Low-Resolution Aerial Hyperspectral Image Processing for Agriculture-related Decision Making

A thesis submitted in fulfillment
of the requirements of the degree of
Doctor of Philosophy

March 2024

Chamika Janith Perera

Department of Electrical and Electronic Engineering
Graduate School of Engineering
Mie University

Thesis Declaration

I, **Kaluarachchige Don Chamika Janith Perera**, declare that this thesis titled **"Low-Resolution Aerial Hyperspectral Image Processing for Agriculture-related Decision Making"** and the work presented in it is my own. I confirm that:

1. This work was done wholly or mainly while in candidature for a Doctor of Philosophy at Mie University.
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at Mie University or any other institution, this has been clearly stated.
3. No part of this thesis has been accepted for any other degree or qualification at Mie University or any other institution.
4. I have acknowledged all sources of information and ideas in the thesis, whether quoted verbatim or paraphrased.
5. I have obtained permission(s) for any third-party material included in the thesis, or I have clearly identified and referenced such material.
6. The thesis is not currently being considered for any other degree or qualification.

I understand that any breach of these declarations may result in disciplinary action by Mie University and/or its partner institutions.

Signed: _____

Kaluarachchige Don Chamika Janith Perera

Date: _____

Dedication

To my parents and wife,

whose unwavering support and love

have been my guiding light

throughout this journey.

Acknowledgement

I would like to express my sincere gratitude to the following individuals and organizations who have contributed to the completion of this thesis:

- My main supervisor, Professor Hiroharu Kawanaka, for providing me with the opportunity to conduct this research. I am grateful for his guidance, support, and valuable insights throughout the research process. Additionally, I appreciate the assistance I received when trying to secure financial support for living in Japan.

- My Co-supervisor, Professor Chinthaka Premachandra, for facilitating the research by securing the main grant. I am thankful for your guidance, support, and direction, which enabled the completion of the research within the allocated time frame. I am also forever grateful for the guidance you provided during the period of travel restrictions due to the coronavirus, preventing me from initiating my research.

- Dr. Takashi Mishima and the staff and Mie University farm for the support provided during the data collection stages, enabling access to the Mie University farm and facilitating the process.

- My family, for their financial support, encouragement, and understanding during the ups and downs of this academic journey.

- The staff of the Graduate School of Engineering and the international relations office for their support in research activities as well as adapting to life in Japan.

- My friends and colleagues, for their encouragement, discussions, and shared experiences that enriched both my academic and personal life.

I am truly grateful for the collective support and inspiration that made this thesis a reality.

Publication List

Journal Publications

- Chamika Janith Perera, Chinthaka Premachandra, and Hiroharu Kawanaka, "Low Pixel Resolution Hyperspectral Image Mosaics Generation Using Learning-Based Feature Matching," in IEEE Access, vol. 11, pp. 104084-104093, 2023
- Chamika Janith Perera, Chinthaka Premachandra, and Hiroharu Kawanaka, "Enhancing Feature Detection and Matching in Low-Pixel-Resolution Hyperspectral Images Using 3D Convolution-Based Siamese Networks" Sensors 23, no. 18: 8004.

Conference Publications

- Chamika Janith Perera, Chinthaka Premachandra, and Hiroharu Kawanaka, "Comparison of light weight hyperspectral camera spectral signatures with field spectral signatures for agricultural applications," in 2023 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–3, 2023.
- Chamika Janith Perera, Chinthaka Premachandra, and Hiroharu Kawanaka, "Feature detection and matching for low-resolution hyperspectral images". 2023 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), PingTung, Taiwan, 2023, pp. 433-434, 2023.

Abstract

Remote sensing in the field of agriculture has become a timely concern to maximize agriculture production from available resources. Information produced from the remote sensing methods has shown immense potential in providing insights that can be directly used in the decision-making processes in agriculture. The use of hyperspectral imaging for agriculture-related decision-making has gained significant attention in recent years due to its ability to provide detailed information on crop health and composition. Even though hyperspectral imaging was started in the domain of satellite-based acquisition systems, the advancement of technology has managed to produce cameras that are capable of mounting in Unmanned Aerial Vehicles such as multi-rotors. These platforms provide localized data acquisition capabilities having high spatial and temporal resolutions. However, the high cost and complexity of hyperspectral cameras have limited their widespread adoption in the agricultural sector. Recent introductions to UAV mountable hyperspectral cameras have managed to reduce this high entry cost and opened up new avenues for many research communities. This thesis was motivated by the potential of the use of low-resolution hyperspectral imaging as a more cost-effective and practical solution for agriculture-related decision-making.

The thesis begins by discussing the use of low-resolution hyperspectral cameras for the initial analysis of spectral signatures. Spectral signatures obtained using the hyperspectral camera were compared with the spectral signatures obtained via a field spectrometer to evaluate the usability of the camera in the domain of agriculture decision-making. However, while interpreting the aerial images captured from the drone, hyperspectral mosaic generation was identified as a significant barrier. Mosaic generation is a key step in interpreting the captured images and deriving information from the captured data. This is because it allows the capability to interpret information relative to the entire captured field rather than relative to each captured image.

Image mosaic generation can be subdivided into several key steps as follows. Feature identification, feature matching, transformation calculation, and finally image blending. However, it was identified that this process failed at the step of feature identification and matching in the low-resolution hyperspectral images. This is because current popular feature detection and matching algorithms fail to either identify features at all or fail to identify features that could be matched across the captured images. Indistinctive features in these low-resolution aerial images such as repeating patterns, low-contrast areas, and uniform textures prevent the traditional feature detection and matching algorithms from identifying any usable matches. To overcome this challenge, a successful stitching pipeline is proposed next, utilizing two learning-based feature-matching methods.

The proposed method employs a stitching pipeline to create a low-resolution hyperspectral image mosaic. It employs two leaning-based feature detection and matching algorithms which have shown promising results in initial experiments with low-resolution hyperspectral images. Matched features were then used to transform the images into one plane and then blended producing the mosaic. Results indicate successful mosaic generations with consistent geometrical features compared to the captured field. The generated mosaic is then evaluated for spectral consistency within the same targets and spectral differentiability within different targets.

However, even with the state-of-the-art feature match detection methods, it was identified that certain landscapes especially at low altitudes struggle to produce any usable outputs. Hence an improved feature match detection method was needed to address the drawbacks. At the same time, it was highlighted that none of the feature detection methods in use utilizes all the spectral information available in a hyperspectral image. To utilize the missing information and increase the robustness of the detected and matched features, a novel feature match identification method based on 3D convolutional neural networks and edge detection was proposed in the second part of the thesis. A set of feature matches were generated using the hyperspectral image's edge maps. Physics-inspired edge detection algorithm was incorporated for this task and a set of SIFT feature matches were obtained for the matched images. The sensitivity of the SIFT detector was reduced so that an initial set of correct and incorrect feature matches were generated. Then the obtained set of feature matches was filtered by using a 3D convolutional Siamese Network trained on matched features and non-matched features. The second half of the thesis discusses the feature set generation, architecture of the Siamese Neural Network, training, and evaluation of the proposed method. The proposed model was evaluated against 4 state of the art feature detectors and the results indicated the superior performance of the proposed method.

Overall, the initial mosaic generation method was capable of successfully creating image mosaics from a given set of hyperspectral images. Spectral and geometrical consistency was observed between the non-stitched image and the stitched image. Secondly, the proposed feature detection and matching method was capable of producing highly accurate feature matches across the evaluated image pairs compared to the state-of-the-art methods. However, a few limitations such as the processing speed were identified during the inference stage of the proposed method.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Hyperspectral imaging is a subdomain of spectral sensing technology that has gained increasing attention in various fields, including agriculture. It involves capturing images of the same scene across a wide range of wavelengths, allowing the acquisition of detailed spectral information from each pixel in the scene. In the agriculture domain, this allows hyperspectral imaging to provide valuable insights into plant physiology parameters such as composition, health, growth, nutrition, and many others. In precision agriculture, hyperspectral imaging has emerged as a promising tool, where it can assist farmers in making informed decisions about crop management. By analyzing the spectral signatures of crops, hyperspectral imaging can provide critical information about plant health, different nutrient contents, stresses due to nutrient content shortages, pest infestations, and other factors that impact crop growth and productivity.

The increasing population, decreasing resources, and climate change have posed a severe threat to food security. There is a pressing need to optimize crop production from available resources and reduce waste. Hyperspectral imaging has the potential to transform the way farmers manage their crops, by enabling them to make data-driven decisions that can improve crop yields, optimize resource utilization, and reduce environmental impact. Furthermore, it can be used as a research tool to determine low-cost multispectral solutions that could reduce the cost while allowing the user the capability of critical decision-making. Recent advances in hyperspectral imaging technology, particularly in the area of low-pixel resolution snapshot cameras, have further enhanced its potential for agriculture. Low-pixel-resolution snapshot cameras can capture hyperspectral images in a single snapshot, allowing for faster data acquisition and reduced motion artifacts. Furthermore, the entry cost of the cameras has decreased allowing them to be a more viable solution in both the research domain and end-application domain. However, reducing the cost of the camera comes with a few disadvantages and this thesis discusses the importance of finding solutions to these problems.

1.2 Motivation

The motivation for this thesis arises from the potential benefits of using hyperspectral imaging in precision agriculture, particularly in the areas of crop monitoring and management. Hyperspectral imaging has recently become more accessible due to the decreasing cost of cameras. However, a significant challenge remains in the generation of high-quality image mosaics from low-resolution images. The creation of mosaics is crucial for interpreting data as a cohesive image of the entire target area instead of multiple individual images.

The image mosaic generation process involves several steps to create a seamless composite image from multiple individual images. The first step in the mosaic generation process is to identify features of each image, which are distinctive points or areas that can be reliably matched across images. These features can include corners, edges, or any other unique structures that can serve as points that can be used to align the images. Once the features are identified, the next step is to identify matching features in between the images. The accuracy of the stitching will depend on the accuracy of the matches. The next step is to calculate the transformation needed to align every image to a common coordinate system. This involves estimating the rotation, translation, and scaling parameters that can transform the images to align them correctly in the mosaic. Once the images are aligned, the last step is to blend the overlapping areas to create a smooth transition without visible seams. In hyperspectral imaging, this process should be repeated for every channel.

However, in low-pixel-resolution hyperspectral imaging, this whole process fails at the feature identification and matching step due to the traditional feature identification and matching algorithms not being able to either identify or match the minimum number of features needed for the successful calculation of the transformation matrices. Apart from the low resolution, narrow optics add to this issue by limiting the number of features captured in one image. One solution to mitigate this problem is to fly at high altitudes, which opens the possibility to capture more features from the scene. However, this results in a low spatial resolution which in turn negates one of the main advantages of UAV-based image hyperspectral image capturing.

In order to overcome this challenge, the motivation behind this thesis is to present an innovative approach to image stitching in low-pixel resolution hyperspectral imaging. This involves leveraging cutting-edge feature detection and matching methods in the initial stage. The proposed solution encompasses the exploration of potential strategies to identify feature matches between images and subsequently employ these matches in the image stitching process. The proposed approach promises to facilitate the creation of high-quality, seamlessly stitched image mosaics, ready for use in subsequent image analysis and segmentation.

The second stage of the thesis delves into the proposition of a novel solution, with the aim of attaining a robust set of feature matches by harnessing all available spectral information within the hyperspectral image. The primary objective of this feature-matching method is to improve the accuracy of detected feature matches, thus enhancing the quality and precision of image mosaic generation. This motivation underscores the significance of advancing current methodologies to address the unique challenges posed by low-pixel resolution hyperspectral imaging, ultimately contributing to the enhancement of image analysis and segmentation processes.

By overcoming the challenges of feature match detection and image stitching in low-resolution hyperspectral imaging, this research has the potential to contribute to the advancement of precision agriculture practices. The proposed solution can provide a practical and effective approach to generating accurate and reliable datasets for crop monitoring, such as assessing crop health, detecting diseases and pests, optimizing irrigation management, and improving nutrient management. The outcomes of this research can have implications for farmers, agronomists, and other stakeholders in the agriculture industry by enabling more precise and data-driven decision-making for sustainable and efficient agricultural practices.

1.3 Contributions of Thesis

This thesis discusses the use of low-pixel resolution hyperspectral imaging in the domain of precision agriculture. Initially, the use of the camera for data acquisition is discussed, and the data are compared with a field spectrometer for spectral similarity among the same targets and spectral differentiability among different targets. Then, a solution for the problem of difficulty in creating low-resolution mosaics is proposed. The proposed method uses the knowledge of the learning-based feature detection method to identify matches in between images and use the matches to generate transformation matrices. Finally, a further improved feature matching method is proposed for more robust feature detection and matching in low-pixel resolution hyperspectral images.

The specific contributions of this thesis are as follows.

- Initial setup of data acquisition setup followed by a data Comparison between spectral signatures obtained using the UAV-based hyperspectral camera and a field spectrometer. Spectral signatures were compared for spectral consistency with the same targets and spectral differentiability with different targets.

- Low-resolution hyperspectral mosaic generation pipeline: Learning-based feature matching methods were investigated and two learning-based feature detectors and matches were adapted to match features between the hyperspectral images. Matched

features were used to obtain the transformation and stitch the images. Problems faced in stitching low-resolution images and solutions taken are discussed in detail. The proposed method was then evaluated using several case studies. The limitations identified during these case studies were addressed, and the final proposed method was evaluated in another case study.

- Enhanced feature match detection using 3D Convolution Neural Network based Siamese network: A novel approach to obtain a robust set of feature matches was proposed using the Siamese network architecture based on 3D convolution. The network was trained to filter out incorrect matches from an initial set of matches generated with converted edge maps of low-pixel-resolution hyperspectral images. The proposed method was evaluated with four state-of-the-art methods and the results indicated robust performance.

Chapter 2

Related Works

The field of remote sensing in agriculture has undergone significant transformations and has emerged as a crucial tool to optimize production and resource utilization. In recent years, the advent of hyperspectral imaging has garnered substantial attention because of its potential to offer detailed insights into crop health and composition. This chapter delves into the existing body of knowledge, tracing the evolution of remote sensing technologies, and exploring the advancements that have shaped the current landscape of agricultural research related to hyperspectral UAV (Unmanned Aerial Vehicles) based decision making.

As remote sensing technology progressed, the transition from traditional satellite-based acquisition systems to the deployment of UAVs equipped with hyperspectral cameras has revolutionized data acquisition capabilities. These UAV platforms present unique opportunities for high-resolution and localized data collection, overcoming spatial and temporal limitations associated with satellite systems. However, challenges such as the high cost and complexity of hyperspectral cameras have historically hindered their widespread adoption within the agricultural sector. Motivated by the need for more cost-effective and practical solutions, this review of the literature critically examines the transition from traditional hyperspectral imaging to the exploration of low-resolution hyperspectral cameras in the context of agriculture-related decision-making.

Furthermore, the literature review will introduce existing literature on feature-matching techniques utilized in the image stitching domain. Later it will explore the following discourse on feature-matching techniques used in UAV-based image mosaic generation. This section will also explore the specific hyperspectral image mosaic generation research literature available in the academic publication domain.

Through a comprehensive exploration of existing literature and methodologies, this chapter lays the foundation for the subsequent discussion on the proposed solutions, their evaluations, and the potential avenues for future research. The aim is to provide a holistic understanding of the challenges and advancements in hyperspectral imaging for agriculture, setting the stage for the unique contributions of the present thesis.

2.1 Introduction to Remote Sensing in Agriculture

Agriculture, as a critical component of global food production, faces numerous challenges such as optimizing crop yields, resource management, and environmental sustainability. Remote sensing technologies offer a transformative solution by providing real-time and accurate information about crop health, soil conditions, and overall field performance.

Remote sensing, a powerful technology that has revolutionized various fields, plays a pivotal role in enhancing agricultural practices and management. By utilizing sensors mounted on satellites, aircraft, or drones, remote sensing enables the collection of valuable data about the Earth's surface without direct physical contact. In the context of agriculture, remote sensing provides a comprehensive and efficient means of monitoring, analyzing, and managing agricultural landscapes. [1, 2]

The most critical step in remote sensing is the data acquisition stage. There are a multitude of sensing methods used in different decision-making domains. These include and not limited to RGB imaging, infrared imaging, RADAR (RAdio Detecting And Ranging), LiDAR(Light Detection and Ranging), GPR (Ground Penetrating Radar), Multispectral imaging, and many more. Hyperspectral imaging is one of these numerous imaging techniques that is becoming popular in the hyperspectral domain.

2.2 Introduction to Hyperspectral Imaging

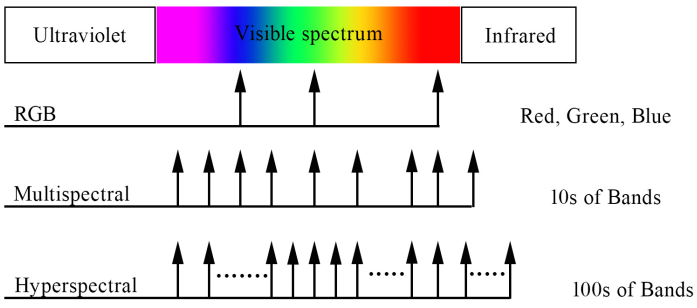


Figure 2.1: Different imaging technology and location where each technology captures information

Hyperspectral imaging is a cutting-edge technology that was first introduced by the National Aeronautics and Space Administration's (NASA) jet propulsion laboratory in the late 1970s [3]. it received widespread attention due to the capability of capturing two-dimensional images from multiple locations of the electromagnetic spectrum. A comparison between popular image acquisition technologies in the precision agriculture

domain is shown in figure 2.1. Black and white (grayscale) imaging can be identified as the most basic form of image capturing. It is a simple representation of the scene presented as variations of intensity at one particular location of the electromagnetic spectrum. RGB imaging, which stands for Red, Green, and Blue, captures information from three locations in the visible range of the electromagnetic spectrum. RGB imaging was considered the most popular sensing technology used in precision agriculture-related sensing tasks.

Multispectral imaging cameras, on the other hand, capture several broad wavelength bands in visible, near-infrared, and short-wave infrared spectrums. However, hyperspectral imaging systems can capture up to hundreds of distinctive wavelength bands in a single capture. Furthermore, the bandwidth of the sampled segment is narrow compared to multispectral imaging systems. The narrowness of the bands provides a continuous spectral measurement across the measured range of the electromagnetic spectrum. Hence making those measurements sensitive to subtle variations in reflected energy.

There are three main hyperspectral acquisition platforms to consider in hyperspectral-based agriculture remote sensing applications. Satellites, Airplane and UAVs. Each platform offers distinct advantages and disadvantages, influencing its suitability for various agricultural monitoring needs. Satellites, while providing global coverage and a consistent overhead view, may face limitations in spatial resolution and susceptibility to atmospheric interference. Airplanes, on the other hand, offer higher spatial resolution and flexible scheduling, but come with challenges related to limited accessibility and higher operational costs.

UAVs, with their unique capabilities, stand out as a promising platform for hyperspectral acquisition in agriculture. These unmanned systems can operate at low altitudes, delivering exceptionally high spatial resolution, which is crucial for detailed analysis of smaller agricultural plots. The rapid deployment capability of UAVs allows for timely data collection, making them well-suited for targeted interventions and responsive agricultural management. Moreover, UAVs offer a cost-effective alternative to airplanes, making frequent and widespread monitoring economically viable.

However, UAVs also come with their own set of challenges, such as limited flight time and susceptibility to adverse weather conditions. Despite these limitations, the potential benefits of UAV-based hyperspectral acquisition in agriculture are substantial. The combination of high spatial resolution, cost-efficiency, and rapid deployment position UAVs as an ideal tool for precision agriculture. Their ability to navigate challenging terrains and access remote areas further enhances their applicability to address the evolving needs of modern agriculture.

As technology continues to advance and regulatory frameworks adapt, UAV-based hyperspectral acquisition is poised to play a transformative role in agriculture. The de-

tailed and timely information provided by UAVs can empower farmers to make informed decisions, optimize resource allocation, and enhance overall crop management. With ongoing developments in sensor technology, data processing capabilities, and regulatory support, UAV-based hyperspectral acquisition holds the potential to revolutionize agricultural practices, fostering sustainability and efficiency in food production. Table 2.1 summarizes an overview of the strengths and weaknesses of each method of acquisition.

Table 2.1: Comparison of main three hyperspectral acquisition platforms.

	Satellite based	**Aircraft Based**	**UAV based**
Coverage area	Broad, large-scale	Moderate, localized	Moderate, Limited, localized
Spatial resolution	Lower	Moderate	Higher
Flexibility	Limited by orbits	Moderate, limited by weather	Higher, limited by weather
Real-time analysis	Limited	Possible	Possible
Response time	Moderate	Moderate	Quick
Weather dependency	Affected by clouds	Affected by weather	Affected by weather
Data acquisition control	Remote operation	Onboard control	Onboard control
Cost	High	Moderate	Moderate, Low

Hyperspectral Sensor Types

There are several types of hyperspectral sensors, each with its unique characteristics and applications. Point and line scan sensors are types of hyperspectral sensors that capture spectral and spatial information of a target line by line Wavelength scan sensors, on the other hand, acquire imagery in hundreds of narrow, contiguous spectral bands, which facilitates fine discrimination between different features on the target surface [4].

In the domain of agriculture remote sensing, push broom and snapshot types of hyperspectral sensors are commonly used [5]. The pushbroom sensor, also known as a line scan hyperspectral camera, collects full spectral and spatial information of the target in a single line. The user has to move the acquisition platform across the target to capture the entire scene [6].

The snapshot hyperspectral sensor is a relatively new hyperspectral image sensor. It offers certain advantages, such as acquiring hyperspectral images at video rate, ultra-portability, and easy handling with a small number of spectral bands. Snapshot hyperspectral imaging is a method for capturing hyperspectral images during a single integration time of a detector array. No scanning is involved with this method, in contrast to push broom and whisk broom scanning techniques. The lack of moving parts means that motion artifacts should be avoided. This instrument typically features detector arrays with a high number of pixels [6, 7].

Hyperspectral Image Cube

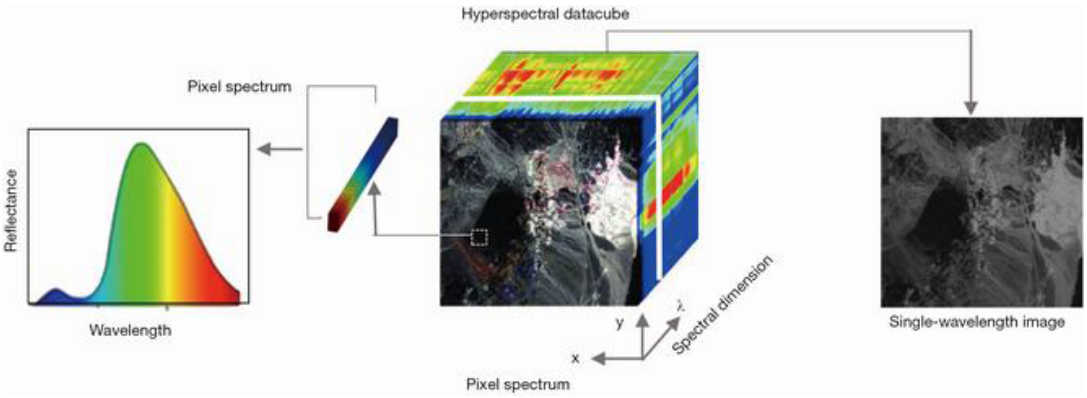


Figure 2.2: Introduction to hyperspectral image cube.

In Figure 2.2, a representation of a typical hyperspectral cube is depicted, illustrating the essential components that convey information about the captured scene. The dimensions along the x and y axes in the cube correspond to the spatial information of the scene, which delineates its layout and structure. These spatial dimensions allow for the characterization of different objects or features within the observed area. Meanwhile, the α direction in the hyperspectral cube encapsulates the spectral information, representing the various wavelengths across the electromagnetic spectrum.

