UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ACCELERATED LABELING OF LARGE SCALE GEO-REFERENCED DATASETS

A thesis submitted in partial satisfaction of the
requirements for the degree of

MASTERS OF SCIENCE

in

ELECTRICAL AND COMPUTER ENGINEERING

by

Nikolaas M. C. Bender

June 2023

The thesis of Nikolaas M. C. Bender
is approved:

Professor Steve McGuire, Chair

Professor Gabriel Hugh Elkaim

Professor Ricardo Sanfelice

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Accelerated labeling of large scale geo-referenced datasets

by

Nikolaas M. C. Bender

The current landscape of modern field robotics confronts a notable expansion challenge, primarily stemming from the proliferation of diverse platforms, vast data volumes, sophisticated algorithms, and powerful computational resources. However, the progress of robotics beyond its conventional role as a mere observational tool is impeded by limited autonomy and undeveloped data inference capabilities. The scarcity of labeled data poses a significant hurdle in the application of machine learning to novel tasks in the space of image comprehension, which incur exorbitant costs for manual labeling. The traditional approach of relying on human annotators at workstations for prolonged periods of time exhibits several limitations, including subjective consistency and human factors concerns. Conversely, the adoption of human-centric labeling tools presents a viable solution to address these challenges by deploying human experts in the field to perform data labeling, thereby enabling swift inspection of the sensed environment while eliminating data representation as a bottleneck. This approach facilitates the rapid annotation of data in a sensor-agnostic manner, culminating in the generation of large-scale, high-quality datasets within feasible timeframes by leveraging comparatively small amounts of expert-derived knowledge to improve data labels over all.

To my family, my friends, and my team.

This work would not have been possible without your support

Acknowledgments

With help from: Morgan Masters, Thomas Luca Altaffer, Adam Korycki, Nick Kuipers, Bobby Gaines, Aram Chemishkian, Conor Guzikowski, Liam Asayag, Tirath Shah, and Steve McGuire "At a scale as small as three acres the ASTRO imaging system generates 180,000 images to be labeled. Labeling this many images by hand is often referred to as cruel and unusual punishment."

Chapter 1

Introduction

Field robotics has emerged as a valuable tool for the task of observation within various domains. Presently, state-of-the-art approaches consider robots as mere extensions of humans leveraging teleoperation or direct human review of data. In the context of infrastructure inspection, the utilization of unmanned aerial vehicles (UAVs) equipped with cameras enables operators to examine the condition of infrastructure for signs of deterioration without the need to send humans into harms way to do the same initial inspections. Notably, robots are increasingly replacing human personnel in tasks such as inspecting wind turbines and power lines. The reliance on robots as remote sensing devices for field technicians is a pragmatic but temporary solution stemming from inadequate investments in automating inspection processes that have traditionally depended heavily on human intervention.

Increased situational awareness and data analysis are particularly valuable for precision agriculture applications. Precision agriculture seeks to take into account

the inherent variability within crop fields that mechanized farming would traditionally treat an entire field based on sparse set of measurements[37]. The averaging behavior of mechanized farming leads to excessive waste and ecological problems. To halt the averaging behavior of mechanized farming, precision agriculture quantifies field health at a scale of meters as opposed to acres, enabling hyper local treatment for crops.

A contemporary diagnostic workflow can be illustrated as follows: an aerial survey utilizing a hyperspectral imaging system captures data across five distinct wavelengths of light including rededge and near infrared. This hyperspectral imagery is subsequently processed using Geographic Information System (GIS) software, which compiles the image data into a high-resolution gigapixel-scale representation of the region of interest. The data is then transformed to derive a health metric known as the Normalized Differential of Vegetation Index (NDVI)[47]. NDVI serves as a summary of plant health based on the light absorption in the red and infrared wavelength bands recorded by the hyperspectral imaging system. Resource allocation decisions are subsequently guided by the NDVI and similar statistical measures.

$$NDVI = (NIR - Red)/(NIR + Red) \tag{1.1}$$

Unmanned aerial vehicle (UAV) imaging systems are increasingly replacing satellite imagery in precision agriculture due to their higher spatial resolution compared to civilian-accessible orbital imaging modes, as well as the ability of UAVs to provide on-demand observations.

As previously stated, NDVI 1.1 is a heuristic for plant health based on the

Imager	M/px
Landsat8	15-30
ASTRO @ 10m	0.003
Miasense @ 60m	0.04

Table 1.1: The ground resolution for comparison. Landsat8 is the most recent earth observation satellite from NASA often being used to investigate deforestation and agriculture. ASTRO is our proposed imaging platform. Micasense is the commercial crop imaging solution with a similar sensor configuration to landast.

observation that healthy plants absorb large quantities of red(668nm) light and reflect large quantities of light in the near infrared(842nm) spectrum. However, when a plant falls ill, less red light is absorbed and less near infrared light is reflected. This difference in leaf color is commonly observed as plants turning brown. From an initial NDVI raster, an expert can tune the NDVI calculation to correctly assess vegetation health, e.g. by creating offsets or using reference areas of known healthy palnts [63]. This expert manipulation is necessary to reject background signals, compensate for the spectral impact of environmental moisture, and otherwise filter out noise. Once corrected, NDVI maps are used to dispatch inspections of the flown area for patches of unhealthy crops. This process for finding unhealthy plants with expert analysis employing NDVI takes days or weeks to compute.

The time-consuming step in the NDVI rastering process lies in the necessary human intervention for corrections, which motivates our proposed approach. Deep learning is a promising technique for capturing expert knowledge and automating plant health assessment and directly inferring plant health without the need of human intervention or interpretation. The effectiveness of a deep learning approach relies on the availability of sufficiently large-scale labeled image datasets for training robust models.

To address this challenge, we propose a novel labeling method where subject matter experts create labels in the field, which can later be projected onto the corresponding image data. This technique offers several advantages over traditional image dataset construction methods1.1, including sensor and representation agnostic, parallelizability, improved speed, and increased human compatibility. In total the proposed system is more efficient than previous methods for labeling large geo-referenced datasets.

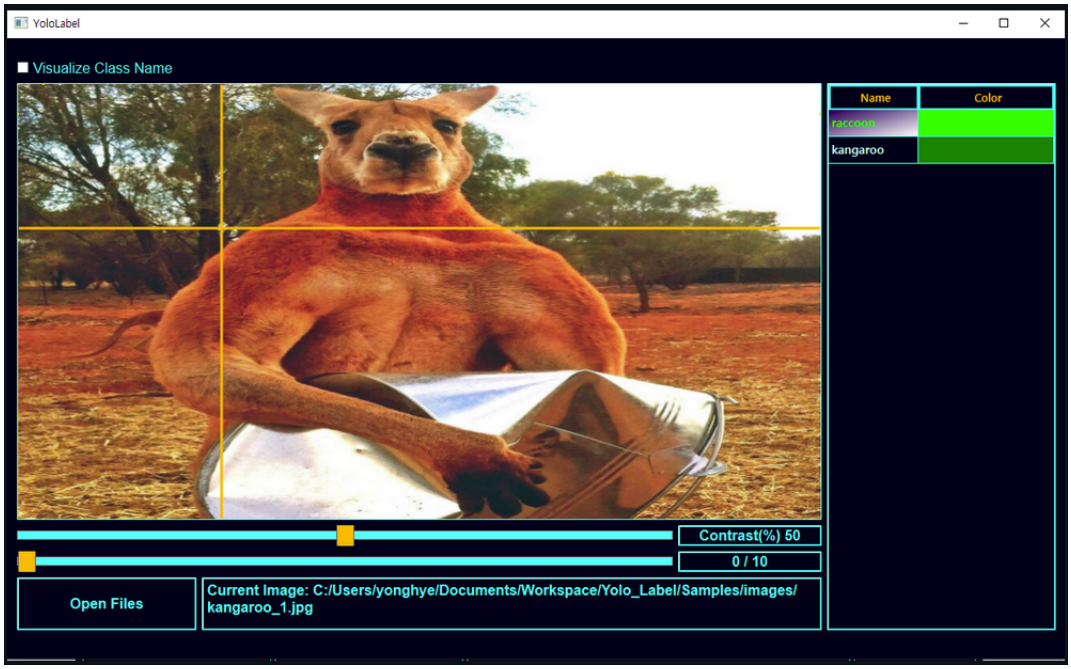

Figure 1.1: This is an example of the "YoloLabel" tool used for labeling images to accomplish the task of object detection on image data. Yololabel prompts humans to draw boxes around objects of the type in question and label the objects according to their class.

The data labeling tool is a small handheld device consisting of a GNSS unit and a data collection unit. The data collected by the hand unit is then correlated to imagery collected by a UAV equipped with a multispectral camera suite4.1. We demonstrate

that it is possible to correlate data between these two devices for the purposes of labeling image data in a format compatible with deep learning image analysis algorithms. The remainder of this manuscript focuses on the application of this labeling paradigm to agriculture, specifically crop heath monitoring, but we stress that the system applies to any context where remote imaging can be used to identify regions of interest within an environment.

In summary, the system we propose facilitates the efficient generation of large-scale, geo-referenced datasets for supervised machine learning, leveraging hardware to expedite the data labeling process.

Chapter 2

Previous Work

To demystify the naming scheme employed for this work: BANNERS and ASTRO are simply names, spawned from jokes in the lab group, which have grown to be familiar callsigns for the projects. The work leading up to the development of BANNERS and ASTRO primarily goes to demonstrate that agriculture technology is a useful candidate for such a system. As applied to agriculture there are vast quantities of data available and a desire to use deep learning to fulfill the promises of precision agriculture. Between these qualities BANNERS supports the creation of large scale geo-referenced and multispectral datasets for the pursuit of UAV based crop disease detection.

2.1 Deep Learning

Supervised learning the simplest format for machine learning; given a compiled dataset, consisting of features and annotations, develop a set of rules to map from

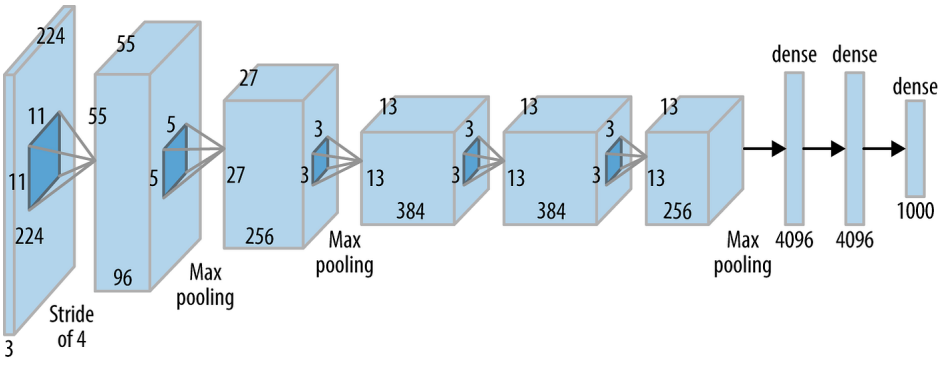

Figure 2.1: Deep learning is the primary approach for image analysis in modern robotics. This is a convolutional neural network used for image classification. Convolutional neural networks are regularly employed for image analysis and processing in machine learning. [24]

data to labels with minimal error. Deep learning is a particularly good paradigm for machines to learn complex patterns through large-scale data analysis, and it is most easily implemented using supervision. However, the process of compiling annotated datasets of sufficient size to support supervised deep learning is prohibitively expensive. Fundamentally, the greater the volume of data available, the richer a set of derived features can be, and the more robust a machine learning algorithm can be to novel data.

2.2 Data Labeling

Advancements are being made to reduce this bottleneck. One promising technique being explored to reduce labeling cost is the use of human-centric techniques. Human-centric techniques pose the data labeling problem in a way where the human does not need to sit behind a computer in order to label data. For example, in the

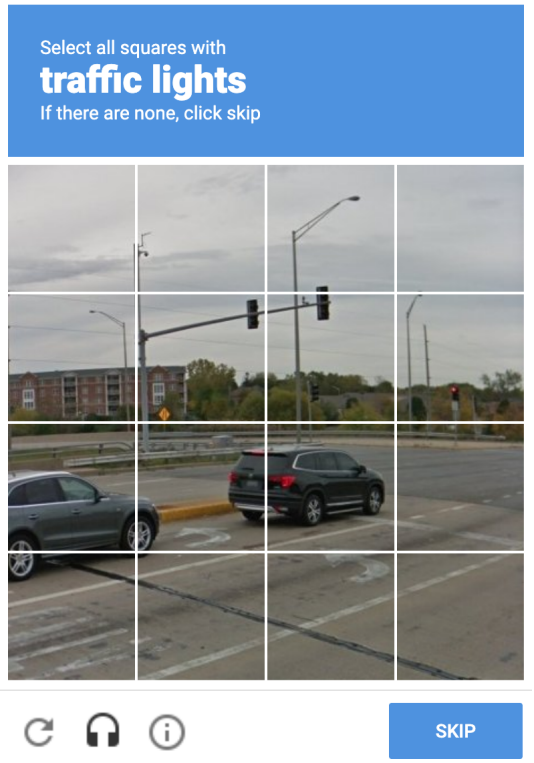

Figure 2.2: This is an example of reCAPTCHA v2 from Google. The reCAPTCHA system redirects human effort for those who are seeking to certify their humanity online to also generate data labels for Google.

context of semantic segmentation, groups are using virtual reality (VR) to paint scene elements with labels [50]. The traditional strategy to overcome the problem of scale when labeling image datasets is to use vast amounts of cheap labor. Universities often employ teams of undergraduates to collect and label data. Corporations will use customers to label data. For example reCAPTCHA is a system by which users click all of the boxes that contain some specific object of interest or type in the letters seen among a noisy image. Other companies will offer free services such as education as a useful byproduct of generating training data from humans. Finally companies will employ cheap labor like Amazon mechanical turk to label data. Each of these approaches seeks to deal with the problem of how to label enormous quantities of data with labels derived from humans.

2.3 Unsupervised Learning

Another approach to mitigating the human cost of data labeling is to circumvent the supervised learning paradigm entirely by using semi-supervised or unsupervised learning. In these contexts, sophisticated loss functions replace concrete annotations to promote the derivation of robust features that differentiate the data accurately[28]. A common loss function for the classification task rewards the creation of a label distribution that is uniform. In semi-supervised learning, an algorithm starts a training regime with enough human-labeled data to establish a predictive baseline, then moves to the unsupervised regime on unlabeled datasets which are decreasingly similar to the initial

training set, in order to learn a more complete representation of the source distribution of the data [59]. Unsupervised training represents the extreme end of this approach, where full responsibility for the training is placed on the loss functions and the dataset size is increased further[22]. For supervised learning a neural network may need 5,000 images with labels such as Cityscapes [9]. Semi-supervised may need 60,000 images such as MNIST where the data is highly curated to give great results for generative models and image classification. Self-supervised learning requires millions of images [58]. Generally self-supervised learning is used for "generative ai" which has become popular as of late and where image labels are difficult to conceive of. There is recent work showing that the scale necessary for "good" datasets may be inflated and that being more intelligent about the distribution of training data may lead to better results [46] in machine learning, however the current wisdom is that more data is better for all deep learning tasks.

2.4 NDVI

Earth sciences are adopting computer vision techniques, to assess larger environments than has been historically possible. One of the most popular techniques, in use by the agricultural community, is the Normalized Differential of Vegetation Index (NDVI) [47] [55] which ranges on $[-1, 1]$. NDVI calculates the percentage of reflected light in specific spectra2.3 to create an index that broadly correlates with the health of a plant. NDVI highlights plants that look sick to humans. A plant that is brown and

dying would have a negative NDVI score. NDVI is a general measure of plant health with applications across a range of agriculture and ecology tasks.

