualitative registration using SIFT Flow. In this test the phase correlation and RIFT completely failed in the registration. For the second consecutive time SIFT Flow attained the best registration. It is worth noting that from the two SIFT methods for this test SAR-SIFT achieved the lowest RMSE. Additionally, to the best of ones knowledge, based on literature in the remote sensing domain, Optical-Flow methods were specifically only utilised for the registration of Optical-SAR images. The known optical-flow algorithms of GeFolki and OS-Flow were created for the registration of Optical-SAR imagery. This might justify why SIFT-Flow was the best performing method for the conducted Optical-SAR tests.

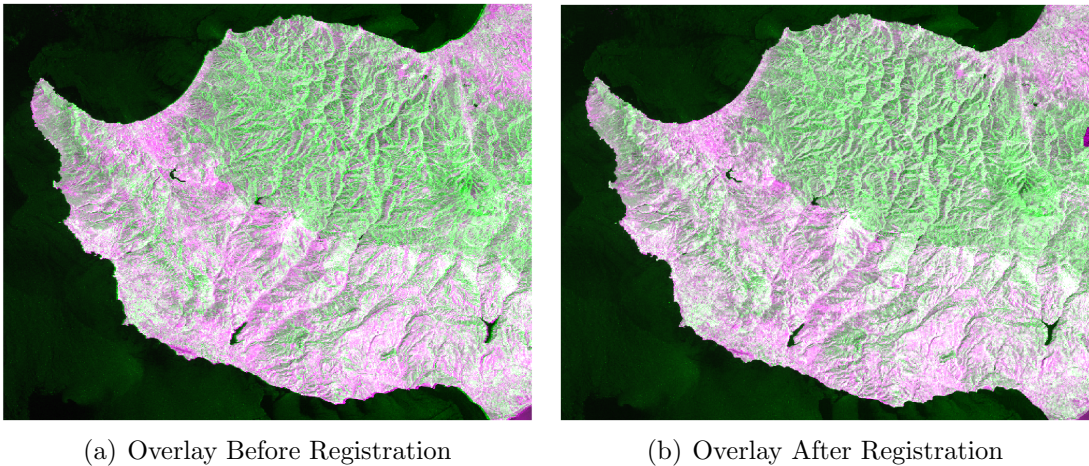


Figure 6.46: Registration before and after of the second Optical-SAR image pair using SIFT Flow. In both images magenta is the reference. Green is the sensed and registered in the left and right images respectively.

### 6.3.3 Optical to Thermal

In this section an optical band of Sentinel-2 is registered with the thermal band of Landsat-8. Landsat-8 image in Fig. 6.47 was resampled to the same spatial resolution as the Sentinel-2 image. Covered location is Samothrace. The reference Sentinel-2 B2 subset was acquired on 2020/07/29. The sensed Landsat-8 B10 was acquired on 2020/07/31. Resampling was done using bicubic interpolation.

Table 6.9 shows the attained RMSE for each method for the registration of the optical-thermal pair of Fig. 6.47. The RMSE after registration did not differ much from the before due to the fact that before registration the two images already had

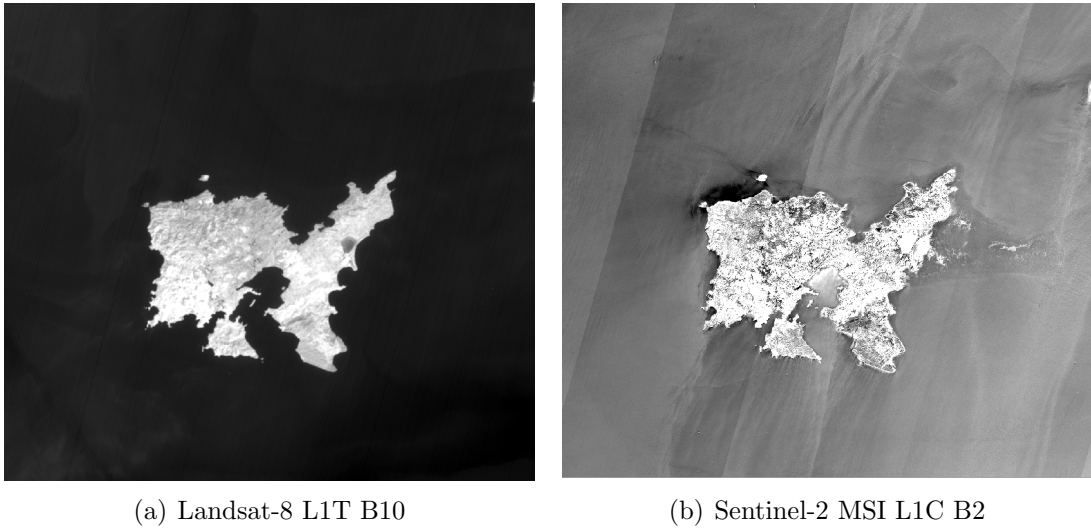


Figure 6.47: Reference and sensed images for the registration of Thermal-Optical images. Size of the images is  $692 \times 767$  pixels.

Image Pair	RMSE Before	POC	ECC	SIFT Flow	SIFT Default	SIFT MC	SIFT GS	RIFT
S2 B2 - L8 B10	15.4627	15.5625	15.5222	15.4911	15.4381	15.4649	15.4617	15.4640

Table 6.9: Table of results for the registration of the optical-thermal pair of Fig. 6.47.

good geocoding as shown in Fig. 6.48(a).

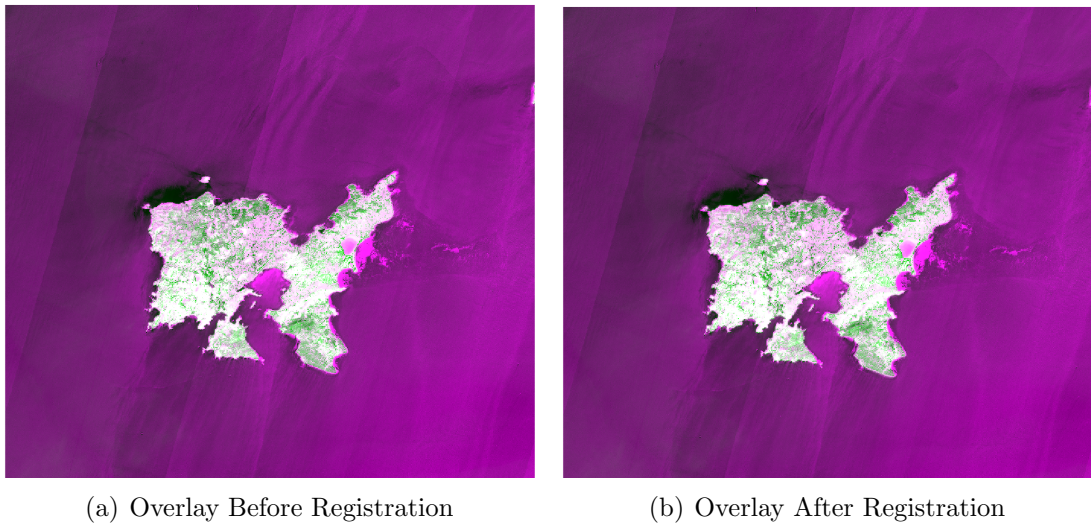


Figure 6.48: Registration before and after of the image pair found in Fig. 6.47 using SIFT with the Monte Carlo parameters.