A noteworthy aspect of hyperspectral data analysis is the concept of a spectral signature, derived from individual pixels in the hyperspectral cube. Each pixel, when plotted, generates a spectral signature that provides detailed information about the spectral reflectance or radiance associated with that particular location. This spectral signature serves as a unique fingerprint, allowing for the identification and differentiation of materials based on their distinctive spectral characteristics.

2.3 Challenges and Advancements in Hyperspectral Cameras

Hyperspectral where sampling was carried out in hundreds of bands in the visible range and infrared range, on the other hand, is still limited to mostly the research domain. Figure 2.3 indicates that the results for the keyword hyperspectral imaging in webofscience indicate that research related to hyperspectral imaging in the remote sensing domain is increasing. However, the cost of entry is still a huge barrier to starting research in this domain. Further, the form factor also plays a key drawback when the sensors are wanted to be used in UAV-based acquisition platforms.

Table 2.2 lists the majority of the hyperspectral cameras used in the research and application domain. Almost all of the listed cameras have a high entry cost except for the

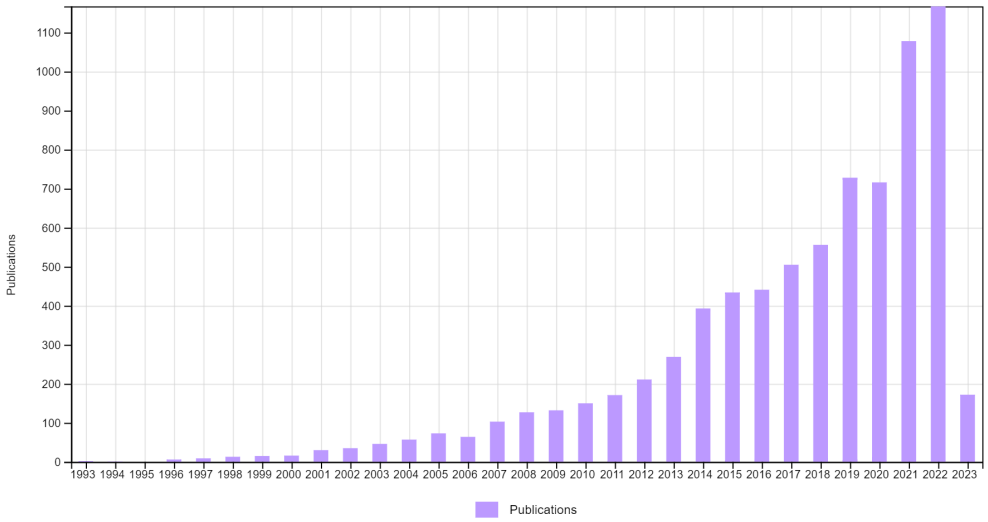


Figure 2.3: Number of publications related to hyperspectral imaging in remote sensing domain

Ximea snapshot camera. Recent introductions of low-resolution snapshot hyperspectral cameras, such as Cubert Ultris 5 [8] and Ximea snapshot cameras [9] can be identified as a solution to remove the entry barrier. Furthermore, these cameras have the added advantage of having a significantly smaller form factor. Which allows the possibility of mounting the cameras in UAV platforms. However, these advantages come at the cost of loosing pixel resolution.

2.4 Image Mosaic Generation

Mosaicing is the process of stitching multiple images into one image [10]. In the remote sensing domain, a mosaic of the captured images is a single image obtained by stitching the images captured using an aerial platform. Interpreting the data and decisions relative to this stitched image is much easier and more convenient than interpreting relative to each captured image separately. The typical work process of image mosaic generation includes identifying the features of each image, matching features, calculating the transformation needed to align every image to one coordinate system, and finally blending the images into one by deciding on a strategy to blend overlapping areas without visible seams. However, this process fails at the feature detection and matching stage for hyperspectral images taken from recently introduced low-cost snapshot hyperspectral cameras. This is due to their low resolution and narrower optics. As an example, the recently introduced Cubert Ultris 5 camera's resolution is limited to 290px $\times$ 270px. [8] and Field of View is limited to 15°. One method to mitigate this problem is by flying at

Table 2.2: Different UAV-based hyperspectral sensors available in the commercial market

Product	**Size**	**Weight**	**Sensing Method**	**Resolution (pixels)**	**Spectral Range (nm)**	**Spectral Bands**
SPECIM FX10	150 x 85 x 71 mm	1.3 kg	Push Boom	1024	400-1000	224
SPECIM AFX10	130 x 152 x 202 mm	2.5 kg	Push Boom	1024	400 - 1000	1024
Headwall Nano HP	132 x 102 x 73 mm	1.05 kg	Push Boom	1020	400 - 1,000	342
HySpex Mjolnir V-1240	250 x 175 x 170 mm	4 kg	Push Boom	1240	400 - 1000	200
BaySpec GoldenEye	60 x 33 x 43 mm	180 g	Snapshot Imaging	648 x 488	400 – 1100	140
BaySpec OCI™-1000	30 x 30 x 75 mm	120 g	Push Boom	2048	470 – 1000	100
Corning microHS 410 SHARK	136 x 88 x 70 mm	700g	Push Boom	704	400 – 1000	155
imec multishot VNIR + RGB	120 x 120 x 150 mm	2.22 Kg	Snapshot	2048 x 1088	460 – 860	33
Cubert firefleye 185	200 x 67 x 60 mm	490 g	Snapshot	1000 x 1000	450 - 950	125
Ximea Snapshot	26 x 26 x 31 mm	32 g	Snapshot	409 x 217	665 - 960	24

relatively high altitudes so that the camera captures more features from the scene that could be identified by traditional feature detection methods. However, doing so will lose the advantage of high spatial resolution in UAV-based hyperspectral image capturing.

Image registration is the process of transforming two or more images into a single plane using the common features of the images and it could be identified as a key step in the mosaic generation process. Feature detection-related research work can be divided into two main subsections focusing on image registration of general image stitching research and aerial image-specific stitching research publications. In general image stitching, image registration can be subdivided into two main areas. Intensity-based methods and Feature-based methods [11]. Under intensity-based methods algorithms such as frequency domain methods [12], spatial domain methods [13], and optical flow methods [14] can be identified. Under feature-based image registration algorithms, geometrical feature identification methods and leaning-based feature detection methods can be identified [11]. However, out of these different methods, most image-stitching applications focus on using area-based registration and feature-based registration [15]. This is true for aerial image-specific stitching research as well. 2022 review of existing UAV image mosaicing research [16] indicates that the most widely used methods used in the registration process are SIFT [13], SURF [17], Harries corner detection [18], FAST [19], BRISK [20] and ORB [21]. All these methods depend on feature-based descriptors that define the point it describes. However, these feature-based detectors perform poorly when there are factors such as poor texture, viewpoint change, illumination variation, repetitive patterns, and motion blur [22]. Poor textures are a common occurrence in low-resolution hyperspectral images. Further, repetitive patterns are a common occurrence in aerial images. Hence the need arises to use alternative feature-detecting methods to overcome the aforementioned problem.

2.5 Feature Identification and Matching in Images

Feature detection and matching in images is a fundamental process in computer vision and image processing. It involves identifying distinctive points or regions in an image, often referred to as key points or interest points, and establishing correspondences between these points in different images. This process begins with feature detection, where every pixel is examined to identify if there is a feature present. Once the key points are identified, a descriptor is calculated for each of them. The final step is feature matching, which aims to establish correspondences between key points in different images. This process is essential for various applications, including image registration, object recognition, and 3D scene reconstruction.

SIFT [13], or Scale-Invariant Feature Transform, is a popular computer vision algo-

rithm that detects and describes key features in images or videos, which are robust to changes in scale, rotation, and illumination. Developed by David Lowe in 1999, SIFT uses a scale-space pyramid and difference of Gaussian (DoG) images to identify stable key points and computes descriptors that capture the local appearance and orientation of each key point. Despite its computational complexity and sensitivity to image noise, SIFT has been widely used in various computer vision applications due to its robustness and versatility. SURF [17], or Speeded-Up Robust Features is also a popular computer vision algorithm that detects and describes key features in images or videos with a focus on efficiency and robustness. It was introduced by Herbert Bay et al. in 2006 as an improvement over SIFT, aiming to address some of its computational limitations. SURF uses a combination of box filters and integral images to accelerate feature detection, and it employs Haar wavelet responses to compute descriptors that are resistant to changes in scale, rotation, and illumination. SURF has gained popularity in real-time and resource-constrained applications due to its speed and robustness, also making it a valuable tool in various computer vision tasks.

FAST [19], or Features from Accelerated Segment Test, is known for its speed and simplicity in detecting key points in images or videos. FAST was proposed by Edward Rosten et al. in 2006 as a fast alternative to existing feature detection methods. It works by comparing the intensity values of pixels in a circle around a candidate key point with a threshold, and it can quickly identify corners or key points based on pixel intensity changes. FAST is particularly efficient for real-time applications due to its minimal computational overhead, making it widely used in scenarios where real-time processing is critical, such as robotics and augmented reality. While FAST may be less robust to scale and rotation changes compared to other methods, it is often used as a fast initial feature detection step in more complex computer vision pipelines. BRISK [20] or Binary Robust Invariant Scalable Keypoints combines speed, efficiency, and robustness in detecting key points in images or videos. BRISK was introduced by Stefan Leutenegger et al. in 2011 as a fast alternative to existing feature detection methods, with a focus on binary descriptors for improved efficiency. BRISK uses a pyramid-based approach to detect key points at multiple scales, and it employs a binary pattern for descriptor computation, making it highly efficient and memory-friendly. BRISK also incorporates rotation invariance and scale adaptability, making it robust to changes in scale, rotation, and illumination.

ORB [21], which stands for Oriented FAST and Rotated BRIEF is also a popular computer vision algorithm that combines the speed and efficiency of FAST and the binary descriptor of BRIEF, with the added capability of handling rotation invariance. ORB was introduced by Ethan Rublee et al. in 2011 as an efficient and robust alternative to existing feature detection methods. ORB uses the FAST algorithm to detect

key points and computes a binary descriptor using BRIEF, while also estimating the orientation of key points for improved rotational invariance. This makes ORB well-suited for applications that require both speed and robustness, such as image matching, object recognition, and simultaneous localization and mapping (SLAM). ORB has gained popularity due to its efficient computation, low memory footprint, and robustness to scale rotation, and illumination changes. AKAZE [23] or Accelerated-KAZE was introduced as an improvement over the KAZE algorithm [23] with the aim of achieving faster processing times without sacrificing performance. AKAZE uses a nonlinear scale space for feature detection, which allows for accurate keypoint localization at multiple scales, and it employs a novel descriptor that captures both local and global information for improved matching. AKAZE is known for its robustness to changes in scale, rotation, and illumination, as well as its ability to handle challenging imaging conditions, such as motion blur and viewpoint changes.

Harris corner detection [18] is another widely used image processing technique that aims to identify and locate important feature points, or corners, in an image. It was developed by Chris Harris and Mike Stephens in 1988 and is based on the concept of detecting regions in an image where intensity changes occur in multiple directions. Harris corner detection utilizes the eigenvalues of a matrix calculated from image intensity gradients to determine if a pixel represents a corner, making it a popular method for applications such as image recognition, image stitching, and object tracking.

In summary, SIFT and SURF are accurate but computationally expensive, Harris corner detection and FAST are fast and efficient but may have limitations in certain scenarios, BRISK and ORB are efficient in terms of storage and matching speed but may have limitations in challenging image conditions, and AKAZE is computationally efficient and robust to aerial image challenges but may have limitations in extreme viewpoint changes or occlusions. The choice of algorithm depends on the specific requirements and constraints of the aerial image stitching task at hand.

2.5.1 Learning-based feature identification and matching

With the recent advancements in deep learning-based technologies and SLAM (Simultaneous Localization And Mapping) technologies, many learning-based feature-matching methods are being introduced. These learning-based methods can be divided into two subgroups, Detector-based feature-matching methods, and detector-free feature-matching methods. Several detector-based and learning-based methods, such as LIFT [24], MagicPoint [25], SuperPoint [26], and SuperGlue [27], have demonstrated superior performance in scenarios with changing illumination compared to traditional feature matching algorithms.

LIFT (LInear Feature Transform) [24] algorithm proposed by Yi et al. in their pa-

per titled "LIFT: Learned Invariant Feature Transform". LIFT aims to improve upon the traditional SIFT (Scale-Invariant Feature Transform) algorithm by learning a linear transformation that is invariant to various image transformations, such as rotation, scale, and illumination changes. In the LIFT algorithm, local image patches are extracted around keypoints of interest, and a descriptor is computed for each patch. Unlike SIFT, which uses handcrafted descriptors, LIFT learns a descriptor transformation matrix using a deep neural network. This transformation matrix is trained to be invariant to different image transformations, allowing the descriptors to be more robust and accurate in matching features across different images. MagicPoint paper [25] aims to address the limitations of traditional feature point detectors, such as sensitivity to image transformations, by proposing a novel approach to detecting stable feature points in images. MagicPoint leverages deep neural networks to learn feature point detectors that are robust to changes in scale, rotation, and illumination. The algorithm uses a convolutional neural network (CNN) to predict both the location and the scale of feature points in an image. It also incorporates a spatial transformer network (STN) that allows the CNN to explicitly model and correct for geometric transformations in the image. One of the key contributions of MagicPoint is its ability to handle challenging imaging conditions, such as extreme scale changes, large rotations, and strong illumination changes, which are often problematic for traditional feature point detectors.

Superpoint [26] is a deep learning-based approach for detecting and describing interest points in images, which can be used for tasks such as image matching, image retrieval, and visual localization. SuperPoint is trained in a self-supervised manner, meaning it does not require labeled data for training. Instead, it leverages image pairs with known camera poses to generate ground truth correspondences, which are then used to supervise the training process. SuperPoint uses a convolutional neural network (CNN) to detect and describe interest points in an image. The detected interest points are represented as pixel-wise heatmaps, indicating the likelihood of each pixel being an interest point, while the descriptors are learned using a separate CNN. Same as MagicPoint, the paper suggests that it can also show a strong ability to handle challenging imaging conditions.

SuperGlue uses graph neural networks (GNNs) to learn the matching process. It first extracts local features from images using a convolutional neural network (CNN) and then constructs a graph representation of the features. The GNN is trained to propagate information across the graph and predict matching scores between pairs of features. The matching scores are used to establish correspondences between features in different images.

"LightGlue: Local Feature Matching at Light Speed" [28] is a paper that introduces LightGlue, a deep neural network designed to match local features across images12. The authors revisit multiple design decisions of SuperGlue, the state of the art in sparse

matching, and derive simple but effective improvements12. LightGlue is more efficient in terms of both memory and computation, more accurate, and much easier to train12. One key property is that LightGlue is adaptive to the difficulty of the problem: the inference is much faster on image pairs that are intuitively easy to match, for example, because of a larger visual overlap or limited appearance change12. This opens up exciting prospects for deploying deep matches in latency-sensitive applications like 3D reconstruction12.

MagicPoint, LightGlue, SuperPoint, and SuperGlue are three state-of-the-art approaches for feature detection, description, and matching in computer vision tasks, especially in the domains of Simultaneous Localization and Mapping (SLAM), and 3d localization. MagicPoint is simple and efficient, achieving high accuracy and computational efficiency. SuperPoint is self-supervised and robust to scale changes and viewpoint changes. SuperGlue uses graph neural networks to capture contextual information and is robust to challenging matching scenarios. However, all three methods rely on handcrafted or learned features, may suffer from ambiguities in feature matching, and can be sensitive to changes in lighting conditions. Especially in the areas of indistinctive regions such as uniform textured areas or low contrast areas.

To overcome these issues, recent research focuses on detector-free leaning-based feature matching methods. NcNet [29], sparse NcNet [30], and DualRC-Net [31] can be identified as some of the models proposed in this area with promising results. Recently the use of Transformers [32] in vision-related tasks has attracted a lot of focus due to the computational efficiency they promise. LoFTR [22], the feature-matching model incorporated in this proposed stitching method is based on Transformers and the paper indicates that it outperformed all of the existing methods even in problematic illumination changing situations, and uniform textured images.

LoFTR- Local Feature TRansformer [22] is a detector-free method that performs local feature matching in image pairs. LoFTR incorporates a detector-free design to extract position and context-dependent local features, which are transformed into feature representations in order to match between images. It uses a Convolutional Neural Network (CNN) to extract multi-level features from the images being compared. A Local Feature Transformer module is then used to extract the position and context-dependent local features. Using linear transformers in the algorithm reduces the computation cost making it suitable for SLAM applications. The use of self-attention and cross-attention layers improves the accuracy of the predictions by allowing the model to attend to important features and patterns within the input data. Furthermore, the model can handle different types of input data and match them more accurately by using the model's ability to use an optimal transport layer or a dual-softmax operator in establishing coarse-level matches.

2.5.2 Feature Detection and Matching in UAV Image Mosaic Generation

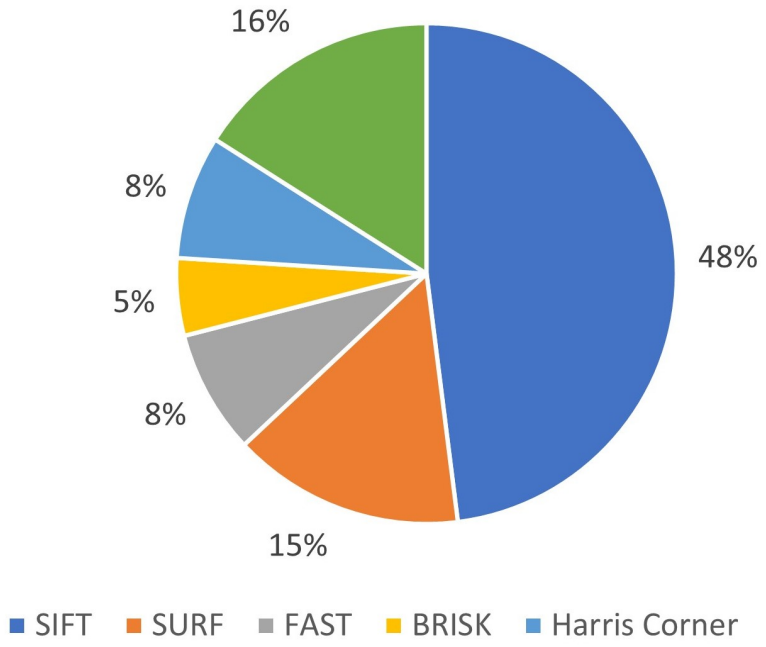


Figure 2.4: Feature detection methods used in aerial image stitching as a percentage of publications.

The examination of various global single transform feature revision methodologies by Gómez-Reyes et al [33] highlights several commonly employed techniques in the field of UAVs, including Harris corner, SIFT, FAST, ORB, SURF, and BRISK. A comparative analysis reveals that the traditional ORB method stands out for its rapid processing capabilities, rendering it suitable for real-time applications. However, it falls short in accuracy compared to alternatives like FAST and BRISK, particularly in scenarios involving rotation and rapid scale changes. Despite this, SURF tends to be favored due to its performance parity with SIFT and its comparable speed to ORB. Nevertheless, when prioritizing accuracy and robustness over speed, SIFT emerges as the preferred choice, as indicated by recent advancements leveraging GPU and computing algorithms, as depicted in Figure 10. These developments signify an enhancement in the speed of SIFT applications. 2.4 summarize the findings in their publication as a percentage.

2.6 Hyperspectral Image Stitching Pipelines

The typical hyperspectral image stitching process differs slightly from the normal RGB imaging process. Depending on the feature-matching algorithms employed, it is

Table 2.3: Summery of the hyperspectral image stitching research.

Paper	Feature Detection and Matching	Resolution of stitched images	Remarks
Mo et al, A robust uav hyperspectral image stitching method based on deep feature matching	SuperGlue	Width 640 to 687 px Height 752 to 7738px	Line scanning, High altitude
Zhang et al , "Hyperspectral panoramic image stitching using robust matching and adaptive bundle adjustment	SuperPoint	960 px x 1057 px	Urban Landscapes 300m altitude
Zang et al, Automatic stitching for hyperspectral images using robust feature matching and elastic warp	SIFT	960px x 1057 px	300m altitude
Peng et al, Hyperspectral Image Stitching via Optimal Seamline Detection	SIFT	960px x 1057 px	300m altitude

crucial to select one or several spectral bands for feature matching. If the matching algorithms support three-channel inputs, such as the SIFT feature detector and matcher, then it is possible to select three spectral bands from the hyperspectral image. Another option is to extract a single channel from the image and use it to match features. Typically, the spectral channel with the highest contrast is chosen due to the greater number of features that detectors can detect from the scene.

Specific research addressing hyperspectral image stitching when this literature survey was carried out is summarized in Table 2.3. There are only a limited number of research available addressing the specific topic. Mo et al [34] discusses the stitching of images obtained with the line scanning sensor, and they incorporate the use of superGlue [27] as their feature detection method. Zang et al [35, 36] discusses the use of the same type of sensor stitching in two separate papers. One using superPoint [26] as the feature

detection method and one using SIFT [13] as the detector and matcher. Finally, Peng et al. [37] discuss another approach using the same SIFT detector and matcher. None of these methods discusses the use of low-pixel resolution snapshot stitching of hyperspectral images. Further, all of the images were taken at higher altitudes which made it easier to capture geometrical features needed in the stitching process. It can be concluded that there is an obvious research gap that needs to be explored to advance hyperspectral imaging technology.

2.7 3D Convolution Neural Network and Siamese Network Architecture

The second part of this thesis proposes a novel approach to obtain a robust set of feature matches. The proposed method utilizes the spectral information available in the hyperspectral image for decision-making. In order to achieve this, a 3D convolution-based Siamese network was utilized. This section of the related work will briefly describe the background on 3D convolution and Siamese networks.

2.7.1 3D Convolution Neural Networks

Convolution is a fundamental operation in neural networks, particularly in Convolutional Neural Networks (CNNs). It involves sliding a filter or kernel over the input data and performing element-wise multiplication followed by a summation, which results in a single output. This process is repeated across the entire input, resulting in a feature map that represents the learned features of the input data. The filters are learned during the training process, allowing the network to adapt to the specific task at hand [38].

While traditional convolution operates on 2D data, such as images, 3D convolution extends this operation to 3D data. In 3D convolution, the kernel is also three-dimensional, and the convolution operation is performed across the width, height, and depth of the input data. This makes 3D convolution particularly suitable for volumetric data, such as medical imaging or video data, where temporal information can be considered as an additional dimension [39].

3D Convolutional Neural Networks (3D-CNNs) leverage 3D convolution to learn spatial and temporal/spectral features simultaneously. They have been successfully applied in various fields, including video analysis, action recognition, and medical image analysis.

Hyperspectral images, which capture a wide spectrum of light, result in a 3D data cube with two spatial dimensions and one spectral dimension. This makes them a perfect candidate for 3D convolution. By applying 3D convolution to hyperspectral images, a model can learn both spatial and spectral features simultaneously, which can lead to

improved performance in tasks such as object detection, classification, and segmentation [39–41].

Moreover, 3D convolution can help to preserve the spatial and spectral correlations present in hyperspectral images, which are often lost when using traditional 2D convolution. This can lead to more accurate and robust models for hyperspectral image analysis. Therefore, 3D convolution and 3D-CNNs hold great promise in advancing the field of hyperspectral image analysis and related machine learning tasks.

2.7.2 Siamese Neural Networks

Siamese Neural Networks (SNNs) represent a specialized class of artificial neural network architectures featuring two or more identical sub-networks. These parallel sub-networks, often termed twin networks, share identical configurations, encompassing the same parameters and weights [42, 43]. Functioning in tandem, they process input data and ultimately compare their respective outputs, typically employing a distance function to gauge similarity. Consequently, the output yielded by a Siamese neural network's execution can be interpreted as the semantic similarity between the projected representations of the two input vectors [42, 43].

The architecture of a Siamese Network comprises two identical sub-networks, each tasked with processing a distinct input sample using the same set of weights. Leveraging convolutional and pooling layers, these networks extract meaningful features from the image samples [44]. Crucially, the final layer in each branch is the comparison layer, responsible for generating an embedding—a compact representation of the data sample—that facilitates the subsequent similarity assessment.