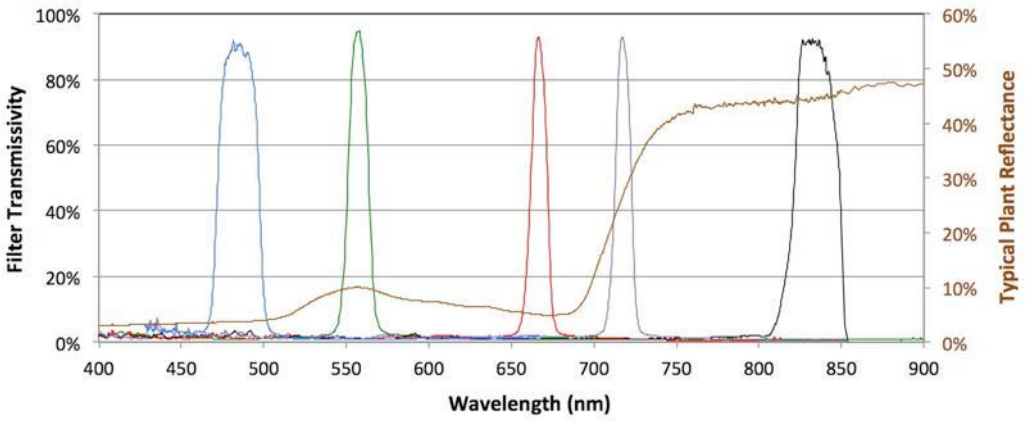

Figure 2.3: This is the quantum response of the Micasense RedEdge MX system which is the industry standard for agricultural imaging.

Satellites are a great platform for generating the data used for NDVI analysis of the Earth; due to their cost, they are equipped with extremely high-fidelity sensors and imaging platforms. These sensors capture detail on the order of meters per pixel [6] and constantly transmit that data to terrestrial receivers. For example, Landsat8 completes a full cycle imaging earth every 16 days. The reliability and fidelity of these sensors complements the unique reach of satellite platforms to enable the routine tracking of ecological and geographical processes in otherwise inaccessible areas of the globe [28].

Multispectral imaging characterized by several band-specific imagers that are tuned to wavelengths of particular interest to ecological and agricultural tasks, the core technology of orbital sensing modalities, has recently become accessible at the consumer or small business scale, where sensors such as the Micasense RedEdge MX can now

be found for prices in the vicinity of $10,000 (USD). Combining these multispectral sensors with low-cost UAV platforms has revolutionized data science, as it applies to the earth sciences. UAVs have a higher spatial resolution compared to satellites as well as a higher deployment frequency. UAVs are limited to a maximum altitude of 400ft above ground level (AGL) even with commercial drone licensing which further benefits the spatial resolution of sub-orbital grade imaging solutions. Between cost, rate, and resolution, UAV-based systems have overtaken satellite-based systems in easily-accessible deployment environments. In the context of agriculture, the high spatial resolution of UAV systems is being used for resource management decisions, such as detecting weeds and ensuring plants receive the correct amount of water [65].

However, these burgeoning applications, summarized well by [34], face the persistent problem of the cost and difficulty of assembling datasets to train machine learning algorithms (see their Section 7). Hereafter, we present our solution to that barrier, embodied as open-source software and a gear shakedown to implement the method.

Chapter 3

Problem Statement

The emergence of widely available aerial robotics has led to the accumulation of extensive datasets covering vast geographical areas, surpassing the practical limits of human annotation in terms of cost and efficiency. A tangible example is that of labeling the imagery produced by a UAV from a flight over a 30 acre farm for the purpose of generating data used in deep learning algorithms.

Chapter 4

Methods

The proposed workflow, shown in Figure4.1 relies on three major components: data collection, data relation, and data projection.

4.1 GNSS labeling tool

In order to label data within a geographic reference frame, the development of a GNSS (Global Navigation Satellite System) system leveraging GPS, GLONASS, BeiDeu, and Galileo constellations with Real-Time Kinematic (RTK) capabilities has been undertaken. GNSS with RTK positioning offers highly accurate positioning and possesses favorable characteristics in terms of Size, Weight, and Power (SWAP).

RTK is a mode of operation for GNSS devices that leverages RTCM (Radio Technical Commission for Maritime Services) data to mitigate ionospheric interference and achieve maximum positional accuracy. These attributes make GNSS positioning the preferred solution for position estimation during data collection. A single base station

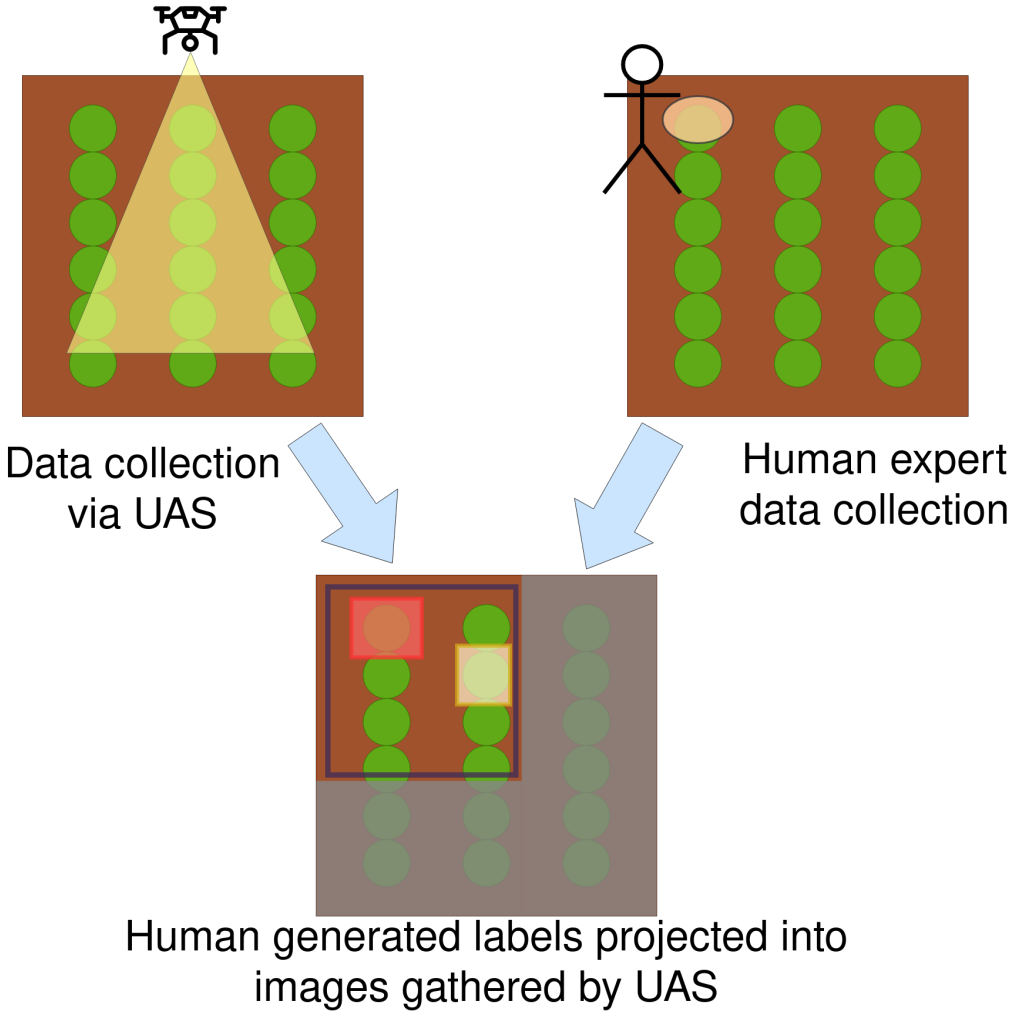

Figure 4.1: This is the high level concept of BANNERS and ASTRO where data is first collected then related in order to accurately generate labels within the desired data.

has been devised to generate RTCM3 corrections, which can be transmitted over LoRa (Long Range radio) to arbitrarily many rover units. The use of LoRa was driven by cost and range. RTCM data can be transmitted over ip with the Networked Transport of RTCM via Internet Protocol (NTRIP) but requires internet connection which incurs data and hardware costs to utilize commercial cellular networks or assembling enormous arrays of WiFi mesh devices to connect back to an NTRIP server on the internet. Using point to point LoRa with a bespoke network and a purpose built base station allowed for low cost and perpetual RTCM message generation. LoRa implements its own communications protocol which allows for broadcast like udp and enables the functionality of a single RTCM base station to broadcast to arbitrarily many rover units be they robots or data collection tools. These rover units are compact and lightweight, making them suitable for deployment on UAVs and handheld devices. The rover units consume minimal power, rendering them ideal for mobile robotics and handheld applications. The rover units offer a high level of accuracy, with a precision of 1.4cm 5.5 using small antennas which will be useful for accurate data projection at later stages.

The most popular data collection and spatial partitioning technique used by experts for field work currently is the quadrat. A quadtrat is a 1m by 1m square made of pvc pipe or some other light weight material. The quadrat becomes the local reference for any data within the quadrat allowing for scientists to break up a large space, such as a field, into $1m^2$ areas where each square meter can be accurately positioned and data collected within each. An example of this would be counting the number of weeds within a field. By first breaking up a field into quadrats scientists can then record the

number of weeds within each quadrat and the position of each quadrat within the field. At the end of counting results can be compiled from data recorded.

The widespread adoption of GNSS positioning, particularly in the agricultural context, is highly valued due to its ease of setup. The use of quadrats to partition a space into a grid, necessitating the construction of a new coordinate frame that must then be related to the photometric data obtained from a UAV with its own reference frame. This aditional reference frame for no benefit in data resolution is why we decided upon directly collecting data in GNSS space.

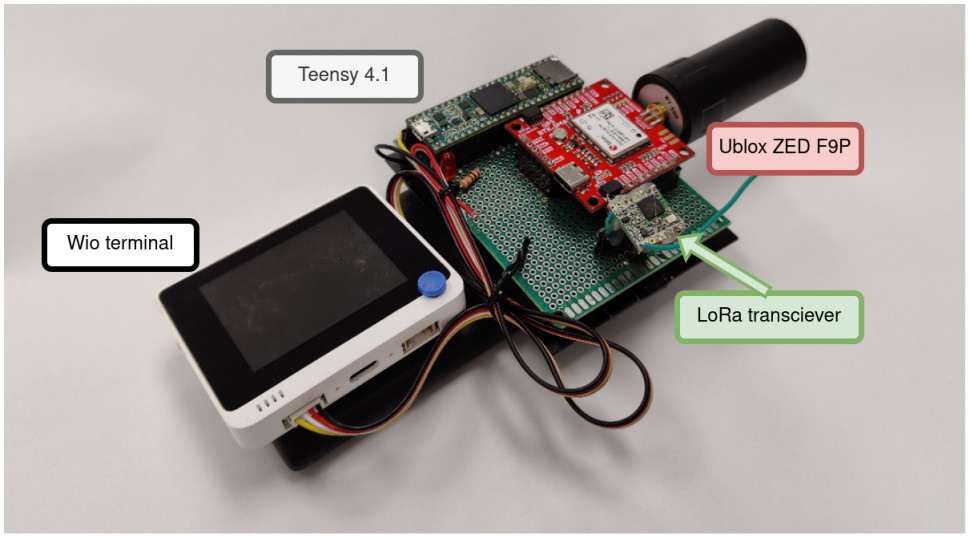

Figure 4.2: The geographic labeling tool utilized in the project is known as the "clicker." This device consists of two primary components: a high-precision GNSS unit based on the Ublox ZED F9P, and a WIO terminal[2]for data collection. The clicker is designed in a compact form, with both components mounted on a common USB battery pack, serving as both a mechanical connection and a power source for the system.

All of the properties outlined collectively support the use of GNSS positioning

[2]ZED module can be found here:`https://www.sparkfun.com/products/16481`
Wio terminal can be found here: `https://www.seeedstudio.com/Wio-Terminal-p-4509.html?queryID=39bdb19891d907f5993cd290cd634075&objectID=4509&indexName=bazaar_retailer_products`

as the fundamental framework for data relation. The ability to collect photometric data and labels within the same reference frame simplifies the task of data correlation in subsequent stages of the project (BANNERS).

4.2 ASTRO

ASTRO is the name given to the system we developed that leverages a DJI Matrice M300 UAV to establish geographic references between ground observations and hyperspectral aerial imagery. ASTRO expands the capabilities of the platform by incorporating sensing and computing functionalities built with ROS (Robot Operating System)[26], thereby enhancing the accessibility of precision agriculture for research entities. The ASTRO system comprises two major components, pose estimation and image sensors.

4.2.1 Platform

The selection of the DJI Matrice M300 as the UAV platform for our study was based on its capacity to effectively support data collection. The Matrice platform is popular for aerial photography, highlighting its reliability and established reputation. Furthermore, the platform boasts compatibility with the ROS to collect data directly from the UAV as well as implement custom autonomous navigation systems, a crucial feature for supporting research.

The quality of the stabilization system built into the Matrice platform is another significant advantage observed during our experiments. The system demonstrates

Figure 4.3: ASTRO: The data collection robot showcasing the integration of DJI Matrice M300 platform. The "roof" section is home to the exposed power and data connections, accompanied by the DJI manifold system facilitating power and data transfer from the UAV. Notably, the high precision GNSS system is also positioned on the roof, enhancing geo-referencing capabilities.

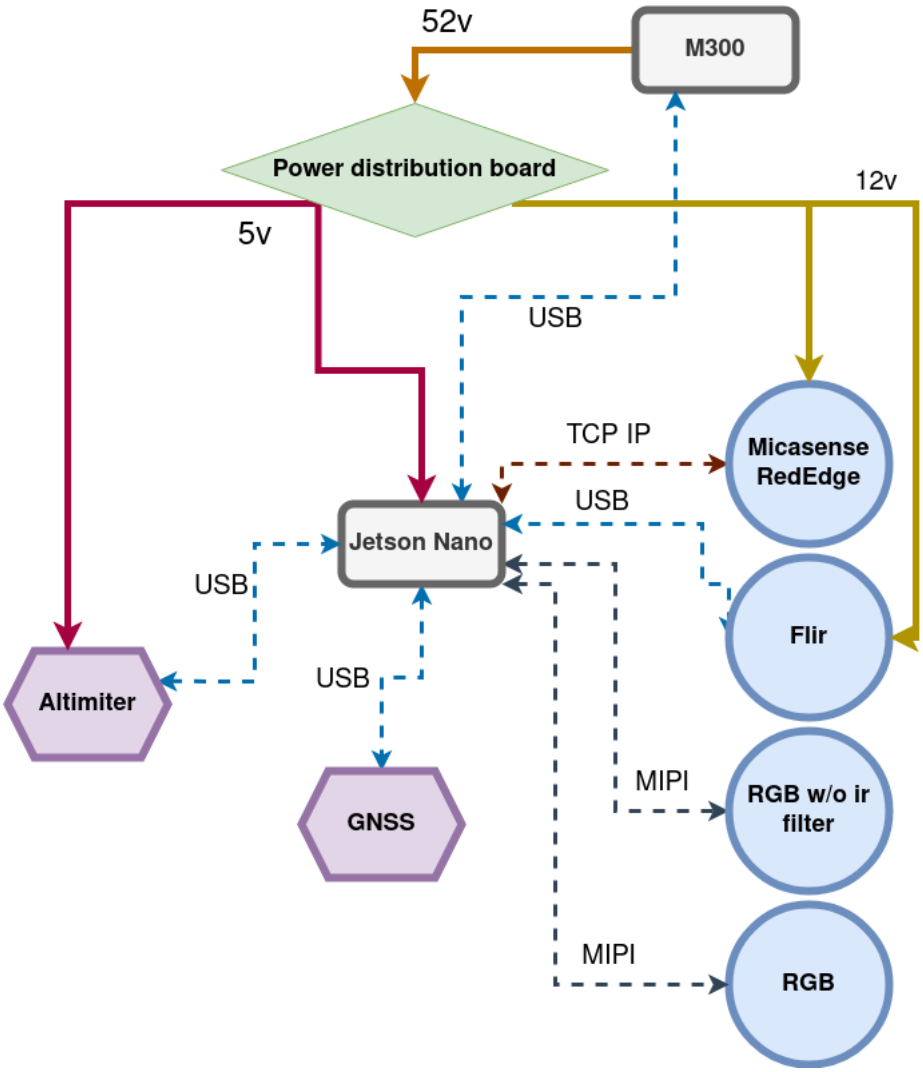

Figure 4.4: This is the electrical and logical wiring diagram of ASTRO. All sensors and systems are connected to the Jetson flight computer which is tasked with recording data from flight. The custom power distribution board has proved to be invaluable for ASTRO by providing stable power at different voltages for each attached system.

nominal performance even in wind speeds of up to 20mph, thanks to the robust features integrated by DJI. This stability enables the utilization of more sensitive imaging systems, such as the Micasense RedEdge, without concerns for image artifacts caused by an unstable platform thus reducing the need to gimbal sensors further simplifying the ASTRO system.