Fig. 6.48 shows the overlay of the registration for the optical-thermal pair before and after. In both images the magenta represents the reference (S2) image. In Fig.

6.48(a), the green represents the sensed image whilst in Fig. 6.48(b) represents the registered image. Fig. 6.48(a) indicates that before registration the images already had good geocoding. Fig. 6.48(a) shows the registration after using SIFT with the Monte Carlo parameters.

## 6.4 Registration of Single Modality Remote Sensing Data

In this section satellite imagery of the same modality are registered using the different algorithms. The chosen single modalities are Optical-Optical and Thermal-Thermal.

### 6.4.1 Optical to Optical

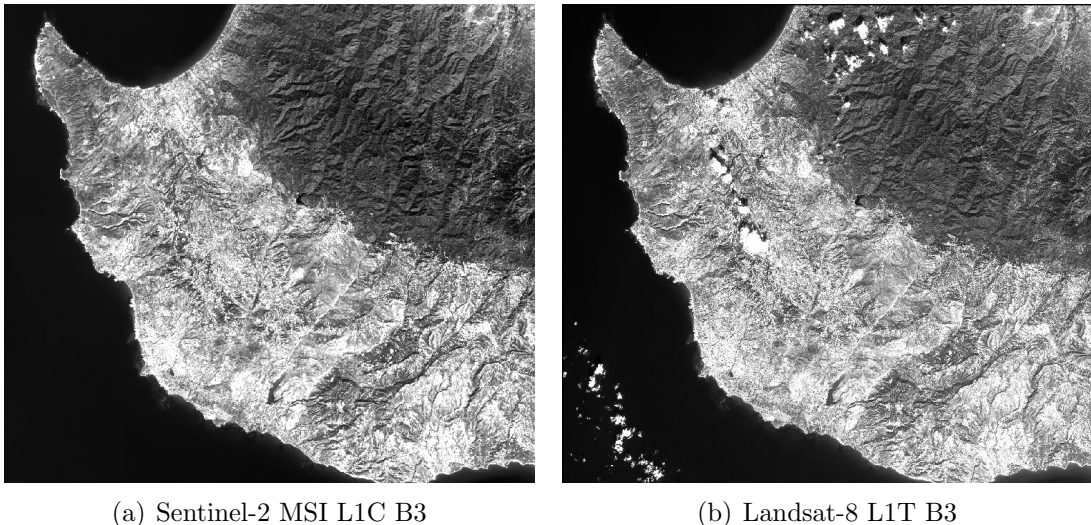


Figure 6.49: Reference and sensed images for the registration of Optical-Optical images. Size of the images is  $692 \times 767$  pixels. Location: Cyprus

The sensed Landsat-8 B3 image was acquired on 2020/10/09 whilst the reference Sentinel-2 B3 images was acquired 2020/10/10. Since the images have good geocoding prior to registration, the sensed image was translated by 2 pixels in both the  $x$  and  $y$  direction. Sensed image was resampled using bicubic interpolation to the same spatial resolution as the reference L8 image i.e. at 30 m.

Image Pair	RMSE Before	POC	ECC	SIFT Flow	SIFT Default	SIFT MC	SIFT GS	RIFT
L8 B3 - S2 B3	8.7293	8.7060	8.3524	8.4476	7.3132	7.3247	7.3196	8.3266

Table 6.10: Table of the RMSE results for the registration of the optical-thermal pair of Fig. 6.47.

In Table 6.10, the SIFT algorithm obtained the smallest RMSE after registration.

SIFT Default		SIFT MC		SIFT GS		RIFT	
$t_x$	$t_y$	$t_x$	$t_y$	$t_x$	$t_y$	$t_x$	$t_y$
2.0393	2.0001	2.0595	2.0326	2.0408	2.0131	2.9885	1.8458

Table 6.11: Table of the translation results for the registration of the optical-thermal pair of Fig. 6.47.

Table 6.11 shows the attained translation results in the  $x$  and  $y$  direction for the SIFT and RIFT algorithms. SIFT attained the closest  $t_x$  and  $t_y$  results compared to the induced shifts of 2 pixels in both directions. Attained  $t_x$  and  $t_y$  for SIFT is approximately 2 pixels for all parameter methods. Whilst in the  $x$  direction, RIFT has a misregistration of almost one pixel. The attained misregistration by RIFT for the  $x$  and  $y$  direction in meters for a spatial resolution of 30 m is 29.6 m and 4.6 m respectively. Highest attained misregistration by SIFT using the Monte Carlo was 1.244 m and 0.978 m.

Fig. 6.50 shows the qualitative registration before and after using SIFT with the GS parameters. From 6.50(b) it can be seen that the green image which represents the registered image was translated. This improved registration is evident in the coastal regions. The magenta in both images represents the reference (S2).

POC	ECC	SIFT Flow	SIFT Default	RIFT
5.5 s	29.6 s	60.9 s	6.7 s	53.3 s

Table 6.12: Table of the runtime results for the registration of the optical-thermal pair of Fig. 6.47.