Let X_1 and X_2 be the inputs to the Siamese network. The network consists of two identical subnetworks that share weights. Let f_θ represent the shared subnetwork with parameters θ. The outputs of the two subnetworks are denoted as $f_\theta(X_1)$ and $f_\theta(X_2)$. The distance between the outputs is typically computed using a distance metric, such as the Euclidean distance or the contrastive loss.

The output of the Siamese network can be represented mathematically as:

$$D(X_1, X_2) = \phi(\|f_\theta(X_1) - f_\theta(X_2)\|),$$

where $\|\cdot\|$ denotes a distance metric, ϕ is an optional transformation (e.g., scaling), and $D(X_1, X_2)$ is the output similarity score between inputs X_1 and X_2.

By combining the two methods it is possible to create 3D Convolution Siamese Networks. These networks can be particularly useful in hyperspectral imaging. They can learn to recognize similar patterns in different hyperspectral images, making them highly effective for tasks such as image matching and object recognition. Moreover, they can

handle the high dimensionality of hyperspectral data, which is often a challenge for traditional machine learning algorithms.

One of the key advantages of 3D Convolution Siamese Networks in hyperspectral imaging is their ability to handle few-shot learning scenarios. In many real-world applications, there is a strong demand for labeled data for hyperspectral classification, but acquiring sufficient labeled data can be a big obstacle. 3D Convolution Siamese Networks can address this problem by learning to recognize patterns in the data with only a few labeled samples. This makes them a promising tool for hyperspectral image analysis.

2.8 Dimentionality Reduction in Hyperspectral Images

In hyperspectral imaging, dimensionality reduction plays a crucial role in managing the vast amount of data collected. This is also an important step when training any of the hyperspectral image-based machine learning techniques. This section discusses what dimensionality reduction entails, its significance in hyperspectral imaging, and common techniques employed for this purpose.

Dimensionality reduction refers to the process of reducing the number of variables or features under consideration. In the context of hyperspectral imaging, where each pixel contains a spectrum with hundreds of bands, dimensionality reduction aims to transform these high-dimensional data into a more manageable and interpretable form while preserving essential information [45].

Hyperspectral imaging systems capture images across numerous narrow and contiguous spectral bands, resulting in data with high dimensionality. However, processing and analyzing such high-dimensional data pose significant challenges, including computational complexity, storage requirements, and the risk of overfitting. Dimensionality reduction techniques alleviate these challenges by extracting the most relevant information from the data while discarding redundant or noisy features [45, 46].

Moreover, dimensionality reduction facilitates data visualization, interpretation, and understanding. By reducing the data to a lower-dimensional space, analysts can explore and interpret the relationships between spectral bands more effectively, aiding in the identification of patterns, anomalies, and features of interest within the hyperspectral imagery.

Dimensionality reduction methods in hyperspectral imaging can be broadly categorized into two types: Supervised and unsupervised. For supervised methods, Supervised Principal Component Analysis (SPCA) [47], Linear Discriminant Analysis (LDA) [48,49], Local Fisher Discriminant Analysis (LFDA) [50,51] can be identified as some of the popular algorithms in literature.

Principal Component Analysis (PCA) [52, 53], Independent Component Analysis

(ICA) [52, 54], Wavelet Dimensionality Reduction [55, 56] could be identified as some of the unsupervised methods using in the hyperspectral domain. Autoencoders are another type of dimensionality reduction technique that could be used in both supervised and unsupervised dimensionality reduction techniques.

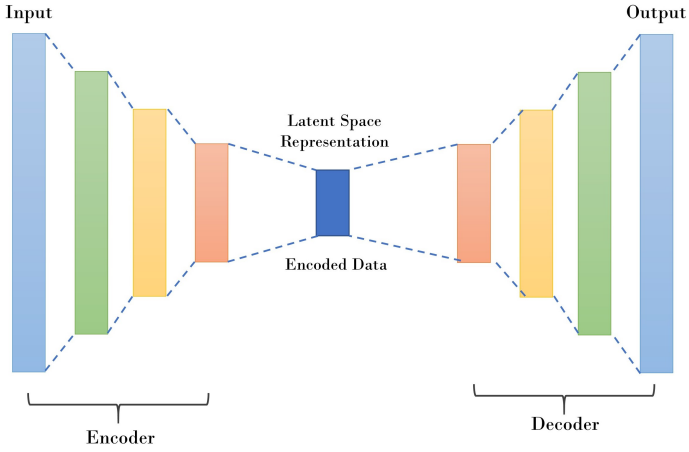


Figure 2.5: Typical encoder architecture.

Autoencoders are a class of artificial neural networks used for unsupervised learning, particularly in the field of deep learning. The fundamental principle behind autoencoders is to learn a compressed representation, or encoding, of input data and then reconstruct the input data from this representation [57]. This is achieved through two main components: an encoder, which maps the input data to a lower-dimensional latent space, and a decoder, which reconstructs the input data from this reduced representation. During training, the autoencoder learns to minimize the reconstruction error, typically using techniques like gradient descent and backpropagation [58]. Typical encorder architecture is depicted in the figure 2.5 where the latent space is taken as the encoded data.

In the context of hyperspectral imaging, where each pixel contains a high-dimensional spectrum of intensities across numerous wavelengths, autoencoders offer an effective means of dimensionality reduction. By compressing the spectral information into a lower-dimensional space, autoencoders can capture the essential features of the data while discarding redundant or less informative components [59].

One significant advantage of using autoencoders for dimensionality reduction in hyperspectral imaging is their ability to capture nonlinear relationships and complex patterns within the data. Unlike traditional linear techniques such as principal component analysis (PCA), which may struggle to effectively capture nonlinearities, autoencoders can learn more intricate representations, leading to potentially better compression and reconstruction of hyperspectral data.

Additionally, autoencoder-based dimensionality reduction often results in more inter-

pretable latent representations compared to other methods. The latent space learned by the autoencoder can provide insights into the underlying structure of the hyperspectral data, potentially revealing meaningful features or clusters that are not readily apparent in the original high-dimensional space [60].

Another notable advantage of autoencoders, particularly in practical applications such as real-time processing or online analysis of hyperspectral data, is their relatively low inference time [61]. Once trained, the encoding and decoding process can be efficiently performed, allowing for rapid compression and reconstruction of hyperspectral images. This is especially beneficial in scenarios where fast processing is essential, such as in remote sensing applications or in environments where computational resources are limited.

Chapter 3

Configuration and Initial Meassurements from the system

Hyperspectral imaging technology has seen significant advancements in recent years, offering unprecedented capabilities in capturing detailed spectral information across a range of wavelengths. In this chapter, we delve into the utilization of the Cubert Ultris 5 Hyperspectral camera, in conjunction with UAV platforms for image acquisition. The chapter provides an overview of the equipment specifications, setup, and methodology employed for data acquisition and preprocessing.

The Cubert Ultris 5 Hyperspectral camera, equipped with 51 spectral bands spanning from 450nm to 850nm, presents an ideal solution for aerial imaging tasks. With a resolution of 290×275 pixels and a compact form factor weighing only 120g, this camera emerges as a promising candidate for integration with UAV systems. Leveraging the capabilities of the DJI M600 Pro drone, equipped with a payload capacity suitable for accommodating the camera and associated hardware, we embarked on UAV missions for hyperspectral imaging.

The chapter describes the image acquisition process, delineating the steps involved in ensuring optimal data quality. From thermal stabilization of the camera to calibration procedures involving white and black calibration images, each stage is outlined. Additionally, the integration of DJI A3 pro flight controller and DJI Ground Station pro software for mission planning and execution is detailed.

Furthermore, the chapter delves into the preprocessing pipeline essential for converting raw image data into radiance images. This encompasses radiometric calibration to account for sensor response characteristics, correction for dark current and flat-field response, and optional atmospheric correction. Noteworthy is the discussion on the spectral signature evaluation, where comparisons between drone-captured hyperspectral images and field spectrometer data unveil insights into the accuracy and reliability of the acquired data.

In conclusion, this chapter sets the stage for subsequent analyses and applications

of drone-based hyperspectral imaging data. The demonstrated capabilities underscore the transformative potential of this technology across diverse domains, ranging from environmental monitoring to precision agriculture and beyond. Moreover, it underscores the importance of methodological refinements and preprocessing techniques in harnessing the full potential of hyperspectral imaging for real-world applications.

3.1 System setup

Table 3.1: Specifications of the camera.

Wavelength Range	450 - 850 nm
Spatial Resolution	290 x 275 pixel
Spectral Bands	51
Spectral Sampling	8 nm
FOV (Field of View)	15°
Weight	126 g
Dimensions	29 x 29 x 65 mm

Hyperspectral image acquisition was carried out using the recently introduced Cubert Ultris 5 Hyperspectral camera [8]. As shown in the specification Table 3.1 camera has 51 spectral bands in the wavelength range of 450nm to 850nm. Further, this camera has a resolution of 290 × 275 pixels with a field of view of 15° and a full-width half maximum of 26nm at 532nm. At a package size of 29 × 29 × 49 mm and weight of only 120g, this is an ideal candidate to be used on most UAV platforms. The camera and the image acquisition computer(Pokini F2 mini PC [62]) sum up to a weight of 500g to 600g depending on the cabling used to connect the interfaces. With the additional circuitry and cabling total weight of the system was about 800g. Hence, any drone with a payload of at least 1kg and has the mounting space for the camera and the single board computer could be used for mounting the camera. Since we already have access to a DJI M600 Pro [63] drone at the laboratory, it was used to carry out the UAV missions.

Figure 3.1 (a) depicts the overall system and 3.1 (b) shows the system configuration. The camera and the Mini PC were powered using the same UAV (Unmanned Aerial Vehicle) battery with a voltage of 18v. A Pololu voltage regulator was used to step down the voltage to 12 for the circuit and to smooth out any power ripples caused by power spikes in the drone. To trigger the camera, one of the DJI PWM (Pulse Width Modulation) outputs was programmed to output the trigger signal. A PWM relay was

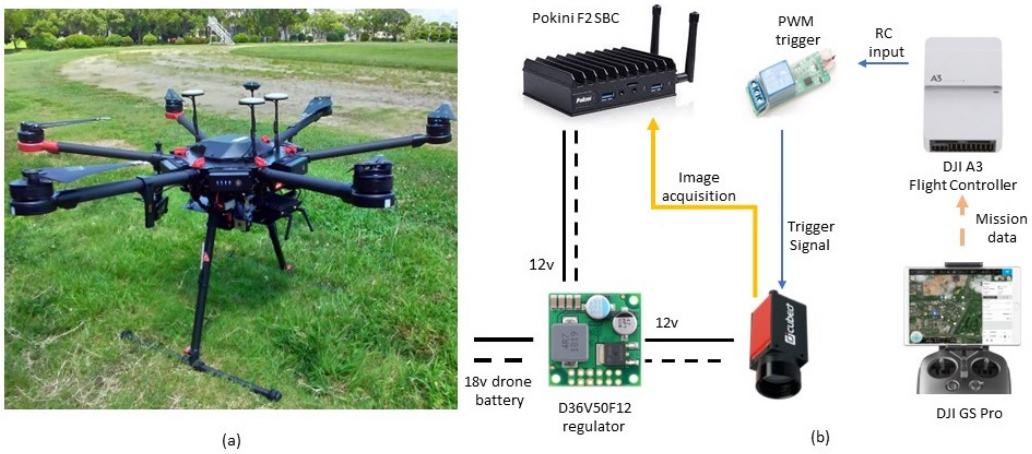


Figure 3.1: (a) Image of the drone (b) System configuration diagram

Table 3.2: Settings used with the DJI GS pro application.

Camera Settings		Application Settings	
Sensor resolution width	290px	Shooting angle	Parallel to main path
Sensor Res height	275px	Capture mode	Hover & capture
Sensor width	0.95mm	Flight course	Scan Mode
Sensor height	0.9mm	Speed	Varying speeds
Focus length	3mm	Front overlap	80%
Min shutter interval	0.1	Side overlap	80%

used to trigger the camera shutter using the camera's trigger interface. The trigger pin on the camera was used to receive the trigger signal and capture the hyperspectral snapshot cube. Once the camera was triggered and the hyperspectral snapshot cube was captured, the camera software running on the Mini PC stored the data on the SSD (Solid State Drive) of the computer.

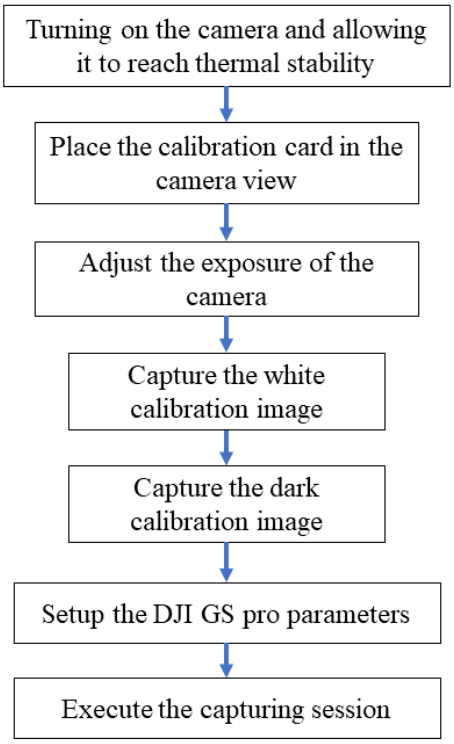


Figure 3.2: Overview of the image acquisition process followed

3.2 Data Acquisition and Preprocessing

3.2.1 Data acqusition

Figure 3.2 presents the overview of the flow followed in order to perform a data acquisition session. Initially, the camera was turned on and kept until it reached thermal stability. Then a white calibration card was placed in front of the camera view at the same angle as the data acquisition was performed. Furthermore, extra care was taken to make sure the card surface was evenly lit by the sunlight and there were no shadows or foreign materials on the surface of the calibration card. To fulfill the next step of adjusting the camera exposure, the flight PC was connected to the laptop via a local wifi network. Virtual Network Client (VNC) was used to connect the laptop to the flight PC. The exposure of the camera was then adjusted using the manufacturer's software. This will ensure the images will not get overexposed. Then a white calibration image was taken followed by a black calibration image. After that flight controller software was used to plan the mission. After planning the mission, the user has the freedom to auto-launch the UAV and start the mission or, manually launch the UAV and start the mission.

The use of the DJI A3 pro flight controller in the drone allows the use of DJI Ground

Station pro [64] software for flight planning. The hyperspectral camera should be configured as a custom camera in the application and the settings used are presented in the Table 3.2 left column. The right column presents the flight path and capturing related settings of the application. An overlap of 80% was used for both front and side overlap. This is to ensure enough coverage between images for feature matching. Furthermore, due to a lack of Real-Time Kinematics (RTK) based positioning in the drone, wind-related disturbances will cause the drone to shift. Due to the narrow FOV of the camera, this shift would cause difficulties in the feature-matching step. The high overlapping percentage will also help to tackle this issue. Capturing was carried out in Hover & capture mode. This will ensure that the camera will be stable during the shutter activation. After selecting the flying area, the DJI GS pro application will calculate the flight path and the number of waypoints it needs to stop and trigger the shutter. DJI A3 flight controller only has the ability to hold 100 waypoints at a time. If the mission has more than 100 data points, it will stop after finishing the first 100d data points, until the remaining data points are loaded in to the memory.

Figure 3.3: Image acquisition location- Mie University farm

All the data acquisition related to this thesis was carried out on the Mie University farm located in the Tsu city of Mie prefecture in Japan. An aerial image of the farm is depicted the Figure 3.3. This farm grows different crops throughout the year including but not limited to rice, wheat, cabbage, onion, potato, sweet potato, tea, mango, orange and leafy vegetables. Before each flight session, the camera was turned on till it reached thermal stability. This is to ensure that the sensor's thermal response was not affecting the measurements. A sequence of images was captured in two-minute intervals and the sensor metadata was checked in order to ensure the temperature stayed the same over several images.

3.2.2 Pre-processing captured images.

Once the image acquisition session was completed, data were transferred to a separate personal computer for further processing. In hyperspectral imaging, the sensor captures the electromagnetic radiation reflected or radiated from the target. Captured data is typically in the form of digital numbers (DN) or radiance values, which represent the intensity of the electromagnetic radiation detected by the sensor. These values are influenced by various factors such as the sensor's response, atmospheric effects, and sensor-specific artifacts such as dark current and flat-field response. A radiometric calibration step should be performed to convert the raw data into reflectance values. Reflectance represents the proportion of incident radiation that is reflected by the target surfaces.

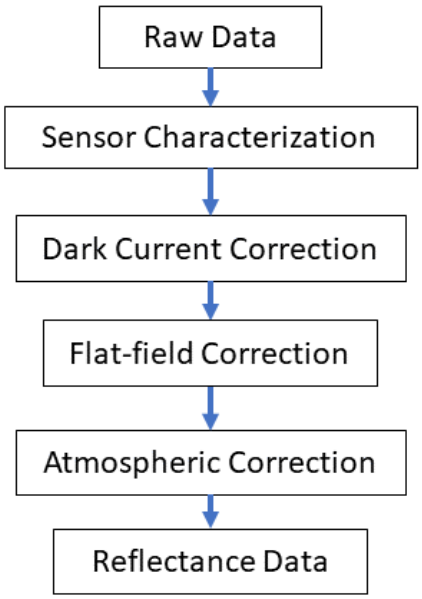


Figure 3.4: Radiometric calibration process

Key steps involved in converting digital numbers to reflectance are shown in Figure 3.4. The first step is to characterize the sensor response. This involves determining its gain, offset, linearity, and spectral response function by using calibrated radiometric sources or reference targets. This step was already carried out by the manufacturer and the provided calibration was used during experiments. The next step is to correct the dark current response of the sensor. Dark current is the signal generated by the sensor even in the absence of any radiation. For this step, the camera was covered with a lens cap to block all the light. An image was captured and stored as a dark image and each measurement was substracted by this image to remove the dark current noise.

$$I_{ref} = \frac{I_{raw} - I_{DARK}}{I_{white} - I_{DARK}} \tag{3.1}$$

The next step of the calibration is to correct the flat-field response of the sensor. In this step, raw data is divided by a flat-field image which is obtained by uniformly illuminating the sensor with a stable radiometric source. In the field, this was achieved by taking an image of a calibration card exposed to the sunlight. Highly reflective, nearly Lambertian surfaces such as Spectralon targets are recommended as the calibration surface. During our experiments, Novoflex Zebra XL [65] was used as the calibration surface. It has a stable reflectance curve in the region from 450 nm to 850nm. Figure 3.5 presents the spectral response of the aforementioned calibration card measured concerning a spectral calibration target using a field spectrometer. The calibration card was placed in a way that would cover the entire image area of the camera and the image was captured after that. Equation 3.1 presents the calculation followed to obtain the corrected image. The dark image captured in the last step I_{DARK} was subtracted from the target image I_{raw} and the white calibration image I_{white}. Then the resultant raw image was divided by the white to obtain the corrected image I_{ref}. This correction accounts for spatial variations in sensitivity across the sensor's field of view. Furthermore, this step removes the vignetting effect from the image.

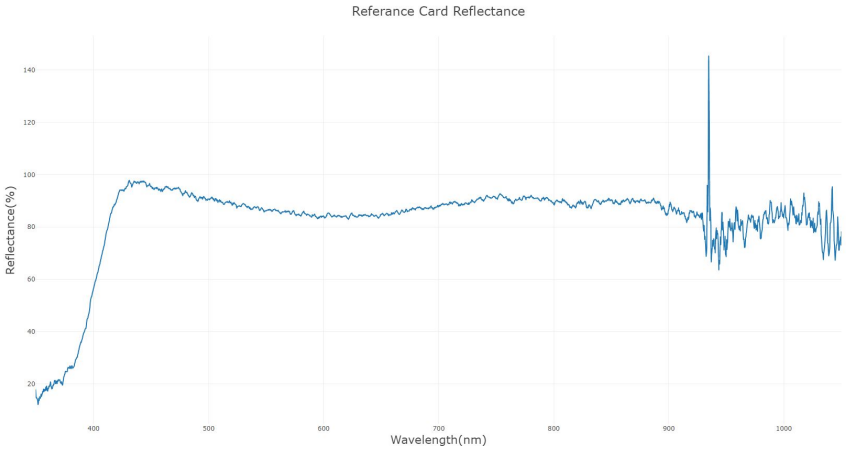


Figure 3.5: Spectral response of the calibration card used during data acquisition sessions.

After flat-field correction, atmospheric correction should be performed to remove the effects of atmospheric scattering and absorption from the raw data. However, this step could be skipped in the UAV-based imaging due to atmospheric influence being negligible in low-flying altitudes [66]. These steps are performed using the manufacturer's Python SDK. Pre-processed images were then saved as numpy 16bit arrays for further processing. Furthermore, Global Positioning System (GPS) information embedded in each image was extracted and saved separately for further processing.

3.3 Spectral Signature Evaluation

After the initial setup of the system, a set of experiments was carried out to evaluate the spectral signatures obtained from the flight with the field-level spectral signatures. The following subsections will discuss the specific process followed, the results obtained, and the conclusion of the performed experiments.

3.3.1 Field Data Collection

In order to compare the hyperspectral camera images with the field spectrometer, two data acquisition sessions were carried out in the Mie University farm located at coordinates 34°48'26.2"N 136°27'13.7"E on two separate days. During session 1, the drone was flown at a height of 80m, resulting in a ground sampling distance of 8.7cm per pixel. The drone was used to capture images of asphalt and rice plants using the steps described in the previous section. At the same time, a field spectrometer (EKO MS-730 [67]) was used to collect spectral signatures of the asphalt and rice plants in the same location.

In session 2, the drone was flown at a lower height of 40m, resulting in a smaller ground sampling distance of 4.9cm per pixel. Hyperspectral images were captured for soil, cabbage, potato, and tea targets using the same procedure and imaging parameters. Spectral signatures of these targets were also collected using the field spectrometer. After data acquisition, one further pre-processing step was carried out by applying a Savitzky-Golay Filter in order to smooth the spectrum. Then a $5pxx5px$ area from the hyperspectral image equivalent to the location of the field spectral meter data was obtained and averaged to get one spectral signature. Spectral values obtained from the hyperspectral images were compared with the spectral values obtained by the field spectrometer for the same locations in this manner. This allows the assessment of the accuracy and reliability of the hyperspectral camera images, as well as the ability to use the drone-based imaging system for applications such as vegetation monitoring, crop classification, and soil analysis.

3.3.2 Results and Discussion

Figure 3.6 and Figure 3.7 present a comparison between the spectrum information obtained from the drone and the field spectrometer data for the same targets in session 1 and session 2, respectively. The first observation from the results shows that field spectrometer data appears to be smoother compared to the spectral signatures obtained using the hyperspectral camera. This difference could be attributed to several factors.

Firstly, the field spectrometer's spectral resolution is much higher than the hyperspectral camera's spectral resolution. So the sampling interval is much higher in the field

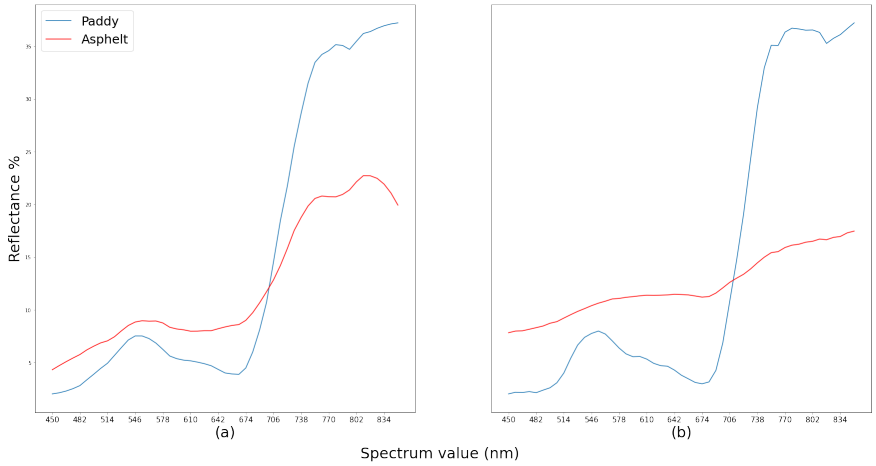


Figure 3.6: Session 1 data (a) Spectral signatures captured from the drone (b) Spectral signatures captured from the field-spectrometer

spectrometer. In this case, the field spectrometer has a spectral resolution of 3 nm and a sample interval of 0.35 nm much higher than the hyperspectral camera's 8nm sampling interval. As a result, the field spectrometer data may show more detailed spectral features and appear smoother when plotted. Secondly, the process of resampling field spectrometer data to match hyperspectral data can also affect the smoothness of the spectral curve. In this case, the 450 nm to 850nm spectral range of the field spectrometer was resampled to match the 51 bands of the hyperspectral camera using an averaging filter. This effectively will smoothen the field spectrometer data.