In addition to stability, the high payload capacity of the Matrice platform played a pivotal role in our mechanical design considerations. This capacity allowed for a less weight-sensitive design, facilitating the integration of many hardware components. As a result, we were able to incorporate diverse imaging systems, including the Micasense RedEdge, FLIR Boson, two Raspberry Pi cameras, a Jetson Nano, networking equipment, power distribution systems, radar altimeter, and a custom GNSS unit. Remarkably, this integration did not significantly impact flight endurance, and the onboard battery supplied sufficient power to operate all sensors effectively in flight.

Lastly, the Matrice platform's compatibility with ROS brought valuable advantages to our data collection process. The utilization of ROS enabled seamless data acquisition, including crucial parameters such as attitude, which is essential for data relation tasks. DJI exposes power and data through the first party "manifold" product which further validates the use of the M300 platform by eliminating the need for be-

[2]Flir: https://www.flir.com/products/boson/
Raspberry Pi cameras: https://www.amazon.com/Arducam-Camera-Raspberry-Interchangeable-M12x0-5/dp/B013JTY8WY
Nvidia Jetson Nano: https://developer.nvidia.com/embedded/jetson-nano-developer-kit
Radar altimeter: https://ainstein.ai/us-d1-all-weather-radar-altimeter/

spoke sensor batteries and enabling communication directly with sensors on the M300 platform.

In summary, the various attributes of the DJI Matrice M300 platform, including its stability, high payload capacity, and compatibility with ROS, significantly facilitated the development of our sensor rig and streamlined data collection. These platform characteristics alleviated unnecessary constraints and expedited our development efforts.

4.2.2 Package

A brief piece of context is necessary to motivate the ASTRO sensor package. The sensor package was initially designed with the overarching objective of supporting precision agriculture research. The intention was to create a sensor package that could encompass a wide range of precision agriculture research aspects, as well as address the general requirements of field robotics research.

As a secondary project goal, to reduce the need for specialized sensor equipment such as the Micasense RedEdge, a sensor package was devised that emulates the current state-of-the-art in imaging technology, but with a more cost-effective set of sensors. The aim was to enable the collection of as much, if not more, data while providing increased computational capabilities on the payload compared to conventional agricultural imaging systems.

The two main components of the sensor package that deserve attention are the imaging systems and the computer. The imaging systems consist of the Micasense

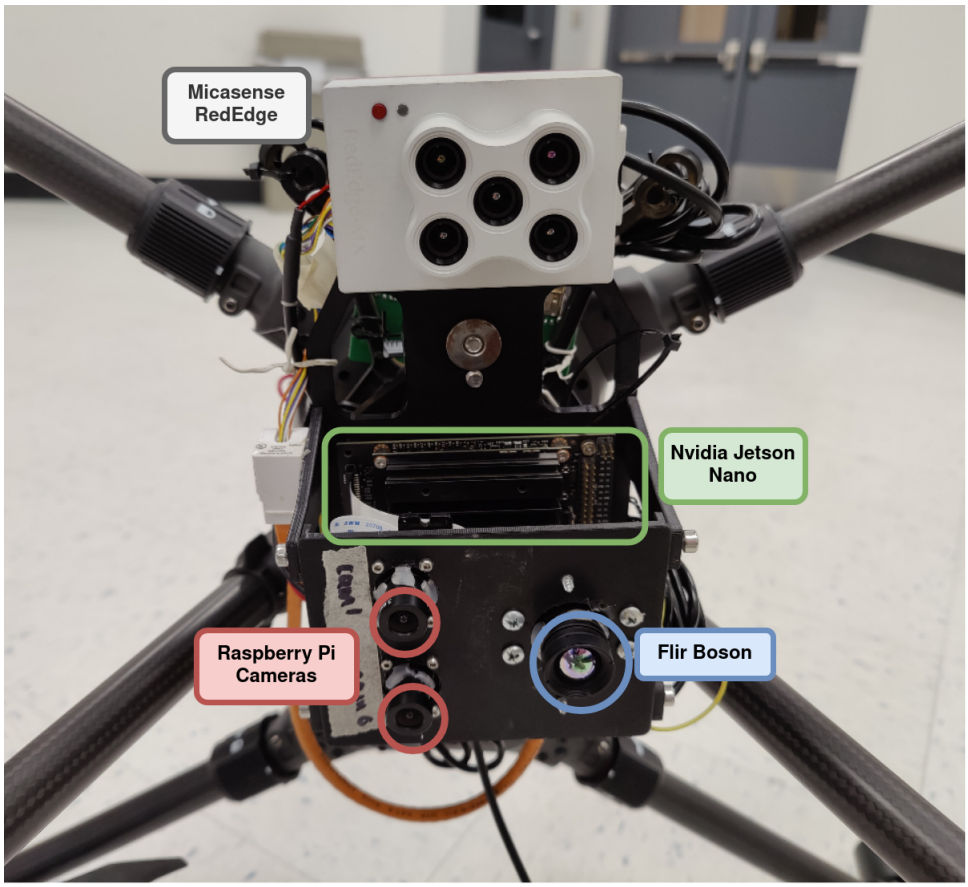

Figure 4.5: The ASTRO package depicted in the image showcases the collection of cameras mounted on the sensor head, emphasizing its role in data acquisition. Positioned at the top of the sensor head are the lenses comprising the MicaSense RedEdge imaging system.

RedEdge mx, which captures light in five spectral bands suitable for agricultural purposes, including near infrared for calculating NDVI. However, the high cost of dedicated agricultural imaging systems hinders their widespread adoption. As an alternative, a custom sensor configuration was developed using raspberry pi cameras and a FLIR Boson, which cover the same electro magnetic spectra as the Micasense RedEdge but at a reduced cost while offering utility for various tasks such as search and rescue and other development. To support the diverse objectives of the ASTRO project, the computer chosen for the sensor package is the Nvidia Jetson Nano. The Jetson Nano was selected for its favorable SWAP characteristics, as well as its hardware compute capabilities, including NV12 image encoding and support for MIPI cameras. The Jetson Nano facilitates image capture from multiple sources, such as MIPI cameras and the FLIR Boson, as well as the Micasense RedEdge. It also enables hardware-accelerated deep learning inference, allowing real-time analysis of photometric data for tasks such as determining plant health during flight. Additionally, the Jetson Nano's Linux and ROS compatibility opens avenues for further research in field robotics without modifications to the established sensor package.

The selection of these sensors and compute components allows for the collection of significant amounts of data that are valuable for generating large-scale, geo-referenced datasets, aligning with the project's overarching objectives.

4.3 BANNERS

BANNERS, in its most basic form is scaling up the classic computer vision problem of projecting points in 3d space to pixel space to create large datasets4.6. On this first iteration BANNERS makes a flat world assumption, using GNSS converted to UTM and neglecting any kind of surface normal estimation during data collection.

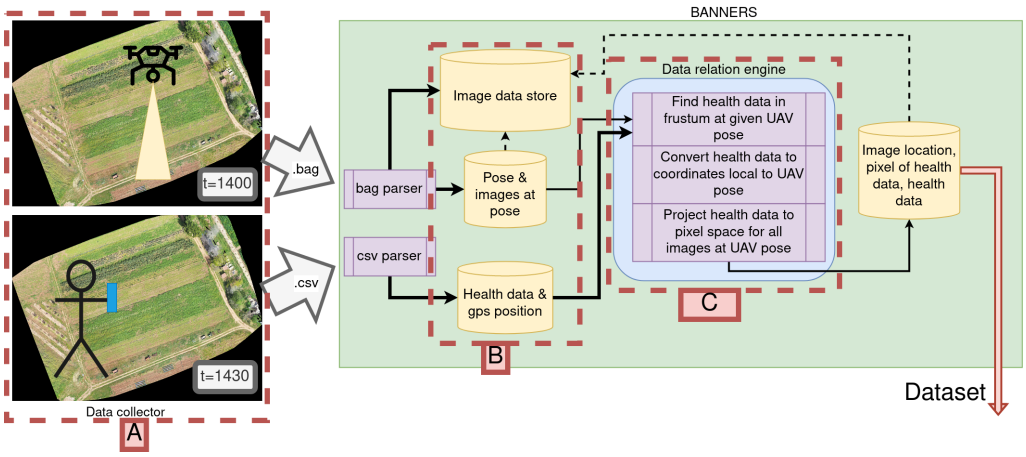

Figure 4.6: This is the full BANNERS pipeline for collecting and projecting data.

4.3.1 State estimation

State estimation for positioning plays a crucial role in the BANNERS project, serving as a hidden key to its success.

We adopted a simplified approach to pose estimation by combining the GNSS position of the robot with the attitude provided by the DJI platform4.8. This simplified scheme proved to be acceptable, owing to the impressive state estimation capabilities of the DJI platform and the high accuracy of our custom RTK GNSS solution. By

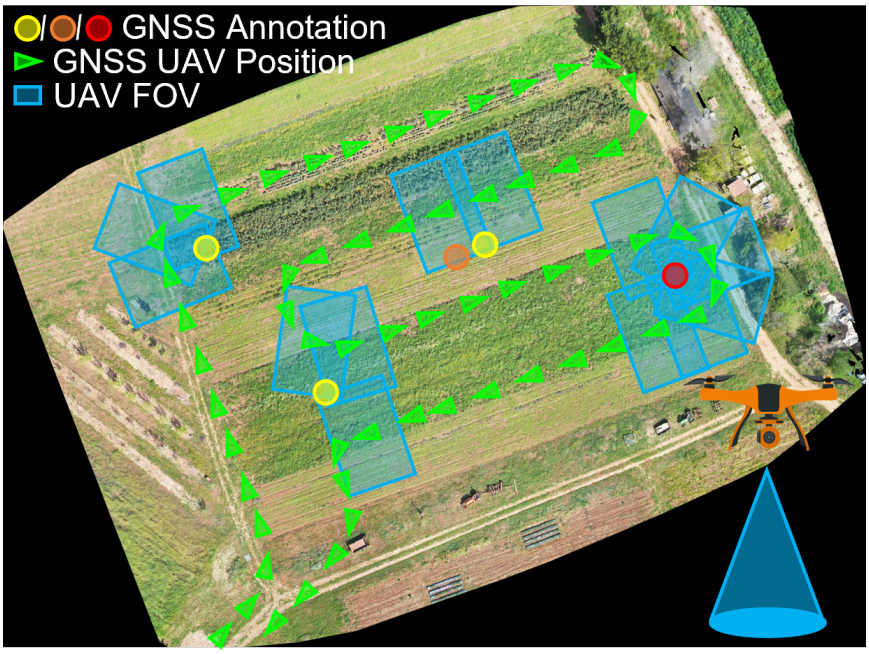

Figure 4.7: The concept of geographic co-location is illustrated here. In the context of agriculture a UAV collects data by flying over a crop field (shown in green). A human then collects data, from that same crop field the UAV flew over, using a data collection device. The data is then related after collection is complete. The annotations generated are common for deep learning algorithms such as Yolo, boxes are placed in image space with labels. The human walks around a field collecting data faster than they would analyzing and labeling data directly from the UAV.

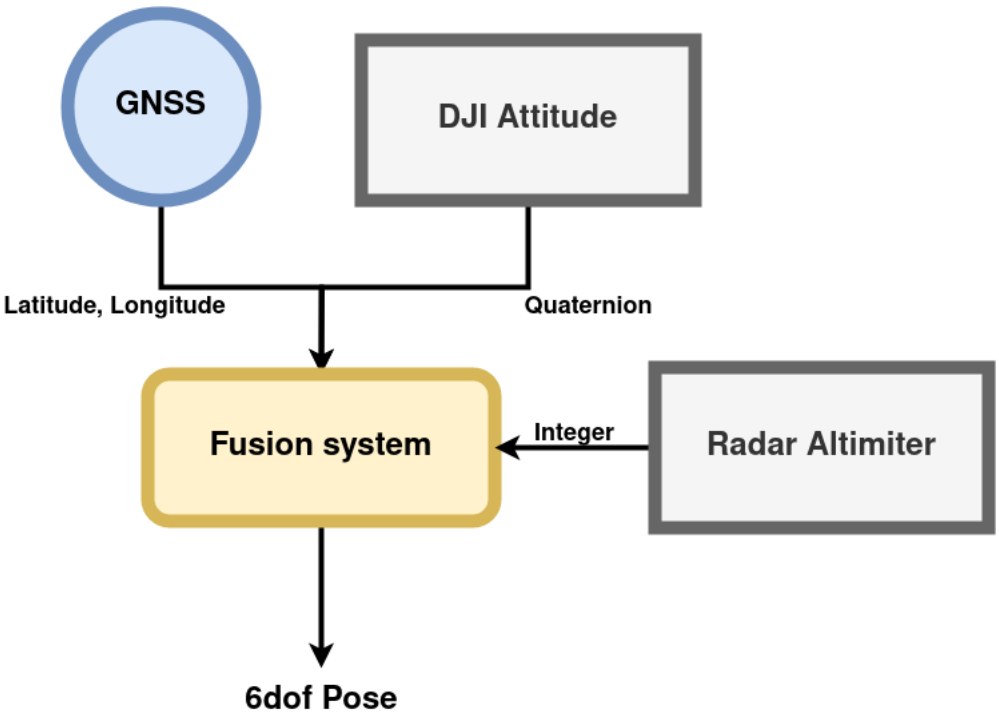

Figure 4.8: This is the simplified pose estimation system used on ASTRO. Data is merely concatenated together to produce high accuracy pose estimates. While these pose estimates are not generated in the traditional sense of raw sensors feeding a singular state estimation system to produce pose estimates, rather each sensor has its own estimation and filtering system internal to the device and our software composes these estimates into a singular or unified pose for use in BANNERS.

concatenating the GNSS position and attitude, we reduced computational overhead with minimal impact to data projection performance.

Time for ASTRO and BANNERS is provided by ROS and is based on system time. Each pose and image is provided with a timestamp by ROS automatically. These timestamps have accuracy to the nanosecond and are used as the basis of time for all time alignment tasks.

Latency issues arose later when working to correlate images and pose estimates in BANNERS, leading to significant errors in target predictions. To address this challenge, we reframed the processing step of BANNERS as a lookup problem, where all pose estimates were read in before processing the images. By doing so, we could find the pose closest in time to each image, ensuring a better match between pose and image. The same technique was applied to group images based on their closest time proximity. This approach proved effective in resolving tracking problems and achieving accurate results particularly during simple and consistent movements.

By emphasizing the importance of state estimation, we underscore its integral role in BANNERS. Accurate estimation of the UAV's state provides a solid foundation for data projection into image space, enabling precise data analysis and interpretation within the project.

4.3.2 Data relation

The data relation component of BANNERS utilizes SQLite, a mature database technology, to efficiently relate large quantities of data analysis. The use of SQL for

this purpose is supported by the simplicity of its interface and its ability to provide high-speed access to vast amounts of data. At commercial scales databases can contain terabytes of data, this scale has forced database systems to evolve and be as efficient as possible.

One of the advantages of using the simple SQL interface, is that BANNERS can perform mathematical calculations necessary to select data within a certain radius of the robot directly within SQL. This system of generating a radius and doing a lookup is not traditionally supported in SQLite but was added with python. It is worth noting that there exists better suited database systems for BANNERS with geospatial query capabilities built in, however those systems were left unexplored in this work. The geospatial query serves as a pre-filtering step to limit the number of matrix operations required later for projecting click data into pixel space, resulting in improved efficiency, especially for larger datasets. Additionally, the versatility of SQL enables easy substitution of the underlying SQL platform, such as replacing SQLite with SQL Server for industrial-scale applications or real-time data relation using internet-connected data collection devices.

SQL is specifically designed for handling large-scale systems and efficiently storing vast quantities of data. In field tests, ASTRO collects approximately 1,000 data points for every three acres surveyed, with flight data comprising 20 images per second between all of the sensors employed with associated poses. To elaborate on the 20 combined frames per second each low cost camera runs at 5 frames per second (fps), the flir also runs at 5fps but produces 16bit images, the Micasense runs at 1fps but each of

its frames comprises of five bands which are stored as individual images for use later. Considering a typical flight duration of 30 minutes for photometric data collection, this translates to approximately 9,000 database rows and 180,000 images. These numbers do not account for time series data, where the same flight and data collection may be performed on a weekly basis for months or years.