Table 6.12 shows the runtime results for the registration of the optical-thermal pair of Fig. 6.47. This attained runtime result accounts for the registration of that image pair for a single time using the default parameters. It is to be noted that the default parameters of the ECC algorithm uses 2 multimode levels whilst SIFT Flow

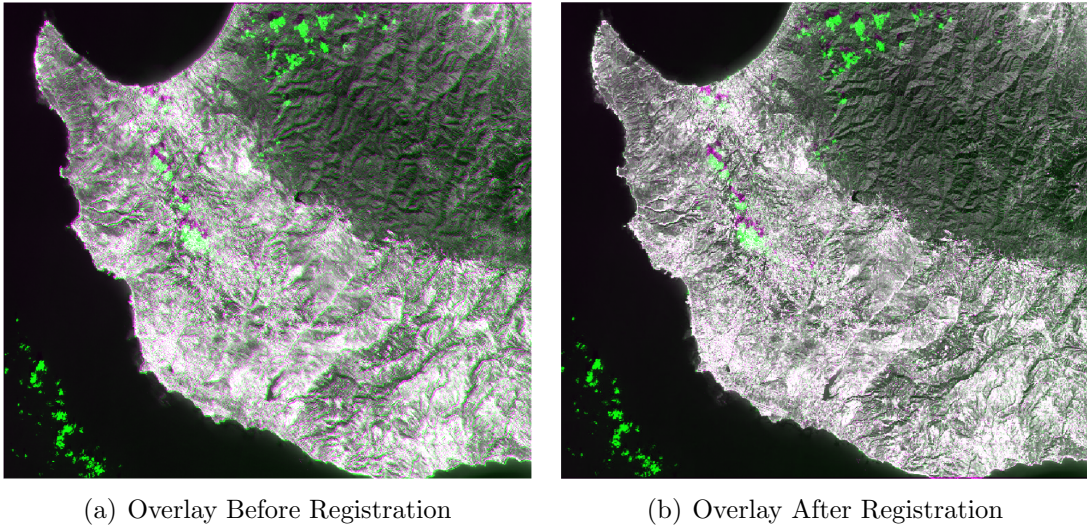


Figure 6.50: Registration before and after of the optical-optical image pair using SIFT with the GS parameters.

uses 4 multimode levels. At 4 multimode levels the runtime for ECC was 34.2s. The two least computationally intensive algorithms are SIFT and POC. This is an aspect that needs to be taken into account if one wants to register many image pairs. The RIFT algorithm takes approximately 7.72 times more computation time than SIFT to register the same pair of images.

### 6.4.2 Thermal to Thermal

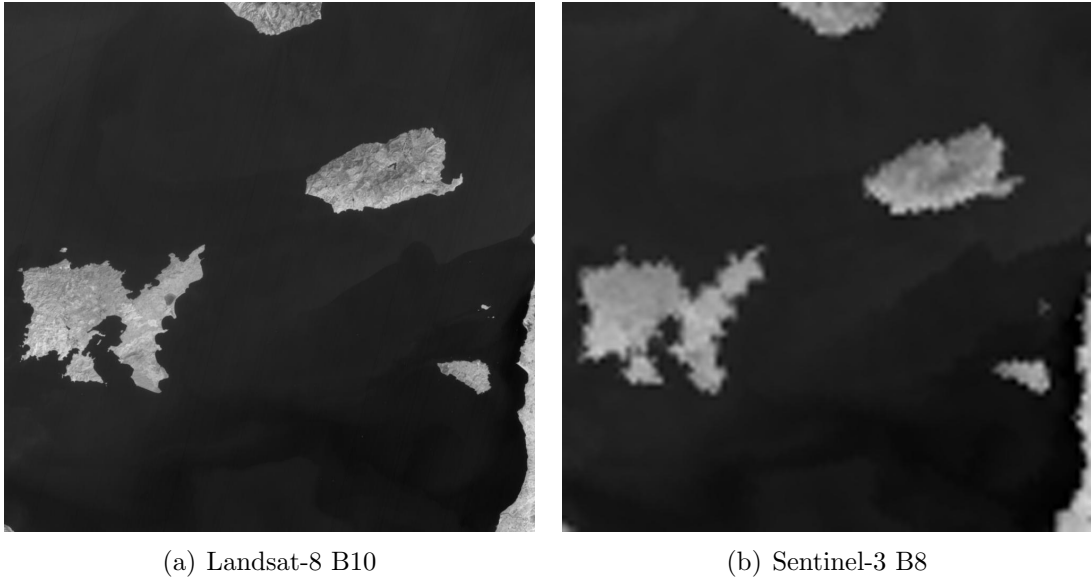


Figure 6.51: Reference and sensed images for the registration of Thermal-Thermal images. Size of the images is  $1000 \times 1000$  pixels. Location: Samothrace

The image pair chosen for the registration of the thermal-thermal image pair is the same location used in Fig. 6.18(d). The sensed image is upscaled by a scale factor of 10 using bicubic interpolation to the same size as the reference image i.e.  $1000 \times 1000$  pixels. Sensed image does not contain any shifts.

Image Pair	RMSE Before	POC	ECC	SIFT Flow	SIFT Default	SIFT MC	SIFT GS	RIFT
L8 B10 - S3 B8 Samothrace	5.8094	✗	5.7915	✗	5.7939	5.8012	5.7819	✗

Table 6.13: Table of the RMSE results for the registration of the thermal-thermal pair of Fig. 6.51.

As shown from Table 6.13, the only two algorithms that registered the thermal-thermal images are ECC and SIFT. The ECC algorithm succeeded because it is somewhat robust to geometric distortions and the two images have similar intensity. The RMSE before and after registration did not vary that much because the two images are already well geocoded before registration.

POC, SIFT Flow and RIFT failed because of the huge scale difference. The phase correlation method failed to find any single peak. The RIFT algorithm failed because the algorithm itself does not use a scale space to cater for a scale differences



between the two images. This limitation was even mentioned in the literature of RIFT by the authors in the respective limitations section [12].

POC	ECC	SIFT Flow	SIFT Default	RIFT
17.5 s	50.533 s	137.97 s	7.76 s	55.89 s

Table 6.14: Table of the runtime results for the registration of the thermal-thermal pair of Fig. 6.51.

Table 6.14, again shows that SIFT was the most computationally efficient for the registration of a  $1000 \times 1000$  image.