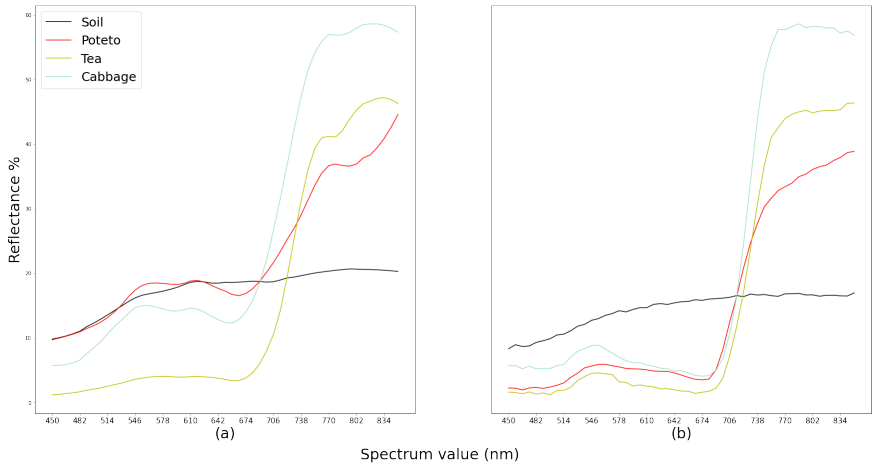


Figure 3.7: Session 2 data (a) Spectral signatures captured from the drone (b) Spectral signatures captured from the field-spectrometer

Another factor that might have been contributing to the observed differences in spectral signatures between the drone and the field spectrometer data can be identified as

the spectra mixing. Spectral mixing is the process by which the hyperspectral camera captures spectral information from multiple materials or targets within a single pixel due to its spatial resolution. For example, session 1 asphalt might contain dirt, or any other signatures from surrounding materials resulting in a different spectral signature compared to the field spectrometer data which localizes the asphalt spectrum directly. This is also evident from the fact that the second session's spectral signatures captured with a much higher spectral resolution compared to session one, displaces a close relationship between the field spectrometer signatures and hyperspectral camera signatures.

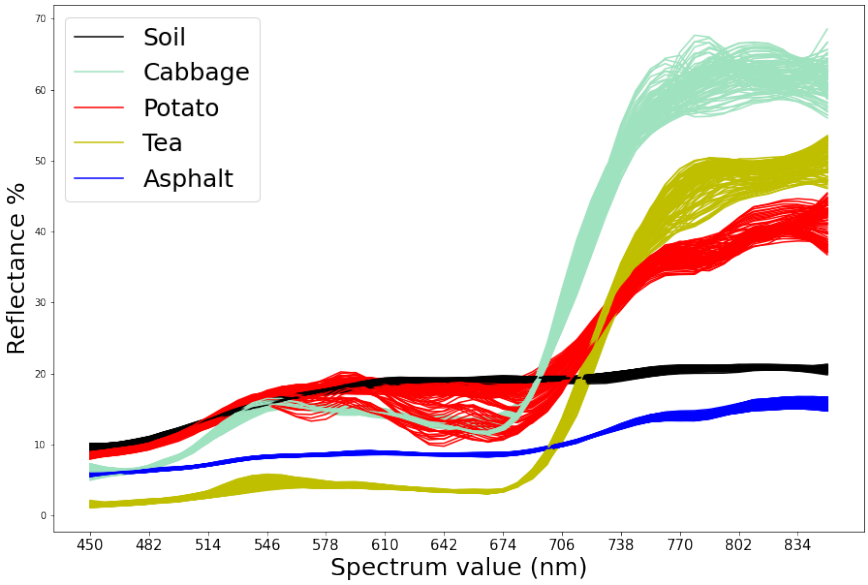


Figure 3.8: Variance of the spectrum of different targets.

However, spectral signatures were still different from each other. Figure 3.8 depicts the variation of each target's spectrum within a 10-pixel by 10-pixel area of the image. All three crops display about 5% to 10% of variance. On potatoes, the same amount of variation can be observed in the 560nm to 700nm spectral range as well. The potato crop selected for this session was nearing its harvest date, which means that it was likely in a stage of mixed maturity with a mix of dead leaves along with green leaves, This mixed state of the potato crop could have contributed to the high variance in the drone spectral signatures.

3.4 Conclusion

The results of this chapter demonstrate the significant potential of drone-based low-resolution hyperspectral cameras in capturing spectral signatures of various targets. The ability of the drone-based camera to provide spectral signatures that closely match those

obtained from the field spectrometer indicates its capability to effectively capture the spectral information of different materials and surfaces. While the spectral signatures may not be a perfect match, the high similarity between them indicates that the drone-based hyperspectral camera is capable of capturing spectral information that is distinguishable and can be further utilized for subsequent analyses.

One of the key implications of this finding is that the drone-based hyperspectral camera can be used for tasks such as the segmentation and classification of different targets or materials. Spectral signatures obtained from hyperspectral images can be used as a basis for developing algorithms and models to classify and identify various objects or materials on the ground, such as vegetation, buildings, water bodies, and geological features. This has significant implications for a wide range of applications, including environmental monitoring, precision agriculture, urban planning, and disaster management.

Furthermore, this section of the thesis highlights the need for case-specific pre-processing methods, such as noise cancellation, to further enhance the quality of the hyperspectral data. Noise or interference in the data can affect the accuracy and reliability of the spectral signatures, and therefore, appropriate pre-processing techniques may need to be employed to mitigate such issues. This underscores the importance of further refining the methodology and processing techniques for drone-based hyperspectral imaging, as well as considering specific environmental or target characteristics in order to improve the quality of the results.

Overall, the findings of this section lay a strong foundation for the subsequent sections of the thesis, which can leverage the drone-based hyperspectral images to address the main research problem of the thesis. The successful use of drone-based hyperspectral imaging opens up new possibilities for future research in the field, including exploring advanced data analysis techniques, refining data acquisition protocols, and integrating hyperspectral data with other remote sensing data sources. The results obtained from this section contribute to the advancement of knowledge in the field of drone-based hyperspectral imaging and pave the way for further research and applications in this exciting and rapidly evolving field.

Chapter 4

Image Mosaic Generation for Low-Resolution Hyperspectral Images

4.1 Introduction

As mentioned in the initial sections, image mosaic generation is a critical step in the decision-making process of any aerial imaging-based application. Hyperspectral imaging, which captures data in a wide range of spectral bands, has gained increasing attention in recent years due to its ability to provide detailed information about the composition and properties of objects and surfaces. However, the image mosaicing process becomes challenging in low-cost hyperspectral cameras with low spatial resolution and limited field of view. Section 2.4 of the thesis discussed the existing literature on image mosaic-generating techniques and approaches, on both normal aerial images and hyperspectral-specific imaging applications. However, the literature survey was unable to find any existing research focusing on the issue of low-resolution hyperspectral image stitching. This chapter specifically focuses on low-pixel-resolution hyperspectral image stitching, which presents unique challenges and requirements compared to traditional image mosaic generation methods.

4.1.1 Feature matching in low-resolution hyperspectral imaging.

This section will explain the research problem in more detail by comparing existing feature detection and matching methods used in image mosaic generation. Detection algorithms were compared for their capability to detect and match features in sample low-resolution hyperspectral images. A set of sample images were selected from different data acquisition sessions which are presented in Figure 4.1 (a) to (h). Each pair of images is adjacent to images from different data acquisition sessions. Each image pair was subjected to feature detection and matching using several selected feature detection and matching algorithms. SIFT [13], SURF [17], FAST [19], BRISK [20] and ORB [21]

AKAZE [23] algorithms were selected as the candidate methods based on their popularity in aerial image stitching [16]. shown images are single-channel images displayed in the Viridis color scheme instead of grayscale for easiness of viewing.

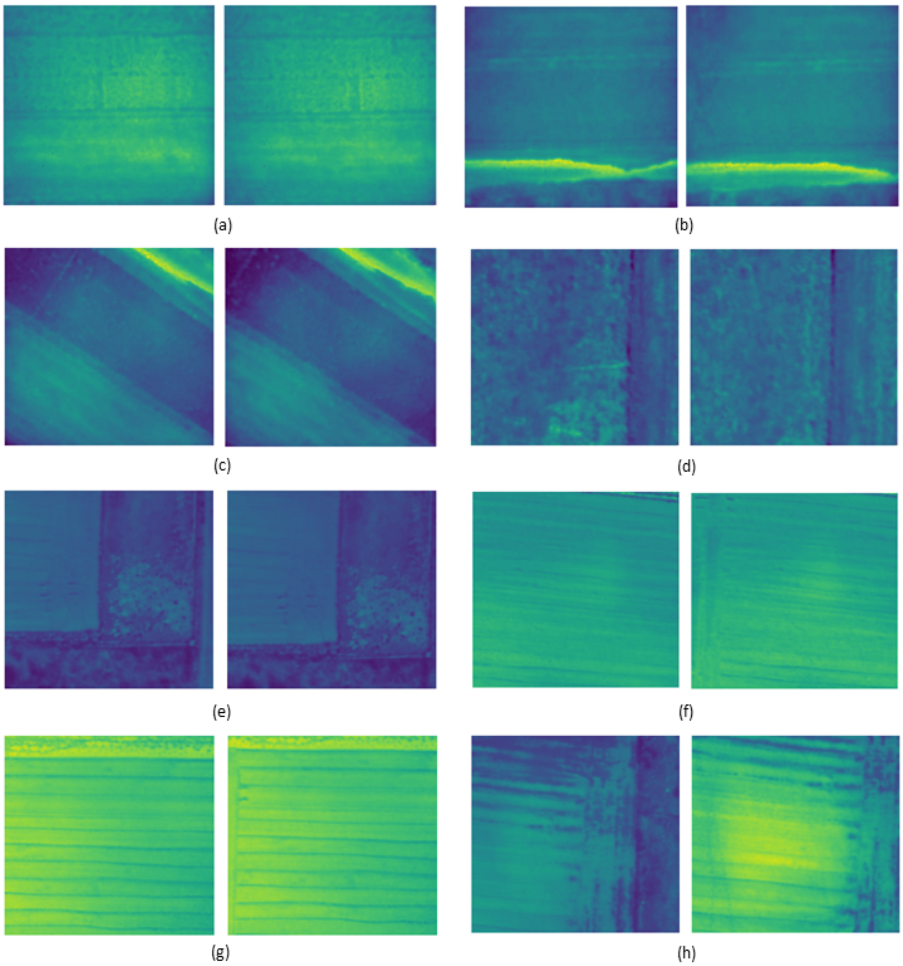


Figure 4.1: Sample images for feature detection and matching

For the comparison image pair was selected bandwise from each hyperspectral image. As an example, input for SIFT was arranged where two input images are each hyperspectral image's channel one. Next channel two was taken and the mean value was calculated at the end. Table 4.1 presents the mean values obtained for each image. Results indicate that certain images struggle to produce any usable amount of matches to calculate the transformation calculation.

4.1.2 Feature Matching Using Learning-based Feature Matching Algorithms

The second set of experiments was carried out to evaluate the performance of modern learning-based algorithms. LIFT [24], MagicPoint [25], SuperPoint [26], SuperGlue [27] and LoFTR [22] algorithms were selected for this set of experiments. similar to the

Table 4.1: Mean features identified and matched over the 51 channels of each hyperspectral image

Image pair	SIFT	ORB	AKAZE	SURF	FAST	BRISK
a	229	17	34	60	32	44
b	0	0	1	0	0	1
c	37	2	3	8	5	6
d	3	2	0	2	0	1
e	156	2	10	31	15	8
f	3	0	0	0	0	3
g	31	3	8	43	5	4
h	Fail	0	0	Fail	0	0

previous experiment number of matches was calculated between each pair of images for each channel of the images. Results indicate the superior performance of the LoFTR and SuperGlue algorithms when identifying matches between images. Here LoFTR feature matches were computed with the pretrained "outdoor" model. DISK feature descriptor was incorporated to be used with the SuperGlue matcher,

4.2 Image stitching Optimization Utilizing GPS Information

The traditional method of image stitching, which involves creating a mosaic by identifying features and descriptors from each image and finding matches between them, can be time-consuming and resource-intensive. This method relies on identifying matches between images and grouping them into subsets to determine the stitching order. A confidence parameter is calculated to determine the most likely adjacent image for stitching [68]. However, this process can be inefficient, especially for hyperspectral images [69].

The GPS-based optimization method utilizes GPS data to aid in the stitching process. The GPS information, such as the coordinates and orientation of the images, is used to determine the relative position and orientation of each image in the mosaic. By using GPS data, the need for extensive feature matching and descriptor calculation is reduced, resulting in a faster and more efficient stitching process [15, 68]. This section proposes a GPS optimization method that would offer a more efficient and accurate approach to image stitching. It reduces the processing time of image stitching by minimizing the need for extensive feature-matching calculation across every image.

Algorithm 1 GPS Line Segmenting Algorithm

```
Initialize first segment [GPS_pnts[0, 1]
for i + 2 ← len(GPS_pnts) do
    Calculate θ
    Calculate d
    Calculate difference between Currentd and lastd
    if d_difference <= d_tresh or θ <= θ_tresh then
        CurrentSegment ← GPS_pnts[i]
    else
        Start a new segment
        Initialize new Segment with GPS_pnts[i, i + 1]
    end if
end for
```

4.2.1 Methodology

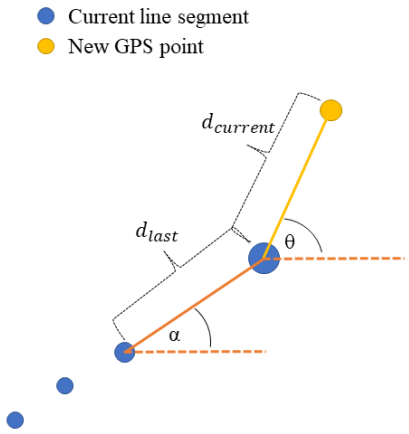


Figure 4.2: Parameters used in identifying segments

Algorithm 1 presents the overall procedure followed to obtain the line segments from the GPS information. This algorithm assumes that the flight planning software created the flight path in straight lines and that images are captured in a sequence. which is always the case for the mission created with the most popular flight planning software such as DJI Ground Station Pro. First, two images are used to initialize the first segment. From the next GPS point onwards θ:tangent angle between the current point and the last point of the segment and α: tangent angle between the last two points of the current segment is calculated. Further, $d_{current}$: distance between the current point and the last

point and d_{last}: distance between the last two points of the current segment is calculated. The aforementioned angles are depicted in figure 4.2. Angle difference $\theta - \alpha$ and the distance difference $d_{current} - d_{last}$ were calculated next. If the angle difference or the distance difference is below the threshold values set current point is added to the current segment. If not current segment will be saved and the new segment will be initialized using the current GPS point and the next one. Figure 4.3 (b) presents segmented lines from the set of GPS points shown in figure 4.3 (a).

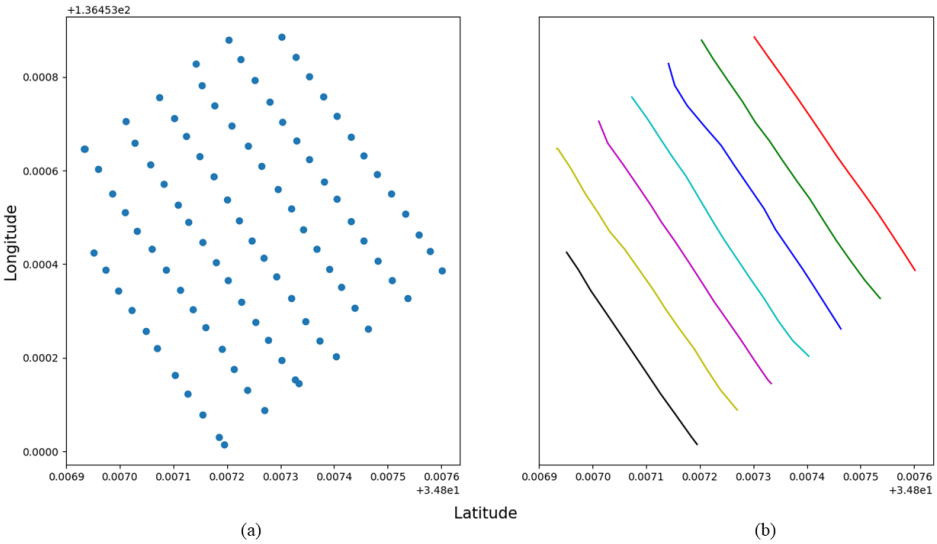


Figure 4.3: (a) GPS points of the recorded images (b) Segmented images subsets

Results show successful segmentation of lines using the aforementioned methodology. Furthermore, parallel processing could be incorporated to speed up the overall process by stitching each segment in separate cores.

4.3 Image Stitching Methodology

Overview of stitching two adjacent images shown in the algorithm 2. Implementation of the LoFTR and SuperGlue for the hyperspectral images was carried out using the Kornia computer vision library [70]. Kornia is an open-source computer vision library that provides a wide range of functionalities for image processing tasks. It is built on top of PyTorch, a popular deep learning framework, and provides a modular and flexible interface for performing various computer vision operations, including image transformations, filtering, feature extraction, and geometric operations. Kornia is designed to be highly efficient, GPU-accelerated, and compatible with both CPU and GPU computations. It is widely used by researchers and practitioners in the field of computer vision for developing cutting-edge applications and solutions. In this section, this thesis utilized Kornia for

Algorithm 2 Stitching Two Adjacent Images

```
Image1[:,:,Selected_band], Image2[:,:,Selected_band]
Convert to 8-bit images
Find matches Keypoints
if Keypoint_confidence > Threshold then
    Keep Keypoints
end if
Calculate Affine Transformation
if Affine Transformation is None then
    Calculate Homography Transformation
end if
Wrap each band using the Transformation matrix
Blend Each Band
```

implementing the LoFTR algorithm and SuperGlue algorithm with DISK descriptors for hyperspectral image stitching, as described in the previous section. The user can select between the two options when initializing the stitching algorithm. The decision between which method to select for feature matching will depend on the captured scene. During the experiments varying performance levels were observed.

Since the LoFTR and SuperGlue functions expect 8-bit images, the selected channel image was converted into 8-bit. Even though this step causes some information loss due to the conversion, it won't affect the final stitched image since two converted images are only used to obtain the matched keypoints. LoFTR algorithm. outputs matched keypoint pairs along with the confidence of each keypoint pair. It was then filtered with a user-defined confidence threshold. Filtered keypoint pairs are then used to calculate the transformation matrix between the two images. Since the hyperspectral camera is mounted perpendicular to the astronomical horizon during flight, and the images are taken with the hover and capture mode, first the algorithm tries to solve for the affine transformation. However, if it fails due to perspective distortions due to unstable flight conditions such as windy conditions, homography transformation is calculated.

The stitching process was carried out using a computer with an AMD Ryzen 9 3900X 12-Core Processor, 64Gb RAM, and an Nvidia RTX 3090 graphics card. GPU was only used in the process of running the feature-matching algorithms.

The calculated transformation matrix is then used to calculate the corners of the output image which is then used to calculate the minimum and maximum x and y coordinates of the output image. The translation matrix is then defined using the minimum

x, and y values, and each channel of the second image was warped using this translation matrix. If the calculation of the affine transformation was successful each channel of the first image was also warped using the translation matrix. Each channel of the translated image was then again warped using the affine transformation matrix. If the affine transformation was unsuccessful, the translation matrix was multiplied by the homography matrix to obtain the full transformation matrix. Which was used to warp the image one to the new image plane. Finely each channel of the warped images was blended separately using a feature blend and stacked to obtain the final stitched image cube with all the channels.

4.3.1 Selection of initial Hyperspectral Band for Stitching

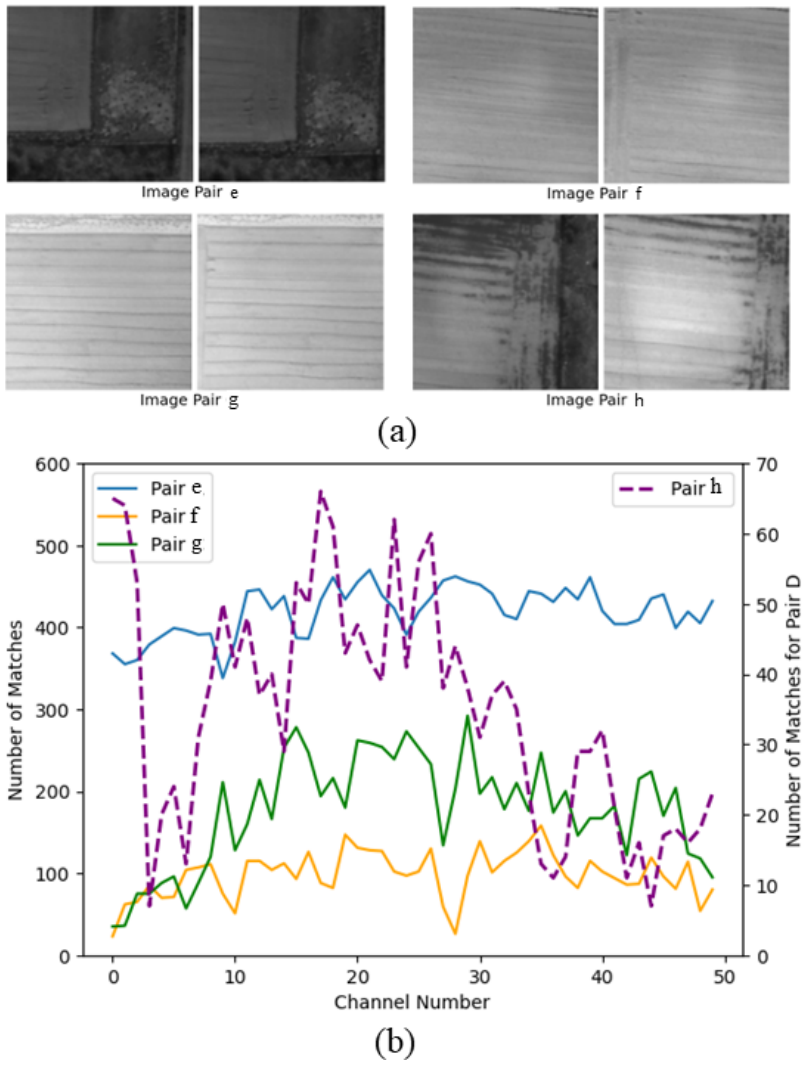


Figure 4.4: (a) Images compared (b) Matches obtained using LoFTR

One of the key factors that significantly impact the accuracy and quality of hyper-

spectral image stitching is the selection of the correct band/bands for feature matching. Hyperspectral imagery typically consists of hundreds of spectral bands, each capturing information about the reflectance or radiance values of the scene at different wavelengths. These bands can be sensitive to different physical properties of the objects in the scene, such as their material composition, vegetation density, and moisture content. Hence, the choice of bands used for feature matching can significantly affect the accuracy and reliability of the stitching process.