The ability of SQL to handle such large quantities of data aligns with the objective of BANNERS to rapidly generate large-scale geo-referenced datasets for supervised machine learning. SQL supports the system's capacity to accommodate the extensive data generated during land inspection at large scales, making it suitable for the proposed work.

4.3.3 Data projection

The process of projecting data from a geographic reference frame to pixel space, as performed by BANNERS, provides annotations for deep learning systems, eliminating the need for manual image labeling by humans, as commonly done with tools like "Yolo Label" [14] 1.1. This projection approach is preferred over direct image labeling due to its compatibility with human interpretation and its well-established nature.

At the scale of three acres, the ASTRO imaging system generates approximately 180,000 images that require labeling. Photographic data, particularly in the context of precision agriculture, often captures subtle information that is challenging to identify directly from the images alone. For instance, accurately identifying a plant with rust disease from UAV photographs can be difficult4.9. On the other hand, hu-

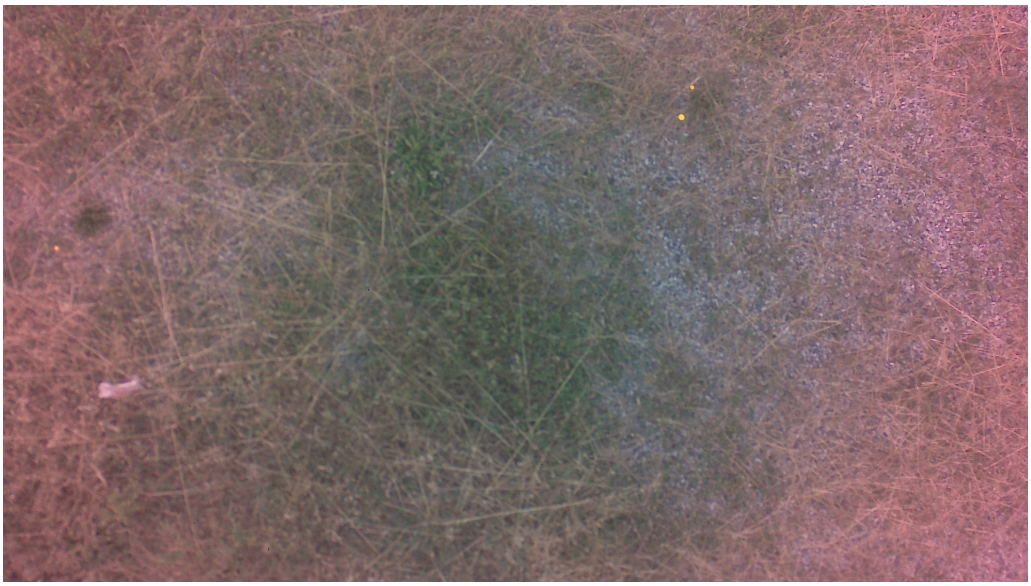

Figure 4.9: This is a sample of data taken from ASTRO viewing a patch of shrubbery. This example image is to highlight the difficulty of inferring plant health from aerial imagery with low cost sensors.

mans are more adept at walking through a field and collecting data on specific plant conditions or diseases, making their input valuable for generating labeled training data. In the initial version of BANNERS, the process of projecting data into pixel space is relatively simple, assuming a locally flat world model4.10. This simplification allows for straightforward mapping of geographic coordinates to pixel coordinates, enabling the generation of annotations for deep learning systems.

The process of projecting data into pixel space is simple in this first version of BANNERS where a flat world assumption is made. The transform from world to camera is captured in $\boldsymbol{A}_{\boldsymbol{wc}}^{4x4}$ which is a 4x4 matrix describing the 6dof pose of the camera. The transform from the world reference frame to the robot reference frame is $\boldsymbol{A}_{\boldsymbol{wr}}^{4x4}$, this again describes the 6dof pose of the robot in the world reference frame. The transform

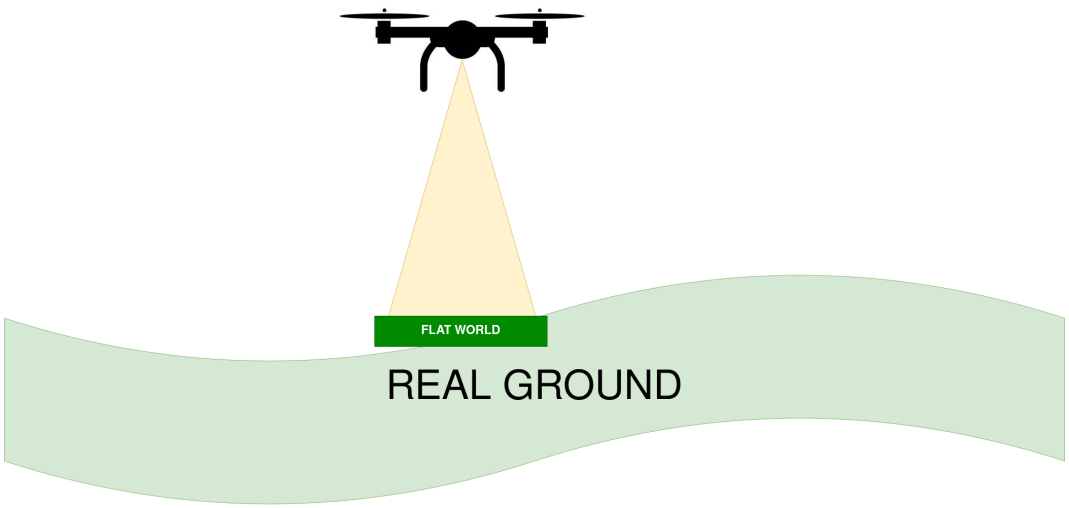

Figure 4.10: With the current data relation system there is no topographical information for the flown area such as a DEM or mesh which means that BANNERS makes a locally flat world assumption or that every viewable pixel is at the same distance from ASTRO.

from the robot reference frame to the camera reference frame is described by $\boldsymbol{E}_{\boldsymbol{rc}}^{4x4}$ also known as the extrinsics of the camera in the robot reference frame.

$$\boldsymbol{A_{wc}} = (\boldsymbol{A_{wr}} \cdot \boldsymbol{E_{rc}})^{-1} \tag{4.1}$$

To simplify the math padding is removed from the world to camera transform to create $\boldsymbol{A}_{\boldsymbol{wc}}^{3x4}$. The world to camera transform is then composed with the camera instrinsics matrix to project world points into pixel space.

$$\boldsymbol{x_s} = \boldsymbol{K} \cdot \boldsymbol{A_{wc}} \cdot \boldsymbol{x}_{\boldsymbol{w}}^{T} \tag{4.2}$$

Click data is collected from the clicker unit and converted to UTM from decimal latitude and longitude for ease of use. Using UTM only describes two dimensions of the click

position and leaves the click data under defined for the technique presented. Padding is used on the click data in order to comply with the rank requirements of the world to camera matrix. A altitude or z value is selected to be near but non zero which fits with the simplified flat world model in use by BANNERS.

$$x_w = \begin{bmatrix} x_0 & x_1 & ... & x_n \\ y_0 & y_1 & ... & y_n \\ 0.001 & 0.001 & ... & 0.001 \\ 1 & 1 & ... & 1 \end{bmatrix}$$

A sample set of points, with padding to support the flat world assumption in BANNERS, is presented. To avoid issues related to infinity during the projection to 2D pixel space, a small but non-zero value is assigned to the altitude or "z" component of each point. These operations are commonly employed in computer vision and involve minimal computational cost, with the matrix inversion step driving the time complexity calculation.

The human compatibility aspect of BANNERS is closely tied to the faster data labeling process. In field tests, two humans each equipped with data collection devices were capable of collecting data across a three-acre space within two hours. Subsequently, the projection and relation of the collected data took only a matter of minutes. This process scales linearly with the number of data collection devices, enabling even faster data labeling. Further reduction in labeling time can be achieved by engaging subject matter experts who are familiar with the technologies and designs commonly used in GIS work. By identifying phenomena of interest, humans can efficiently label data,

irrespective of the format or sensor type. For instance, the same data collection tool can be used to simultaneously label photometric and LiDAR data. As a result, georeferenced datasets can now encompass diverse forms of data and corresponding labels.

The speed improvements achieved through efficient data projection, coupled with the opportunities presented by this process, represent the primary practical innovation introduced by BANNERS. The ability to label large volumes of arbitrary georeferenced data efficiently, leveraging the expertise of subject matter experts, marks the advent of modern data labeling. This advancement significantly enhances the generation of labeled datasets, unlocking new possibilities for supervised machine learning in various domains.

4.3.4 Validation tests

To validate the functionality and performance of BANNERS and ASTRO, comprehensive end-to-end testing was conducted, focusing on key components that are fundamental to their operation. The following components were tested:

Camera intrinsics: The camera intrinsics were verified by utilizing the PNP (Perspective-n-Point) algorithm to estimate the camera pose based on known world points captured in pixel space. This test helped identify any inaccuracies in the estimation of camera intrinsics.

BANNERS coordinate frame: The expected coordinate frame of BANNERS, based on the Universal Transverse Mercator (UTM) system, was validated. This involved constructing a known coordinate frame using a flat reference surface with cal-

ibration points at known locations within the coordinate system. By correlating the estimated pose of ASTRO with the observed projection of world points, the accuracy of the coordinate frame was assessed.

GNSS RTK: The noise characterization of the GNSS RTK system was performed through a random walk test. The GNSS rover unit was subjected to controlled movements while RTK was engaged, aiming to evaluate the level of variance and predictability in the system. A lightweight GNSS antenna designed for aircraft was used for this test.

Camera transform chain: The chain of transforms between different coordinate frames, including the transformation from world to robot to camera, was tested. This involved examining how points in the real world would be projected into the BANNERS system as a whole.

Extrinsics estimation: Estimating the extrinsics, which involves determining the transform between the UAV's pose and the camera's pose, required a common reference frame. An April tag was utilized to provide both the position and orientation of the camera, with an offset defined by collecting a GNSS position measurment at the center of the April tag. The drone's state estimation directly provided its position and orientation. By aligning both components in the GNSS reference frame, extrinsics estimation was carried out.

End-to-end testing: End-to-end testing simulated field operations to evaluate the overall performance of BANNERS. Although a high-visibility target was used for validation instead of real-world scenarios, the test aimed to assess the operation of visual

detection and numerical prediction (projection).

These tests were designed to systematically validate the core components of BANNERS and ASTRO, ensuring the accuracy and reliability of the mathematical models at each stage of the process. Given that BANNERS operates within a global-scale reference frame, all subsystems, regardless of their size, must be properly related to ensure effective operation.

Chapter 5

Results & Discussion

5.1 Results

5.1.1 Camera intrinsics

The assessment of camera intrinsics aimed to validate the accuracy of projecting coordinates from the world reference frame to pixel coordinates.

To accomplish the estimation of camera intrinsics, the widely recognized camera calibration tool, Kalibr, developed by ETH Zurich[42], was employed. Kalibr facilitated the estimation of the camera intrinsic matrix, denoted as $\boldsymbol{K}$, along with the camera distortion coefficients. A comprehensive accuracy report is generated by Kalibr, providing insights into the precision of the intrinsic parameter estimation using a radial tangential distortion model and a pinhole projection model. The radial tangential distortion model was chosen to more accurately model and correct the lens distortions introduced by the low cost lenses. In addition to kalibr, a secondary test was con-

ducted to verify the success of the calibration process. This evaluation involved the employment of a large wall as a reference, a rigid mount for ASTRO 5.1, and an independently estimated pose of ASTRO. The purpose of this test was to confirm that the calibration procedure yielded satisfactory results in terms of accurately projecting world coordinates onto the pixel coordinates of the camera5.7. The necessity for good camera intrinsic estimation seems obvious within the robotics community, good camera calibration supports every other computer vision task but camera calibration has new stakes on ASTRO and BANNERS. Computer vision is like any other data processing in the respect that sub-optimal input data leads to sub-optimal processing results and can break entire systems. In BANNERS, this dependence on calibration quality is compounded by the relatively vast distances the system is intended to operate over. A minor error in camera calibration propagates on top of errors in position estimation and imprecision of coordinate transform estimates. We calculate the error that a single pixel makes for the accuracy of BANNERS at an altitude of 10m without any other systematic errors. ASTRO's low cost imagers have a horizontal resolution of 1920 pixels. At 10m agl we find that the horizontal area seen in an image is 5.14015m. This converts to 0.26cm per pixel of resolution along the horizontal and vertical axis. Along a 45° the distance is 0.36cm per pixel. In an example case with a bad calibration where projection error is 1.7 pixels, which is a realistic projection error with low cost cameras, an error of 0.612cm is incurred when BANNERS projects data. This error is further compounded by any inaccuracy in the positioning estimates and any errors in the attitude estimate.

Attitude errors are even more pronounced at 10m agl.

$$tan(1^\circ) = d/10$$

$$10 * tan(1^\circ) = d$$

$$0.17 = d$$

With a 1° error in attitude estimation we see5.1.1 that there is a minimum horizontal displacement of 0.17m when the projected target is directly underneath ASTRO, as the data moves to the edges of the field of view of ASTRO's sensors the error increases further. This error becomes an issue when estimating the transform from the camera body frame to the robot frame where single centimeters matter. By improperly calibrating the camera intrinsics errors compound through multiple avenues both in other calibrations as well as in final data projection. In the pursuit of perfect data projection we worked to minimize camera intrinsics estimation error which was made particularly difficult with the low cost imaging systems employed by ASTRO.

The decision to validate the output of Kalibr through an independent test proved to be a prudent choice, as it revealed an error in the initial calibration. Despite Kalibr being widely accepted as the standard camera calibration tool in the academic community, our prior experience with the low-cost cameras utilized in ASTRO led us to conduct an additional verification step. The estimated intrinsics for the raspberry pi cameras often produce flawed but plausible estimates which makes verifying the estimates of intrinsics that much more difficult. To verify the accuracy of the camera intrinsics estimated by Kalibr, we employed a PNP solver on a target with known

Figure 5.1: This test was part of confirming the camera intrinsics estimates from Kalibr. The robot is held rigidly within a well defined coordinate frame to remove run to run variance. PNP was then used to solve for the pose of the drone given well defined world points as reference. The mount also allowed us to confirm the pose estimated by PNP by making the pose of the drone obvious.

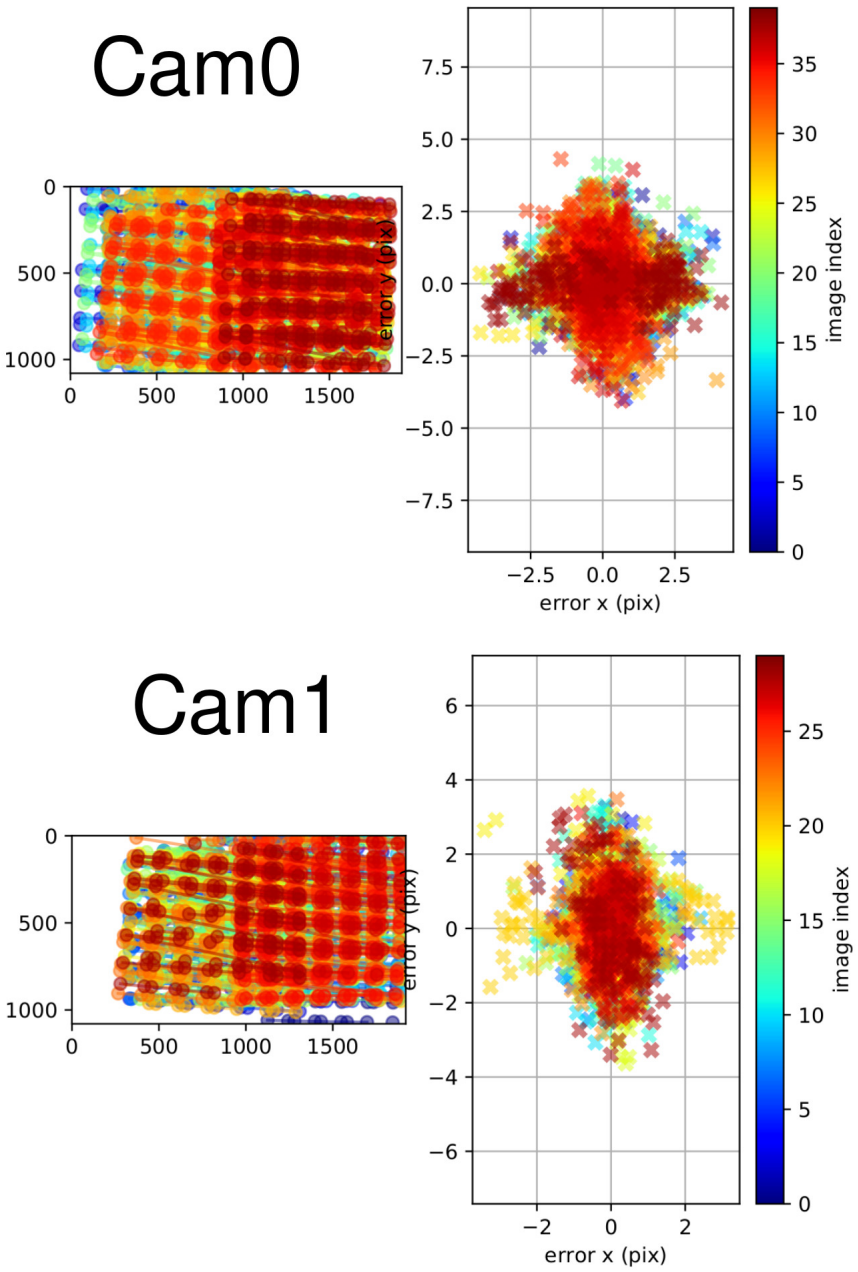

Figure 5.2: This is the report generated by Kalibr as a result of the calibration process describing the accuracy of the estimates. It is worth establishing that the calibration result shown is taken from the ASTRO package but is not the calibration in use. Rather this image further demonstrates the challenges of utilizing low cost imaging system for robotics applications. The main challenge they pose is one of quality where each calibration attempt delivers vastly different results ranging from acceptable to terrible.