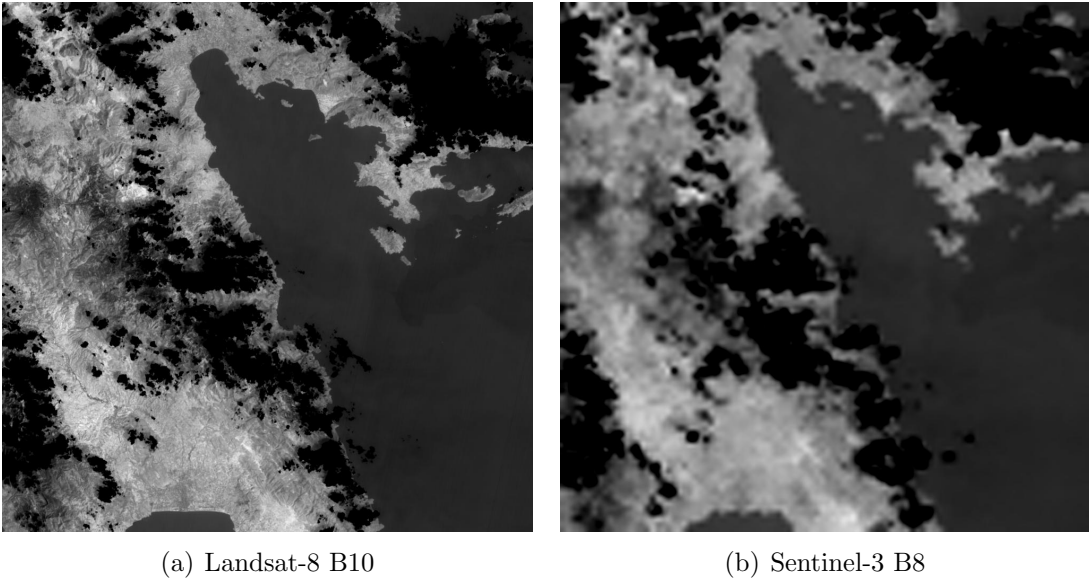


Figure 6.52: Reference and sensed images for the registration of second thermal-thermal image pair. Size of the images is  $1000 \times 1000$  pixels. Location: Athens

The image pair chosen for the registration of the second thermal-thermal image pair is the same location used in Fig.6.18(c). The sensed image is upscaled by a scale factor of 10 using bicubic interpolation to the same size as the reference image i.e.  $1000 \times 1000$  pixels. Sensed image does not contain any shifts.

Image Pair	RMSE Before	POC	ECC	SIFT Flow	SIFT Default	SIFT MC	SIFT GS	RIFT
L8 B10 - S3	7.7859	✗	7.7900	✗	7.8007	7.8243	7.7967	✗

Table 6.15: Table of the RMSE results for the registration of the second thermal-thermal pair of Fig. 6.52.

Similarly to the first test, POC, SIFT Flow and RIFT failed because of the huge scale difference. The phase correlation method failed to find any single peak and the RIFT algorithm failed because it does not use a scale space.

# Chapter 7

## Conclusion

### 7.1 Findings

- The first objective was to investigate the gap in literature concerning the image registration of remote sensing images for multi-modal and multi-sensorial satellite imagery. It was found that the most common type of registration is between *Optical-Optical* and *SAR-Optical* registration. To the author's knowledge, till now no literature considered the use of thermal satellite imagery. In addition, few literature considered the aspect of registering images with a high scale difference.

In the work by Y. Zhou et al. [24] where they proposed a fusion algorithm to register and fuse hyperspectral and multispectral images they highlight this aspect. Despite the fact that image registration is a major application in remote sensing there is little work that aimed at the image registration of images with significant scale differences [24].

- In relation to the registration of thermal satellite imagery using the feature-based approach explained in Chapter 5 the findings are as follows:
  1. In relation to the registration of Landsat-8 with Landsat-8, results showed that for this single-modality the highest attained misregistration was 1.2063 and 1.5352 pixels in the  $x$  and  $y$  direction respectively. This occurred when the sensed Landsat-8 image was downsampled and upsampled by a scale factor of 10 using  $\mathbf{H}_{R,T,S}$ . At a scale factor of 9 the attained

misregistration was of 0.6410 and 0.808 pixels in the  $x$  and  $y$  direction respectively.

2. For the registration of Landsat-8 with Sentinel-3, when the Landsat-8 was downscaled to a spatial resolution of 1 km the highest attained misregistration was 0.1999 and 0.3 pixels in the  $x$  and  $y$  direction respectively. When the Sentinel-3 image was upsampled by a scale factor of 10 to a spatial resolution of 100 m the highest attained misregistration was 2.584 and 1.2376 pixels in the  $x$  and  $y$  direction respectively.
3. For both the Landsat-8 to Landsat-8 and Landsat-8 to Sentinel-3 the general pattern was that as one upscales, the misregistration and RMSE increases due to a higher scale difference. Upscaling with bicubic interpolation causes the addition of false information and more blur which are causes of this increased misregistration.
4. The effect of clouds was another observed aspect in the image registration of satellite imagery. From the obtained results it was found that as the cloud coverage increased, for a feature-based method the number of matches decreased and the RMSE increased. Misregistration was not affected that much until there are no salient features that can be matched due to the predominant cloud occluded images. Once the cloud cover covered most of the salient features in both images, registration was no longer possible. This was evident in Fig. 6.12(h) with 61.65% cloud cover.

Image Pair	RMSE Before	POC	ECC	SIFT Flow	SIFT Default	SIFT MC	SIFT GS	SAR-SIFT	RIFT
Night-Day	10.9705	✗	✗	✗	8.9586	8.9199	8.8567	N/A	9.0077
Depth-Optical	6.2562	✗	✗	5.6783	✗	✗	✗	N/A	6.1291
Infrared-Optical	12.7390	12.5521	10.9320	✗	✗	✗	✗	N/A	10.9018
Map-Optical	0.3660	✗	0.1802	0.1744	0.1874	0.1478	0.2025	N/A	0.2374
Optical-Optical	9.4798	8.7781	8.2049	8.9239	8.2134	8.7065	8.2111	N/A	8.9228
SAR-Optical	14.0206	✗	14.0502	14.1113	✗	✗	✗	✗	13.9702
Optical-SAR 1	12.9106	✗	✗	12.9009	✗	✗	✗	✗	✗
Optical-SAR 2	8.2695	✗	8.2011	8.0615	8.1752	8.1689	8.1622	8.1590	✗
Optical-Thermal	15.4627	15.5625	15.5222	15.4911	15.4381	15.4649	15.4617	N/A	15.4640
Optical-Optical	8.7293	8.7060	8.3524	8.4476	7.3132	7.3247	7.3196	N/A	8.3266
Thermal-Thermal 1	5.8094	✗	5.7915	✗	5.7939	5.8012	5.7819	N/A	✗
Thermal-Thermal 2	7.7859	✗	7.7900	✗	7.8007	7.8243	7.7967	N/A	✗

Table 7.1: Table of the RMSE results for the registration of multi-modal and single modality remote sensing imagery. A number in red, blue and green denotes the first, second and third smallest RMSE respectively. The SAR-SIFT algorithm was only utilised for the registration of SAR-optical images. First six rows are the attained results for the images of the RIFT dataset. Remaining rows are the attained results for the chosen satellite imagery.