The selection of bands for feature matching is crucial because it directly affects the ability to find corresponding features or points in overlapping images. The choice of bands for feature matching should consider several factors. Firstly, the selected bands should have distinctive and stable features that can be accurately detected and matched across different images. Bands with low contrast or those that are affected by atmospheric conditions, illumination changes, or noise may not provide reliable and consistent feature matches. Secondly, the selected bands should capture relevant information about the scene or objects of interest. For example, in agricultural applications, bands sensitive to vegetation properties may be crucial for accurate stitching of images showing crop fields.

Figure 4.4 presents an experiment with the previously cited image pairs of e, f, g, and h. LoFTR algorithm was run on each image pair for each channel. Figure 4.4 (b) plot shows the resultant graph. The number of features of each channel is the sum of features that had more than 0.9 confidence level from the LoFTR algorithm. In most of the cases, LoFTR manages to find features more than 100 features on average where image pairs e, f, and g had mean feature identifications of 419, 105, and 178 respectively. Where the minimum number of features was identified in the bands of 522nm, 450nm, and 450nm respectively and the maximum number of features was identified in the bands of 618nm, 730nm, and 682nm for e, f, and g image pairs. On image pair h on the other hand, matching algorithms showed a significant reduction in producing matches higher than the 0.9 confidence threshold. The mean value of feature identification was 36 where the maximum value of 82 at the 586nm band and minimum feature identification of 7 in the 802nm band was observed.

The reason for this change is the contrast level of each band changes depending on the composition of the image contents. Hence it is important to select a suitable hyperspectral band having a high contrast value depending on the specific application area. In the agriculture domain, from the experiments, it was identified that the spectral ranges from 650nm to 750nm produce high-contrast images because of the contrast between vegetation and the ground. However, foreign objects like metal surfaces and building surfaces can change this contrast. To tackle this problem when any one of the pairs failed to identify enough features for the stitching, other bands were scanned to verify whether any high-contrast bands were producing enough features above the 0.9 confidence value. In cases

where even other bands did not produce enough features, the confidence threshold value was reduced to 0.7 in order to increase the number of features.

4.4 Results and Discussion

4.4.1 Introduction

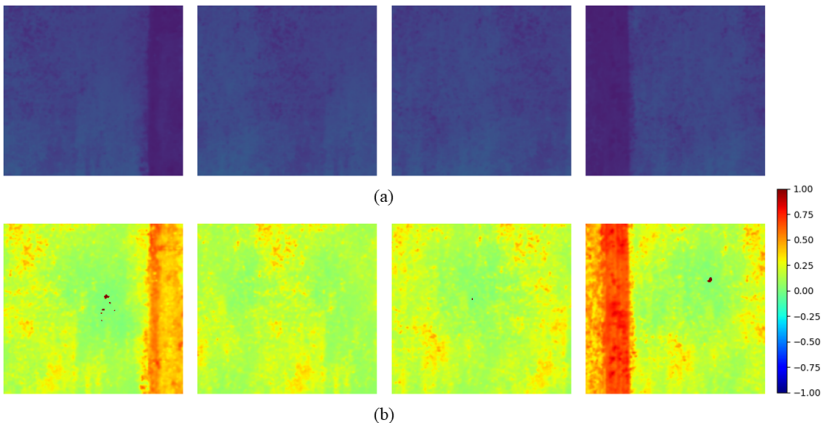


Figure 4.5: Sample set from the image set selected for case study 1

The process of stitching line segments to obtain a final result is complex and involves multiple sub-processes. To provide a comprehensive and easy-to-understand overview of the process, the results are presented through several case studies in this section. Each case study focuses on a specific data acquisition session, providing an introduction to the data, presenting intermediate results, and discussing the outcomes in terms of both positive and negative aspects at each step.

By adopting a case study approach, the thesis aims to provide a detailed and in-depth analysis of the stitching process for different scenarios. This approach allows for a thorough examination of the methodology's performance under various conditions, such as different environmental settings, target characteristics, and data acquisition sessions. It also enables the identification of potential strengths and limitations of the methodology in different contexts.

In each case study, the data acquisition session is carefully selected to represent a specific scenario, and the data used for stitching is introduced, including details on the initial GPS points captured and the segmentation of line segments. Intermediate results, such as images or visual representations, are presented to provide a clear understanding of the progress of the stitching process. These intermediate results serve as a basis for evaluating the performance of the methodology at each step of the process.

Furthermore, the case studies provide a comprehensive analysis of the outcomes in

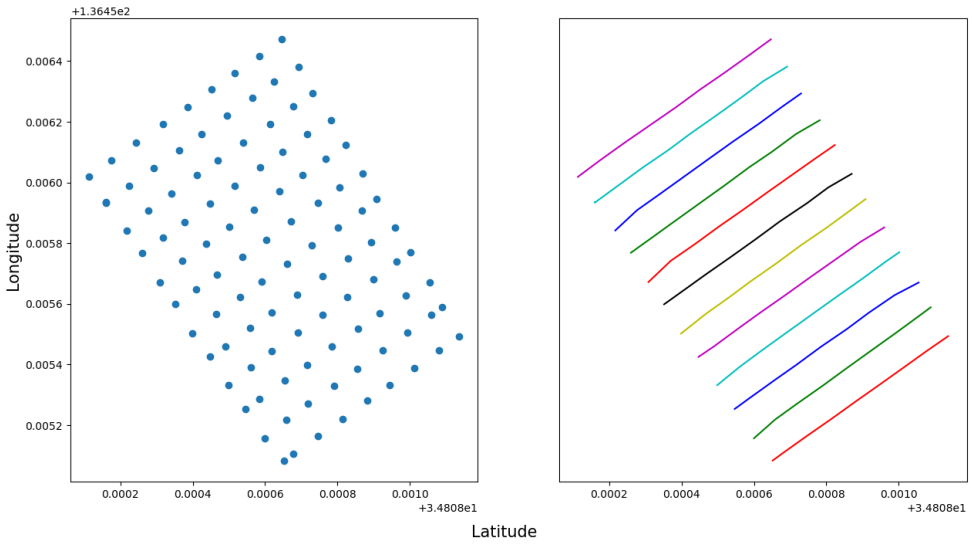


Figure 4.6: Case study 1 (a) GPS points of the selected image set (b) Segmented lines

terms of positives and negatives. This allows for a critical evaluation of the methodology's performance, highlighting the strengths and advantages, as well as the limitations and challenges encountered during the process. This analysis provides valuable insights for further refinement of the methodology and for identifying potential areas for improvement in future applications.

Overall, the case study approach in presenting the results of the stitching process allows for a detailed and comprehensive assessment of the methodology's performance, providing a clear understanding of the outcomes at each step of the process. The analysis of positives and negatives in each case study contributes to a holistic evaluation of the methodology's effectiveness and limitations and forms the basis for recommendations and further research in the field of line segment stitching for image stitching using drone-based low-resolution hyperspectral data.

4.4.2 Case Study 1

The first dataset selected was obtained from the same location described in previous sections (Mie University Farm) on a clear day without cloud cover. At the start of the data acquisition session Altitude angle: the angle between the horizon and the center of the sun was 56.69° and the Azimuth angle: the angle between the meridional plane of the earth and the vertical plane of the sun was 138.91°. At the end of the data acquisition session, the altitude angle was 60.03° and the azimuth angle was 150.53°. During this time solar irradiance was changed from 1002 W/m^2 at the start to 1012 W/m^2 at the end of the session.

UAV was flown at a height of 120 meters. At this height, each image had a spatial

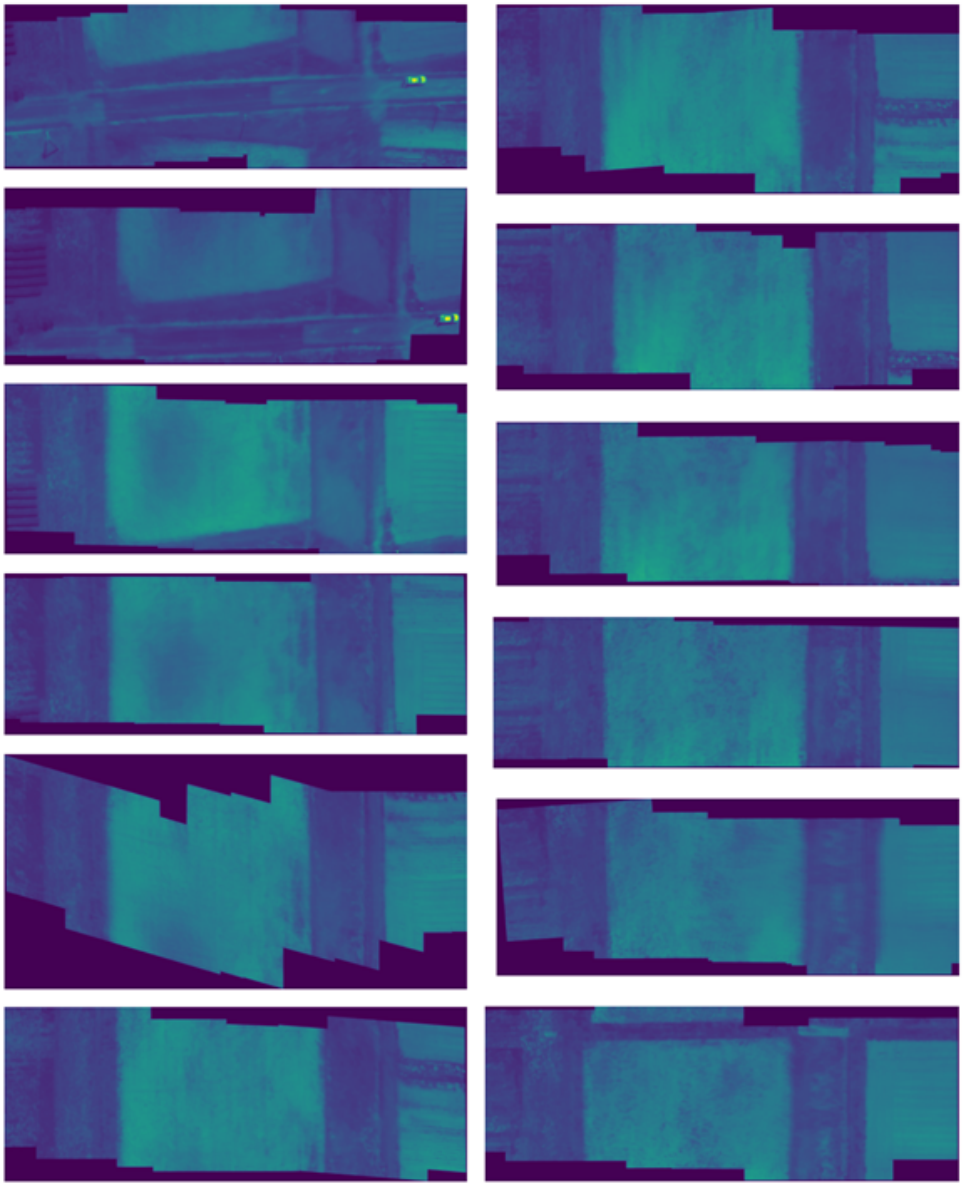


Figure 4.7: Intermediate results after stitching each line segment using band 29

resolution of 13.1 cm/px. A total of 110 images were captured to cover the desired plantation area. The whole of the selected farming plot was already sown in order to prepare the soil for the next season's cultivation. Because of this, the contrast level of most images was homogeneous. Figure 4.5 (a) shows a sample of images selected for this case study. Figure 4.5 (b) shows the NDVI values of the selected images. NDVI or Normalized Difference Vegetation Index is used to interpret vegetation health and vigor. NDVI is calculated using the reflectance values of near-infrared (NIR) and red bands of the image. NDVI index is represented in the range from -1 to 1 where higher values indicate healthier vegetation and lower values indicate less healthy or sparse vegetation. In this case, values close to 0 mean little or no vegetation was present in the sown plot.

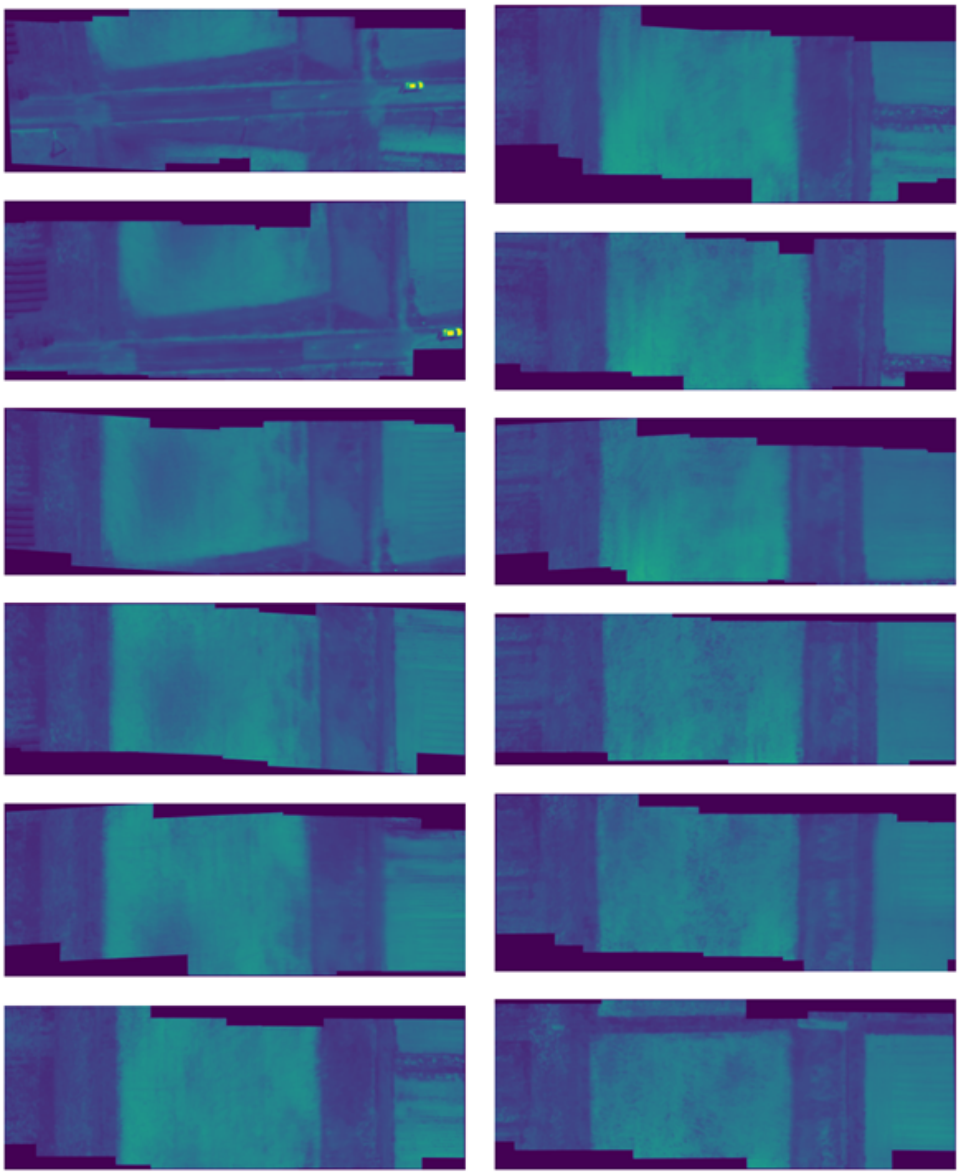


Figure 4.8: Intermediate results after stitching each line segment using band 35

However, the surrounding area of the selected plot contained a healthy weed population.

Figure 4.6 demonstrates the initial GPS points captured, and in (b), segmented lines from the GPS points are represented in different colors. As mentioned previously these segments are then stitched separately using parallel processing. The initial feature selection band was selected as 29 which captures the information from 682 nm. Total line segments were stitched in under 11 minutes. Figure 4.7 presents the intermediate line stitching obtained from the aforementioned methodology. The image sequence is represented starting from the top left following downward and shifting to the right side from the seventh line segment. All of the images were stitched as intended except for the fifth line segment. Further investigation is discussed in a later section about the particular

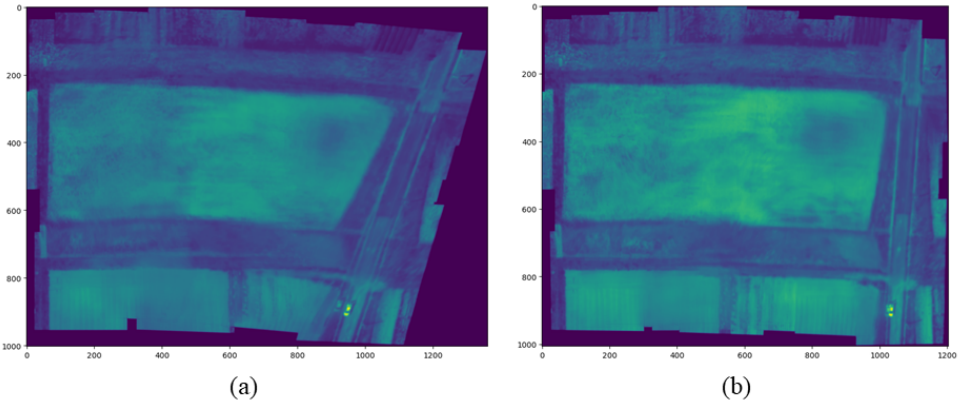


Figure 4.9: Stitched image of case study 1 using the band spectral 682nm

cause of the issue and ways to mitigate the issue.

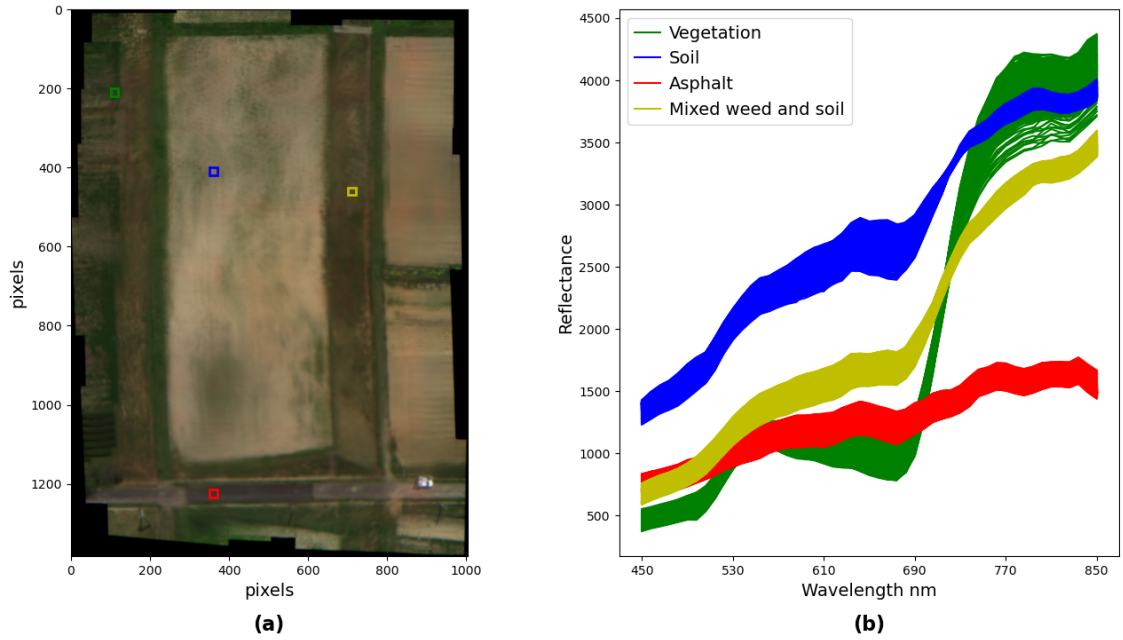


Figure 4.10: Spectral analysis of the stitched image case study 1. (a) Stitched image and the selected patches (b) spectral signatures associated with each bounding box

Figure 4.9 (a) depicts one channel image of the final stitched image using channel no 29 as the initial feature detection band. As evident from the result, stitching quality is poor. Geometrical lines have shifted by a considerable margin and severe ghosting effects can be visible in different areas. Figure 4.8 presents the intermediate results for the same dataset by initial feature identification with the 35th band or the 730nm band of the image. From the visual inspection, some intermediate results are considerably improved over the band 29th intermediate results. Further Figure 4.9 (b) depicts the final stitching

result using band 35 as the initial band selection. It is also an improvement over the stitch done using band 29 in terms of geometrical consistency and ghosting effects. This raises the need to improve the accuracy of feature matches.

Figure 4.10 presents a spectral plot that sheds insight into the spectral consistency of stitched images. The first plot (a) of the figure indicates the stitched image in RGB format. To evaluate the spectral consistency of the stitched image, a set of bounding boxes from the image that represents several targets of hyperspectral images was selected. These bounding boxes were set to a size of 20 pixels by 20 pixels.

From each pixel within the bounding boxes, hyperspectral signatures were extracted and plotted, as shown in Figure 4.10 (b). The selected targets for this analysis were vegetation, soil, asphalt, and a mixture of soil and weed. Results indicate that the spectral signatures of each pixel within the bounding box are consistent and can be differentiated, similar to the signatures discussed in Chapter 3. This finding suggests that stitching images do not significantly impact their spectral consistency, which is an important consideration in remote sensing applications.