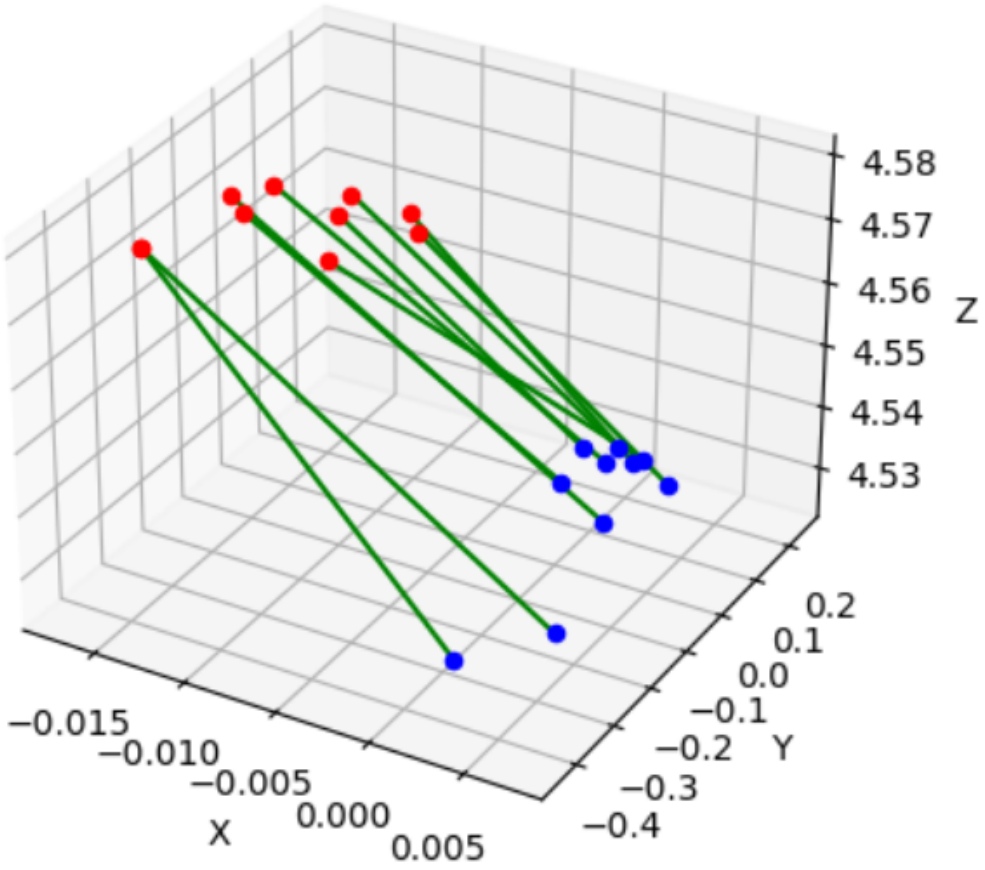

Figure 5.3: Pictured is the pose of ASTRO as estimated by PNP using an april tag as the reference geometry. The altitude (z) has 5mm of variance through 0.5m of translation. The red and blue dots signify the two separate m12 cameras in use while the green lines show correspondence between estimated poses at the same time step. The displacement in X and Y is within 3mm of measurements made by hand with a caliper. While vertical discrepancy between each camera looks significant and should be zero the difference is systematic and within 5mm at an altitude of 4.5m which we deemed acceptable. This was a plot produced during camera to robot transform estimation also known as extrinsics.

geometric properties. Through the PNP test, we successfully confirmed that Kalibr's estimation of the camera intrinsics was accurate after multiple rounds of calibration. The results obtained from the PNP test aligned closely with the ground truth pose estimation, which was determined using a tape measure. The discrepancies observed were within the acceptable margin of error 5.4.

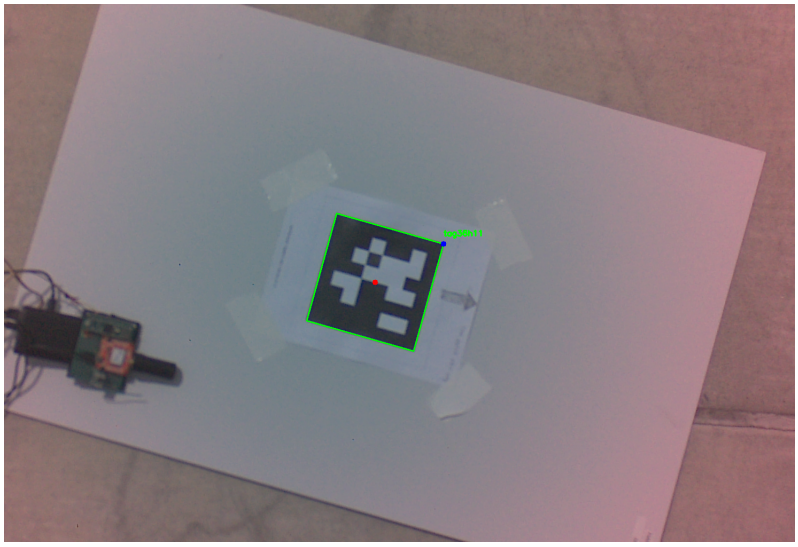

Figure 5.4: This is a sample of feature extraction performed on an april tag, useful for automatrically and reliably generating reference points that can be fed into PNP for estimating camera pose. The use of april tags for landmark generation and to feed PNP based camera pose estimation was reached after other attempts to use landmarks on a wall yielded highly variable results.

By establishing the reliability of the camera calibration, we have established a solid foundation for subsequent pose estimation and projection tasks in the context of BANNERS. This confidence in the accuracy of camera intrinsics estimation allows us to proceed with further development and application of the project.

5.1.2 GNSS RTK

The precise positioning afforded by RTK allows for projection of data into image space with minimal error.

RTK GNSS serves as the foundational pillar for the BANNERS system, providing high accuracy positional information for precise data projection and relation later in BANNERS. To achieve RTK, a base station was employed to gather ionospheric observations and generate RTCM (Radio Technical Commission for Maritime Services) data. RTCM data is indispensable for the successful implementation of RTK. However, the RTCM data that enabels RTK also poses its own difficulties while while transmitting to the rover unit at optimal data quantity and minimal latency. The quantity and latency of RTCM data significantly impact the RTK fix times and reliability. Achieving improved RTK performance necessitates a balance between the requirements of RTCM and the capabilities of the network solution employed to transmit RTCM data from the base station to the rover. In RTK systems, a minimum data rate of 300 bits per second (bps) with a maximum latency of 3 seconds is typically specified. However, in the context of robotics applications, our experimental findings indicate that better results are achieved with a minimum data rate of 625 bps with a maximum delay of 1 second. Through extensive testing, we discovered that these experimentally determined minimum data requirements align well with the limitations of the LoRa communication protocol operating at a bandwidth of 125 kHz, a spreading factor of 7, and a coding rate of 5. This combination of settings on the LoRa transceiver modules offer a suitable range

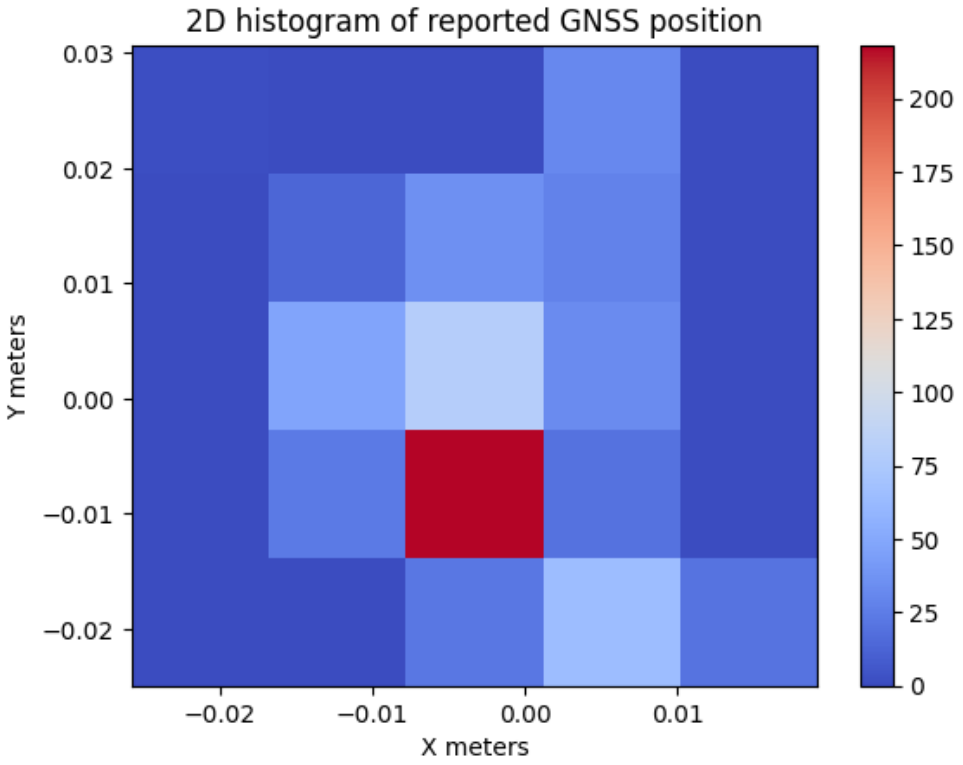

Figure 5.5: This histogram illustrating the density of recorded locations. The distribution exhibits a unimodal pattern, which aligns with the accuracy reports provided by the GNSS receiver, indicating consistent and reliable location data. The unit reports 1.4cm of variance on its position estimation, this is confirmed by a random walk test with RTK engaged. Left on the ground with a wide sky view the GNSS unit is allowed to collect data, we see a high density of position estimates at a single location. We calculate the standard deviation of the data to be $6.8E-6$ by first interpolating the recorded data with a mesh grid in 2d, then estimating the gaussian kde. We then derive the standard deviation from the covariance matrix associated with the kernel given by the gaussian kde.

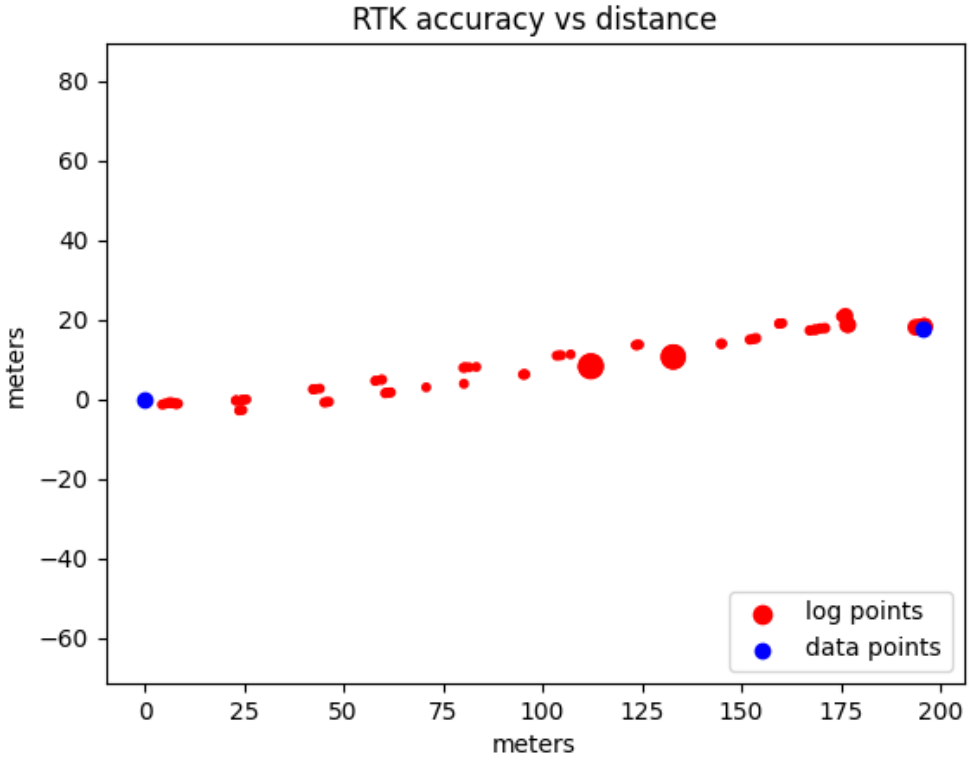

Figure 5.6: This plot illustrates the positional data and corresponding accuracy of the rover unit during RTK positioning. The plot is presented on a meter scale with the origin positioned at 0,0. Notably, an observation of RTK positioning loss is discernible at a distance of 200 meters, even during consistent walking speed conditions.

for LoRa communication while ensuring desirable RTK characteristics from the GNSS system. By optimizing the RTCM data flow with higher data rates and reduced latency, BANNERS can leverage the benefits of RTK positioning for enhanced performance in robotics applications.

By means of extensive experimentation, we have successfully accomplished RTK GNSS positioning via a LoRa communication link tailored for field robotics applications. The attained positioning results exhibit a high degree of accuracy while preserving optimal usability within the context of robotics.

5.1.3 Camera chain

The camera chain test was employed to characterize the coordinate system internal to BANNERS. This testing generated insights to the egocentric design assumptions of the math within BANNERS.