- In relation to the comparison between area-based, feature-based and optical flow-based methods for the registration of multi-modal and multi-sensorial satellite imagery based on the attained smallest RMSE as shown in Table 7.1. The findings are as follows:
  1. For the registration of SAR-Optical the overall best performing was SIFT-Flow. Coincidentally, to the author's knowledge, till now based on literature in the remote sensing domain, Optical-Flow methods were specifically only utilised for the registration of Optical-SAR images. The known optical-flow algorithms of GeFolki and OS-Flow were created for the registration of Optical-SAR imagery.
  2. For the registration of single-modality, the overall best were SIFT followed by ECC.
  3. For the registration of multi-modal data, the overall best was SIFT (including all three parameter methods).
- General findings for all results include:
  1. For the SIFT based method with default, Monte Carlo and Grid Search parameters, the biggest conclusion was that different parameter sets can lead to the same/identical minima and identical registration.
  2. An salient aspect to be mentioned is the *Lack of Ground Truth* when utilising satellite and remote sensing data. Since there is a lack of ground truth one does not know the actual misregistration. It is for this reason that the results started off with a Landsat-8 to Landsat-8 comparison as one knows that for that same modality the misregistration and RMSE were zero. Additionally, this aspect and problem of having a lack of ground truth is also noted in the literature such as the fusion paper proposed by Y.Zhou et al. [24] and in RIFT [12]. In these two examples, they highlight this encountered challenge.

To cater for this challenge both in the proposed fusion algorithm and in RIFT, the authors used an *approximate ground truth*. For the approximate ground truth in RIFT, they selected five evenly distributed correspondences in the chosen reference image and then they utilised these correspondences to estimate an accurate affine transformation as

the approximation of the ground truth for the geometric transformation [12]. The authors first performed feature matching on the image pair and then incorrect matches/outliers were removed using the NBCS method [66]; then they calculated the residuals of these image correspondences under the estimated affine transform. The correspondences with residuals less than 3 pixels were deemed as correct matches [12].

## 7.2 Limitations and Future Work

In the registration of multi-modal and single modality satellite imagery found in Sections 6.3 and 6.4 respectively, more data products and locations could have been utilised. Similarly, in Section 6.1.3 more locations could have been chosen to analyse the effect of cloud cover and the cloud cover percentage could have been quantified in a more systematic manner.

SIFT was not designed or created to be utilised for the registration of remote-sensing satellite imagery. In the review paper proposed by Tondewad and Dale where they presented a review of different registration methodologies [49] with their involved challenges, in their findings they stated that in literature concerning image registration in the remote sensing domain, the most commonly used methodology are feature-based algorithms [49]. SIFT was found to be the most utilised in feature detection algorithms [49]. Nonetheless, their review also stated that utilising SIFT alone is not enough since SIFT tends to suffer from insufficient feature points and a high outlier ratio under severe appearance change.

A suggested improved approach is to combine traditional based methods such as SIFT with neural networks such as a CNN. Hence, to improve this proposed work, SIFT could be combined with a neural network scheme or even an area-based/optical flow-based method to create a hybrid scheme. The choice of combination also heavily depends on the type of satellite images to be registered.

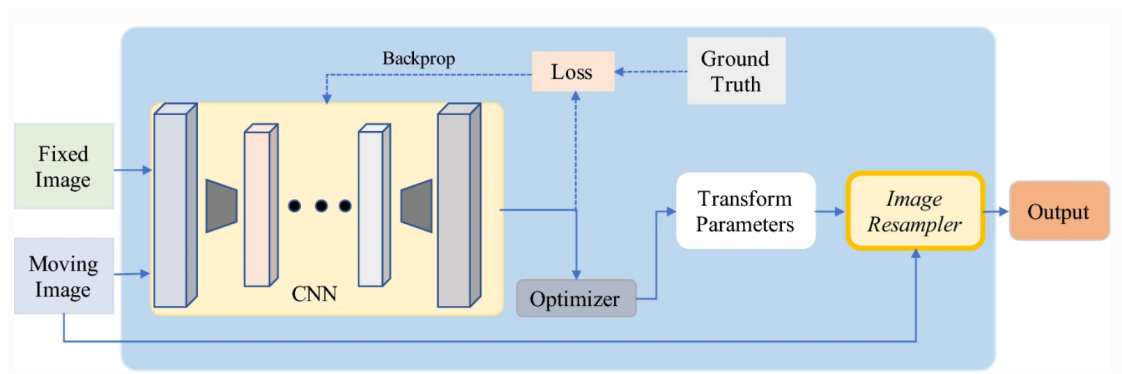


Figure 7.1: Visualisation of a deep learning intensity-based registration network [15].

Fig. 7.1 shows the workflow of a deep learning intensity-based registration network. The aim of Fig. 7.1 is to show how one would utilise deep learning in conjunc-

tion with image registration. The Fixed Image is the reference image whilst the Moving Image is the sensed image to be registered. The neural network utilised in this workflow is a CNN. An intensity-based registration requires both a metric which quantifies the similarity between the reference and sensed image and an optimisation based algorithm which updates the transformation parameters such that the similarity amongst the two images is maximised. For an area-based image algorithm examples of such metrics include the cross-correlation, normalised cross-correlation, mutual information, normalised mutual information and the sum of squared-differences. For a feature-based approach the metric would consist of the distance between the matched features such as Euclidean distance. The solid lines represent the work flow required during training and testing whilst the dashed lines represent the workflow required during training alone. For the training data, the Ground Truth images consist of reference images whilst the testing data consists of reference images and sensed images. The Loss function would refer to the attained RMSE after registration.



# References

- [1] *Radiometric Resolution of Sentinel-3 SLSTR*, accessed: 01/09/2020. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-slstr/resolutions/radiometric>
- [2] G. Brigot, E. Colin-Koeniguer, A. Plyer, and F. Janez, “Adaptation and evaluation of an optical flow method applied to coregistration of forest remote sensing images,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 7, pp. 2923–2939, 2016.
- [3] *Electromagnetic Spectrum*, accessed: 02/09/2020. [Online]. Available: <https://hubpages.com/education/Teachers-Guide-for-Radiation-beyond-Visible-Spectrum>
- [4] E. S. Agency, “Sentinel-1: ESA’s Radar Observatory Mission for GMES Operational Services,” Tech. Rep. SP-1322/1, 03 2012.
- [5] *Sentinel-3 SLSTR Data Access and Products*, accessed: 03/09/2020. [Online]. Available: [https://sentinel.esa.int/documents/247904/1848151/Sentinel-3\\_SLSTR\\_Data\\_Access\\_and\\_Products.pdf](https://sentinel.esa.int/documents/247904/1848151/Sentinel-3_SLSTR_Data_Access_and_Products.pdf)
- [6] *Explaining Ground Control Points*, accessed: June 24, 2020. [Online]. Available: [https://www.pcigeomatics.com/geomatica-help/COMMON/concepts/GCP\\_explainGCPs.html](https://www.pcigeomatics.com/geomatica-help/COMMON/concepts/GCP_explainGCPs.html)
- [7] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [8] *Histogram of Gradient*, accessed: 09/09/2020. [Online]. Available: <https://donghwa-kim.github.io/hog.html>
- [9] *Generating a Feature*, accessed: 09/09/2020. [Online]. Available: <https://aishack.in/tutorials/sift-scale-invariant-feature-transform-features/>