4.4.3 Case Study 2

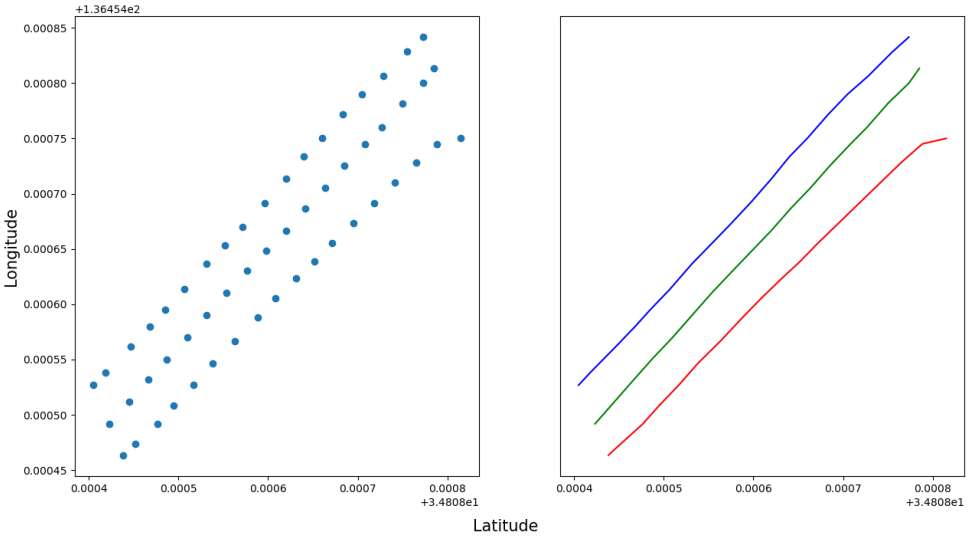
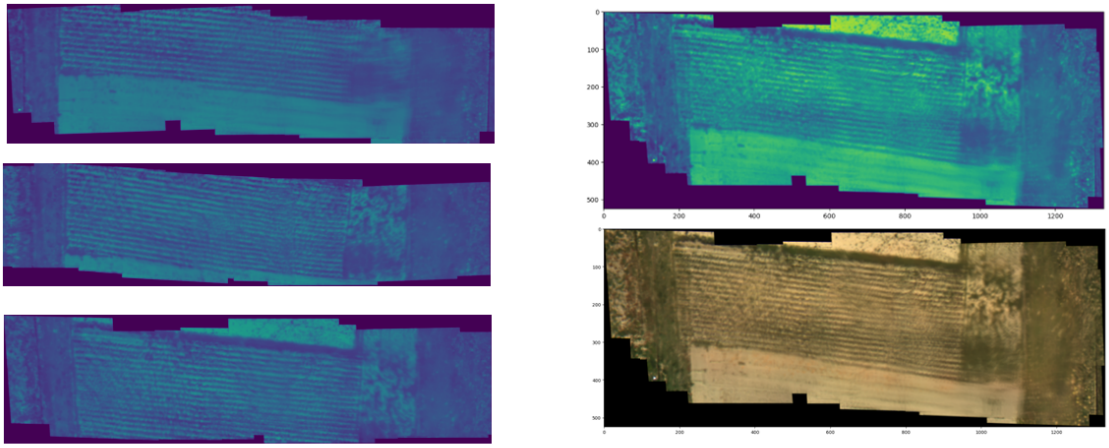


Figure 4.11: Case study 1 (a) GPS points of the selected image set (b) Segmented lines



(a) (b)

Figure 4.12: Case study 2 (a) Intermediate results of the line segment stitching. (b) Stitched image in monochrome and RGB representations

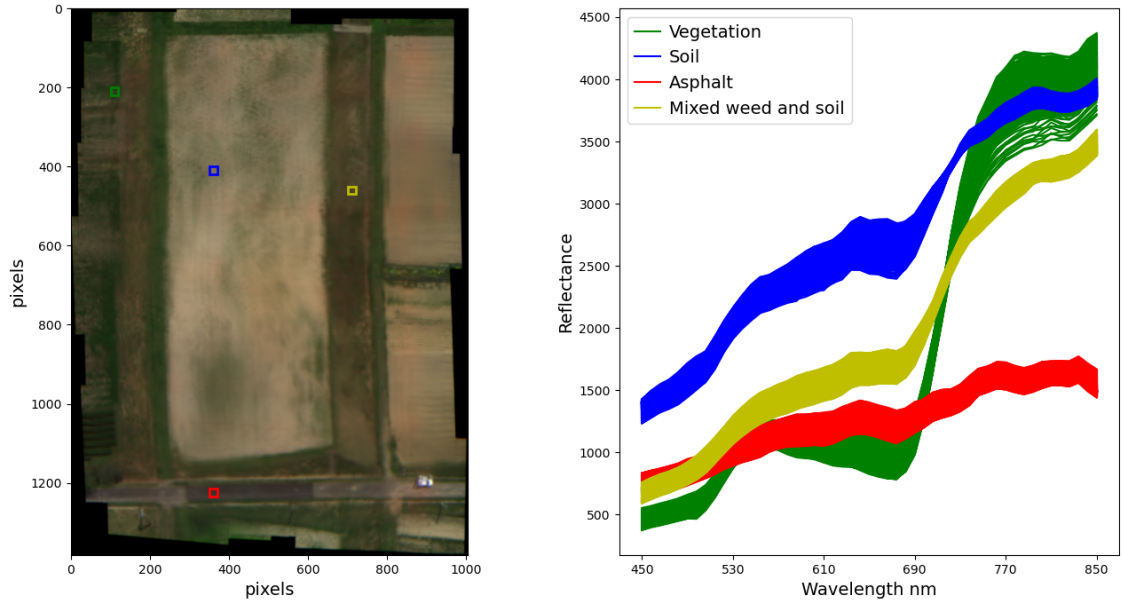


Figure 4.13: Spectral analysis of the stitched image case study 2. (a) Stitched image and the selected patches (b) spectral signatures associated with each bounding box

A second case study was carried out on an image set where the acquisition height was much lower than in the previous study. This session was carried out at an altitude of 50 meters from the ground and at that altitude, it had a spatial resolution of 5.5 cm/px. Which is more than double the spatial resolution of case study one. This session only

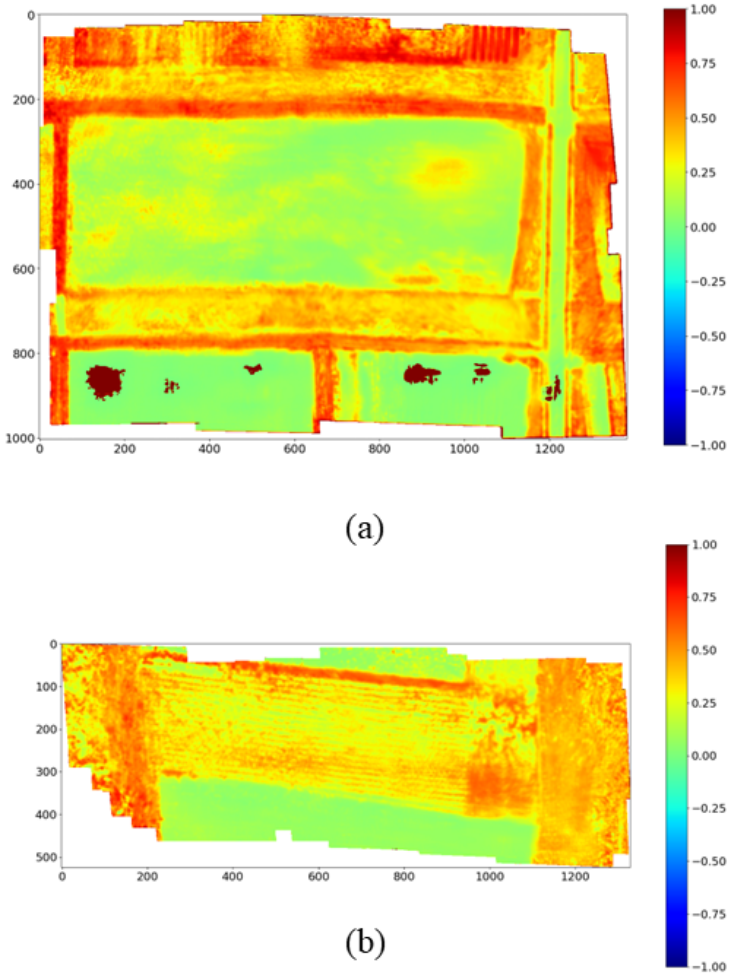


Figure 4.14: NDVI plots (a) Stiched image of case study 1. (b) Stiched image of case study 2

contained 54 images. Figure 4.11 plots the GPS pots of this session and the corresponding line segments obtained. Apart from the flight height, every other flight parameter was kept the same as in case study 1. It can be observed that the GPS points are plotted a lot closer than the case study one due to the drone flight height being much lower than the study one.

Figure 4.12 (a) presents the intermediate results of the stitched line segments and (b) presents the final stitched image in monochrome and RGB format. It can be observed that the proposed stitching process succeeded in producing a stitched image with some geometrical imperfections.

In the previous spectral analysis, three locations were chosen from the stitched image. These locations included plots containing soil and onions, soil alone, and plots containing

weeds. The resulting plots are shown in Figure 4.13. In (a), the RGB image displays the selected samples with their bounding boxes, while in (b), the spectral signatures within each bounding box are shown.

Similar to the findings in Case Study 1, the spectral signatures obtained from the selected plots in Case Study 2 were consistent within their respective bounding boxes. However, in cases where multiple materials are present in a single pixel, such as in mixed pixel situations, spectral unmixing might be necessary to separate the spectral components of each material in that pixel.

4.5 Chapter Conclusion

In conclusion, the initial stitching process of hyperspectral images obtained via a UAV, as described in this section, has shown successful results that can be utilized in various decision-making processes in agriculture and other fields. The generation of NDVI maps, growth estimation, crop monitoring, precision agriculture, crop stress detection, yield prediction, crop phenotyping, and field boundary mapping are some of the potential end uses of stitched hyperspectral mosaics.

However, it is important to note that the selection of the initial feature detection band is critical for achieving geometrically consistent stitching. Some regions of the spectrum may pose challenges for the LoFTR algorithm in identifying matches accurately, resulting in low-accuracy transformation matrices and poor performance of the proposed method. This limitation needs to be carefully considered when applying the proposed method in practical applications.

Further research and experimentation are needed to refine the method and overcome this limitation for more reliable and accurate stitching results. Despite this limitation, the proposed method represents a significant advancement compared to previous methods, where the entire process failed at the feature identification step. The successful results achieved so far using the proposed method offer a promising approach for stitching hyperspectral images from UAVs and have the potential to enhance the analysis of remote sensing data for various applications in agriculture, environmental monitoring, and other fields. Such an example of a possible use case is depicted in the FIgure 4.14 where NDVI images were generated using the stitched images described in case study 1 and case study 2.

In summary, while the proposed method has limitations related to the selection of feature detection bands, it presents a promising approach for stitching hyperspectral images obtained via UAVs, and further research and improvement can contribute to its wider application in diverse remote sensing applications, benefiting agriculture and other domains.

Chapter 5

Enhancing Feature Detection and Matching in Low-Pixel-Resolution Hyperspectral Images

5.1 Introduction

Although chapter 4 proposed the mosaic generation pipeline based on the LoFTR and SuperGlue feature detection and matching methods, several limitations of the current state-of-the-art methods were identified. Feature detection was not performing adequately on some image pairs even with the state-of-the-art methods. One main drawback of these methods is that they are reliant on a single band of the hyperspectral image. None of the state-of-the-art methods are designed to be used with hyperspectral images. As explored in the previous chapter, these algorithms take an image of one spectral band selected by the user and use it to detect and match features between adjacent images. Hence it was hypothesised that if an algorithm was proposed to utilize all the information available in all of the hyperspectral bands feature detection could be improved further. The use of all the available bands in a hyperspectral image allows the algorithms to better make predictions on correct and incorrect features. Based on this, a 3D Convolution Neural Network-based Siamese network was proposed where an initial feature match set was generated using edge maps and then the Siamese network was used to filter out the non-matches from the matches. This chapter of the thesis describes the design, training, and evaluation steps of the proposed model.

5.2 Proposed Model

Figure 5.1 provides an outline of the proposed approach, which involves a sequence of four main steps for generating a collection of highly accurate filter matches from a

given pair of images I_A and I_B. The initial step selects a spectral band, extracts edges from each image pair, and creates a binary edge map using Phase Stretch Transformation (E_A and E_B). Subsequently, SIFT feature detection is applied to each edge map, and a brute-force matcher is employed to derive a set of matches from the two edge map images. Following this, 32 by 32 patches are extracted around the identified matched keypoints, capturing all the spectral band information from each selected patch (P_{A1}, P_{A2}, ... P_{Ai} and P_{B1}, P_{B2}, ... P_{Bi}). In the concluding step, a 3D convolution-based Siamese network is employed to filter out inaccurate feature matches from the identified list of feature matches. The subsequent sub-sections in the methodology section will delve into the specifics of each of these processes.

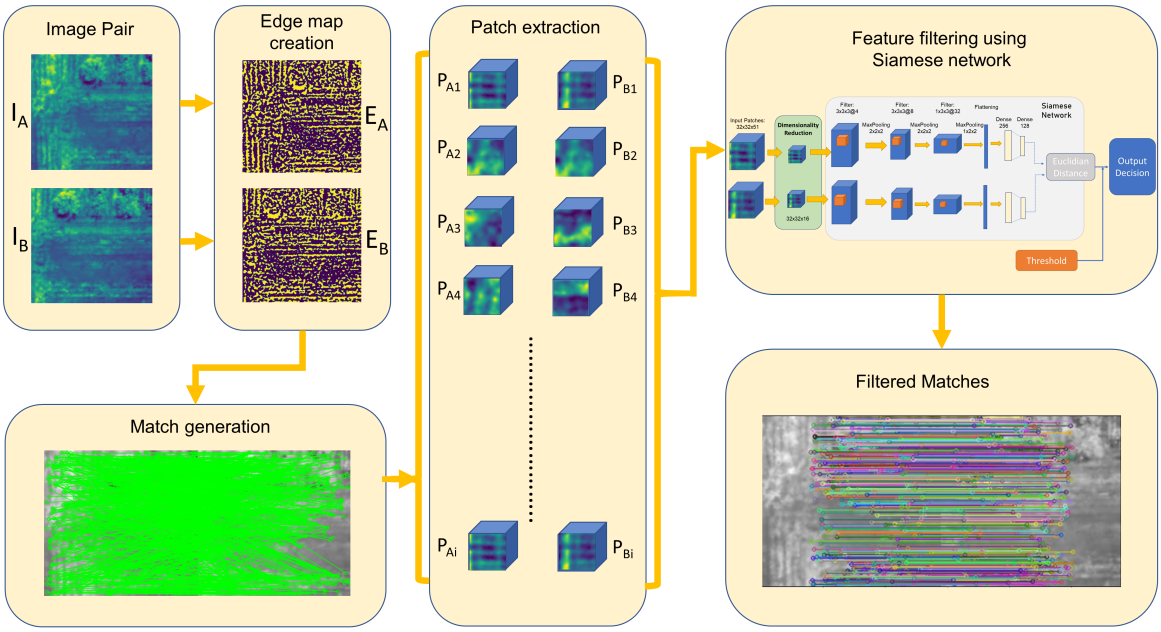


Figure 5.1: Overview of the proposed method

5.3 Data acquisition and preprocessing

All hyperspectral images utilized in this experiment were captured using the Cubert Ultris 5 hyperspectral camera [8]. This camera boasts a resolution of 290 pixels by 275 pixels and comprises 51 spectral bands sampled across the 450 nm to 850 nm range, with an 8nm spectral resolution. Its compact dimensions of 29mm by 29mm by 40mm and a weight of 120 grams make it an ideal solution for mounting on a variety of UAVs. The complete acquisition system, including the provided single-board computer, triggering circuit, and wiring-mounting hardware, weighed under 800g. The DJI M600 Pro drone, along with its accompanying SDK, was employed for data acquisition sessions, and captured irradiance images were converted to radiance. Figure 5.2 illustrates the data

acquisition platform. Additional details about the setup and the initial data conversion process can be found in the initial chapter 3.

Image acquisition was carried out using the DJI Ground Station Pro software, employing the "hover and shoot" mode to trigger the hyperspectral camera. An 80% front and side overlap was chosen between images to compensate for the accuracy limitations of traditional GPS systems. This deliberate overlap ensures a substantial common area between consecutive images, facilitating the detection of feature matches. The selected fields for data acquisition encompassed various scenes, including freshly plowed soil, a mix of weed and soil, watered paddy fields, asphalt roads, fully grown weed, and orange plants. This diverse selection aimed to showcase feature detection challenges encountered in agricultural plots, particularly scenarios involving early-stage vegetation where a significant portion of soil is visible. Furthermore, selecting a wide variety of targets will aid the model's robustness.

Figure 5.2: DJI M600 pro drone with the camera mounted

5.4 PST-based feature match generation

Training the Siamese network to discern matches from non-matches, based on a set of potential feature matches, constitutes the primary methodology employed in this paper to acquire a robust set of filter matches. The basic approach involves selecting a pixel region from the first image and obtaining all possible combinations from the second image with the same region size, iterating until the entire first image area is covered. However, this method is computationally demanding. An alternative strategy involves transforming the image into a different domain and extracting features from it. After exploring various approaches such as hyperspectral band fusion and image enhancement, it was determined that generating edge maps from hyperspectral images and subsequently identifying features within these maps yielded a sufficiently accurate dataset for filtration by the Siamese network.

The Phase Stretch Transformation (PST) emerges as a noteworthy edge and texture

detection method, offering commendable performance on visually impaired images with computational efficiency. Initially introduced by Asghari et al [71, 72] based on the concept of photonic time stretch [73], PST was chosen as the preferred edge map creator over other popular algorithms like Canny edge detection [74], Sobel edge detection [75], and Laplacian of Gaussian (LoG) edge detector [76], as indicated by the evaluation results presented in Zhou et al.'s work [77]. The preference for PST was further motivated by the unavailability of trained algorithms, discrepancies in data size, and computational burden associated with learning-based edge detection methods.

Initially, one band from each hyperspectral image underwent PST edge detection using the Python library [77]. The PST algorithm, with user-tunable parameters including phase strength, warp strength, Gaussian low-pass filter standard deviation (sigma), lower and upper thresholds, and a boolean for binary/analog edge detection, was applied with phase and warp strengths set at 0.2 and 50, and a sigma of 0.15 for the low pass filter. The threshold range spanned 0 to 0.65, producing a binary mask as the resultant edge map. Subsequently, the converted edge image underwent SIFT feature detection and knn matching using OpenCV, yielding a list of matches for the image pair. The sensitivity of the SIFT detector was enhanced by adjusting contrast threshold, sigma, and octaveLayers parameters, with specific values set for this experiment.

By tuning the SIFT detector's sensitivity, more than 10,000 feature matches were generated, providing substantial data for the Siamese network. Notably, the spectral band selection for edge detection played a crucial role in this step. Although feature matches remained high across various selected bands, opting for spectral bands within the 600nm to 800nm range produced considerably higher set of accurate matches. Further details on specific outcomes will be presented in the results section. Figure 5.3 (a) illustrates the original images, (b) showcases the edge map created via the PST method, and (c) depicts 100 randomly selected feature matches from the designated image pair. Following this, 32x32-pixel size patches were extracted, centered around the matched keypoints, each containing spectral information. This process resulted in two 32x32x51 image patches for every matched keypoint in the first and second images.

5.5 Dimentionality reduction using Autoencorder

Employing a neural network to train on the 32x32 pixel patches extracted with the 51 spectral bands is not advisable due to challenges associated with high dimensionality, notably the curse of dimensionality [78]. Consequently, this research implements a straightforward autoencoder-based dimensionality reduction network to address these challenges, reducing the hyperspectral image dimensions from 51 bands to 16 bands. The choice of an autoencoder-based approach stems from considerations of execution time.

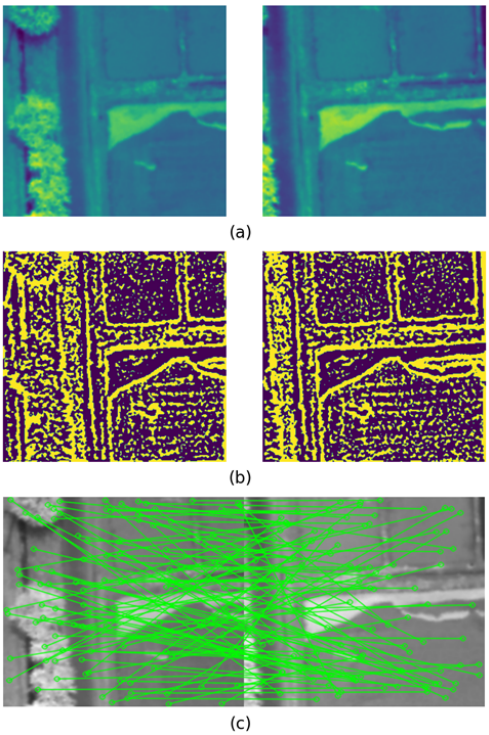


Figure 5.3: Intermediate results of feature match generation method. (a) selected image pair from the 750 nm band. (b) PST edge map. (c). Ransom sample 100 detected matches.

While training the autoencoder incurs computational expenses, the subsequent inference of the trained model proves significantly faster than traditional methods like Principal Component Analysis.

The dimensionality reduction is accomplished through a feed-forward neural network-based encoder and decoder, each comprising three hidden layers. Rather than representing the entire patch in a lower-dimensional space, a 1D autoencoder is trained to reduce the dimensions of each spectral signature at each pixel. Consequently, the input for the autoencoder is of size (1, 51). Each hidden layer is followed by a batch normalizing layer and a leaky ReLU layer. Post-training, the encoder portion of the autoencoder is utilized to reduce the 51 dimensions to 16 at each pixel within the 32x32 patches. Regenerated pixel spectrums from the autoencoder exhibit a high correlation with the original data, affirming the effectiveness of the autoencoder approach.

5.6 Feature filtering using a 3D convolution Siamese Network

The feature matches produced encompass two subsets: correct matches and incorrect matches. The inclusion of the Siamese network serves the purpose of discriminating

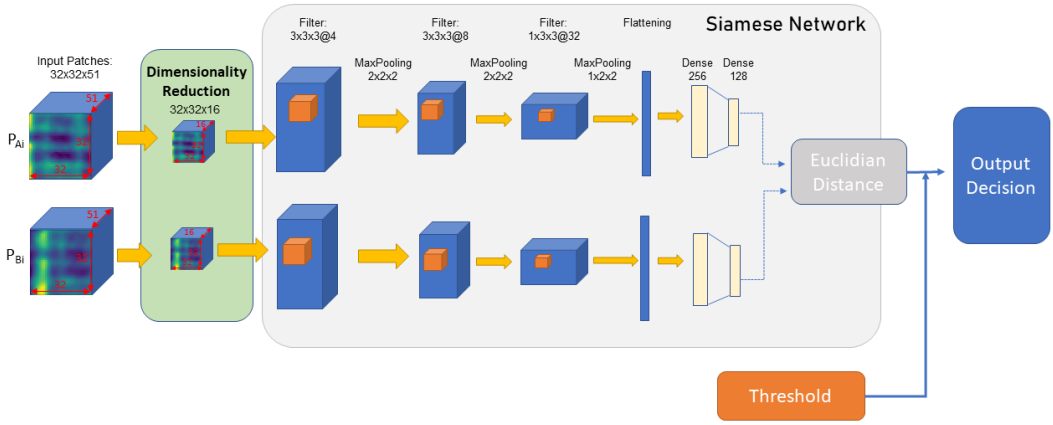


Figure 5.4: Proposed network architecture

between matches and non-matches. To leverage both spectral and spatial information in the filtering process, this research introduces a 3D convolution-based Siamese network. Figure 5.4 illustrates the architecture of the proposed Siamese model. The network undergoes training on both matched and non-matched features, with each patch having an input size of 32x32x16.

5.6.1 Network architecture

The architecture comprises three 3D convolution layers, where the first two layers use 3D convolution kernel sizes of 3x3x3, and each subsequent layer utilizes a 1x3x3 kernel. Following each convolution layer, batch normalization is applied, and activation is achieved through a LeakyReLU function with a negative slope set at 0.2. Following the activation layers, three max-pooling layers are employed, with sizes of 2x2x2 for the initial two layers and 1x2x2 for the final max-pooling layer. The output is then flattened and fed into a dense network featuring two layers with sizes of 256 and 128. Ultimately, the Euclidean distance is computed from the output of each network.

To enhance the distance between output embedding for non-matched images and reduce the distance for matched images, a contrastive loss function is employed during the training process. This loss function penalizes the network when it predicts that two inputs are the same, and vice versa. The proposed network architecture was determined after experimenting with various shapes and sizes, considering the accuracy it yielded.

The contrastive loss function is computed as follows. Let z_1 and z_2 denote embeddings of two patches in an embedding space, y represent the binary label indicating whether the pair of patches is a match or not, and $margin$ be the hyperparameter defining the minimum desired separation between similar and dissimilar points in the embedding space. The calculation of the contrastive loss function can be expressed as follows.

First, compute the Euclidean distance (L2 distance) between the two embeddings by calculating the element-wise difference between z_1 and z_2. Then, calculate the squared L2 distance for each pair of embeddings. Finally, obtain the Euclidean distance by taking the square root of the squared distances, as represented in equations 5.1 and 5.2.

$$difference = z_1 - z_2 \tag{5.1}$$

$$Euclidean\_Distance = \sqrt{\sum_{i=1}^{n}(\text{difference}_i)^2} \tag{5.2}$$

Then calculate the difference between the margin and the calculated distance. Clamp the negative values to ensure that negative values are set to zero, indicated in equation 5.3. This enforces that dissimilar points are only penalized if they are closer than the desired margin. Finally, calculate the loss for the two patches by the equation 5.4.

$$negative\_distance = min(max((margin - Euclidean\_Distance), 0.0), max\_value_i) \tag{5.3}$$

$$\text{loss} = \frac{y \cdot \text{distance\_squared} + (1 - y) \cdot \text{negative\_distance}^2}{2.0} \tag{5.4}$$

The training process utilized a dataset generated by traversing a 32x32 pixel window across an image. As depicted in Figure 5.5 (a), dataset matches were formed by replicating the same image patch while shifting the P_r window in 5-pixel strides. This 32x32 window was then systematically moved to cover the entire image, and duplicated patches were preserved as matches. For non-matches, the P_r window remained stationary while the P_y window moved in 5-pixel strides to encompass the entire image. Each non-match comprised patches P_r and P_y. Once the P_y window covered the entire image, the P_r window advanced in 5-pixel strides, and the motion of P_y was repeated to avoid duplications. This combination of movements resulted in 2548 matches and 1685502 non-matches from a single image. However, non-matches were randomly selected, and only 5000 non-matches were retained from one image. Three hundred hyperspectral images were chosen from distinct acquisition sessions to generate the data samples for training and validation.