To ensure the accuracy and integrity of our mathematical computations and data flow, we conducted a simulation of the complete chain of transforms, encompassing the conversion from world coordinates of test points and the robot to image coordinates. While this test did not constitute a full end-to-end evaluation, its purpose was to validate the correctness of our mathematical operations and data formatting. By utilizing well-defined coordinates for the target and drone poses, we focused on ensuring the accurate functioning of each component, correct mathematical calculations, and proper data formatting. This test served as a valuable means of identifying and resolving issues such as matrix inversion and vector transpose errors that had arisen during earlier

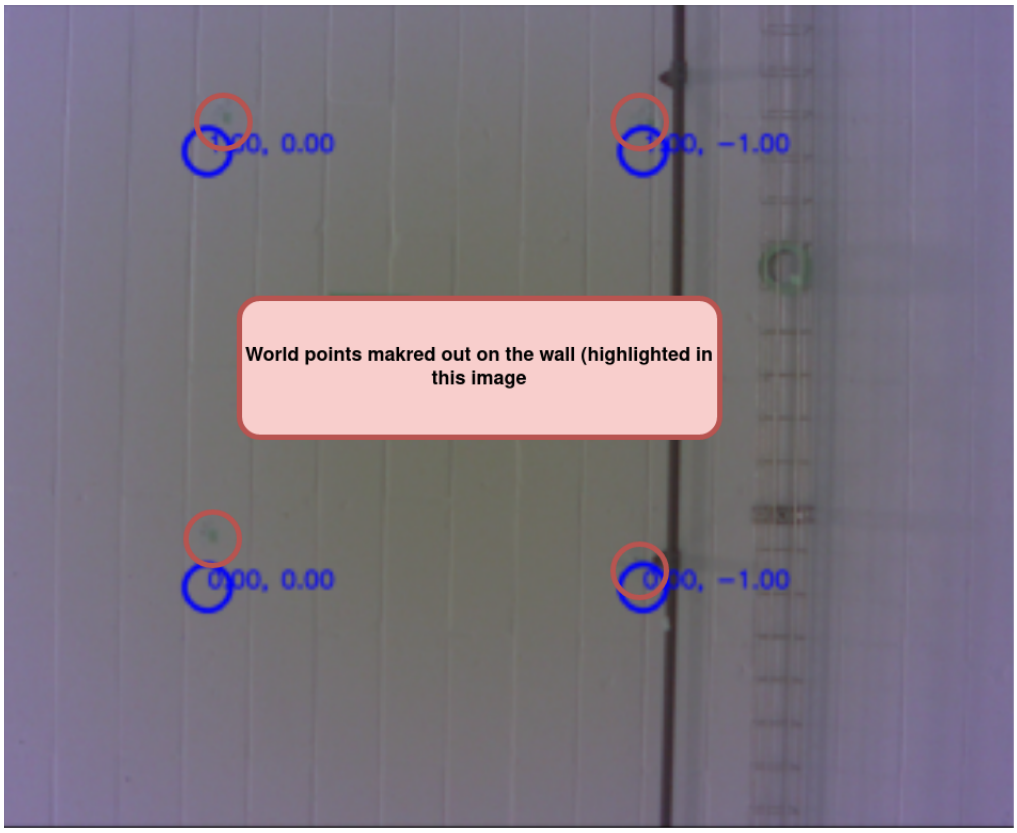

Figure 5.7: This is a wall test used to confirm assumptions and math used in BANNERS. By rigidly controlling the robot relative to a known coordinate frame test points could be easily projected into the image space. This basic test illuminated incorrect assumptions made by BANNERS pertaining to orientation.

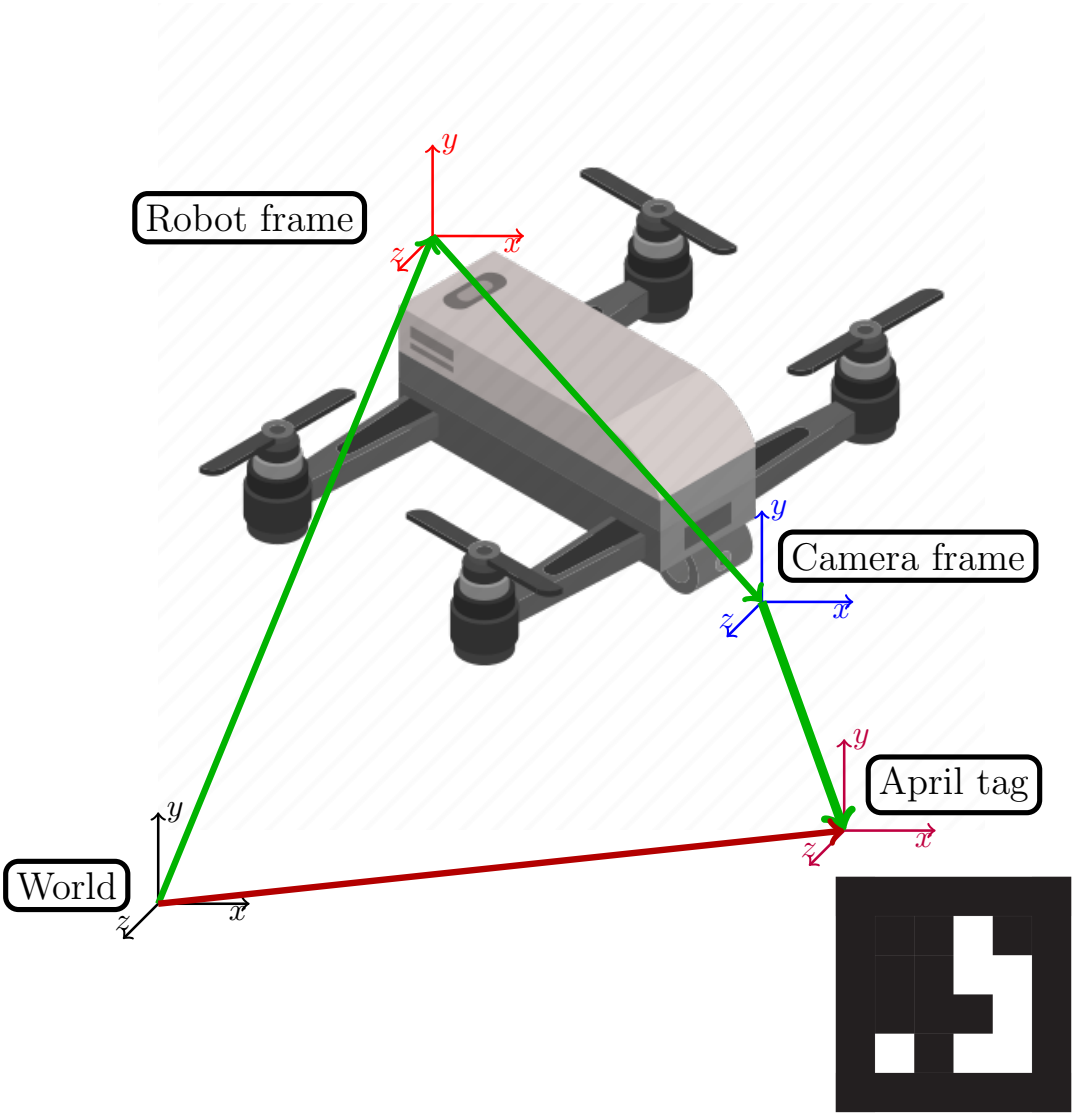

Figure 5.8: This is the functional transform chain to translate from world coordinates to pixel coordinates. The method tested here is highlighted with green arrows. The point of interest are also well known in the world coordinate frame as denoted by the red arrow.

stages of development.

By fully comprehending and refining the camera chain, we establish a solid foundation for the subsequent processes of data relation and projection. Through the verification of the correctness and effectiveness of our transform chain and projection mechanisms, we can confidently assert the integrity and accuracy of BANNERS' data piping for data relation and projection tasks.

5.1.4 Extrinsics estimation

Accurately determining the transformation between the sensor frame and the robot frame is of utmost importance to ensure accurate projection of data between image coordinates and world coordinates.

To validate the proper estimation of extrinsics, the calibration process involved solving for the transformation between the 6-degree-of-freedom (6DoF) pose of the robot and the 6DoF pose of the target. This involved aligning the coordinate systems of the robot and the target by estimating the appropriate transformation parameters. By establishing this precise relationship, the extrinsics estimation step contributes to the overall accuracy and reliability of the data projection process.

The camera-to-robot transform estimation process yields results that fall within the expected range, corroborating the accuracy of the obtained measurements compared to manual measurements. This extrinsics calibration procedure shares similarities with those employed in spacecraft missions, where a prominent target is positioned at a precisely known location and orientation. In a similar manner to the method employed

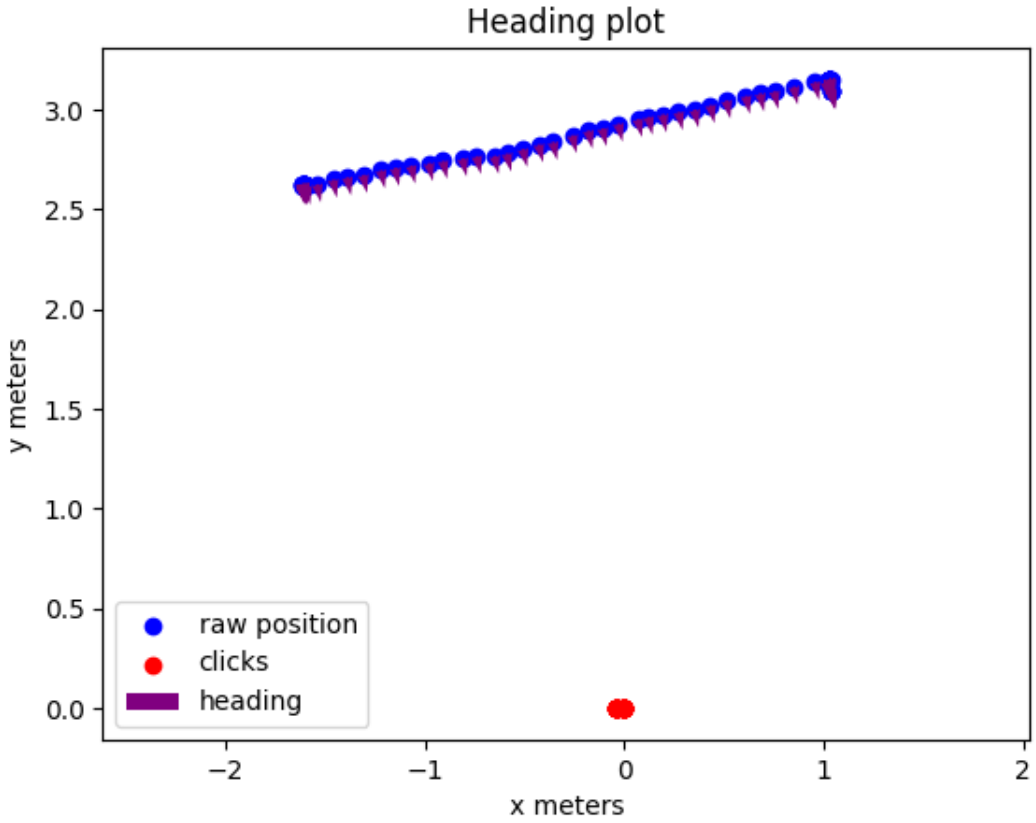

Figure 5.9: This plot illustrates the heading and position of ASTRO during the data collection process, specifically employed to validate the accuracy of heading estimates provided by the DJI system. The plot showcases the trajectory and orientation of ASTRO as it captures data, allowing for a comprehensive assessment of the consistency and reliability of the heading estimates generated by the DJI system.

Figure 5.10: The provided image showcases the test rig employed to comprehensively simulate the flight of ASTRO, serving purposes such as calibration and debugging. ASTRO is securely affixed to the side of the mast lift, enabling it to be maneuvered in close proximity to a designated target, thereby facilitating the collection of consistent and reliable data.

for estrinsics estimation, a known target geometry placed in a well-defined position and orientation facilitates the estimation of the camera-to-robot transform. This approach ensures a reliable and consistent calibration process that later enables accurate projection of data in BANNERS.

$$\boldsymbol{P}_{\boldsymbol{cw}}^{4x4} = \boldsymbol{T}_{\boldsymbol{rc}}^{4x4} \cdot \boldsymbol{P}_{\boldsymbol{rw}}^{4x4} \tag{5.1}$$

In the context of estimating the pose of the camera in the world coordinate frame ($\boldsymbol{P_{cw}}$), the camera pose is determined using the PnP solution. The obtained camera pose is then offset by the known coordinates collected at the center of the target. This adjustment ensures that the pose of the camera is directly estimated in the world coordinate frame.

To calculate the transform from the robot coordinate frame to the camera coordinate frame, a simple matrix inverse operation can be applied to both sides of the equation 5.1.

$$\boldsymbol{P_{cw}} \cdot \boldsymbol{P}_{\boldsymbol{rw}}^{-1} = \boldsymbol{T_{rc}} \cdot \boldsymbol{P_{rw}} \cdot \boldsymbol{P}_{\boldsymbol{rw}}^{-1} \tag{5.2}$$

$$\boldsymbol{P_{cw}} \cdot \boldsymbol{P}_{\boldsymbol{rw}}^{-1} = \boldsymbol{T_{rc}} \tag{5.3}$$

The transformation matrix $\boldsymbol{T_{rc}}$ between the robot frame and the camera frame is directly solved for using the pose of the robot ($\boldsymbol{P_{rw}}$). To mitigate errors arising from positional measurements of the known geometry and inaccuracies in the PnP estimation of the full 6DOF position, multiple measurements are utilized. The computed $\boldsymbol{T_{rc}}$ from

multiple views is averaged to establish a reasonable baseline, which is later reflected in the accuracy of data projection and the overall performance of BANNERS. A simple averaging approach was taken due to the small quantities of calibration data available while other techniques should be explored in future work.

By calibrating the extrinsics in the world coordinate frame, BANNERS achieves the crucial task of accurately relating the pose of the drone to the pose of the camera. This relationship facilitates precise data projection into image coordinates, thereby supporting the generation of accurate and reliable datasets.

5.1.5 End to end

The conducted end-to-end tests serve the purpose of integrating all the previously discussed systems and verifying the accurate projection of a point in the world onto an image.

The end-to-end testing conducted on well-structured data, featuring simple movements and precisely defined target geometry, provides evidence of the acceptable accuracy achieved under such conditions.

By subjecting the system to this comprehensive test, we validate the quality and accuracy of various components, including pose estimation, extrinsics estimation, camera intrinsics estimation, and the correct integration of mathematical operations to combine multiple reference frames for data projection. The successful outcome of this test supports the claim that BANNERS, in conjunction with ASTRO, has the capability to rapidly and automatically generate large-scale datasets of geo-referenced data.

Figure 5.11: The figure illustrates the targeting precision of BANNERS during a simplified test scenario. The circular target accurately traces the center of the April tag throughout a primarily linear translation motion.

The result of the end-to-end testing shows near perfect projection performance 5.11. The data projected by BANNERS almost perfectly tracks the center of the april tag in this test. This test was the very first validation of the BANNERS system.

Furthermore, the results of the end-to-end test establish the system's proficiency in interpreting the data acquired by ASTRO, leading to accurate projection of the collected data into image space. This achievement reinforces the thesis statement that BANNERS can effectively process and project data, facilitating the generation of precise and reliable geo-referenced datasets.

5.1.6 Accuracy characterization

The accuracy of BANNERS is measured to characterize the accuracy of the system and inform its usefulness as a tool for generating labeled data for supervised machine learning. BANNERS meets our criteria for accuracy which allows for precise label generation.

The goal of being within 1m of accuracy was picked for two reasons, the first being that 1m is convenient. The second is that 1m of error at the scale of 1 acre represents an error of 0.0008% which only shrinks as scale increases. BANNERS and ASTRO were designed with a 30acre farm in mind, this brings an entire meter of inaccuracy to be 0.00003% of inaccuracy at the scale of a small farm.

BANNERS was tested on multiple scales of data collection. Starting with a short flight to demonstrate the general concept. This flight served as a calibration and validation flight internally5.13. The medium test was performed to show that ASTRO could report its position accurately over long periods of time thus enabling proper data relation for BANNERS5.12. Finally a long test flight was performed to test data relation, pose estimation over time, ability to compensate for different view angles, and finally the computational run time of BANNERS5.14. The longest flight takes a considerable time to process with banners, however this longest flight took over sixteen minutes to fly.

Test length (m)	mean error (m)	median error (m)
16	0.117	0.10
305	0.41	0.39
932	0.60	0.51

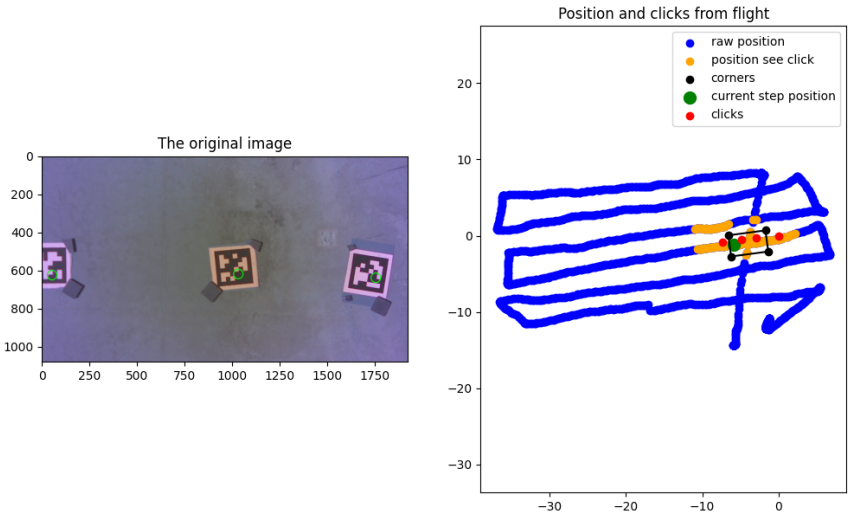

Figure 5.12: This is sample data from the medium test flight. A simple raster flight pattern is flown over a set of targets. This raster flight pattern is used for agricultural flights as well.