## References

---

- [10] *Project 2: Local Feature Matching*, accessed: 09/09/2020. [Online]. Available: [https://www.cc.gatech.edu/classes/AY2016/cs4476\\_fall/results/proj2/html/scalhoun9/index.html](https://www.cc.gatech.edu/classes/AY2016/cs4476_fall/results/proj2/html/scalhoun9/index.html)
- [11] D. I. R. A. Farrugia, *Lecture 3 - Feature Detection*, accessed: 09/09/2020.
- [12] J. Li, Q. Hu, and M. Ai, "Rift: Multi-modal image matching based on radiation-variation insensitive feature transform," *IEEE Transactions on Image Processing*, vol. 29, pp. 3296–3310, 2020.
- [13] Y. Xiang, F. Wang, L. Wan, N. Jiao, and H. You, "Os-flow: A robust algorithm for dense optical and sar image registration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6335–6354, 2019.
- [14] R. C. Gonzalez and R. E. Woods, "Digital image processing, 3rd ed." 2008.
- [15] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *Machine Vision and Applications*, vol. 31, no. 1, p. 8, 2020.
- [16] *Sentinel-2 Product Types*, accessed: 02/09/2020. [Online]. Available: <https://hubpages.com/education/Teachers-Guide-for-Radiation-beyond-Visible-Spectrum>
- [17] *Sentinel-2 Spatian Resolution*, accessed: 03/09/2020. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial>
- [18] *Landsat-8 level-1 Processing Details*, accessed: 03/09/2020. [Online]. Available: <https://www.usgs.gov/land-resources/nli/landsat/landsat-level-1-processing-details>
- [19] *Landsat-8 Bands*, accessed: 03/09/2020. [Online]. Available: [https://www.usgs.gov/land-resources/nli/landsat/landsat-8?qt-science\\_support\\_page\\_related\\_con=0#qt-science\\_support\\_page\\_related\\_con](https://www.usgs.gov/land-resources/nli/landsat/landsat-8?qt-science_support_page_related_con=0#qt-science_support_page_related_con)
- [20] *Sentinel-3 Radiometric Resolution*, accessed: 03/09/2020. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-slstr/resolutions/radiometric>

## References

---

- [21] S. Skakun, E. Vermote, J.-C. Roger, and C. Justice, “Multispectral misregistration of sentinel-2a images: Analysis and implications for potential applications,” *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2408–2412, 2017.
- [22] J. Storey, M. Choate, and D. Moe, “Landsat 8 thermal infrared sensor geometric characterization and calibration,” *Remote Sensing*, vol. 6, no. 11, pp. 11 153–11 181, 2014.
- [23] J. Storey, M. Choate, and K. Lee, “Landsat 8 operational land imager on-orbit geometric calibration and performance,” *Remote sensing*, vol. 6, no. 11, pp. 11 127–11 152, 2014.
- [24] Y. Zhou, A. Rangarajan, and P. D. Gader, “An integrated approach to registration and fusion of hyperspectral and multispectral images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3020–3033, 2020.
- [25] E. S. Agency, “Sentinel-2: ESA’s Optical High Resolution Mission for GMES Operational Services,” Tech. Rep. SP-1322/2, 03 2012.
- [26] M. Folkman, J. Pearlman, L. Lushalan, and P. Jarecke, “EO-1/Hyperion hyperspectral imager design, development, characterization, and calibration,” *Proceedings of SPIE*, pp. 40–51, 2001. [Online]. Available: [http://link.aip.org/link/?PSI/4151/40/1{&}Agg=doi{%}5Cnhttp://eo1.gsfc.nasa.gov/new/validationReport/Technology/TRW{\\_{}}EO1Papers{\\_{}}Presentations/07.pdf](http://link.aip.org/link/?PSI/4151/40/1{&}Agg=doi{%}5Cnhttp://eo1.gsfc.nasa.gov/new/validationReport/Technology/TRW{_{}}EO1Papers{_{}}Presentations/07.pdf)
- [27] T. Cudahy, R. Hewson, J. Huntington, M. Quigley, and P. Barry, “The performance of the satellite-borne Hyperion hyperspectral VNIR-SWIR imaging system for mineral mapping at Mount Fitton, South Australia,” *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, vol. 1, no. C, pp. 314–316, 2002.
- [28] E. S. Agency, “Sentinel-3: ESA’s Global Land and Ocean Mission for GMES Operational Services,” Tech. Rep. SP-1322/3, 10 2012.
- [29] S. N. Goward, J. G. Masek, D. L. Williams, J. R. Irons, and R. Thompson, “The landsat 7 mission: Terrestrial research and applications for the 21st century,” *Remote Sensing of Environment*, vol. 78, no. 1-2, pp. 3–12, 2001.