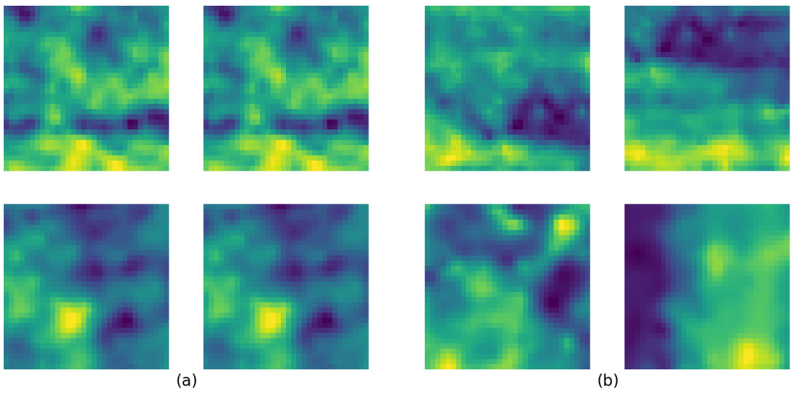


Figure 5.6: Four samples of created dataset. (a) Matches. (b) Non Matches.

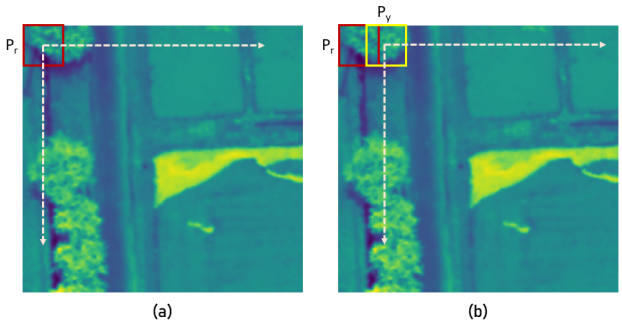


Figure 5.5: Training Data creation (a) Matches creation (b) Non-matches creation

5.6.2 Training dataset creation and training process

From the established dataset, 7,000,000 matches and non-matches were randomly chosen for each class, contributing to the training, validation, and testing phases. Figure 5.6 (a) illustrates two matched samples from the dataset, while (b) displays non-matched samples. The selected data was distributed into training, validation, and test sets with a ratio of 0.8, 0.1, 0.1. Model training concluded at epoch 18, ceasing when five consecutive validation losses failed to exhibit improvement. The model was trained with a learning rate of 0.0001 using Adam optimization [79]. Additionally, dynamic mini-batch sizing was implemented during training, progressively increasing the batch size from 256 to 2048 over the initial 5 epochs. This strategy, known to offer advantages such as swift convergence, a regularization effect, and efficient memory and computational resource utilization in large-scale models [80, 81], was employed. Despite the proposed model having a relatively small number of tunable parameters, specifically 202,568, compared to some larger models, dynamic sizing contributed to the model's rapid convergence. Figure 5.7 (a) presents the training and validation loss values for the model.

Following the training process, the inference for predictions involved calculating the

same Euclidean distance and thresholding the distance to derive the prediction. Figure 5.7 (b) showcases the Matthews Correlation Coefficient (MCC) curve for the test dataset. The MCC is a performance metric for binary classification tasks, factoring in true positives, true negatives, false positives, and false negatives. The resulting score falls within the range of [-1, 1], where 1 signifies a perfect prediction, 0 represents a random prediction, and -1 denotes a complete disagreement between prediction and observation. The MCC curve is obtained by varying the threshold value and calculating the corresponding MCC values. The x-axis in Figure 5.7 (b) represents threshold values ranging from 0 to 3, while the y-axis presents the MCC values obtained at each threshold. The curve indicates that threshold values below 0.9 yielded nearly perfect results for the test set. However, when applying the trained model to adjacent image pairs, thresholds above 0.6 introduced a significant number of incorrect matches. This discrepancy arises from slight variations between the trained data and real data. While the trained data comprised identical matches from the same images, applying the trained algorithm to adjacent images introduced slight illumination and perspective changes due to differences in image acquisition locations. The method employed for dataset generation eliminates the labor-intensive manual labeling required when generating a dataset with adjacent images. It is believed that the features learned from the dataset created with the proposed method effectively imparted relevant information to CNN, a claim that will be substantiated in the results section.

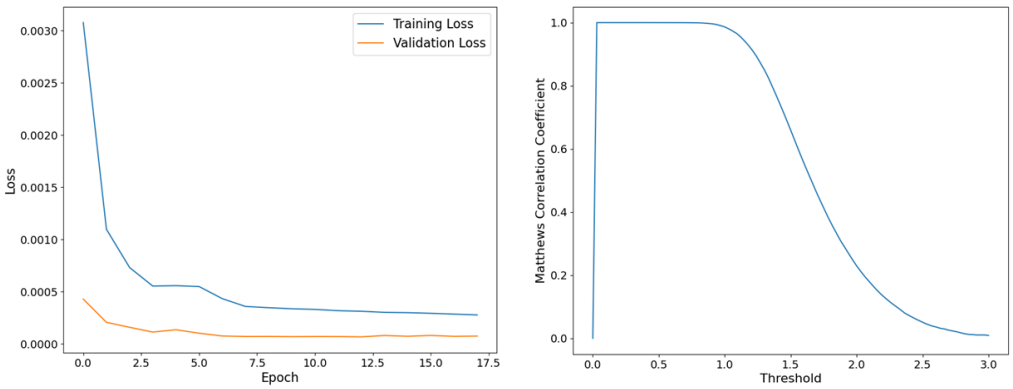


Figure 5.7: (a) Training and validation loss curve (b) Matthews Correlation Coefficient curve

5.7 Evaluation Procedure

To assess the proposed method, several hyperspectral image pairs were chosen to showcase both the superior performance of the proposed model and instances where it

does not perform as well. The 762nm (38^{th}) spectral band was selected from each image pair, and the edge image was generated as detailed in Section 5.4. Subsequently, a set of feature matches was acquired using SIFT detectors as outlined in the same section. Following this, 32x32 patches were extracted in pairs for each match, and dimensions were reduced from 51 to 16 using the encoder. Each pair of patches was then processed through the Siamese network to obtain the Euclidean distance. A threshold value was chosen for binary classification. The filtered feature matches set was then employed to evaluate the proposed method against both traditional and state-of-the-art methods.

As evaluation metrics for method comparison, an inlier percentage was determined by computing homography with RANSAC. Additionally, a common region was extracted from the two overlapping regions after the homography transformation. Subsequently, the Structural Similarity Index Metric (SSIM) [82] and the correlation coefficient were calculated for the two regions. SSIM involves three sub-components: luminance comparison, contrast comparison, and structural comparison, making it indicative of the accuracy of identified feature matches. On the other hand, the correlation coefficient measures the linear relationship between image pixels in the two regions, with values ranging from -1 to 1. A value of -1 indicates a perfect negative linear correlation, while 1 denotes a perfect positive correlation. However, it does not account for the perceptual qualities of the two regions.

To compare the proposed method, the previously mentioned LoFTR, and LightGlue methods were selected. LightGlue was employed in conjunction with the DISK feature detector [83] during the experiments, leveraging the superior performance observed with this combination of feature detector and matcher. Additionally, a combination of Good Features to Track (GFTT) detector [84] + OpenGlue [85], and a combination of KeyNet keypoint detector [86] + OpenGlue were utilized as state-of-the-art methods for performance comparison with the proposed method. The second nearest neighbor matcher (SNN) was employed to filter out non-matches from the matches in each of the two aforementioned detectors and descriptors. OpenGlue employs Convolutional Neural Networks to generate descriptors for the keypoints identified using GFTT and KeyNet. These methods were selected after an initial evaluation and demonstrated comparable results with the LoFTR and LightGlue algorithms.

5.8 Results and discussion

The presentation and subsequent discussion of the results will be presented in two distinct subsections. The first section will discuss the utilization of spectral bands in PST-based feature match generation, examining the influence of selecting a specific band on the performance of the proposed method. Additionally, a discussion will be under-

taken to understand how the choice of a particular spectral band affects the accuracy and reliability of the feature matches generated by the proposed approach. Subsequently, the second subsection will center on the performance evaluation of the proposed method. This evaluation will gauge its effectiveness using carefully selected image pairs derived from diverse datasets acquired in varied locations. This assessment aims to showcase the method's robustness across different scenarios and datasets, offering a comprehensive perspective on its capabilities and potential limitations. The insights and findings derived from these analyses will contribute to the understanding of the proposed method's efficacy and its practical applicability in real-world scenarios involving hyperspectral image matching.

5.8.1 Selection of the spectral band

As mentioned in section 5.4 it was identified that the performance in the 600 to 800nm range, specifically in band 18 to band 42, was comparatively better than in the other spectral bands. Figure 5.8 presents graphs obtained in order to evaluate this statement. Ten image pairs were randomly selected, and the feature matching was carried out for each band of the hyperspectral image pair. Figure 5.8 (a) presents all the matches obtained for each image pair at each spectral band while figure 5.8 (b) presents the filtered features obtained from the proposed Siamese model. Derived feature matches were then used to calculate the homography between the two images at each band and the inlier ratio and the mean reprojection ratio were calculated for each band. Figures 5.8 (c) and (d) present the respective plots. It could be highlighted that the high number of matches identified in the initial few bands did not translate into accurate match predictions. This is because the contrast between pixels in the initial bands is lower, and this produces noise rather than accurate matches when the image is subjected to PST-based edge detection. Furthermore, bands 32 and 33 suffered from the same situation as well, where there was a drop in filtered features. In terms of inliers, most of the bands from all image pairs were able to produce more than 80% inliers. However, image pair H had a fluctuating performance where it produced a respectable inlier ratio and a re-projection error within the band 20 to 30 (618nm to 706nm). Hence, it is suggested that in order to get the maximum performance from the proposed method, one needs to consider the contrast between pixels within each band. When incorporating edge-based feature detection in an application area other than agriculture plots, such as biomedical hyperspectral imaging, it is necessary to consider the band suitable for the specific application.

5.8.2 Evaluation of the proposed method

To assess the proposed method, four state-of-the-art feature detectors and matches commonly used in Simultaneous Localization and Mapping (SLAM) and other visual tracking tasks were chosen, in addition to the traditional SIFT method. Eighteen image pairs were carefully selected from three distinct datasets captured on different dates and at varying locations, with respective flight altitudes of 100m, 110m, and 80m. Figure 5.9 presents samples from each dataset, where the primary number indicates the dataset number, and the subscript denotes the image pair. The selection of image pairs was strategically done to underscore the strengths and weaknesses of the proposed method, facilitating a comparative analysis with state-of-the-art methods.

Each image pair underwent processing through the proposed method, as outlined in Section 5.7, to derive the Structural Similarity Index (SSIM) and correlation coefficient. Precisely, if the identified matches are accurate, higher SSIM and correlation values would be expected, signifying similarity in the two overlapping regions and the success of each method in generating high-quality feature matches. In addition to SSIM and the correla-

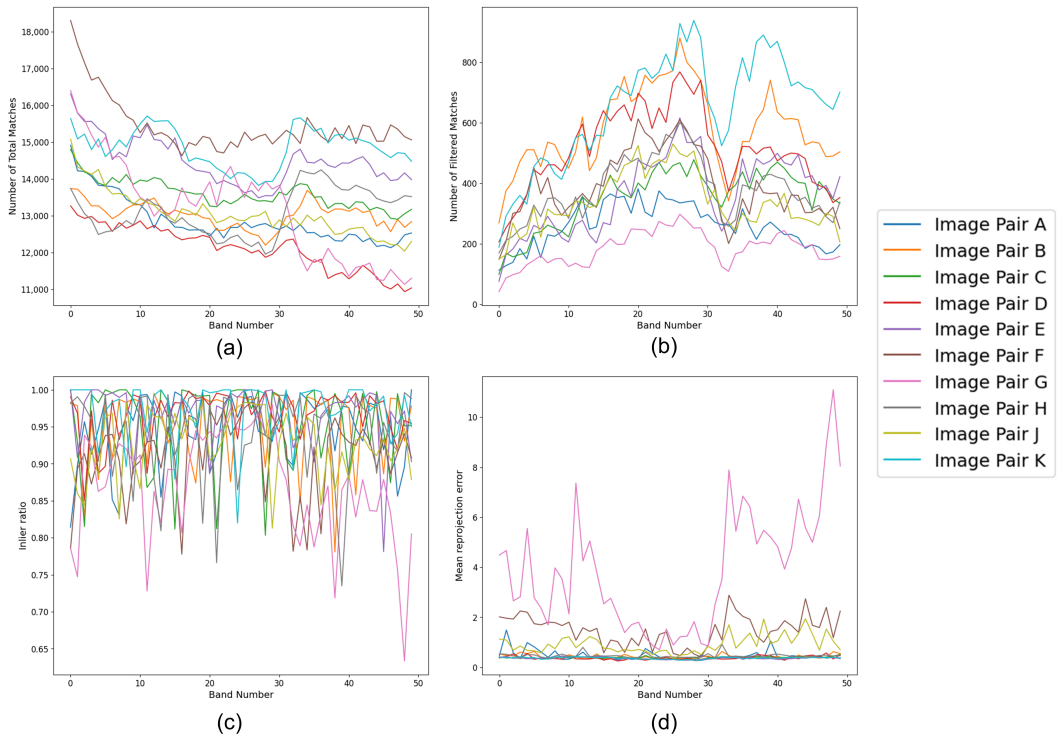


Figure 5.8: Evaluation of selecting a band for initial feature matching for 10 image pairs. (a) Total matches produced for each band. (b) Total matches produced from the proposed methods. (c) Inlier ratio each band produced. (b) Mean reprojection error for each band

tion coefficient, the inlier percentage, corresponding to the homography calculation, was also computed. This metric provides insight into the proportion of matches that align with the calculated homography, offering further evaluation criteria for the proposed method and facilitating comprehensive comparisons with other methodologies.

The results obtained from the evaluation experiments are presented in Table 5.1. In instances where the match number is specified, and the remaining columns indicate "Fail," it signifies that the algorithm generated matches; however, it either failed to calculate a homography using the identified matches or the computed homography did not yield meaningful results. When the term "Fail" is mentioned without indicating any matches, it suggests that the algorithm could not generate any matches for the respective image pairs.

Overall, the proposed method, the LoFTR method, and LightGlue with the DISK descriptor-matcher exhibited significantly superior performance in terms of both the num-

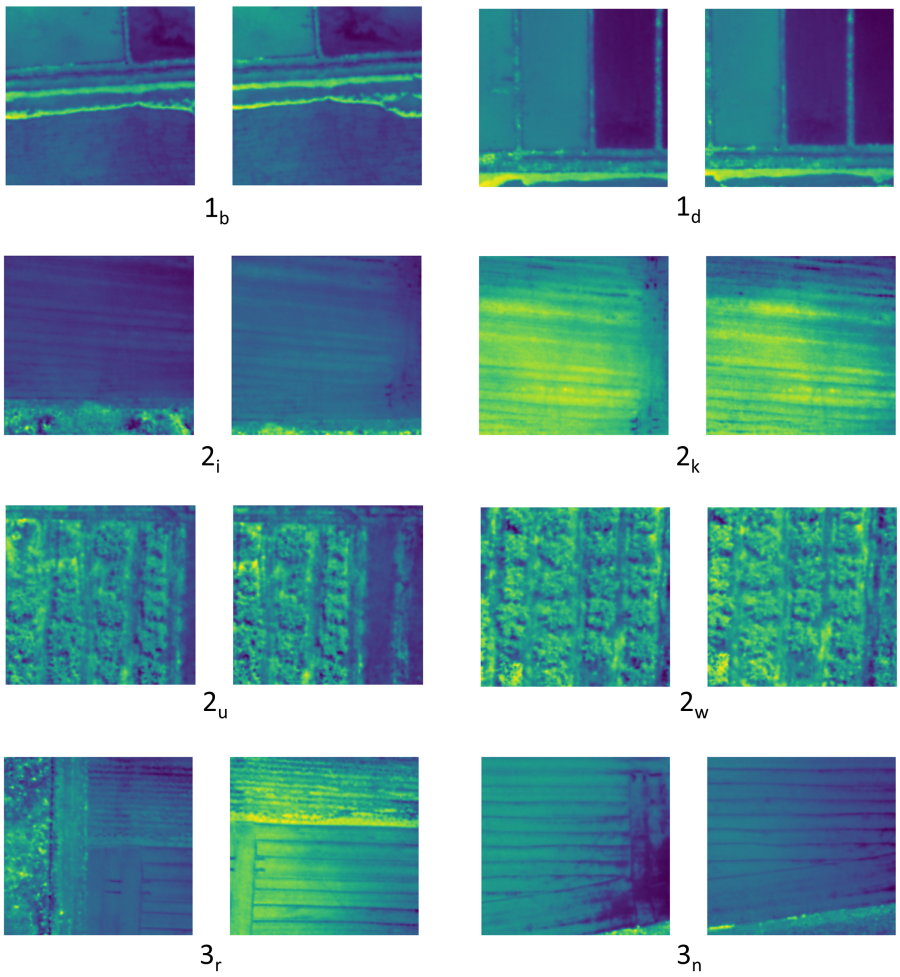


Figure 5.9: Sample of the selected data, Each figure name represents the corresponding dataset and the image pair in accordance with the Table 5.1

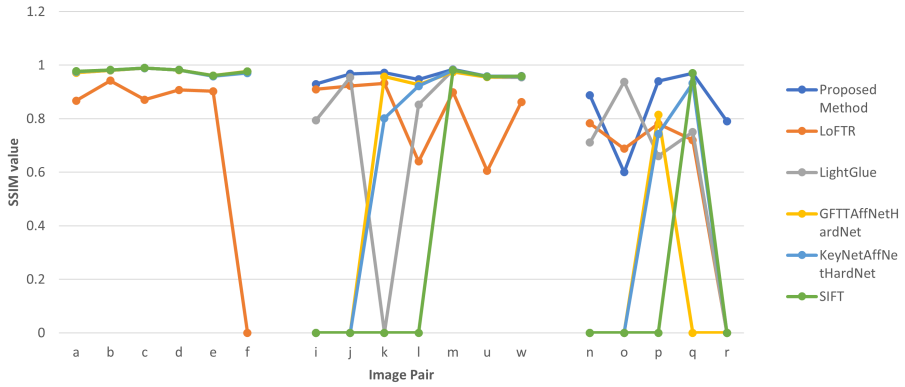


Figure 5.10: Line graphs for SSIM values obtained for each image pair.

ber of matches produced and the number of inliers falling under the calculated homography. In the first dataset, all methods yielded respectable metrics for the identified feature matches, except for matches obtained from LoFTR for image pair f, which failed to produce any usable homography from the detected matches. However, concerning the number of feature matches, traditional SIFT, GFTT + OpenGlue, and KeyNet + OpenGlue generated fewer results compared to the other three methods. This diminished performance was more evident in datasets 2 and 3, where the produced results were subpar compared to the other three methods.

Observations revealed that in scenes characterized by complexity and a diverse range of shapes and pixel intensities, most evaluated methods demonstrated satisfactory performance. However, an exception was noted in the case of the LoFTR method applied to image 2_u in Figure 5.9. Such scenarios typically involve complex vegetation, as depicted in images 2_u and 2_w, or scenes like 1_b and 1_d, where the radiometric characteristics create contrasting pixel regions. This underscores the importance of developing methods specifically tailored for application in agricultural plots, where such characteristics may not be prevalent. The need for methodologies adept at handling agricultural scenes, characterized by specific challenges, becomes apparent in ensuring reliable and accurate feature matching under diverse and complex environmental conditions.

Among the assessed methods, LightGlue consistently yielded the highest number of feature matches in most instances. LoFTR also demonstrated comparable results with LightGlue, although its performance was suboptimal for specific image pairs, such as i and r from datasets 2 and 3, as illustrated in Figure 5.9. The proposed method exhibited a performance level between LightGlue and LoFTR concerning the number of feature matches. Figure 5.10 visually represents the SSIM values, showcasing that, overall, the proposed method provided the highest SSIM values for the datasets, except in 5 instances out of 18, where SIFT delivered the highest SSIM, and two occasions where LightGlue and GFTT+OpenGlue produced superior SSIM.

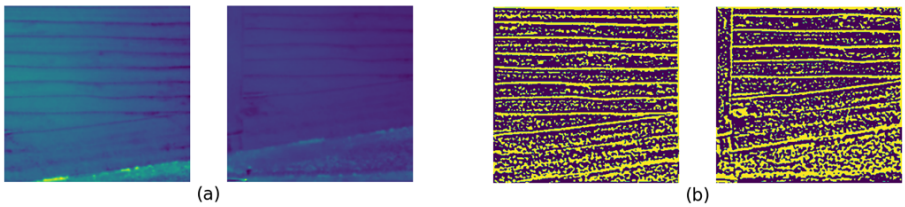


Figure 5.11: (a) Image pair *o* in dataset 3. (b) After the PST edge detection

Despite LoFTR generating a substantial number of matches, it did not achieve as high accuracy as the proposed method and the LightGlue method in terms of SSIM values. Among the feature matches obtained, the proposed method secured 9 out of the top 18 inlier ratios. Additionally, in the remaining 9 cases where our method did not achieve the highest inlier ratio, 7 instances were within 10% of the highest value. This underscores the highly accurate nature of the proposed method. However, image pair *o* depicted in Figure 5.11 exhibited suboptimal performance with the proposed method. This can be attributed to two primary reasons. A) There exists an illumination difference between the two images, causing the Siamese network to classify correct matches as non-matches. This is a consequence of the training data comprising identical pairs from the same image. B) Due to the low contrast of the image, the initial edge-based feature detection method struggled to identify correct matches. Furthermore, a consistent correlation was observed between correlation coefficient values and SSIM values across the results for each algorithm.

Nevertheless, a notable drawback of the presented approach lies in its computational speed. The proposed algorithm exhibits a performance slowdown of up to a factor of 10 in comparison to alternative methods, attributed to the substantial number of feature matches requiring filtration. It is crucial to emphasize that, throughout the experiments, trials were performed with room for further code optimization, particularly in the area of parallel processing and GPU utilization. We advocate employing the current model as an adjunctive feature detection method to complement state-of-the-art approaches. In practical scenarios, such as image mosaicing tasks, users can integrate evaluation metrics like SSIM to determine if the identified matches meet a specified threshold. If not, the proposed method could be invoked to discern feature matches.

5.9 Chapter Conclusion

In conclusion, Chapter 5 introduced a novel approach to hyperspectral image feature matching using a 3D Convolution Neural Network-based Siamese network. While the previous chapter proposed a mosaic generation pipeline based on the LoFTR feature detection and matching method, it critically assessed the limitations of existing state-of-

Table 5.1: Results from the proposed method vs others

		Proposed Method				LoFTR				LightGlue + DISK			
Dataset	Image Pair	Matches	SSIM	Correlation	Inliers	Matches	SSIM	Correlation	Inliers	Matches	SSIM	Correlation	Inliers
1	a	282	0.973	0.935	94.68	13	0.867	0.851	92..0	677	0.971	0.999	93.83
	b	254	0.980	0.992	91.73	719	0.942	0.978	80.80	834	0.980	0.992	84.77
	c	412	0.989	0.997	95.63	499	0.870	0.964	77.35	838	0.989	0.997	98.56
	d	394	**0.982**	0.998	96.70	63	0.907	0.979	85.71	873	0.981	0.998	96.44
	e	132	0.959	0.974	77.20	42	0.902	0.926	92.85	334	0.960	0.974	84.43
	f	231	0.975	0.980	81.38	5	Failed			438	0.973	0.977	98.18
2	**i**	74	**0.929**	0.622	89.18	240	0.909	0.717	87.08	97	0.793	-0.025	82.47
	j	88	**0.967**	0.896	87.50	287	0.922	0.833	75.95	387	0.954	0.943	76.74
	k	36	**0.971**	0.941	72.22	91	0.932	0.925	69.23	Failed			
	l	107	**0.946**	0.983	85.04	7	0.640	0.578	85.71	241	0.853	0.911	68.87
	m	215	**0.983**	0.968	86.04	143	0.899	0.880	74.12	212	0.982	0.965	88.21
	u	753	**0.958**	0.978	92.69	58	0.605	0.712	93.10	960	0.955	0.976	98.43
	w	1057	**0.959**	0.980	99.81	841	0.862	0.929	93.10	1042	0.954	0.977	91.26
3	**n**	14	**0.887**	0.874	100.0	410	0.782	0.747	71.21	6	Failed		
	o	52	0.600	0.693	94.23	234	0.687	0.931	75.21	285	0.710	0.958	92.63
	p	131	**0.940**	0.947	94.02	32	0.780	0.514	65.62	507	0.937	0.819	81.65
	q	36	0.969	0.980	75.00	45	0.720	0.805	66.66	118	0.660	0.497	33.89
	r	23	**0.790**	0.761	91.30	Failed				41	0.750	0.657	60.97

		GFTTAffNetHardNet +snn				KeyNetAffNetHardNet +snn				SIFT			
Dataset	Image Pair	Matches	SSIM	Correlation	Inliers	Matches	SSIM	Correlation	Inliers	Matches	SSIM	Correlation	Inliers
1	**a**	51	0.971	0.927	98.11	125	0.975	0.982	83.20	38	**0.977**	0.956	13.47
	b	52	0.979	0.992	100.0	126	0.981	0.992	88.09	29	**0.981**	0.993	11.41
	c	87	0.988	0.997	87.35	143	0.989	0.997	95.10	44	**0.989**	0.997	10.67
	d	88	0.981	0.998	95.45	150	0.981	0.998	86.66	38	0.981	0.998	8.12
	e	23	0.960	0.974	86.95	88	0.958	0.974	96.59	33	**0.961**	0.974	18.18
	f	23	**0.960**	0.974	86.95	88	0.958	0.974	96.59	33	0.961	0.974	18.18
2	i	Failed				Failed				34	Failed		
	j	Failed				Failed				Failed			
	k	7	0.957	0.932	100	10	0.801	0.774	80.00	Failed			
	l	4	0.927	0.971	100	18	0.922	0.968	88.88	12	Failed		
	m	27	0.973	0.958	92.59	73	0.981	0.964	84.93	66	0.981	0.963	8.37
	u	192	0.955	0.976	24.70	131	**0.958**	0.978	87.78	357	0.957	0.977	83.75
	w	259	0.958	0.980	24.50	131	0.957	0.979	83.21	456	0.958	0.979	96.49
3	n	Failed				5	Failed			26	Failed		
	o	8	Failed			8	Failed			Failed			
	p	14	0.814	0.510	78.57	41	0.742	0.452	75.60	19	Failed		
	q	1	Failed			26	0.932	0.956	69.23	7	**0.970**	0.978	19.44
	r	1	Failed			6	Failed			Failed			

the-art methods. One significant drawback identified was the reliance on a single band of hyperspectral images, which led to inadequate feature detection performance in certain image pairs.

Motivated by this limitation, the chapter presented a solution that leverages information from all available hyperspectral bands. The proposed Siamese network, integrated into the feature matching pipeline, demonstrated improved feature detection accuracy by using 32x32 patches extracted from different spectral bands. The methodology involved generating initial feature matches using edge maps obtained through Phase Stretch Transformation (PST), followed by dimensionality reduction using a 1D autoencoder. The Siamese network was then employed to filter out inaccurate matches, resulting in a collection of highly accurate filter matches.

The evaluation process included data acquisition and preprocessing steps, highlighting the use of the Cubert Ultris 5 hyperspectral camera mounted on a DJI M600 Pro drone. The evaluation criteria encompassed the Structural Similarity Index (SSIM), correlation coefficient, and inlier percentage based on homography calculations. The proposed method was compared with state-of-the-art algorithms, including LightGlue and LoFTR, across diverse datasets and scenarios.

The results demonstrated the efficacy of the proposed method, showcasing its ability to generate highly accurate feature matches in challenging agricultural scenes. The analysis of spectral bands revealed that selecting specific bands within the 600-800 nm range significantly contributed to improved performance. While the proposed method exhibited slightly slower computational speeds, it excelled in accuracy and robustness, especially in complex agricultural environments.

Despite achieving remarkable results, the proposed method faced challenges in scenarios with illumination differences and low-contrast images. Ongoing efforts for code optimization were emphasized, emphasizing the potential for further improvements in computational efficiency.

In summary, this Chapter presented a comprehensive and innovative approach to hyperspectral image feature matching, offering a valuable contribution to the field. The proposed method's strengths in accuracy and robustness position it as a promising tool for applications in agriculture and other domains where hyperspectral image analysis is crucial.

Chapter 6

Conclusion

This thesis intended to develop an approach to generate an image mosaic generation pipeline that would work for low-resolution hyperspectral images. Furthermore, a novel feature match generation method was also proposed to be incorporated with hyperspectral image mosaic generation pipelines. This chapter summarizes the work carried out and presented in this thesis, a list of contributions, and a discussion of potential future work.

6.0.1 Summery

In Chapter 2, the background of the research problem was presented and the current work in the relevant field was reviewed. Chapter 3 presented the initial setup of the acquisition UAV platform, the data acquisition process followed, initial measurement comparison between the hyperspectral camera and a field spectrometer. In Chapter 4, further investigation into the limitations of the current feature-matching algorithms, performance of the learning-based methods in the domain of hyperspectral low-resolution images, and image stitching pipeline were discussed. Further, several case studies were presented in order to better deliver the proposed method and intermediate results were discussed. Moreover, modification to the presented pipeline was discussed in order to mitigate the identified issues. Chapter 5 presented a novel approach to utilize all the spectral information available for hyperspectral image feature matching. The use of a 3D convolution Siamese network was proposed to filter out incorrect feature matches from a set of initial feature match sets, obtained using edge images of the hyperspectral images. Results indicated robust performance of the proposed method.

6.0.2 Contributions

Hyperspectral imaging is a powerful remote sensing technology that captures spectral information of a scene at a high level of detail. By collecting information across many narrow and contiguous spectral bands, hyperspectral imaging can reveal the chemical

and physical properties of the imaged objects, leading to a wide range of applications in various fields such as agriculture, mineralogy, and environmental monitoring. However, due to the large amount of data generated by hyperspectral sensors and their limited spatial resolution, it is often necessary to mosaic multiple low-resolution images to obtain a complete view of the scene. The importance of mosaic generation in aerial hyperspectral images lies in the fact that the quality of the mosaics directly affects the accuracy and effectiveness of subsequent data analysis and interpretation, making it a crucial step in the overall workflow of hyperspectral image analysis. The motivation for this thesis came from the use of low-pixel-resolution hyperspectral imaging in agriculture applications. It was identified that one of the main challenges in low-resolution hyperspectral imaging for agricultural applications is feature matching and subsequently in generating image mosaics from the captured hyperspectral images.

In the initial stage, this thesis investigated the problem of hyperspectral mosaic generation from low-resolution hyperspectral images. One of the major challenges in this task is identifying features from images, which is crucial for the accurate registration and alignment of images. Traditional feature detection methods used in already proposed methods are not sufficient for hyperspectral images due to their limited resolution. Furthermore, indistinctive features like uniform textures and repeating patterns present in hyperspectral low-resolution images taken via drones, further add to this problem. In the initial steps, various learning-based approaches were explored to identify a suitable method that would provide features with high accuracy. A learning-based approach named LoFTR and LightGlue was identified as a promising candidate for the feature identification and matching task.

Image stitching methodology incorporating the aforementioned learning algorithms was then proposed to identify the feature matches necessary in the stitching process. The proposed stitching method was evaluated using several case studies. Initial results from the proposed method indicated significant improvements over the traditional methods where it was able to stitch the image sets with some geometrical errors. This is a significant improvement because the previous methods were unable to produce any stitching results. However, the identified limitations of this initial method were then discussed and a modified method was proposed.

Subsequent chapter 5 introduces an innovative hyperspectral image feature matching method using a 3D Convolution Neural Network-based Siamese network. Addressing limitations from the previous chapter, the approach leverages information from all hyperspectral bands, improving feature detection accuracy. The methodology includes using 32x32x51 patches from the hyperspectral image obtained via Phase Stretch Transformation (PST) based edge maps to detect matching features. A 1D autoencoder was utilized in the dimensionality reduction process and the subsequent patch was then classified into

match or nonmatch by using a 3D convolution-based Siamese network.

Evaluation against state-of-the-art algorithms demonstrates the method's efficacy in challenging agricultural scenes, particularly when specific bands within the 600-800 nm range are selected. Despite slower computational speeds, the method excels in accuracy and robustness. Challenges include scenarios with illumination differences and low-contrast images, with ongoing efforts focused on code optimization for improved computational efficiency. In summary, the proposed method constitutes a considerable contribution to hyperspectral image analysis, showing promise for applications in agriculture and related domains.

6.1 Future Works

In the proposed mosaic generation pipeline, some ghosting effects were observed from the final stitched images. It has been observed that the blending method used was causing the ghosting effect. This is due to the misalignment or inconsistencies between overlapping regions of the images. This is especially observed around distinctive objects like buildings and structures. Identifying seams or edges of the objects and blending the images so that edges would not fall on the distinctive features could be researched to improve the blending accuracy. Furthermore, georectification of the mosaics was not considered in this stage of the research.

In the proposed 3D Siamese network-based feature filtering method, one of the major drawbacks was the processing speed. For filtering features for one image pair proposed model took 5 to 10 seconds of processing time. This is not ideal hence code optimization and GPU-based processing should be considered as future works in this research.

Reference

[1] A. S. Ray, "Remote sensing in agriculture," *International Journal of Environment, Agriculture and Biotechnology*, vol. 1, no. 3, p. 238540, 2016.

[2] J. B. Campbell and R. H. Wynne, *Introduction to remote sensing.* Guilford press, 2011.

[3] A. F. Goetz, "Three decades of hyperspectral remote sensing of the earth: A personal view," *Remote sensing of environment*, vol. 113, pp. S5–S16, 2009.

[4] A. Al-Hourani, S. Balendhran, S. Walia, and T. Hourani, "Line scan hyperspectral imaging framework for open source low-cost platforms," *Remote Sensing*, vol. 15, no. 11, 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/11/2787

[5] M. Teke, H. S. Deveci, O. Haliloğlu, S. Z. Gürbüz, and U. Sakarya, "A short survey of hyperspectral remote sensing applications in agriculture," in *2013 6th International Conference on Recent Advances in Space Technologies (RAST)*, 2013, pp. 171–176.

[6] J. J. Sousa, P. Toscano, A. Matese, S. F. Di Gennaro, A. Berton, M. Gatti, S. Poni, L. Pádua, J. Hruška, R. Morais *et al.*, "Uav-based hyperspectral monitoring using push-broom and snapshot sensors: A multisite assessment for precision viticulture applications," *Sensors*, vol. 22, no. 17, p. 6574, 2022.

[7] B.-H. Cho, Y.-H. Kim, K.-B. Lee, Y.-K. Hong, and K.-C. Kim, "Potential of snapshot-type hyperspectral imagery using support vector classifier for the classification of tomatoes maturity," *Sensors*, vol. 22, no. 12, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/12/4378

[8] ""gmbh – real-time spectral imaging," cubert. [online]. available: https://www.cubert-hyperspectral.com/products/ultris-5. [accessed: 15-nov-2022]."

[9] XIMEA, "Hyperspectral snapshot usb3 camera 24 bands 665-960nm," 2024, accessed: 2024-02-06. [Online]. Available: https://www.ximea.com/en/products/hyperspectral-cameras-based-on-usb3-xispec/mq022hg-im-sm5x5-nir

[10] D. Capel, "Image mosaicing," in *Image Mosaicing and super-resolution.* Springer, 2004, pp. 47–79.

[11] R. Feng, H. Shen, J. Bai, and X. Li, "Advances and opportunities in remote sensing image geometric registration: A systematic review of state-of-the-art approaches and future research directions," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 4, pp. 120–142, 2021.

[12] Y. Li, J. Wang, and K. Yao, "Modified phase correlation algorithm for image registration based on pyramid," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 709–718, 2022.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[14] H. Lee, S. Lee, and O. Choi, "Improved method on image stitching based on optical flow algorithm," *International Journal of Engineering Business Management*, vol. 12, p. 1847979020980928, 2020.

[15] A. Moussa and N. El-Sheimy, "A fast approach for stitching of aerial images." *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 41, 2016.

[16] J. K. Gómez-Reyes, J. P. Benítez-Rangel, L. A. Morales-Hernández, E. Resendiz-Ochoa, and K. A. Camarillo-Gomez, "Image mosaicing applied on uavs survey," *Applied Sciences*, vol. 12, no. 5, p. 2729, 2022.

[17] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[18] C. Harris, M. Stephens *et al.*, "A combined corner and edge detector," in *Alvey vision conference*, vol. 15, no. 50. Citeseer, 1988, pp. 10–5244.

[19] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2. Ieee, 2005, pp. 1508–1515.

[20] 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.

[21] 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.

[22] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "Loftr: Detector-free local feature matching with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8922–8931.

[23] P. F. Alcantarilla and T. Solutions, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," *IEEE Trans. Patt. Anal. Mach. Intell*, vol. 34, no. 7, pp. 1281–1298, 2011.

[24] K. M. Yi, E. Trulls, V. Lepetit, and P. Fua, "Lift: Learned invariant feature transform," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*. Springer, 2016, pp. 467–483.

[25] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Toward geometric deep slam," *arXiv preprint arXiv:1707.07410*, 2017.

[26] ——, "Superpoint: Self-supervised interest point detection and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 224–236.

[27] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4938–4947.

[28] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, "Lightglue: Local feature matching at light speed," *arXiv preprint arXiv:2306.13643*, 2023.

[29] I. Rocco, M. Cimpoi, R. Arandjelović, A. Torii, T. Pajdla, and J. Sivic, "Ncnet: Neighbourhood consensus networks for estimating image correspondences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 2, pp. 1020–1034, 2020.

[30] I. Rocco, R. Arandjelović, and J. Sivic, "Efficient neighbourhood consensus networks via submanifold sparse convolutions," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*. Springer, 2020, pp. 605–621.

[31] X. Li, K. Han, S. Li, and V. Prisacariu, "Dual-resolution correspondence networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17 346–17 357, 2020.

[32] N. Kitaev, Ł. Kaiser, and A. Levskaya, "Reformer: The efficient transformer," *arXiv preprint arXiv:2001.04451*, 2020.

[33] J. K. Gómez-Reyes, J. P. Benítez-Rangel, L. A. Morales-Hernández, E. Resendiz-Ochoa, and K. A. Camarillo-Gomez, "Image mosaicing applied on uavs survey," *Applied Sciences*, vol. 12, no. 5, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/5/2729

[34] Y. Mo, X. Kang, P. Duan, and S. Li, "A robust uav hyperspectral image stitching method based on deep feature matching," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[35] L. Zhang, H. Su, and J. Shen, "Hyperspectral dimensionality reduction based on multiscale superpixelwise kernel principal component analysis," *Remote Sensing*, vol. 11, no. 10, p. 1219, 2019.

[36] Y. Zhang, X. Mei, Y. Ma, X. Jiang, Z. Peng, and J. Huang, "Hyperspectral panoramic image stitching using robust matching and adaptive bundle adjustment," *Remote Sensing*, vol. 14, no. 16, p. 4038, 2022.

[37] Z. Peng, Y. Ma, X. Mei, J. Huang, and F. Fan, "Hyperspectral image stitching via optimal seamline detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[38] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[39] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, no. 1, p. 67, 2017.

[40] C. Wang, N. Ma, Y. Ming, Q. Wang, and J. Xia, "Classification of hyperspectral imagery with a 3d convolutional neural network and jm distance," *Advances in space research*, vol. 64, no. 4, pp. 886–899, 2019.

[41] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3d full convolutional neural network," *Remote Sensing*, vol. 9, no. 11, p. 1139, 2017.

[42] D. Chicco, "Siamese neural networks: An overview," *Artificial neural networks*, pp. 73–94, 2021.

[43] N. Serrano and A. Bellogín, "Siamese neural networks in recommendation," *Neural Computing and Applications*, pp. 1–13, 2023.

[44] I. Melekhov, J. Kannala, and E. Rahtu, "Siamese network features for image matching," in *2016 23rd international conference on pattern recognition (ICPR)*. IEEE, 2016, pp. 378–383.

[45] J. Khodr and R. Younes, "Dimensionality reduction on hyperspectral images: A comparative review based on artificial datas," in *2011 4th international congress on image and signal processing*, vol. 4. IEEE, 2011, pp. 1875–1883.

[46] J.-L. Xu, C. Esquerre, and D.-W. Sun, "Methods for performing dimensionality reduction in hyperspectral image classification," *Journal of Near Infrared Spectroscopy*, vol. 26, no. 1, pp. 61–75, 2018.

[47] A. Mehta and O. Dikshit, "Spca assisted correlation clustering of hyperspectral imagery," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 2, pp. 111–116, 2014.

[48] M. Imani and H. Ghassemian, "Two dimensional linear discriminant analyses for hyperspectral data," *Photogrammetric Engineering & Remote Sensing*, vol. 81, no. 10, pp. 777–786, 2015.

[49] S. D. Fabiyi, P. Murray, J. Zabalza, and J. Ren, "Folded lda: extending the linear discriminant analysis algorithm for feature extraction and data reduction in hyperspectral remote sensing," *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 14, pp. 12 312–12 331, 2021.

[50] H. Wu and S. Prasad, "Semi-supervised dimensionality reduction of hyperspectral imagery using pseudo-labels," *Pattern Recognition*, vol. 74, pp. 212–224, 2018.

[51] R. Zaatour, S. Bouzidi, and E. Zagrouba, "Class-adapted local fisher discriminant analysis to reduce highly-dimensioned data on commodity hardware: application to hyperspectral images," *Multimedia Tools and Applications*, vol. 78, pp. 17 113–17 134, 2019.

[52] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE transactions on geoscience and remote sensing*, vol. 44, no. 6, pp. 1586–1600, 2006.

[53] D. Fernandez, C. Gonzalez, D. Mozos, and S. Lopez, "Fpga implementation of the principal component analysis algorithm for dimensionality reduction of hyperspectral images," *Journal of Real-Time Image Processing*, vol. 16, pp. 1395–1406, 2019.

[54] C. Jayaprakash, B. B. Damodaran, V. Sowmya, and K. Soman, "Dimensionality reduction of hyperspectral images for classification using randomized independent

component analysis," in *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2018, pp. 492–496.

[55] E. El-Araby, T. El-Ghazawi, J. Le Moigne, and K. Gaj, "Wavelet spectral dimension reduction of hyperspectral imagery on a reconfigurable computer," in *Proceedings. 2004 IEEE International Conference on Field-Programmable Technology (IEEE Cat. No. 04EX921)*. IEEE, 2004, pp. 399–402.

[56] E. T. Gormus, N. Canagarajah, and A. Achim, "Dimensionality reduction of hyperspectral images using empirical mode decompositions and wavelets," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 6, pp. 1821–1830, 2012.

[57] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *Machine learning for data science handbook: data mining and knowledge discovery handbook*, pp. 353–374, 2023.

[58] G. Dong, G. Liao, H. Liu, and G. Kuang, "A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images," *IEEE Geoscience and Remote Sensing Magazine*, vol. 6, no. 3, pp. 44–68, 2018.

[59] J. Zabalza, J. Ren, J. Zheng, H. Zhao, C. Qing, Z. Yang, P. Du, and S. Marshall, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 185, pp. 1–10, 2016.

[60] L. Windrim, R. Ramakrishnan, A. Melkumyan, R. J. Murphy, and A. Chlingaryan, "Unsupervised feature-learning for hyperspectral data with autoencoders," *Remote Sensing*, vol. 11, no. 7, p. 864, 2019.

[61] S. Pande and B. Banerjee, "Dimensionality reduction using 3d residual autoencoder for hyperspectral image classification," in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2020, pp. 2029–2032.

[62] "pokini f2 — lüfterloser mini-pc für die industrie." [Online]. Available: https://pokini.de/pokini-f2/

[63] "Matrice 600 Pro - Product Information - DJI — dji.com," https://www.dji.com/jp/matrice600-pro/info, [Accessed 17-Mar-2023].

[64] DJI, "Dji gs pro - dji." [Online]. Available: https://www.dji.com/ground-station-pro

[65] NOVOFLEX, "Novoflex zebra xl extra large zebra card (gray/white, 21 × 30 cm)." [Online]. Available: https://novoflex.com.au/product/novoflex-extra-large-zebra-card-gray-white-8-x-12/

[66] R. Booysen, R. Jackisch, S. Lorenz, R. Zimmermann, M. Kirsch, P. A. Nex, and R. Gloaguen, "Detection of rees with lightweight uav-based hyperspectral imaging," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[67] "Ms-730 portable spectroradiometer." [Online]. Available: https://eko.co.jp/products/solar_radiation_measurement/spectroradiometer/ms-730.html

[68] L. A. Weber and D. Schenk, "Automatic merging of split construction plans of hydraulic structures," *Engineering*, vol. 99, no. 5, pp. 330–340, 2022.

[69] R. Xiang, M. Sun, C. Jiang, L. Liu, H. Zheng, and X. Li, "A method of fast mosaic for massive uav images," in *Land Surface Remote Sensing II*, vol. 9260. SPIE, 2014, pp. 725–733.

[70] D. P. E. R. E. Riba, D. Mishkin and G. Bradski, "Kornia: an open source differentiable computer vision library for pytorch," in *Winter Conference on Applications of Computer Vision*, 2020. [Online]. Available: https://arxiv.org/pdf/1910.02190.pdf

[71] M. H. Asghari and B. Jalali, "Physics-inspired image edge detection," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 293–296.

[72] ——, "Edge detection in digital images using dispersive phase stretch transform," *Journal of Biomedical Imaging*, vol. 2015, pp. 6–6, 2015.

[73] F. Coppinger, A. Bhushan, and B. Jalali, "Photonic time stretch and its application to analog-to-digital conversion," *IEEE Transactions on microwave theory and techniques*, vol. 47, no. 7, pp. 1309–1314, 1999.

[74] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

[75] I. Sobel, "History and definition of the so-called "sobel operator", more appropriately named the sobel-feldman operator. 2014," *URL https://www. researchgate. net/publication/239398674_An_ Isotropic_3x3_Image_Gradient_Operator. First presented at the Stanford Artificial Intelligence Project (SAIL)*, 1968.

[76] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.

[77] Y. Zhou, C. MacPhee, M. Suthar, and B. Jalali, "Phycv: the first physics-inspired computer vision library," *arXiv preprint arXiv:2301.12531*, 2023.

[78] M. Köppen, "The curse of dimensionality," in *5th online world conference on soft computing in industrial applications (WSC5)*, vol. 1, 2000, pp. 4–8.

[79] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[80] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM review*, vol. 60, no. 2, pp. 223–311, 2018.

[81] T. Takase, "Dynamic batch size tuning based on stopping criterion for neural network training," *Neurocomputing*, vol. 429, pp. 1–11, 2021.

[82] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[83] M. Tyszkiewicz, P. Fua, and E. Trulls, "Disk: Learning local features with policy gradient," *Advances in Neural Information Processing Systems*, vol. 33, pp. 14 254–14 265, 2020.

[84] J. Shi *et al.*, "Good features to track," in *1994 Proceedings of IEEE conference on computer vision and pattern recognition*. IEEE, 1994, pp. 593–600.

[85] O. Viniavskyi, M. Dobko, D. Mishkin, and O. Dobosevych, "Openglue: Open source graph neural net based pipeline for image matching," *arXiv preprint arXiv:2204.08870*, 2022.

[86] A. Barroso-Laguna, E. Riba, D. Ponsa, and K. Mikolajczyk, "Key.net: Keypoint detection by handcrafted and learned cnn filters," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.