BANNERS and ASTRO project data into image space properly with sufficient accuracy. With this sufficient accuracy data labeling can be built effectively.

BANNERS and ASTRO meet the specified objectives set out at the start of this work and are now tested systems capable of supporting the rapid generation of datasets.

5.1.7 Example labeled data

BANNERS has already been developed to encompass the two primary labeling techniques used in image data analysis, namely pixel-wise classification and object detection.

The manifestation of semantic data labels presented by BANNERS represents

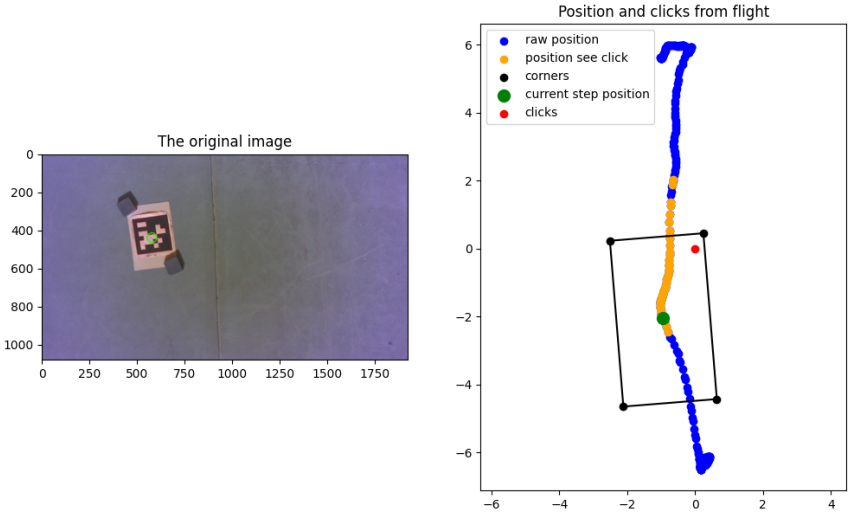

Figure 5.13: This is sample data from the short test flight. ASTRO merely flew directly over an april tag and collected data. We see near perfect tracking on this short and simple flight.

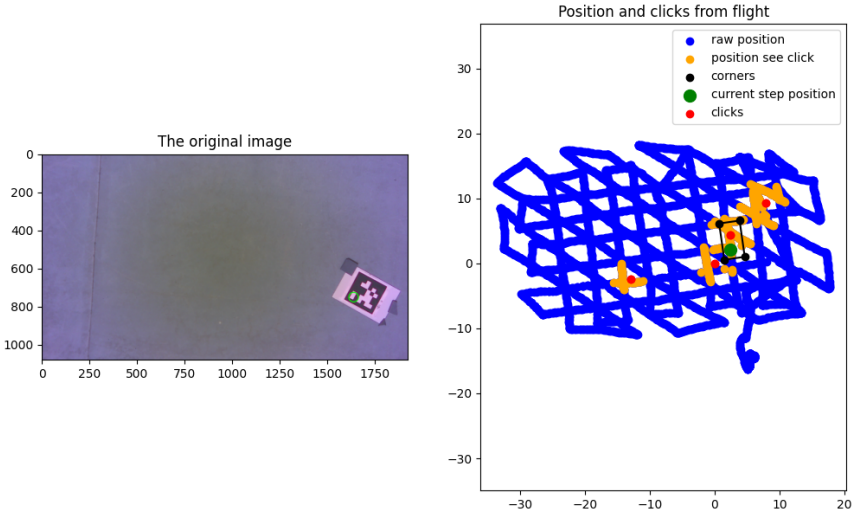

Figure 5.14: This is sample data from the long test flight. On the right the flightpath shows the raster pattern on multiple axis.

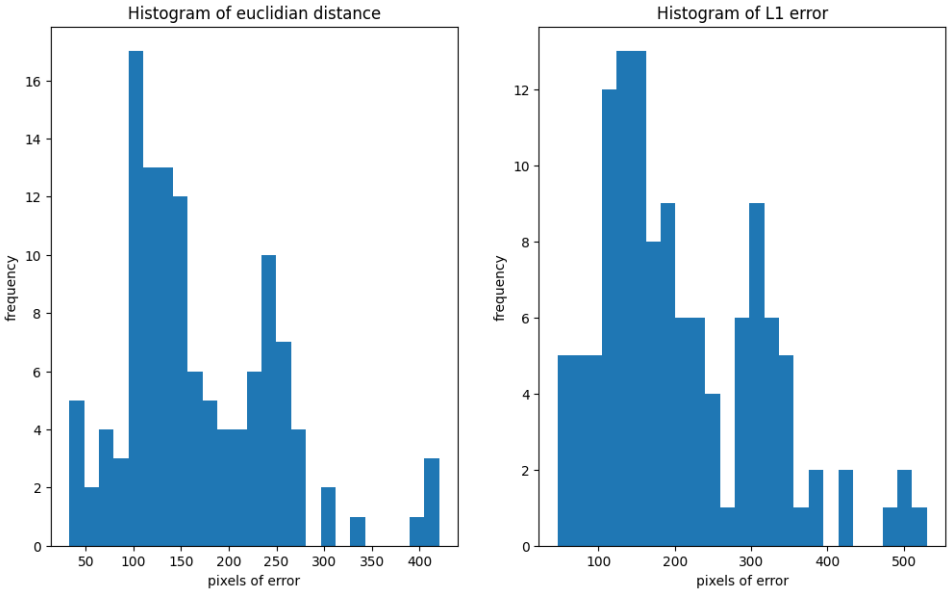

Figure 5.15: This is sample report from the longest test. The BANNERS testing software analyzes each image where a data point is seen and computes the distance from the projected point to the center of the nearest april tag. This method was selected for its reliance entirely on image data and the ease of april tag comprehension.

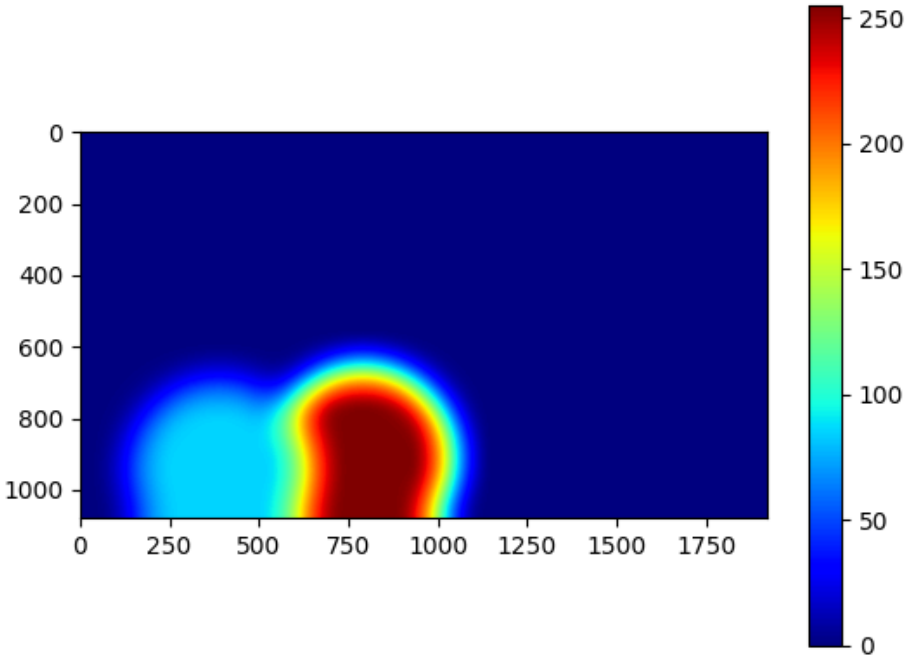

Figure 5.16: The depicted image demonstrates the utilization of a pixelwise mask to represent plant health as a continuous spatial attribute. The assumption is made that plant health is evaluated based on an epicenter and gradually diminishes from a sick state to a healthy state. This particular data sample has been selected to showcase two prominent modes of plant health.

a single instance of the potential applications for this framework. In this specific illustration, the focus lies on the semantic labeling of crop health data. The underlying assumption in this labeling context is the continuity of crop health, which is most suitably characterized by ensembles of Gaussian distributions. This utilization of Gaussian-based disease representation serves to reinforce the observation that crop stress exhibits a continuous nature and emanates outward from distinct focal points.

Pixel-wise classification involves assigning class labels to individual pixels in an image. BANNERS has the capability to perform this labeling technique, allowing it to categorize each pixel based on the specific classes or categories defined within the labeling task. This enables detailed and precise classification of image regions at the pixel level.

In addition to pixel-wise classification, BANNERS also supports object detection, which involves identifying and localizing specific objects within an image. With object detection, BANNERS can detect and outline the boundaries of objects of interest, providing valuable information about their positions and extents in the image. This enables more comprehensive analysis and understanding of the image content.

By encompassing both pixel-wise classification and object detection, BANNERS offers a versatile labeling framework that can handle a wide range of image analysis tasks. These techniques provide the foundation for accurate and detailed annotation of image data, facilitating subsequent machine learning algorithms and applications that rely on labeled datasets.

5.2 Discussion

5.2.1 Human factors

An aspect that is highly subjective but highly consistent in its reporting is that of human factors. Mentioned earlier in this work, the goal of BANNERS was to be less taxing for the human experts to generate data labels. It seems frivolous to avoid the most simple and straightforward data labeling technique like that found in yololabel 1.1 but from a human factors perspective this change makes an enormous difference in the quality of life for the human experts employed to label data. It is generally regarded within the deep learning community that data labeling is akin to back breaking manual labor. The work of labeling image data is not technically hard, draw a box around something within an image or paint some sections of an image. The task is simple but when repeated for hundreds or thousands of images the tedium takes its tole. The mind numbing mental labor and need to sit focusing on a screen for hours is often regarded as a right of passage for those within the robotics field. The shared misery is a badge of honor demonstrating ones love for the pursuit of robotics over riding the pain induced by data labeling. Divorcing the community from this self-flagellation and instead working within fields collecting data is a marked improvement. It is worth noting that labeling the data is still tedious and time consuming, however the approach demonstrated here is an improvement.

Time taken to label data is reduced taking only 8 total hours of labor to label three acres of data equating to nearly 200,000 images from ASTRO. This labeling time

Figure 5.17: This image was taken from a field test for future work of labeling agricultural data. The image is taken at the research farm at the University of California Santa Cruz. This image was taken as the team was labeling winter rye with a different form-factor revision of the data labeling tool than what is presented in this manuscript. The team can be seen closely inspecting crops for rust, a common fungal infection for cereal grains.

is split across teams of two each with their own clicker, collecting data5.17. Besides the time requirements it is healthier to label this data where humans are tromping through fields instead of sitting behind a desk. Finally the ability for humans to closely inspect the plants being labeled is a benefit to the accuracy of the labels. The ability to inspect plant diseases or stress in person is a marked improvement from attempting to discern plant health from aerial imagery 4.9. This also bleeds into the ability to label data regardless of the sensors being used. This becomes pronounced when asking a question such as "what does a sick plant look like in the thermal spectra?" By labeling directly in the field any data that is related to GNSS position can be labeled unlike before with yololabel where a human would have to be able to identify a sick plant as seen through spectra other than visible light such as thermal.

5.2.2 Interpret validation

BANNERS is a viable system for annotating large scale geo-referenced datasets. We see that because of the reprojection accuracy of BANNERS it is possible to generate labels for photographic data collected from ASTRO.

The margin of error for BANNERS was set up to reflect the scale of the phenomena that we are tracking. In the agricultural context, one meter is about the size of three or four plants such as strawberries. This makes one meter a convenient metric by which we can judge performance. To go beyond the work presented here, BANNERS leads into future work in agriculture where a system can alert a human to dying crops. A system that can detect and direct a human to some agricultural

problem needs to be accurate to within a meter for a human to find the dying plants. Similarly for search and rescue, directing humans to within a meter of the target is plenty useful. The one meter metric is further supported by the DARPA Subterranian Challenge where scoring accepted up to one meter of error in the reported position of any artifact including survivors. On the scale of acres and kilometers of flight one meter of error is acceptable.

Average projection error falling well under 1m between prediction and ground truth with the worst test showing an average error of 0.6m5.1.6 despite demonstrating even better accuracy in other tests. This demonstrates the viability of BANNERS as a data labeling tool with sufficient accuracy to be deployed for use generating the labels for deep learning systems.

We see that error quickly grows with with flight length5.1.6, we are unsure of the cause of increases in projection error correlating to increases in flight duration. It is assumed that there is error in the extrinsics calculated earlier but a transform error such as the one found in the extrinsics calibration should produce consistent error over time. The pose estimates of the robot are highly consistent5.5 and no drift in orientation is observed5.9 which leaves error or drift in the mounting solution between the UAV and the payload. Further testing with a larger sample size will be needed to fully diagnose and correct the error growth with increases of flight duration. The error between prediction and detection appears guassian5.15 which supports the hypothesis that error is being introduced by sub-optimal camera to robot extriniscs estimation. A systematic error in orientation would generate a gaussian where an error in translation

makes the guassian non-zero mean. Despite this error BANNERS still shows usefulness as a label generation tool and still shows marked improvements in the human factors aspect of its implementation over traditional data labeling methods.

5.3 Future work

BANNERS, as a novel system, demonstrates its core functionality and meets the necessary requirements. However, there are areas for improvement that can enhance its performance and capabilities.

One aspect that can be improved is the incorporation of depth information for precise data and label projection into images. Currently, BANNERS lacks depth data, which affects its accuracy on steep slopes or at high attitude angles due to the foreshortening of camera frustums. Although depth data was omitted to simplify BANNERS and validate the concept of automated labeling, it is essential to consider depth information for accurate operations.

Moreover, the development of calibration hardware and software compatible with the GNSS reference frame for UAVs is crucial. While GNSS calibration is traditionally associated with satellites, the increasing application of UAVs, exemplified by systems like BANNERS, necessitates calibration in a GNSS frame. One potential approach is utilizing dual high-precision GNSS systems rigidly mounted to a calibration target placed on the ground, allowing UAVs to fly over it. This setup resembles the ground control points used in orbital-based imaging systems but on a scale suitable for

UAVs. Building out small scale calibration systems that utilize GNSS as the common reference frame is similar to using a motion capture system like optitrack to estimate the transform between sensors on a robot within a lab. The technology that has unlocked this as a possibility for non-governmental entities is the development of low cost RTK enabled GNSS positioning. While systems like optitrack are more accurate than 1.4cm a fused AHRS or INS pose estimation system can reliably achieve the accuracy observed by a system like optitrack. Furthermore a calibration target with dual GNSS receivers gets heading estimation while being left motionless on the ground can use non-linear least squares or even averaging to precisely determine its position in GNSS coordinates again acting like optitrack where the accuracy could be measured in single digit millimeters. A calibration target for imagers and other sensors would be the most difficult part of the proposed apparatus to manufacture for all of the problems presented. Making a calibration target that can be used for all of the sensors included on a system like ASTRO may be easy. An aptil tag works well for visual cameras and can work for thermal cameras. For lidar or radar however there would need to be more thought put into a calibration target appropriate for those sensing modalities. The need to move out of a laboratory and outdoors for calibration is driven by the optics chosen for the research being performed. Moving outdoors enables incredible amounts of positional freedom while calibrating sensors intended for use over long distances.