## References

---

- [30] D. P. Roy, M. A. Wulder, T. R. Loveland, C. Woodcock, R. G. Allen, M. C. Anderson, D. Helder, J. R. Irons, D. M. Johnson, R. Kennedy *et al.*, “Landsat-8: Science and product vision for terrestrial global change research,” *Remote Sensing of Environment*, vol. 145, pp. 154–172, 2014.
- [31] C. Fratter, M. Moulin, H. Ruiz, P. Charvet, and D. Zobler, “The spot 5 mission,” in *52nd International Astronautical Congress*, 2001.
- [32] E. Stoll, H. Konstanski, C. Anderson, K. Douglass, and M. Oxford, “The rapid-eye constellation and its data products,” in *2012 IEEE Aerospace Conference*, March 2012, pp. 1–9.
- [33] F. De Lussy, P. Kubik, D. Greslou, V. Pascal, P. Gigord, and J. P. Cantou, “Pleiades-hr image system products and quality, pleiades-hr image system products and geometric accuracy,” in *Proceedings International Society for Photogrammetry and Remote Sensing International Conference*, 2005, pp. 17–20.
- [34] R. Werninghaus, “TerraSAR-X mission,” in *Proc. SPIE 5236, SAR Image Analysis, Modeling and Techniques VI*, 2004.
- [35] L. C. Morena, K. V. James, J. Beck, “An introduction to the RADARSAT-2 mission,” *Canadian Journal of Remote Sensing*, vol. 30, no. 3, pp. 221–234, 2004.
- [36] *Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) 15- to 30- meter multispectral data from Landsat 8*, accessed: 03/09/2020. [Online]. Available: [https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-8-oli-operational-land-imager-and?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-8-oli-operational-land-imager-and?qt-science_center_objects=0#qt-science_center_objects)
- [37] J. Storey, D. P. Roy, J. Masek, F. Gascon, J. Dwyer, and M. Choate, “A note on the temporary misregistration of landsat-8 operational land imager (oli) and sentinel-2 multi spectral instrument (msi) imagery,” *Remote Sensing of Environment*, vol. 186, pp. 121–122, 2016.
- [38] *Sentinel-3: Datation and Geolocation*, accessed: 04/09/2020. [Online]. Available: <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-3-altimetry/level-0-algorithms-products/datation-and-geolocation>

- [39] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” in *European conference on computer vision*. Springer, 2006, pp. 404–417.
- [40] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, “Kaze features,” in *European Conference on Computer Vision*. Springer, 2012, pp. 214–227.
- [41] P. Alcantarilla, J. Nuevo, and A. Bartoli, “Fast explicit diffusion for accelerated features in nonlinear scale spaces british machine vision conference (bmvc),” 2013.
- [42] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: An efficient alternative to sift or surf,” in *2011 International conference on computer vision*. Ieee, 2011, pp. 2564–2571.
- [43] S. Leutenegger, M. Chli, and R. Y. Siegwart, “Brisk: Binary robust invariant scalable keypoints,” in *2011 International conference on computer vision*. Ieee, 2011, pp. 2548–2555.
- [44] S. A. K. Tareen and Z. Saleem, “A comparative analysis of sift, surf, kaze, akaze, orb, and brisk,” in *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2018, pp. 1–10.
- [45] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [46] G. D. Evangelidis and E. Z. Psarakis, “Parametric image alignment using enhanced correlation coefficient maximization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1858–1865, 2008.
- [47] B. D. Lucas, T. Kanade *et al.*, “An iterative image registration technique with an application to stereo vision,” 1981.
- [48] C. Liu, J. Yuen, and A. Torralba, “Sift flow: Dense correspondence across scenes and its applications,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 5, pp. 978–994, 2010.
- [49] M. P. S. Tondewad and M. M. P. Dale, “Remote sensing image registration methodology: Review and discussion,” *Procedia Computer Science*, vol. 171, pp. 2390–2399, 2020.

- [50] S. Skakun, J.-C. Roger, E. Vermote, C. Justice, and J. Masek, “Automatic co-registration of multi-temporal landsat-8/oli and sentinel-2a/msi images,” in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2017, pp. 5272–5274.
- [51] A. Stumpf, D. Michéa, and J.-P. Malet, “Improved co-registration of sentinel-2 and landsat-8 imagery for earth surface motion measurements,” *Remote Sensing*, vol. 10, no. 2, p. 160, 2018.
- [52] Z. Zhu, S. Wang, and C. E. Woodcock, “Improvement and expansion of the fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images,” *Remote Sensing of Environment*, vol. 159, pp. 269–277, 2015.
- [53] A.-M. Rosu, M. Pierrot-Deseilligny, A. Delorme, R. Binet, and Y. Klinger, “Measurement of ground displacement from optical satellite image correlation using the free open-source software micmac,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 100, pp. 48–59, 2015.
- [54] R. Behling, S. Roessner, K. Segl, B. Kleinschmit, and H. Kaufmann, “Robust automated image co-registration of optical multi-sensor time series data: Database generation for multi-temporal landslide detection,” *Remote Sensing*, vol. 6, no. 3, pp. 2572–2600, 2014.
- [55] D. Scheffler, A. Hollstein, H. Diedrich, K. Segl, and P. Hostert, “Arosics: An automated and robust open-source image co-registration software for multi-sensor satellite data,” *Remote Sensing*, vol. 9, no. 7, p. 676, 2017.
- [56] L. Yan, D. P. Roy, H. Zhang, J. Li, and H. Huang, “An automated approach for sub-pixel registration of landsat-8 operational land imager (oli) and sentinel-2 multi spectral instrument (msi) imagery,” *Remote Sensing*, vol. 8, no. 6, p. 520, 2016.
- [57] R. Ambati, C. Reddy, S. Patel, S. Tadepalli, and U. Kumar, “Automatic feature selection for landsat-8 and sentinel-2 image co-registration using surf-fann,” in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2018, pp. 1884–1889.
- [58] K. K. Rai, A. Rai, K. Dhar, J. Senthilnath, S. Omkar, and K. Ramesh, “Sift-fann: An efficient framework for spatio-spectral fusion of satellite images,”

- Journal of the Indian Society of Remote Sensing*, vol. 45, no. 1, pp. 55–65, 2017.
- [59] F. Eugenio and J. Marcello, “Featured-based algorithm for the automated registration of multisensorial/multitemporal oceanographic satellite imagery,” *Algorithms*, vol. 2, no. 3, pp. 1087–1104, 2009.
- [60] M. Wahed, G. S. El-tawel, and A. G. El-karim, “Automatic image registration technique of remote sensing images,” *Int. J. Adv. Comput. Sci. Appl*, vol. 4, no. 2, 2013.
- [61] X. Liu, Z. Tian, C. Chai, and H. Fu, “Multiscale registration of remote sensing image using robust sift features in steerable-domain,” *The Egyptian Journal of Remote Sensing and Space Science*, vol. 14, no. 2, pp. 63–72, 2011.
- [62] M. Ding, L. Li, C. Zhou, and C. Cai, “A multi-sensor image registration method based on harris corner matching,” in *International Conference on Virtual Systems and Multimedia*. Springer, 2006, pp. 174–183.
- [63] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, “Sar-sift: A sift-like algorithm for sar images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 453–466, 2015.
- [64] Y. Xiang, F. Wang, and H. You, “Os-sift: A robust sift-like algorithm for high-resolution optical-to-sar image registration in suburban areas,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3078–3090, 2018.
- [65] Y. Wu, W. Ma, M. Gong, L. Su, and L. Jiao, “A novel point-matching algorithm based on fast sample consensus for image registration,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 1, pp. 43–47, 2014.
- [66] J. Li, Q. Hu, and M. Ai, “Robust feature matching for geospatial images via an affine-invariant coordinate system,” *The Photogrammetric Record*, vol. 32, no. 159, pp. 317–331, 2017.
- [67] S. Baker and I. Matthews, “Lucas-kanade 20 years on: A unifying framework,” *International journal of computer vision*, vol. 56, no. 3, pp. 221–255, 2004.

- [68] J. Chen, L. Luo, C. Liu, J.-G. Yu, and J. Ma, “Nonrigid registration of remote sensing images via sparse and dense feature matching,” *JOSA A*, vol. 33, no. 7, pp. 1313–1322, 2016.
- [69] J. Rabin, J. Delon, Y. Gousseau, and L. Moisan, “Mac-ransac: a robust algorithm for the recognition of multiple objects,” 2010.
- [70] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, “Sar-sift: a sift-like algorithm for sar images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 453–466, 2014.
- [71] T. Chai and R. R. Draxler, “Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature,” *Geoscientific model development*, vol. 7, no. 3, pp. 1247–1250, 2014.
- [72] H. Foroosh, J. B. Zerubia, and M. Berthod, “Extension of phase correlation to subpixel registration,” *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–200, 2002.
- [73] *Sentinel-2 L1C Data Quality Report Issue 55 (September 2020)*, accessed: 13/09/2020. [Online]. Available: [https://sentinel.esa.int/documents/247904/685211/Sentinel-2\\_L1C\\_Data\\_Quality\\_Report](https://sentinel.esa.int/documents/247904/685211/Sentinel-2_L1C_Data_Quality_Report)
- [74] H. S. Stone, B. Tao, and M. McGuire, “Analysis of image registration noise due to rotationally dependent aliasing,” *Journal of Visual Communication and Image Representation*, vol. 14, no. 2, pp. 114–135, 2003.
- [75] *satpy*, accessed: 19/09/2020. [Online]. Available: <https://pypi.org/project/satpy/>
- [76] *rasterio*, accessed: 19/09/2020. [Online]. Available: <https://pypi.org/project/rasterio/>
- [77] *OpenCV SIFT Class Reference*, accessed: 19/09/2020. [Online]. Available: [https://docs.opencv.org/3.4.9/d5/d3c/classcv\\_1\\_1xfeatures2d\\_1\\_1SIFT.html](https://docs.opencv.org/3.4.9/d5/d3c/classcv_1_1xfeatures2d_1_1SIFT.html)
- [78] H. Singh, *Practical Machine Learning and Image Processing*. Springer, 2019.
- [79] A. Abela, *GitHub Repository for the Image Registration using the SIFT Algorithm*, accessed: 30/10/2020. [Online]. Available: <https://github.com/AARS07/SIFT-Image-Registration>



## References

---

- [80] G. Evangelidis, “Iat: A matlab toolbox for image alignment,” <https://sites.google.com/site/imagealignment/>, 2013.
- [81] A. Vedaldi and B. Fulkerson, “Vlfeat: An open and portable library of computer vision algorithms,” in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1469–1472.
- [82] *RIFT GitHub Repository*, accessed: 22/09/2020. [Online]. Available: <https://github.com/lh9171338/RIFT>
- [83] *GitHub Repository for the SAR-SIFT Algorithm*, accessed: 24/10/2020. [Online]. Available: <https://github.com/ZeLianWen/Image-Registration>
- [84] *KOMPSAT-2 (Korea Multi-Purpose Satellite-2) / Arirang-2*, accessed: 22/09/2020. [Online]. Available: <https://directory.eoportal.org/web/eoportal/satellite-missions/k/kompsat-2>
- [85] *SNAP*, accessed: 22/09/2020. [Online]. Available: <https://step.esa.int/main/toolboxes/snap/>
- [86] *Landsat-8 Data Users Handbook*, accessed: 26/09/2020. [Online]. Available: <https://www.usgs.gov/media/files/landsat-8-data-users-handbook>
- [87] *Least Squares Derivation*, accessed: 14/03/2021. [Online]. Available: <http://ltcconline.net/greenl/courses/203/matrixonvectors/leastSquares.htm>
- [88] *Non-Linear Least Squares, Scipy Curvefit Function*, accessed: 30/09/2020. [Online]. Available: [https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve\\_fit.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html)
- [89] Y. Ri and H. Fujimoto, “Drift-free motion estimation from video images using phase correlation and linear optimization,” in *2018 IEEE 15th International Workshop on Advanced Motion Control (AMC)*, 2018, pp. 295–300.
- [90] *POC GitHub Repository implemented by Y. Ri et al.*, accessed: 19/10/2020. [Online]. Available: <https://github.com/YoshiRi/ImRegPOC>
- [91] J.-S. Lee, J.-H. Wen, T. L. Ainsworth, K.-S. Chen, and A. J. Chen, “Improved sigma filter for speckle filtering of sar imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 202–213, 2008.

# Appendix A

## Numeric Example

This appendix contains a numeric example for the derivation of the Affine Transformation with rotation, scale and translation as explained in Section 5.3.5.1.

Aim is to derive the Affine transformation suitable to register a sensed image to a reference image. Assume we have two vectors that contain the pixel locations of the matched keypoints resulting after using SIFT and RANSAC.

- **Sensed Points**  $\mathbf{s} = [[1373, 1430, 1382, 1366, 1383, 1409], [314, 316, 337, 380, 376, 379]]$
- **Reference Points:**  $\mathbf{r} = [[30, 98, 64, 48, 64, 80], [30, 30, 64, 98, 98, 98]]$   
where both  $\mathbf{s}$  and  $\mathbf{r}$  are of size (2, 6)

Aim is to solve (5.3)

- Size of  $\mathbf{A}$ : (12, 6)
- Size of  $\mathbf{b}$ : (12, 1)
- Size of  $\mathbf{x}$ : (6, 1)
- $\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$
- Reshape  $\mathbf{x}$  of size (6, 1) to matrix  $\mathbf{X}$  of size (3,3):

$$\mathbf{X} = \begin{bmatrix} 0.961947439 & -0.0610228279 & 0 \\ 0.200522694 & 1.03187293 & 0 \\ -1343.83770 & -206.980576 & 1 \end{bmatrix}$$

•

$$\mathbf{H} = \mathbf{X}^T = \begin{bmatrix} 0.961947439 & 0.200522694 & -1343.83770 \\ -0.0610228279 & 1.03187293 & -206.980576 \\ 0 & 0 & 1 \end{bmatrix}$$