BANNERS would also benefit enormously from including full state estimation on the sensor payload or a more rigid sensor mount. The decision to make the sensor payload easily detachable from the UAV body was an intelligent decision overall. The

detachable design aided in repair and debugging during development but also introduced flex into the mounting solution which decreases the accuracy of BANNERS through ambiguity in the transform from the camera reference frame to the UAV inertial frame. This problem is purely a mechanical one which can be solved by iterating further on the design of the payload and the mount on the UAV. Increasing the rigidity of the mount between the sensors and the UAV body would substantially improve the projection accuracy of BANNERS. The alternative would be to include more sensors such as an altitude heading and reference system directly on the sensor payload. This would be a more rigid mount between the image sensors and the state estimation sensors. Higher navigation rates or synchronising the pose estimates with the camera frames would help even further but may make less of an impact than increasing rigidity would.

Chapter 6

Conclusion

BANNERS and ASTRO demonstrate the ability to label arbitrary data through a relation in a GNSS reference frame. The initial phase of constructing and verifying the systems as outlined involves the construction of a suitable payload that corresponds to the specific work and flight platform. In the case of ASTRO, this entailed the creation of a payload capable of accommodating a single board computer, multiple imaging systems, and a GNSS receiver mounted on a DJI Matrice M300. Subsequently, the validation and characterization of ASTRO were conducted concurrently with the development of BANNERS. Each transform computed for ASTRO played a crucial role in influencing the design choices made for BANNERS, enabling the establishment of a complete sequence of transforms for projecting data from GNSS space into images.

Our findings reveal opportunities for enhancement, particularly concerning the calibration of systems like ASTRO that rely on GNSS positioning to accurately estimate the coordinate transform from the robot body frame to the camera body

frame. Additionally, we propose future research directions aimed at improving ASTRO's performance through enhanced mechanical design. Furthermore, we advocate for the development of depth estimation techniques to compute surface normals, as this would extend the applicability of BANNERS and ASTRO to more rugged terrains with steeper inclines.

This study presents an innovative approach to address the cost issue of data annotation for supervised deep learning frameworks by employing conventional image projection methods. In order to showcase this method, the implementation of a reliable LoRa-based RTK GNSS system was required for field robotics applications. Additionally, unconventional calibration techniques were investigated within the realm of sub-orbital UAV-based imaging systems. The research findings demonstrate the viability and practicality of a streamlined pose estimation system in the context of field robotics, thereby highlighting its potential benefits.

The presented research represents a significant advancement in the field of deep learning for field robotics, characterized by an evolutionary trajectory. The findings illustrate the feasibility of labeling diverse datasets within a GNSS reference frame to facilitate the training of supervised deep learning algorithms. This achievement showcases the potential for substantial progress in the integration of deep learning techniques in large scale field robotics applications.

Bibliography

[1] Case Studies Archives.

[2] Nemra Abdelkrim, Nabil Aouf, Antonios Tsourdos, and Brian White. Robust nonlinear filtering for INS/GPS UAV localization. In *2008 16th Mediterranean Conference on Control and Automation*, pages 695–702, June 2008.

[3] Telmo Adão, Jonáš Hruška, Luís Pádua, José Bessa, Emanuel Peres, Raul Morais, and Joaquim João Sousa. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. *Remote Sensing*, 9(11):1110, November 2017. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[4] Alireza Ahmadi, Michael Halstead, and Chris McCool. Towards Autonomous Visual Navigation in Arable Fields. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6585–6592, October 2022. ISSN: 2153-0866.

[5] Benjamin Allen, Michele Dalponte, Hans Ole Ørka, Erik Næsset, Stefano Puliti, Rasmus Astrup, and Terje Gobakken. UAV-Based Hyperspectral Imagery for

Detection of Root, Butt, and Stem Rot in Norway Spruce. *Remote Sensing*, 14(15):3830, January 2022. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

[6] Cody Anderson, Steve Labahn, Dennis Helder, Greg Stensaas, Christopher Engebretson, Christopher Crawford, Calli Jenkerson, and Christopher Barnes. The U. S. Geological Survey's Approach to Analysis Ready Data. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 5541–5544, July 2019. ISSN: 2153-7003.

[7] Peter M. Bach and Jayantha K. Kodikara. Reliability of Infrared Thermography in Detecting Leaks in Buried Water Reticulation Pipes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(9):4210–4224, September 2017. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

[8] M. Dian Bah, Adel Hafiane, and Raphael Canals. Deep learning with unsupervised data labeling for weed detection in line crops in uav images. *Remote Sensing*, 10(11):1690, November 2018. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[9] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] Jun Dai, Songlin Liu, Xiangyang Hao, Zongbin Ren, and Xiao Yang. UAV Localization Algorithm Based on Factor Graph Optimization in Complex Scenes. *Sensors*, 22(15):5862, January 2022. Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

[11] Frank Dellaert. Factor Graphs and GTSAM: A Hands-on Introduction. September 2012. Accepted: 2012-11-01T15:35:07Z Publisher: Georgia Institute of Technology.

[12] Frank Dellaert and GTSAM Contributors. borglab/gtsam, May 2022.

[13] Frank Dellaert and Michael Kaess. *Factor Graphs for Robot Perception.* Foundations and Trends in Robotics, Vol. 6, 2017.

[14] 0hye developer0hye. Gui for marking bounded boxes of objects in images for training neural network yolo, Jan 2023.

[15] Herman Eerens, Dominique Haesen, Felix Rembold, Ferdinando Urbano, Carolien Tote, and Lieven Bydekerke. Image time series processing for agriculture monitoring. *Environmental Modelling & Software*, 53:154–162, 2014.

[16] Christopher Engebretson. CHRISTOPHER Digitally signed by.

[17] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza. *IMU preintegration on Manifold for Efficient Visual-Inertial Maximum-a-Posteriori Estimation.* 2015.

[18] Eliakim Hamunyela, Jan Verbesselt, and Martin Herold. Using spatial context to

improve early detection of deforestation from Landsat time series. *Remote Sensing of Environment*, 172:126–138, 2016.

[19] J. L. E. Honrado, D. B. Solpico, C. M. Favila, E. Tongson, G. L. Tangonan, and N. J. C. Libatique. UAV imaging with low-cost multispectral imaging system for precision agriculture applications. In *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, pages 1–7, October 2017.

[20] Markus Isser, Hannah Kranebitter, Andreas Kofler, Gernot Groemer, Franz J. Wiedermann, and Wolfgang Lederer. Rescue blankets hamper thermal imaging in search and rescue missions. *SN Applied Sciences*, 2(9):1486, August 2020.

[21] Karlsruhe Institute of Technology, Institute of Information Systems and Marketing (IISM), Karlsruhe, Germany, Merlin Knaeble, Mario Nadj, and Alexander Maedche. Oracle or Teacher? A Systematic Overview of Research on Interactive Labeling for Machine Learning. In *WI2020 Zentrale Tracks*, pages 2–16. GITO Verlag, March 2020.

[22] Memoona Khanum, Tahira Mahboob, Warda Imtiaz, Humaraia Abdul Ghafoor, and Rabeea Sehar. A Survey on Unsupervised Machine Learning Algorithms for Automation, Classification and Maintenance. *International Journal of Computer Applications*, 119(13):34–39, June 2015.

[23] David M. Kim, Hairong Zhang, Haiying Zhou, Tommy Du, Qian Wu, Todd C. Mockler, and Mikhail Y. Berezin. Highly sensitive image-derived indices of water-

stressed plants using hyperspectral imaging in SWIR and histogram analysis. *Scientific Reports*, 5(1):15919, November 2015. Number: 1 Publisher: Nature Publishing Group.

[24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.

[25] Meng Lu, Edzer Pebesma, Alber Sanchez, and Jan Verbesselt. Spatio-temporal change detection from multidimensional arrays: Detecting deforestation from MODIS time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117:227–236, 2016.

[26] Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. Robot operating system 2: Design, architecture, and uses in the wild. *Science Robotics*, 7(66):eabm6074, 2022.

[27] Guoqiang Mao, Sam Drake, and Brian D. O. Anderson. Design of an Extended Kalman Filter for UAV Localization. In *2007 Information, Decision and Control*, pages 224–229, February 2007.

[28] Ignacio Martinez-Alpiste, Gelayol Golcarenarenji, Qi Wang, and Jose Maria Alcaraz-Calero. Search and rescue operation using UAVs: A case study. *Expert Systems with Applications*, 178:114937, 2021.

[29] Gaetano Messina and Giuseppe Modica. Applications of UAV Thermal Imagery in Precision Agriculture: State of the Art and Future Research Outlook. *Remote Sensing*, 12(9):1491, January 2020. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

[30] Jeremy P. Mondejar and Alejandro F. Tongco. Near infrared band of Landsat 8 as water index: a case study around Cordova and Lapu-Lapu City, Cebu, Philippines. *Sustainable Environment Research*, 29(1):16, April 2019.

[31] Érika Akemi Saito Moriya, Nilton Nobuhiro Imai, Antonio Maria Garcia Tommaselli, Adilson Berveglieri, Guilherme Henrique Santos, Márcio Augusto Soares, Marcelo Marino, and Thiago Tiedtke Reis. Detection and mapping of trees infected with citrus gummosis using UAV hyperspectral data. *Computers and Electronics in Agriculture*, 188:106298, 2021.

[32] A. Moura, J. Antunes, A. Dias, A. Martins, and J. Almeida. Graph-SLAM Approach for Indoor UAV Localization in Warehouse Logistics Applications. In *2021 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 4–11, April 2021.

[33] Abdelkrim Nemra and Nabil Aouf. Robust INS/GPS Sensor Fusion for UAV Localization Using SDRE Nonlinear Filtering. *IEEE Sensors Journal*, 10(4):789–798, April 2010. Conference Name: IEEE Sensors Journal.

[34] Krishna Neupane and Fulya Baysal-Gurel. Automatic Identification and Monitor-

ing of Plant Diseases Using Unmanned Aerial Vehicles: A Review. *Remote Sensing*, 13(19):3841, January 2021. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

[35] Thanh Tam Nguyen, Thanh Dat Hoang, Minh Tam Pham, Tuyet Trinh Vu, Thanh Hung Nguyen, Quyet-Thang Huynh, and Jun Jo. Monitoring agriculture areas with satellite images and deep learning. *Applied Soft Computing*, 95:106565, October 2020.

[36] Thien Minh Nguyen, Abdul Hanif Zaini, Kexin Guo, and Lihua Xie. An Ultra-Wideband-based Multi-UAV Localization System in GPS-denied environments. 2016.

[37] noauth. Precision Agriculture in Crop Production.

[38] noauth. Raspberry Pi NoIR camera marker tracking | DreamOnward.

[39] noauth. NDWI: Index Formula, Value Range, And Uses In Agriculture, September 2021.

[40] noauth. Kalman filter, January 2023. Page Version ID: 1135377424.

[41] Roberto Opromolla, Giancarmine Fasano, Giancarlo Rufino, Michele Grassi, and Al Savvaris. LIDAR-inertial integration for UAV localization and mapping in complex environments. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 649–656, June 2016.

[42] Luc Oth, Paul Furgale, Laurent Kneip, and Roland Siegwart. Rolling Shutter Camera Calibration. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1360–1367, June 2013. ISSN: 1063-6919.

[43] Issouf Ouattara, Vesa Korhonen, and Arto Visala. LiDAR-odometry based UAV pose estimation in young forest environment. *IFAC-PapersOnLine*, 55(32):95–100, January 2022.

[44] Fariz Outamazirt, Li Fu, Yan Lin, and Nemra Abdelkrim. A new SINS/GPS sensor fusion scheme for UAV localization problem using nonlinear SVSF with covariance derivation and an adaptive boundary layer. *Chinese Journal of Aeronautics*, 29(2):424–440, April 2016.

[45] Rakhel Kumar Parida, Vidhya Thyagarajan, and Sreeja Menon. A thermal imaging based wireless sensor network for automatic water leakage detection in distribution pipes. In *2013 IEEE International Conference on Electronics, Computing and Communication Technologies*, pages 1–6, January 2013.

[46] Mansheej Paul, Surya Ganguli, and Gintare Karolina Dziugaite. Deep learning on a data diet: Finding important examples early in training. *CoRR*, abs/2107.07075, 2021.

[47] Nathalie Pettorelli, Jon Olav Vik, Atle Mysterud, Jean-Michel Gaillard, Compton J. Tucker, and Nils Chr. Stenseth. Using the satellite-derived NDVI to as-

sess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9):503–510, September 2005.

[48] Jorge Peña Queralta, Carmen Martínez Almansa, Fabrizio Schiano, Dario Floreano, and Tomi Westerlund. UWB-based System for UAV Localization in GNSS-Denied Environments: Characterization and Dataset. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4521–4528, October 2020. ISSN: 2153-0866.

[49] Panagiotis Radoglou-Grammatikis, Panagiotis Sarigiannidis, Thomas Lagkas, and Ioannis Moscholios. A compilation of UAV applications for precision agriculture. *Computer Networks*, 172:107148, May 2020.

[50] Pierluigi Zama Ramirez, Claudio Paternesi, Luca De Luigi, Luigi Lella, Daniele De Gregorio, and Luigi Di Stefano. Shooting Labels: 3D Semantic Labeling by Virtual Reality. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 99–106, December 2020.

[51] Christopher Dahlin Rodin, Luciano Netto de Lima, Fabio Augusto de Alcantara Andrade, Diego Barreto Haddad, Tor Arne Johansen, and Rune Storvold. Object Classification in Thermal Images using Convolutional Neural Networks for Search and Rescue Missions with Unmanned Aerial Systems. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2018. ISSN: 2161-4407.

[52] Piotr Rudol and Patrick Doherty. Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery. In *2008 IEEE Aerospace Conference*, pages 1–8, March 2008. ISSN: 1095-323X.

[53] Lars Schmarje, Monty Santarossa, Simon-Martin Schröder, and Reinhard Koch. A Survey on Semi-, Self-and Unsupervised Learning for Image Classification. *IEEE Access*, PP:1–1, May 2021.

[54] P. Serra and X. Pons. Monitoring farmers' decisions on Mediterranean irrigated crops using satellite image time series. *International Journal of Remote Sensing*, 29(8):2293–2316, April 2008. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01431160701408444.

[55] Lakesh K. Sharma, Honggang Bu, Anne Denton, and David W. Franzen. Active-Optical Sensors Using Red NDVI Compared to Red Edge NDVI for Prediction of Corn Grain Yield in North Dakota, U.S.A. *Sensors*, 15(11):27832–27853, November 2015. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[56] Yang Song and Li-Ta Hsu. Tightly coupled integrated navigation system via factor graph for UAV indoor localization. *Aerospace Science and Technology*, 108:106370, January 2021.

[57] Ximena Tagle. *Study of radiometric variations in Unmanned Aerial Vehicle remote sensing imagery for vegetation mapping.* PhD thesis, June 2017.

[58] Atharva Tendle, Andrew Little, Stephen Scott, and Mohammad Rashedul Hasan.

Self-supervised learning in the twilight of noisy real-world datasets. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 461–464, 2022.

[59] Jesper E. van Engelen and Holger H. Hoos. A survey on semi-supervised learning. *Machine Learning*, 109(2):373–440, February 2020.

[60] Jeffrey Walton, David Nowak, and Eric Greenfield. Assessing Urban Forest Canopy Cover Using Airborne or Satellite Imagery. *Arboriculture & Urban Forestry*, 34(6):334–340, November 2008.

[61] Jeffrey T. Walton, David J. Nowak, and Eric J. Greenfield. Assessing urban forest canopy cover using airborne or satellite imagery. *Arboriculture & Urban Forestry. 34(6): 334-340.*, 34(6), 2008. Number: 6.

[62] Dan Wang and Yi Shang. A new active labeling method for deep learning. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 112–119, July 2014. ISSN: 2161-4407.

[63] Qiaoyun Xie, Jadu Dash, Wenjiang Huang, Dailiang Peng, Qiming Qin, Hugh Mortimer, Raffaele Casa, Stefano Pignatti, Giovanni Laneve, Simone Pascucci, Yingying Dong, and Huichun Ye. Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index Retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1482–1493, May 2018.

Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.

[64] Jibo Yue, Jia Tian, Qingjiu Tian, Kaijian Xu, and Nianxu Xu. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154:216–230, August 2019.

[65] Yangchengsi Zhang, Jiaqiang Du, Long Guo, Zhilu Sheng, Jinhua Wu, and Jing Zhang. Water Conservation Estimation Based on Time Series NDVI in the Yellow River Basin. *Remote Sensing*, 13(6):1105, January 2021. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute.