Toward Image-Based Three-Dimensional Reconstruction from Cubesats: Impacts of Spatial Resolution and SNR on Point Cloud Quality

by

Jordyn Stoddard

B.S. U.S. Air Force Academy, 2012

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology

August 12th, 2014

Signature of the Author _____

Accepted by _____
Coordinator, M.S. Degree Program Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

M.S. DEGREE THESIS

The M.S. Degree Thesis of Jordyn Stoddard
has been examined and approved by the
thesis committee as satisfactory for the
thesis required for the
M.S. degree in Imaging Science

Dr. David Messinger, Thesis Advisor

Dr. Carl Salvaggio

Dr. Michael Gartley

Date

Toward Image-Based Three-Dimensional Reconstruction from Cubesats: Impacts of Spatial Resolution and SNR on Point Cloud Quality

by

Jordyn Stoddard

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Master of Science Degree
at the Rochester Institute of Technology

Abstract

The adoption of cube-satellites (cubesats) by the space community has drastically lowered the cost of access to space and reduced the development lifecycle from the hundreds of millions of dollars spent on traditional decade-long programs. Rapid deployment and low cost are attractive features of cubesat-based imaging that are conducive to applications such as disaster response and monitoring. One proposed application is 3D surface modeling through a high revisit rate constellation of cubesat imagers. This work begins with the characterization of an existing design for a cubesat imager based on ground sampled distance (GSD), signal-to-noise ratio (SNR), and smear. From this characterization, an existing 3D workflow is applied to datasets that have been degraded within the regime of spatial resolutions and signal-to-noise ratios anticipated for the cubesat imager. The fidelity of resulting point clouds are assessed locally for both an urban and a natural scene. The height of a building and normals to its surfaces are calculated from the urban scene, while quarry depth estimates and rough volume estimates of a pile of rocks are produced from the natural scene. Though the reconstructed scene geometry and completeness of the scene suffer noticeably from the degraded imagery, results indicate that useful information can still be extracted using some of these techniques up to a simulated GSD of 2 meters.

Acknowledgements

I would like to thank Dr. David Messinger, Dr. Michael Gartley, and Dr. Carl Salvaggio for their support and guidance during this research. I would also like to thank Dr. Derek Walvoord and Adam Rossi for introducing me to image-based 3D reconstruction and helping with troubleshooting. Thank you to Dr. David Nilosek for assisting in applying his 3D workflow to my data sets. Thank you to Jason Faulring for being an invaluable resource on the data that were used in this research and for his continuous assistance with running code on Cyclone. Finally, thank you to the STARE team at Lawrence Livermore National Laboratory for providing specifications on their cubesat imagers.

Disclaimer

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background

With recent encouragement and outreach through NASA's Educational Launch of Nanosatellites (ELaNa) and the Colony II CubeSat Bus Program sponsored by the National Reconnaissance Office (NRO), the implementation of space applications on specifically tasked nanosatellites has surged. This work reports the initial investigation of the viability of a novel cubesat imager for the application of 3D reconstruction. The primary challenges of developing a cubesat tasked with imaging are designing a quality optical system that can fit into such a small volume and downlinking the data to ground stations. This research focuses on the former. The optical payload is constrained by the small physical size of the cubesat, which drastically limits the image quality of the system. The anticipated performance of a particular design is characterized according to ground sampled distance (GSD), signal-to-noise ratio (SNR), and image smear, or blur caused by spacecraft motion.

The structure from motion (SfM) concept has been successful in geo-accurately reconstructing the third dimension from a set of multi-perspective, two-dimensional images captured with a high-resolution airborne framing camera [2]. If this application could be extended to cubesats, a small constellation or train of cubesat imagers could provide a set of images with varying perspective over an area of interest for 3D reconstruction. This concept would be an inexpensive tool for surface modeling,

damage assessment following a natural disaster, and volumetric change detection. Advantages to a spaceborne SfM implementation also include the capability of any of these applications in areas with denied airborne access. This work examines the viability of a novel cubesat imager for this application based on the anticipated performance of the payload with respect to spatial resolution and SNR.

1.1 Objectives

The primary objective of this research is to assess the impacts of spatial resolution and radiometric limitations of the proposed cubesat design on the 3D reconstruction process. In order to fulfill this objective, several tasks must be accomplished. The imager has not yet launched, so data are unavailable for use in this work. Therefore, data must be simulated from existing data or generated with a tool such as the Digital Imaging and Remote Sensing Image Generation (DIRSIG) model [3]. The former is presented in this research. In order to simulate data, the spatial resolution and radiometric limitations of the payload must first be determined. The design is detector-limited, so GSD is used to characterize the spatial resolution. SNR summarizes the radiometric performance of the imager and is bound by blur caused by spacecraft motion. Next, existing data are degraded to simulate the cubesat data according to the aforementioned limitations.

A secondary objective is to demonstrate differences in point cloud quality. The simulated data are input to an existing 3D workflow, however, the community lacks metrics that communicate point cloud quality. This objective is accomplished by isolating small regions of the point clouds for which calculations are made to extract scene information and geometry from basic principles. These calculations are compared across degraded versions of the same scene to illustrate the link between image quality and point cloud quality.

1.2 Background

The remainder of this chapter presents the background information that is necessary to understand the methods implemented and the results produced in this work. An overview of cubesats and image quality is presented, followed by the Empirical Line Method (ELM) which is used for image simulation in this research. Then a popular feature extraction algorithm often used in the first stage of SfM is discussed. This feature extraction algorithm is approximated in the implementation of the 3D workflow used, the details of which are also outlined in this section and accompanied by an alternate means of obtaining 3D models via laser scanning which will serve as truth data in this study.

1.2.1 Cube Satellites

Cubesats are a subclass of nanosatellites that are composed of stackable 10x10x10 cm units (1U), shown in Fig. 1.1, that weigh approximately 1 kg. These units can be stacked along one dimension to create larger 2U, 3U, or 6U cubesats.

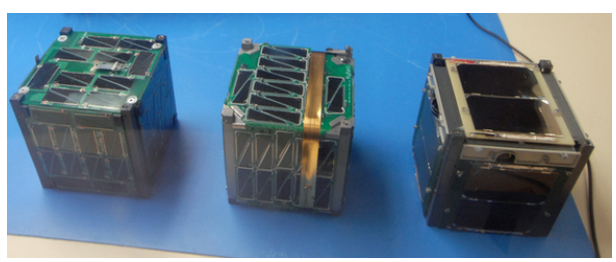


Fig. 1.1: Three 1U cubesats awaiting integration to a P-POD. Photo credit: NASA.

The cubesat form factor was established at Stanford from lessons learned following the successful operations of the pioneering OPAL picosatellite [4]. Stanford partnered with California Polytechnic Statue University (Cal Poly) to produce a cubesat deployer that could be attached to the launch vehicle, known as the Poly Picosatellite Orbital Deployer (P-POD). Each P-POD holds three cubesats and can be stacked

together with additional P-PODs. The P-PODs are attached to the exterior of a launch vehicle, as shown in Fig. 1.2.

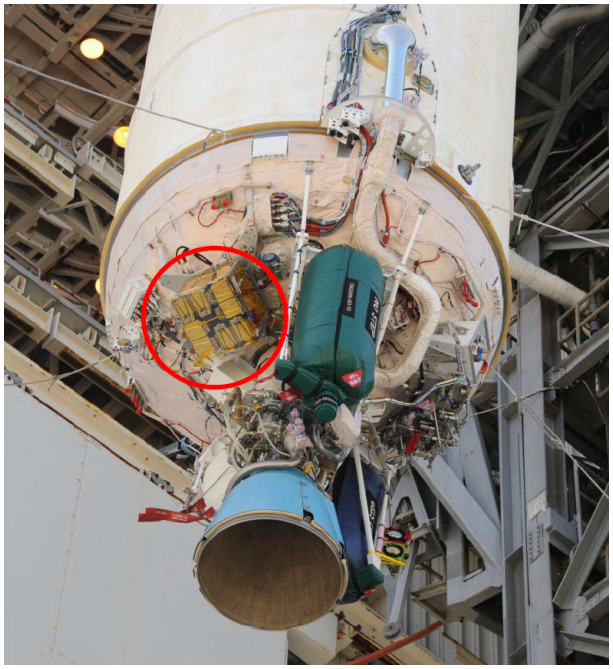


Fig. 1.2: The cubesats are attached near the engine, so they must endure severe vibration and heat exposure before deployment. From LeGaux [5]

While the small volume available for a payload poses an engineering design problem to fit the best possible system in the small volume, the advantages of the cubesat versus a larger satellite are many. The small form factor and light weight of the cubesats allow several to be launched as auxiliary payloads to larger satellites. This ride-share significantly lowers the cost of deploying cubesats to space. The standard size allows for use of the P-POD to deploy the cubesats from the launch vehicle. Organizations such as NASA and the NRO have further facilitated the process by offering launch opportunities and providing a standard cubesat bus and housing structure. This outreach allows research teams to simply develop their desired payload and integrate it with the standard bus. This culture rapidly accelerates the

timeline from initial concept development to launch and drastically lowers the cost of sending payloads to space, effectively lowering the barrier of entry in order to foster experimentation and growth in space technologies.

The development of cubesats has spurred space experiments in a variety of fields. The most recent ELaNa launch included a space biology experiment and a technological demonstration of deploying hundreds of sprites from a cubesat. SPORESat was developed by Purdue and NASA to investigate the gravitational sensing of plant spores [6]. The mission of the 3U KickSat was to validate their design for deploying hundreds of single-board, 3.5 x 3.5 cm spacecraft, termed sprites [7]. Implementing imaging systems onboard cubesats is also an active area of research. In the commercial sector, Planet Labs and Skybox imaging have led the charge towards Earth observation from small satellites. In February 2014, Planet Labs launched Flock 1, consisting of 28 3U cubesats, from the International Space Station (ISS) that delivers imagery with 3-5 meter resolution [8]. Two of these cubesats are shown in Fig. 1.3.

Fig. 1.3: Two Dove satellites from Flock 1 are ejected from the ISS [9].

Skybox Imaging launched SkySat-1 in 2013 which was successful in capturing high-resolution, HD video of Earth from space [10]. SkySat-1 touts sub-meter resolution from a small satellite the size of a mini-fridge. In other imaging from small satellites, LLNL has developed pathfinder cubesats capable of imaging other objects in orbit [11]. The most recent 3U pathfinder has been redesigned for Earth observation. These pathfinders are the inspiration for this research and will be discussed in depth, in the next section.

1.2.1.1 STARE

The Space-Based Actionable Refinement of Ephemeris (STARE) program originated with the objective of refining orbital projections of space debris in order to decrease the false alarm rate of collision warnings to operational satellites [11]. The program proposed to do this with a constellation of 3U cubesat imagers in low-Earth orbit (LEO) that captured debris streaking across the field-of-view in imagery. The primary challenges to accomplishing this mission from cubesats are designing a quality optical system whose resolution is constrained by the small payload volume and communicating with the platform through the bus electronics, the quality of which is also limited by the available space.

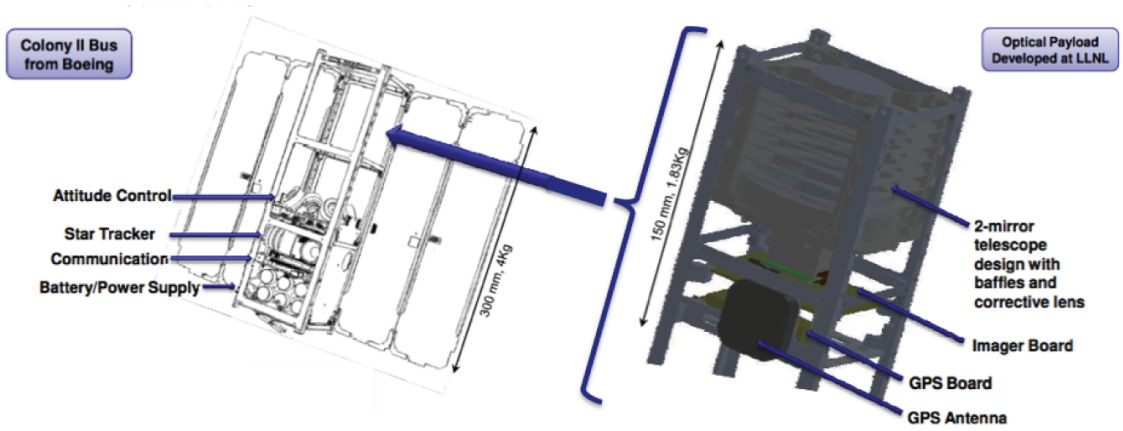


Fig. 1.4: The STARE optical payload (right) as it fits in to the provided Colony II Bus (left). From Simms, et al. [11].

The small space for the payload and bus inside the cubesat is shown in Fig. 1.4. The optical resolution of imagery provided by the optical payload is limited by the diameter of the aperture that can fit inside the cubesat. The Colony II Bus electronics limit the realization of the STARE mission by the slow, 57.6 kbps rate at which data can be downlinked [5]. Due to the constraints of one ground station and the low data rate, only 1 MB of data could be downloaded per day [11]. With images on the order of 1MB in size, this downlink constraint necessitated on-board processing for the mission. Each imager was made capable of processing the imagery on-board with an algorithm that refined orbital information of the debris by detecting the streak endpoints and calculating its trajectory relative to known stars in the background. Each cubesat imager would calculate this information for any object crossing its field-of-view and transmit the information to a master cubesat that downlinked the relevant information to a ground station.

In addition to the anticipated downlink limitation, there were three major problems with the Colony II bus: the reaction wheels were not functional therefore the bus had to rely on weak magnetic torque coils for slew, there was a parasitic drain on one of the battery cells that was predicted to likely cause the batteries to discharge before use on-orbit, and the solar panels and antenna did not deploy [5]. The NRO, Naval Postgraduate School (NPS), and LLNL are developing the CubeSat Next-

Generation Bus (C-NGB) framework to remedy such issues [12]. This program is also aiming to accomodate form factors larger than 3U and increase the data rate. Improved cubesat communication continues to be an active area of research, with designs such as inflatable antennae that will bridge the gap between current and desired data rates for cubesats [13].

Additional issues encountered by the pathfinders include severe defocus and misalignment of the Cassegrain optical payloads during launch. To remedy this problem, a novel compact optical sensor was developed that was robust to the forces experienced by the cubesats during launch [14]. This design has been optimized for Earth observation, but has not yet launched. Key design parameters for this imager are shown in Table 1.1. This research will focus on the expected image quality of

Tab. 1.1: STARE V4 Design Parameters

Focal length	650 mm
Aperture diameter	85 mm
Pixel pitch	2.2 μm
Well depth	7,641 electrons
Read noise	4.7 electrons
Bit depth	12

the constrained optical design for the V4 pathfinder, and its implications for 3D reconstructions derived from images of such quality.

1.2.2 Image Quality Metrics

Various image quality metrics have been developed to characterize imagery in two major categories: spatial resolution and radiometric fidelity. Common measures of spatial resolution include the ground-sampled distance (GSD), the system modulation transfer function (MTF), and the relative edge response (RER) which is a simplified metric of the MTF. GSD is the projection of one detector element from the imaging system altitude onto the ground, and is calculated for nadir image

viewing as

$$GSD = p\frac{h}{f} \tag{1.1}$$

where p is the detector pitch, h is the altitude of the system, and f is the effective focal length [15]. In other words, GSD maps the area on the ground that will be represented in one pixel of the image. Equation 1.1 assumes that the detector pitch/spacing and the detector width are equal. The GSD is often an estimate of the spatial resolution of a detector-limited system, but it does not account for resolution effects induced by blur of the optical system. The system MTF is a more comprehensive metric that includes effects from the optical system and the detector sampling. The RER is the slope of the edge-spread function (ESF), the spatial equivalent of the MTF. This metric is typically determined empirically. A popular measure of radiometric fidelity is the SNR, which has many forms, the simplest of which is

$$SNR = \frac{S}{N}, \tag{1.2}$$

where S is the signal and N is the noise. All of these metrics are related to image quality, yet none of them, alone, can be used to quantify image quality.

The remainder of this section will address historical developments to describe image quality such as the National Imagery Interpretability Rating Scale (NIIRS) and the General Image Quality Equation (GIQE). Then a review of the more complex image quality metrics such as SNR and Q will be presented as they pertain to this research.

1.2.2.1 NIIRS

The NIIRS was developed for the government in the early 1970's in order to encapsulate the effects of numerous image quality variables into a one-dimensional scale [16]. Multiple scales were developed for different types of imagery: visible, infrared, and multispectral. The scale was meant to rate the utility of an image based on its information potential, or the extent of information that could be gleaned from the

image by an analyst. The scale ranges from 0 to 9, where a rating of 0 signifies that none of the higher level tasks can be accomplished from the image and the highest rating of 9 corresponds to the ability to identify individual barbs on a barbed wire fence from the image. Intermediate examples of corresponding NIIRS levels and criteria for civilian uses of visible imagery are listed in Table 1.2. The NIIRS allows

Tab. 1.2: Examples of NIIRS Levels and Criteria

NIIRS Level	Task
1	Distinguish between major land use classes (e.g., urban, agricultural, forest, water, barren).
3	Detect individual houses in residential neighborhoods.
5	Detect large animals (e.g., elephants, rhinoceros, giraffes) in grasslands.
7	Detect individual steps on a stairway.
9	Identify individual bunches of pine needles.

for decimal values for images with criteria that fall between the specified levels. A difference in NIIRS of 0.2 between two images is considered noticeable. The NIIRS has become a useful tool for communication among image analysts and between remote sensing system manufacturers and their customers as a general metric of imaging system capabilities.

1.2.2.2 GIQE

Traditionally, remote sensing sytems are very expensive to produce. Therefore, a prediction of image quality is useful before investing on the order of a billion dollars to build a particular system. The GIQE was developed to predict the image quality of remote sensing systems based on physical collection parameters [17]. These parameters include GSD, the system MTF by means of the RER, and SNR. While these parameters account for system design choices that affect image quality, additional consideration was made to account for post-processing of the imagery that makes it more appealing to the eye. Modulation transfer function compensation

(MTFC) is typically performed on EO imagery to offset the blur induced by the remote sensing sytem. This sharpening amplifies edges and noise in the image. These effects are represented by the edge overshoot term (H) and the noise gain (G). This combination of design variables attempts to relate the anticipated scale and resolution of the imagery to the NIIRS. The most recent form of the GIQE is

$$\begin{aligned} NIIRS &= 10.251 - a \cdot log_{10} GSD_{GM} + b \cdot log_{10} RER_{GM} - (0.656 \cdot H_{GM}) \\ &\quad -(0.344 \cdot G/SNR) \end{aligned} \tag{1.3}$$

where the GM subscript denotes the geometric mean of that variable, a is 3.16 and b is 2.817 if RER <0.9, otherwise a is 3.32 and b is 1.559 [17]. The noise gain is calculated as

$$G = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (MTFCkernel_{ij})^2}. \tag{1.4}$$

and the edge overshoot term is calculated as the magnitude of the normalized edge response of the MTFC filter.

The original GIQE was ascertained from a regression analysis on a set of images rated by trained NIIRS analysts for which the system collection parameters were known. It was determined that the original equation did not account for the complex interactions of the variables on the NIIRS rating, i.e., the GSD term became less significant as the RER decreased. Therefore, the GIQE was updated with another set of images, half of which were used to perform regression analysis while the other half were used to validate the model. The coefficient of determination for the set of images used for validation was 0.934 with a standard error of 0.307, slightly more than a perceptible change in image quality. The characteristics of the input data from which the equation was regressed are summarized in Table 1.3. The NIIRS prediction is dominated by the GSD term, as shown in Fig. 1.5.

Tab. 1.3: Range of Variables in Complete GIQE 4 Data Set

Variable	Minimum	Maximum	Average
GSD	3 in. (0.08m)	80 in. (2.03m)	20.6 in.
RER	0.2	1.3	0.92
H	0.9	1.9	1.31
G	1	19	10.66
SNR	2	130	52.3

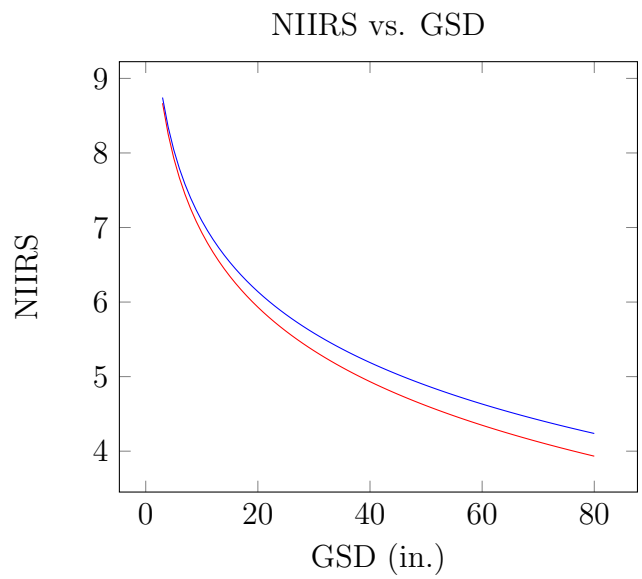


Fig. 1.5: The GIQE is dominated by the GSD term of the regression equation, shown for both possible coefficients of a: 3.32 if RER >0.9 (red), 3.16 otherwise (blue).

Due to the large dependence of the NIIRS level on GSD, the best-case prediction can be calculated from just the first term of the GIQE.

1.2.2.3 SNR

SNR is a common image-quality metric that measures the ability of a remote sensing system to accurately translate the target scene of interest to an image, on the photon level. This metric affects the contrast, and effectively the resolution, of an image.

The SNR of a system is often represented as

$$SNR = \frac{mean\ target\ signal}{standard\ deviation\ of\ signal}[18]. \tag{1.5}$$

The mean target signal is calculated by propagating the radiance of the target through the remote sensing system. The spectral radiance of a target in the visible spectrum, assuming a Lambertian target, at the entrance aperture of a remote sensing system is

$$L_{target} = \tau_{atm}(\lambda)\frac{\rho_{tgt}(\lambda)}{\pi} \cdot E_{tgt}(\lambda), \tag{1.6}$$

where τ_{atm} is the atmospheric transmission from the target to the sensor aperture, ρ is the reflectance of the target, and E_{tgt} is the irradiance at the target [18]. Additional sensor-reaching radiance includes radiance scattered off of the background and reflected towards the entrance aperture, this will be referred to as the upwelled radiance. Both the target radiance and the upwelled radiance are often calculated with MODTRAN due to the rigor of the calculations based on the collection geometry [19]. The number of photons that arrive at the detector after passing through a Cassegrain design is

$$n_{det} = \frac{A_{det}\pi ff t_{int}}{4(F\#)^2 hc}\int_{\lambda_1}^{\lambda_2} L_{target}(\lambda) \cdot \tau_{optics}(\lambda) \cdot \lambda d\lambda, \tag{1.7}$$

where A_{det} is the area of an individual detector element, ff is the fill-factor of the Cassegrain, $F\#$ is the f-number of the optical system, h is Planck's constant, c is the speed of light, and τ_{optics} is the wavelength-dependent optical transmission [18]. The signal in electrons can be calculated as

$$S_e = n_{det} \cdot QE(\lambda). \tag{1.8}$$

The most dominant sources of noise tend to be shot noise (if the signal is large), read noise, and quantization noise. Shot noise arises from the random fluctuation in

arrival of photons at the detector. This type of noise follows a Poisson distribution, the standard deviation of which is the square root of the expected value. In other words, the shot noise from the target signal, alone, can be expressed as

$$\sigma_{shot} = \sqrt{n_{det}} \tag{1.9}$$

The shot noise includes the effect of the entire sensor-reaching radiance, specifically it includes contributions from both the target radiance and the upwelled radiance. Noise terms add in quadrature, so the total shot noise is

$$\sigma_{shot} = \sqrt{n_{det,tgt} + n_{det,upwelled}} \tag{1.10}$$

where $n_{det,tgt}$ is calculated as in Eq. 1.7 and $n_{det,upwelled}$ is calculated by replacing L_{target} with $L_{upwelled}$ in the same equation. The sensor read noise is typically measured in the laboratory, and can be incorporated into the system noise calculation as follows.

$$\sigma_{system} = \sqrt{\sigma_{shot}^2 + \sigma_{read}^2} \tag{1.11}$$

1.2.2.4 Q

Q is a measure of how finely the detector samples the point spread function and is sensitive to GSD, MTF, and SNR. It is defined as

$$Q = \lambda \frac{F\#}{p} [15], \tag{1.12}$$

where λ is the band-average wavelength and p is the detector pitch. Specifically, Q is the ratio of the detector sampling frequency and the optical bandpass limit for a diffraction-limited system. Therefore, a Q = 2 system would have a detector sampling frequency that is twice the highest frequency that will be passed by the optical transfer function. This means that the detector sampling frequency meets the

Nyquist criterion and no frequencies that pass through the optics will be aliased. This ratio is an attractive characteristic that suggests the optimization of both optical and detector resolution. However, due to satellite motion during integration time, it is very difficult to design a remote sensing system with Q = 2. Smear induced on the focal plane as a satellite passes over a target while integrating limits the allowable integration time. This constraint also limits the SNR. The smaller pixel size required to meet the Nyquist criterion for frequencies passed by the optics adversely affects the smear, and thus the SNR of the system. Fiete demonstrates that in some cases lowering Q from 2 results in noticeable aliasing yet the overall image quality improves because edges appear sharper [15]. Therefore, increasing Q from 1 to 2 will only improve image quality if the change does not cause a noticeable drop in SNR.

1.2.3 ELM

ELM is used to estimate surface reflectances from observed radiances using known ground truth reflectances in the scene [20]. The method is a regression technique used to calculate unknown reflectances with the aid of ground truth reflectances. Typically, ground truth is acquired by producing panels of known reflectances and placing them in the scene to be imaged. Measuring the reflectance spectra of objects already in the scene is another method of acquiring ground truth. In its simplest form, if the sensor's radiance can be expressed as a linear function of digital counts, the ground truth digital count values can be related to the reflectance values as follows.

$$DC = mr + b \tag{1.13}$$

If at least two reflectances in the scene are known, they can be used to calculate m and b to define the linear relationship in Eq.1.13. Ideally, the known reflectances would correspond to extrema in DC. Eq. 1.13 can be rearranged for reflectance and applied to the entire image to generate a reflectance image. This reflectance image

can be used with the equations discussed in Section 1.2.2.3 to simulate images taken from a particular system.

1.2.4 Feature Extraction

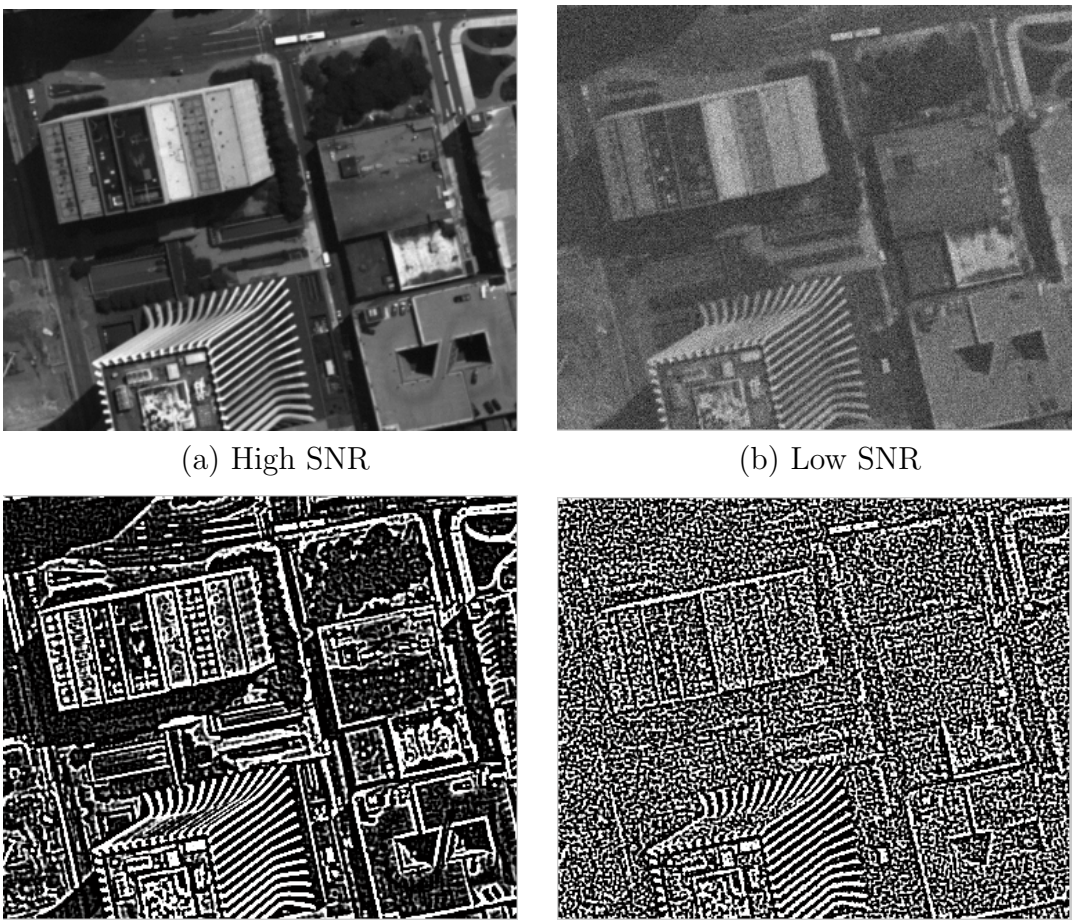
1.2.4.1 SIFT

The Scale-Invariant-Feature Transform (SIFT) is a popular and well-proven method of feature extraction developed for the computer vision community. It was created for the purpose of object recognition between images that have different perspectives [21]. Images taken of the same object at different times may have changes in illumination, perspective and noise; thus SIFT was designed to be robust to these changes. In short, the algorithm attempts to detect stable keypoints and their locations in the images, then a keypoint descriptor is generated based on the computed orientation of the keypoint and its surroundings for each keypoint that is deemed stable. In the first step, the algorithm implements a difference-of-Gaussian (DoG) filter as an approximation to the Laplacian filter, which has been proven to identify scale-invariant features [21]. During the second step of keypoint orientation and descriptor assignment, rotational invariance is achieved by including the orientation of neighboring pixels in the descriptor computation. The result is a rotation and scale invariant keypoint descriptor that is robust to partial affine distortion.

The first step involves searching through a scale-space, or a series of the same image blurred with Gaussian kernels of varying standard deviation [21]. Features are filtered out by differencing adjacent scales; this is the DoG filter. The filter essentially acts as a bandpass filter, which attenuates the low frequencies of homogenous regions and the high frequencies caused by noise, while passing frequencies representing edges. The downside to this filter is that the overall contrast of the image is reduced in the filtering process. Each filtered pixel is compared to its neighbors within the image, and to the corresponding pixel and its neighbors of other images in the scale space in order to identify extrema as keypoint candidates.

Additional steps are taken to eliminate weak edge responses from the pool of candidate keypoints produced by the filtering process. Once the weaker edge responses have been removed from the set, the keypoints are assigned descriptors. SIFT descriptors are 128 elements long, and are composed of the estimated orientations of the pixels surrounding the keypoint location. The magnitudes and directions of the gradient surrounding the keypoint are calculated, and weighted by a Gaussian window. This window functions to assign lower weights to gradients farther from the keypoint location. These gradient orientations are binned into orientation histograms. A 4x4 array of orientation histograms is computed, each of which contains 8 bins. The vector describing these histograms is the 128-element (4x4x8) descriptor.

Lowe demonstrated that the orientation assignment to keypoints was accurate 95% of the time in the face of (plus or minus) 10% pixel noise. The major cause of error in matching descriptors in the presence of noise was empirically determined as the initial keypoint location and scale detection [21]. Examples of applying the same DoG filter to images with different SNR are shown in Fig. 1.6. The filtered images have been scaled between 0 and 1 for display purposes.



(a) High SNR (b) Low SNR

(c) High SNR DOG Filtered (d) Low SNR DOG Filtered

Fig. 1.6: Two clips of the same image are shown with (a) high SNR and (b) low SNR. The difference-of-Gaussian filtered images are shown (c) for the high SNR clip and (d) for the low SNR clip. Note that the features on the rooftop of the upper left building are clearly visible in (c) but they are obscured by noise artifacts in (d).

As seen in Fig. 1.6(c), the filter highlighted features from Fig. 1.6(a). However, applying the same filter to the noisy image in Fig. 1.6(b) resulted in noise artifacts being detected as strong features in Fig. 1.6(d). Because noise levels were added to the same original image in order to produce the images with different SNRs, any extra features indicated in Fig. 1.6 (d) are artifacts.

1.2.5 Capturing Structure Geometry with 3D Models

Technology developments such as LIDAR (LIght Detection and Ranging) enable large-scale 3D mapping of the environment that can be useful for performing environmental assessments that would take much more time to perform without the model. LIDAR has been used to assess vegetation health, to model floodlines, and to evaluate carbon dioxide pollution[22]. An alternative to the active sensing of LIDAR systems is 3D reconstruction from multiple images providing various perspectives of the same scene, known as structure from motion (SfM). An advantage to this method is that passive imaging systems are more common, and thus, less expensive than LIDAR platforms. However, the method relies on geometric calculations to construct the model, which results in more room for error in the resulting point cloud. Additionally, areas of the scene that are occluded in shadow or that lack texture are omitted from the reconstructed model. This section will describe LIDAR and the SfM process. In this work, a SfM-based 3D reconstruction technique will be used and compared against LIDAR data collected over the same scene.

1.2.5.1 LIDAR

Light Detection and Ranging (LIDAR), alternatively known as laser altimetry, is the optical version of radar. A laser pulse is emitted from the platform, and the height of the surface below it is calculated by the elapsed time between the generation and return of each laser pulse [22]. The result is a 3D point cloud that can be geo-referenced using information of the platform position and orientation. Airborne LIDAR platforms are capable of capturing sub-meter detail [23], [24].

In a billion-dollar program, this 3D mapping technique has been extended to the spaceborne platform, the Ice Cloud and Elevation Satellite (ICESat), whose mission is to monitor the topography of polar ice sheets in order to make predictions of future changes in ice volume and sea-level [25]. To accomplish this mission, ICESat contains a telescope with a 1-meter aperture diameter, a dimension which is much larger than

what can be accomodated by a 3U cubesat. LIDAR has not yet been implemented on a cubesat, however, research on using a similar method for generating 3D point clouds of other spacecraft in orbit with a laser range finder from a 6U cubesat is underway [26].

1.2.5.2 Structure-from-Motion

Structure-from-motion (SfM) is a computer vision application that generates a 3D model from multiple views of the same scene. The SfM workflow typically can be broken down into two general stages: determining the relative camera pose, and triangulating 3D points to build the model. The relative camera pose, or the positions representing the camera center in 3D space where each image was collected, is calculated with projective geometry after finding point correspondences between each image pair. This stage of SfM requires application of a feature extraction algorithm like SIFT to each image, feature matching between images to identify point correspondences, and typically some sort of optimization to minimize any errors in the feature matches and/or camera parameters. With the camera pose estimated, corresponding points in the images can be located in 3D space. This can be done directly from corresponding points, or more complex algorithms can be used to reduce error in point correspondences and generate dense 3D models [27], [28]. The result is a point cloud that is positioned in an arbitrary coordinate system.

1.2.5.3 3D Workflow

The 3D workflow used in this work was created by David Nilosek [**?**]. It utilizes SiftGPU feature extraction and matching, bundle adjustment for the first stage of the SfM process. To actually prodcue the 3D model, a combination of the software packages Cluster-based Multi-View Stereo (CMVS) and Patch-Based Multi-View stereo (PMVS) are used [29], [27]. A seven degrees-of-freedom similarity transform is used to position the relative SfM point cloud in the desired Earth-fixed coordinate

system.

1.2.5.3.1 SiftGPU Developed by Chanchang Wu, SiftGPU is an implementation of SIFT that utilizes the graphics processing unit (GPU) in addition to the central processing unit (CPU) in order to reduce the processing time of the algorithm [30]. The use of the GPU allows processing to occur in parallel, which rapidly speeds up the process when high-resolution imagery is used.

1.2.5.3.2 Bundler This workflow uses Bundler, a software package developed by Noah Snavely, to estimate the optimized relative camera pose of the input imagery[31]. Bundler takes a set of images, image features, and point correspondences as input and produces a 3D model of the cameras and sparse scene geometry. In this process, Bundler estimates the intrinsic and extrinsic camera parameters to include: the focal length, radial distortion coefficients, and the transformation that translates a 3D point to a point on the camera's image plane (the camera projection matrix). The algorithm iteratively calculates camera pose, triangulates points, then optimizes the calculation by minimizing the reprojection error, which is defined as the sum of distances between the projections of each triangulated point onto the image plane of the camera in which it appears and its corresponding image features. This cycle begins with the image pair with the most feature matches and repeats with the addition of the camera with the next highest number of feature matches. These models are often too sparse to be useful, an example of which is shown in Fig. 1.7.

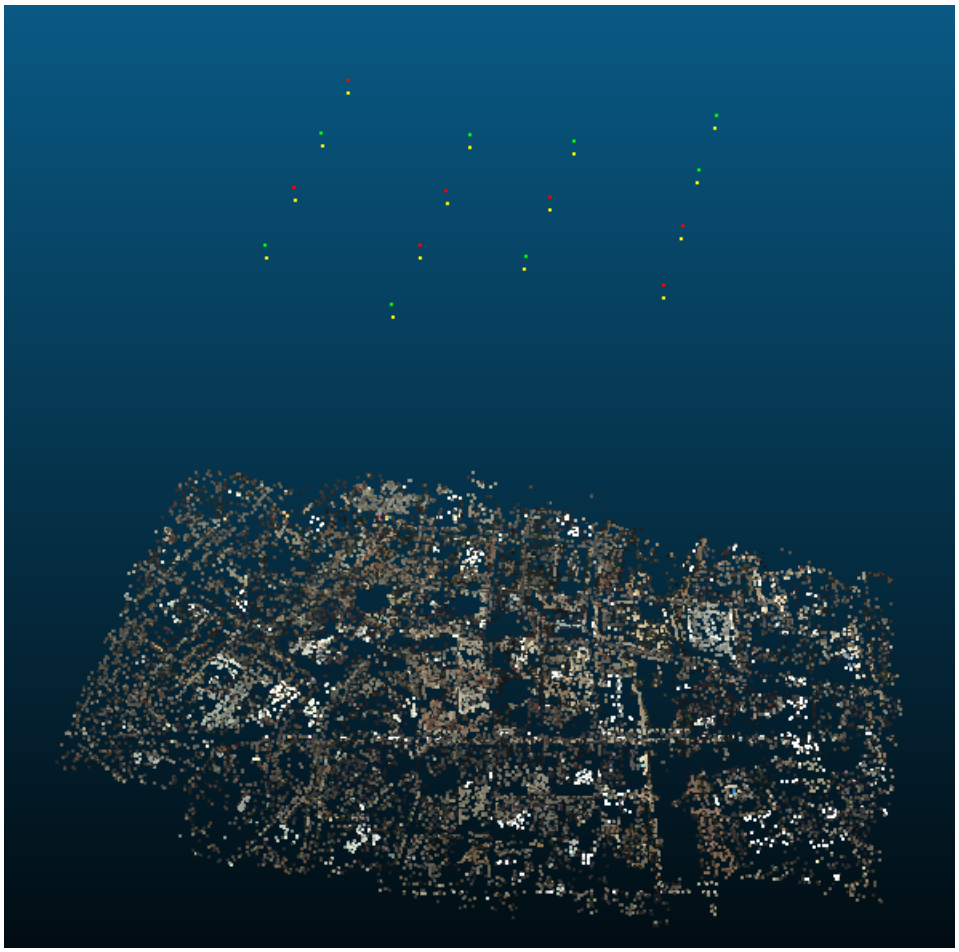


Fig. 1.7: A sparse point cloud produced by Bundler is shown below the estimated camera pose. The pairs of yellow and green/red points indicate each of the camera centers calculated for the input imagery.

1.2.5.3.3 CMVS/PMVS Cluster-based Multi-View Stereo (CMVS) is a software package that optimizes the multi-view stereo reconstruction algorithm by dividing the reconstruction into clusters that can be processed in parallel [29]. The software package also removes redundant images that hardly contribute to the 3D model and are better off excluded from the image set in order to reduce processing time. This process is particularly desirable for large image sets numbering hundreds or thousands of images, where optimizing the reconstruction can save days to weeks of processing time.

The Patch-Based Multi-View Stereo (PMVS) software package takes as input the camera pose calculated from bundle adjustment and the image clusters from CMVS and outputs a dense 3D point cloud of the entire scene [27]. The procedure uses a match, expand, and filter approach to iteratively produce a geometrically accurate, dense reconstruction. The algorithm employs its own feature extraction step using DoG and Harris corner detection to pick out robust features. Each image is divided into a user-specified grid and only the strongest features in each grid cell are retained. For each detected feature in a cell, a set of features is collected in the remaining images that fall within two pixels from the corresponding epipolar line. These points are triangulated to find the 3D point that is termed the patch center. The algorithm iterates through these possible patch centers until a good patch is found, which is determined by minimizing the angular difference between the patch center and the optical axis of the corresponding camera and the patch center and the optical axis of the original image. There may be cells in the image where no patches were reconstructed. If this is the case, neighboring patches are used to to initialize new patches in the empty image cells. Filtering is performed to remove erroneous patches by enforcing visibility constraints, and the expansion and filtering process are repeated to optimize the model. The result is a dense, 3D point cloud positioned in an arbitary coordinate system.

1.2.5.3.4 Georegistration The georegistration transform attempts to calculate the seven degrees of freedom similarity transform between the relative point cloud output from PMVS and the desired Earth-fixed coordinate system [32]. The two sets of points between which the transform is calculated are the GPS camera centers recorded onboard while the imagery was collected and the relative camera centers produced by the SfM workflow. Three components must be determined to compute this transform: scale, rotation, and translation. The scale is calculated by the ratio of the mean vectors of the two sets of points. The rotation matrix, R, is calculated with the Kabsch algorithm which finds the rotation matrix between two sets of

points with a least-squares method [33]. Finally the translation vector is calculated by the difference of the GPS centroid and the rotated and scaled relative centroid. Then, the components can be simultaneously applied to the relative point set, as follows

$$T_s = \begin{bmatrix} sR & T \\ 0 & 1 \end{bmatrix} [32]. \tag{1.14}$$

The resulting point cloud is anchored in the Earth-fixed coordinate system specified by the GPS camera centers and can be compared to known locations. Nilosek demonstrated that this transformation is more sensitive to error in the camera centers, than more complex georegistration approaches in terms of geoaccuracy [32]. However, it is used in this work to bring point clouds into the same coordinate system, so that relative measurements on different point clouds can be compared.

Chapter 2

Methodology

This chapter will cover the methods used to produce the results. First, the NIIRS prediction of V4 based on GSD is presented. Next, the SNR and smear calculations used to determine the operating SNR of the system are discussed. Because there is no on-orbit data from the cubesat, the next section describes the images that are altered to better represent the expected performance of the V4 sensor. Finally, the metrics that will be used to assess the quality of the point clouds are outlined.

2.1 Evaluate System Parameters

The GIQE provided a starting point for the expected performance of the V4 imager at Earth-observing altitudes [17]. To characterize the spatial resolution of the system dependent on altitude and detector performance, the nadir GSD of the system was evaluated from Eq.1.1. Altitudes between 200 and 700 km are considered. Orbits below 400 km experience greater drag which causes the spacecraft to fall out of orbit within months. However, these altitudes are considered with applications in mind such as one-time, on-demand intelligence over a general area. Recall from Section 1.2.2.2 that the dominant term in the GIQE is the GSD term. An approximation

of the GIQE was used based only on the nadir GSD in order to produce a rough estimate for the best-case scenario, written as

$$NIIRS \approx 10.251 - 3.32 * log_{10}(GSD)[17]. \tag{2.1}$$

The resulting NIIRS prediction for the altitude range of interest is illustrated in Fig. 2.1. Another factor that will affect spatial resolution is spacecraft motion

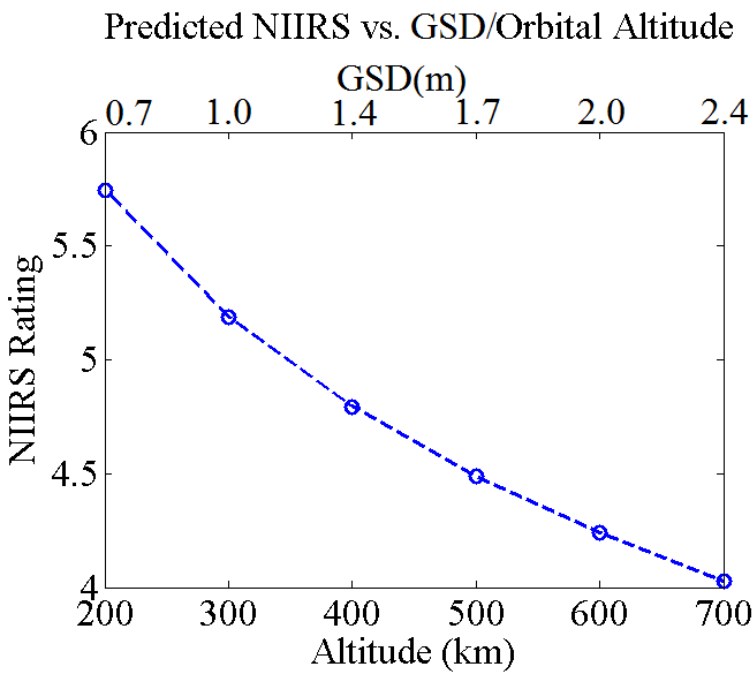


Fig. 2.1: The nadir GSD range of 0.7 to 2.4 m results in an approximate NIIRS prediction ranging from 5.8 to 4, respectively.

during the integration time which induces blur on the focal plane, or smear. Because satellites travel at such high speeds (on the order of 7 km/s in LEO), smear poses a potential threat to the image quality of these systems. The bus electronics for cubesats are currently not sophisticated enough to rotate the optical payload to stare at the same point on Earth throughout the integration time. In this case,

where the look angle of the platform is constant throughout the integration time, the smear can be calculated from the Eq. 2.2.

$$smear = \frac{v \cdot t_{int}}{GSD} \tag{2.2}$$

The mean orbital speed of the cubesat can be approximated as

$$v = \sqrt{\frac{GM}{r}} \tag{2.3}$$

where the gravitational constant (G) is equal to 6.67×10^{-11} $m^3kg^{-1}s^{-2}$, M is the mass of Earth, and r is the sum of Earth's radius and the orbit altitude. Equation 2.3 assumes that the mass of the cubesat is negligible compared to that of Earth and that the orbit is circular. The smear calculation bounds the functional integration time of the system; beyond one pixel of smear the image quality will suffer noticeably to the human eye.

Another important factor that will affect image quality is the signal-to-noise ratio (SNR). The SNR of the system is calculated for a 10% reflector with the two most dominant sources of noise: shot noise and read noise. The upwelled and target radiance values are generated using MODTRAN with multiple imaging satellite elevation (ISEL) and sun elevation angles. Then the SNR is calculated as

$$\frac{S_e}{\sigma_{system}} \tag{2.4}$$

where S_e and σ_{system} are calculated according to Eq. 1.7 through Eq. 1.11. With the combination of the smear analysis and the SNR predictions, an acceptable range of integration times is derived. These integration times and corresponding signal-to-noise ratios for V4 will be used in calculations for image-based simulation.

2.2 Data and Alteration

The V4 pathfinder has not launched yet, so imagery is not available for this study. Instead, two datasets were used for this evaluation: 15 images collected over downtown Rochester in 2011 and 31 images from the SHARE 2012 campaign that were collected over a quarry in Avon, NY [34], [35]. The former provided an urban scene while the latter provided a more natural scene for analysis in this work. These sites were captured with the Wildfire Airborne Sensor Program (WASP) which includes an RGB sensor. LIDAR data were also collected over the same sites, and will be used as truth data. The following subsections will specify how the data are degraded both spatially and with noise.

2.2.1 Spatial Resolution

To degrade the spatial resolution of the WASP imagery to better approximate the expected spatial resolution of V4, a Gaussian image pyramid was used to downsample the full-resolution imagery. The GSD of the imagery over the urban scene was approximately 25 cm, while the GSD of the natural scene was approximately 15 cm. The images from each scene were downsampled to four additional resolutions, totaling 5 distinct resolutions per dataset. The cv2.pyrdown function was used from the Open Computer Vision (OpenCV) library to implement this step [36]. This function blurs the image with a Gaussian kernel, then rejects even rows and columns. The levels of image resolution created with this function and their corresponding GSDs are summarized in Table 2.1.

Tab. 2.1: Data Sets for Point Cloud Generation

Level	Image Dimensions	Urban GSD (m)	Natural GSD (m)
1	4000 x 2672	0.25	0.15
2	2000 x 1336	0.50	0.30
3	1000 x 768	1.0	0.60
4	500 x 384	2.0	1.2
5	250 x 192	4.0	2.4

These datasets have GSDs that fall within the expected range for the V4 design and the higher resolution images provide a reference for comparison. Level 5 for each dataset was excluded from analysis for different reasons. In the urban scene, a point cloud was reconstructed at this resolution, however, the content geometry was so warped that it was unrecognizable. In the natural scene, less overlap between the input images resulted in CMVS excluding too many images from the reconstruction to reproduce the entire area of interest in the point cloud. More details on this occurrence are included in Appendix A.

2.2.2 Signal-to-noise Ratio

The workflow for SNR degradation of the imagery is illustrated in Fig. 2.2. The

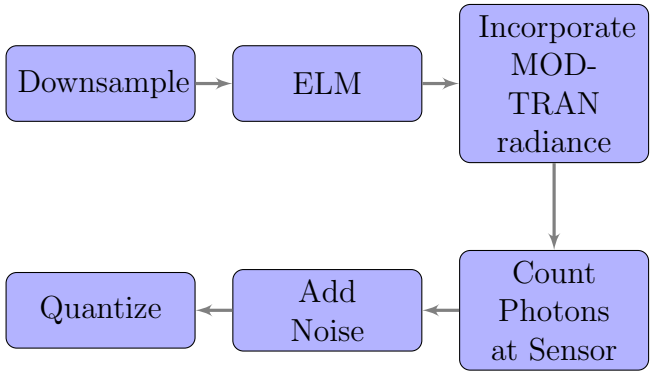


Fig. 2.2: SNR Degradation Workflow

images are downsampled first due to memory constraints. This step also minimizes noise acquired from the actual sensor that produced the images. ELM is applied assuming that the maximum reflectance of the scene is 0.60, corresponding to metallic

building surfaces, and assuming a minimum reflectance of zero. After these assumed reflectances are used to identify the variables in Eq. 1.13, the equation can be rearranged for reflectance and applied to the entire image to generate a reflectance image. Then, MODTRAN is used to provide ground-reflected and upwelled radiances for a 100% reflector. A sensor-reaching radiance image is predicted by the product of the ground-reflected radiance with the reflectance image plus an added bias of the upwelled radiance. An image at the focal plane of a particular system can be predicted with the remainder of the equations discussed in Section 1.2.2.3 and the system specifications. Signal-dependent shot noise is simulated by performing a random draw from a Poisson distribution whose mean value is the pixel value in photons, at every pixel of the image. Conversion to electrons and quantization complete the SNR degradation workflow. This method of SNR degradation was chosen as opposed to simple addition of Gaussian noise in order to simulate the signal-dependence of sensor noise.

2.3 3D Workflow

The 3D workflow has multiple parameters that are user-specified. Parameters that were set to non-default values are summarized in Table 2.2.

Tab. 2.2: 3D Workflow Options [1]

Command Line Option	Parameter	Default
-f	Focal length in pixels	6111.11
-x	Maximum dimension for SiftGPU	2000
-y	Maximum number of features detected by SiftGPU	8000
-l	Pyramid Level for PMVS	1

The default focal length in pixels (focal pixels) was calculated for the WASP RGB

sensor as

$$focal\ pixels = \frac{focal\ length}{pixel\ size}[1]. \tag{2.5}$$

For each downsampled case, the focal pixels must be scaled by the downsample factor, e.g., the focal pixels for Level 1 should be

$$\begin{aligned} focal\ pixels &= 6111.11 * 0.5 \\ &= 3055.56 \end{aligned} \tag{2.6}$$

Both SiftGPU and PMVS have the option to downsample the input imagery internally, however, to maintain control of the downsampling function, this was performed external to the 3D workflow. To prevent both algorithms from downsampling the externally downsampled imagery, the maximum image dimension parameter was set to 4000 for the full resolution cases of both scenes. Because the number of columns in the remaining levels are less than or equal to 2000, as shown in Table 2.1, the default parameter was sufficient for those levels. For all cases, the PMVS pyramid level was set to 0, again to prevent additional downsampling in this stage. Finally, the maximum number of SiftGPU features detected was changed as needed. The default value causes unreliable features to be detected at Levels 3-5, which ultimately causes the reconstruction to fail. At these levels, the number of features was incrementally reduced from the default until a reconstruction was successful. Note that the selection of this parameter may contribute to error in the point clouds, as it may cause bad matches that do not cause the reconstruction to fail completely.

2.4 Metrics

2.4.1 Building Metrics

Two tasks were performed on the resulting point clouds in order to characterize the degradation of the reconstructions due to the loss of spatial resolution. The first

task was estimating the height of a building and the second was estimating normals to a few of the building surfaces. Surface normal calculation must be performed in order to produce a facetized model from a point cloud, a common post-processing step. Before these measurements were made, the Clinton Square building from the urban scene was manually segmented from the point cloud using CloudCompare, an open-source point cloud processing software package [37]. As shown in Fig. 2.3, this building is not aligned with the axes.

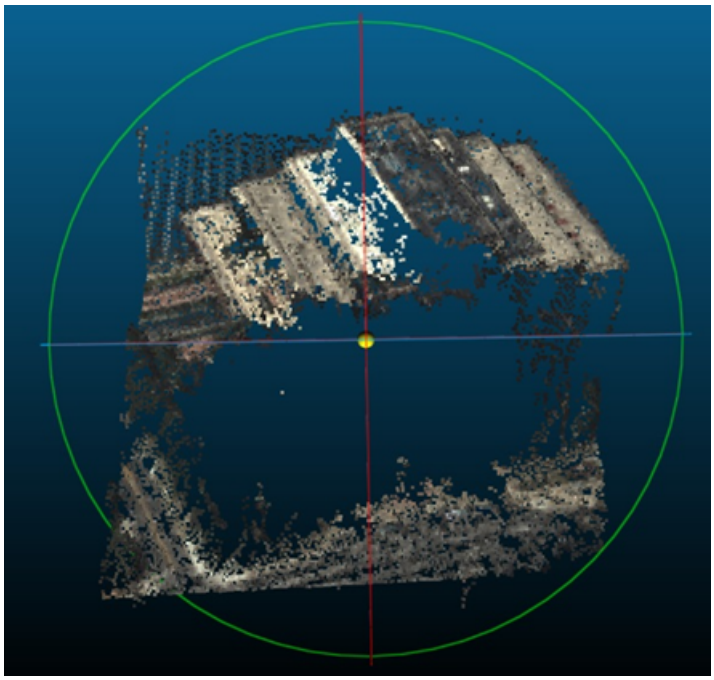


Fig. 2.3: The Clinton Square building cropped out of the reconstructed point cloud is tilted with respect to the axes.

The building was rotated until the sides, top, and bottom surfaces appeared to be aligned with the axes. Next, the points were projected onto the xz plane and thresholds were visually selected to isolate points on the top surface of the building and points on the ground plane, as shown in Fig. 2.4.

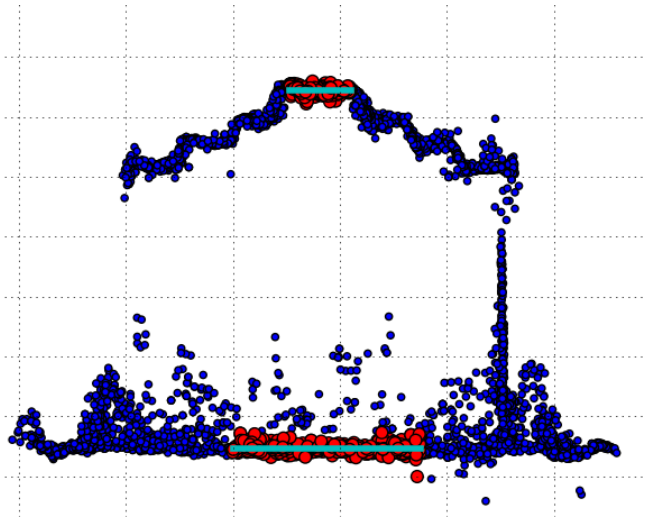


Fig. 2.4: Red points representing the roof and ground of the aligned and projected point cloud indicate those points thresholded for use in the height calculation.

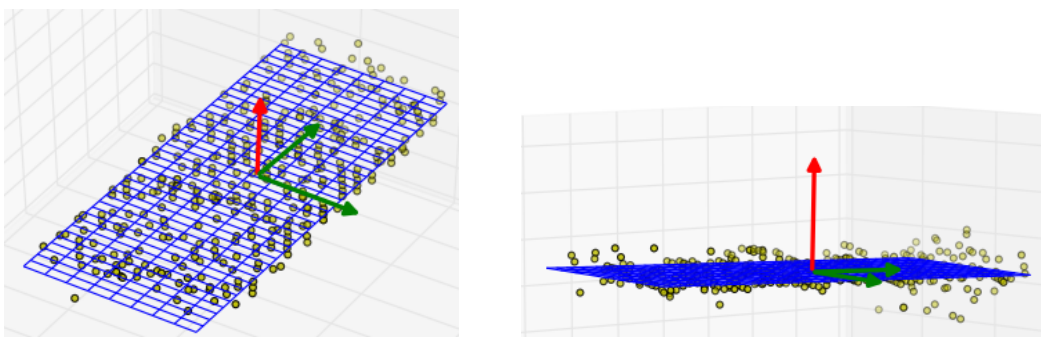
The mean z-coordinate of the ground points was subtracted from the mean z-coordinate of the roof points to yield the height estimate.

Principal Component Analysis (PCA) was used to estimate the normal to each planar surface indicated in Fig. 2.5.

Fig. 2.5: The numbered surfaces, outlined on a subset of one of the WASP images used for the reconstructions, will be used for surface normal calculation.

While there is some error in the reconstruction of the surfaces that prevents the

points from falling on the same plane, the points vary much less in the z-direction than in the other two dimensions. Therefore, the first two principal component vectors should capture the directions along which the points vary most, or two vectors that fall within the near-planar surface described by the points. As principal component vectors are chosen orthogonal to the preceding vectors, the third indicates the direction normal to the plane. This application of PCA is illustrated for a set of points in Fig. 2.6.



(a) (b)

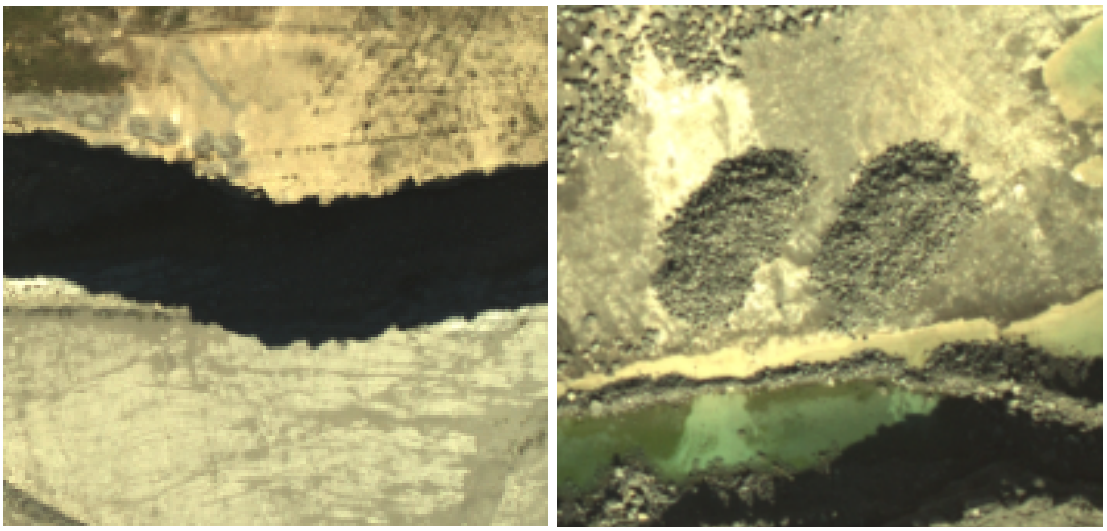
Fig. 2.6: The first two principal component vectors (green) indicate the directions of maximum variability in the 3D points, while the third principal component vector (red) is the normal to the surface defined by the first two principal component vectors.

The eigenvalue corresponding to the third eigenvector captures the variance of the points in the normal direction, or a sense of how well the points fit to a plane, where a lower value represents a closer planar fit.

2.4.2 Quarry Metrics

Two metrics were developed for the quarry data: the vertical depth of one section of the quarry and the volume of a pile of rocks and dirt. The areas of the point clouds where these measurements were made are shown in the WASP imagery in

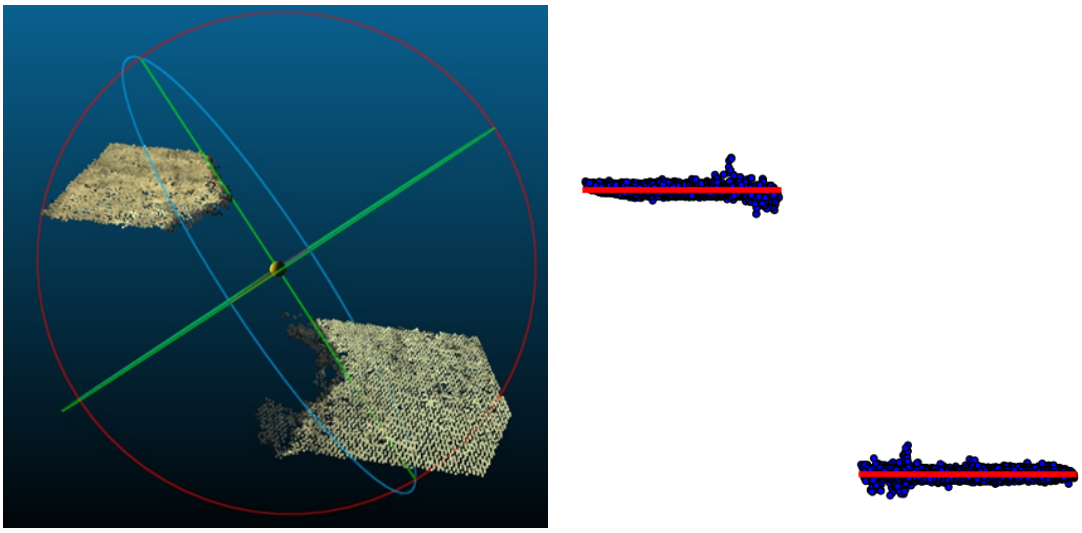
Figure 2.7.



(a) (b)

Fig. 2.7: The depth of the quarry was measured from the region in (a) and the volume of the left rock/dirt pile in (b) was calculated.

Similar to the process in the height estimation of the Clinton Square building, the flat surfaces at the top and bottom of the quarry wall in Fig. 2.7 (a) were aligned with the xy plane. This allowed a simple calculation of the depth through the difference of the mean z-coordinates of the two surfaces. The region of the quarry cropped with CloudCompare and the aligned points projected onto the yz plane are shown in Fig. 2.8.



(a) Quarry wall in CloudCompare *(b)* Aligned and projected points

Fig. 2.8: The points in (a) were aligned and projected onto the yz plane as shown in (b).

The second task was more complicated because the ground surrounding the two rock piles was not entirely flat. This complication made it difficult to threshold points comprising the rock pile from points comprising the ground. The z-coordinate used for thresholding was selected through visual inspection. Thresholds were also imposed in the x and y directions to isolate the left rock pile in Fig. 2.7 (b), resulting in the separated points shown in Fig. 2.9.

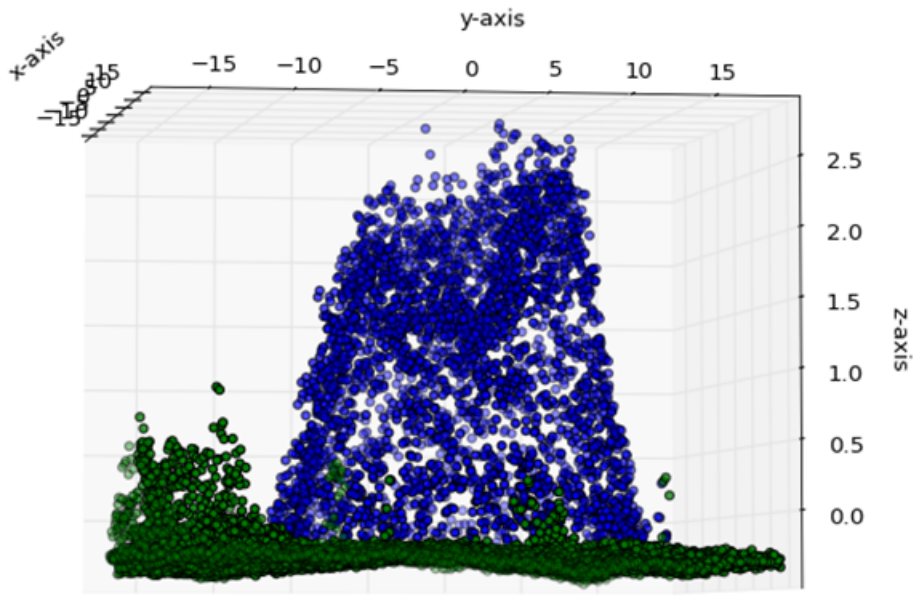
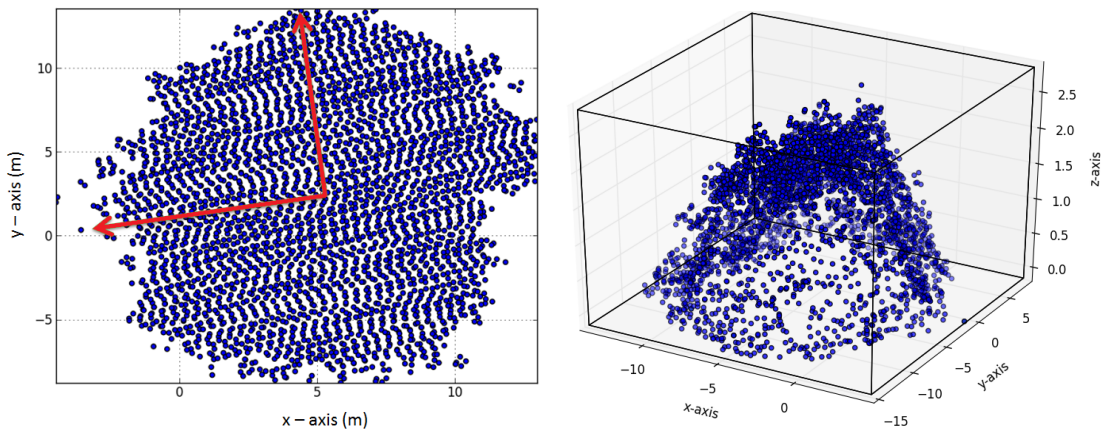


Fig. 2.9: The rock pile from the LIDAR data thresholded to show the identified ground points (green) and the points that comprise the pile (blue).

For the first volume approach, the rock points were projected onto the xy, or the ground plane, as shown in Fig. 2.10 (a). The principal component vectors of these points were calculated to find the directions of maximum variability in the ground plane of the rock pile. The bounding volume of the rock pile was identified by the location of the extrema in the directions of the 2 PCA vectors and the original z-axis of the aligned points. Using the first two principal component vectors of the 3D set of points resulted in the primary and secondary principal component vectors falling outside of the ground plane, since the variability in the z-dimension was not sufficiently smaller than that in the other two dimensions. For this reason, the former method was used to keep the z-axis fixed. This resulting bounding box is shown in Fig. 2.10 (b).

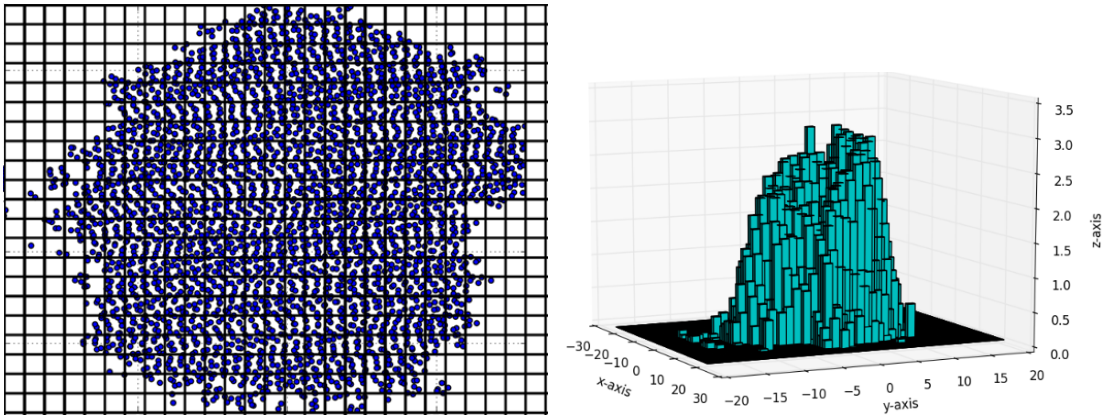


(a) PCA vectors in ground plane *(b)* Bounding Box

Fig. 2.10: The PCA vectors of the projected points in (a) are used to identify the limits of the bounding volume shown in (b).

After constructing the bounding box, the volume of the box was calculated and compared across the multi-resolution point clouds. The volume of the bounding box is not an ideal estimate of the rock pile, rather some fraction of the volume would provide a more accurate estimate. However, the bounding box is still a reasonable measurement to capture the extent of the rock pile dependent on GSD.

A second approach to volume estimation was implemented based on a pseudo-integration of the pile surface. To estimate the surface, the space was divided into equal grid elements on the ground plane. The highest point in each grid element, or bin, was retained and used to calculate the volume of that grid element. All volumes were summed to complete the volume estimate, as shown in Fig. 2.11



(a) Grid elements on ground plane *(b)* Volume Approximation

Fig. 2.11: For each grid element, or bin, in (a) the point with the highest z-coordinate is used to calculate the volume of that element. The final volume is the sum of each of the volumes in (b).

The most accurate volume calculation of such an asymmetrical group of points is beyond the scope of this research and is a topic of research in itself. Therefore, this analysis was restricted to volume calculations founded on basic principles that should still provide insight into how the structure of the rock pile varies with GSD.

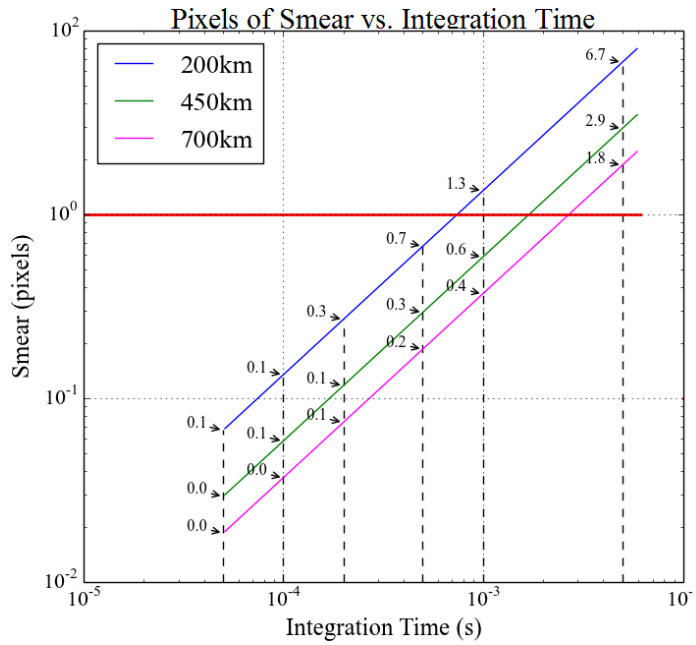
Chapter 3

Results and Analysis

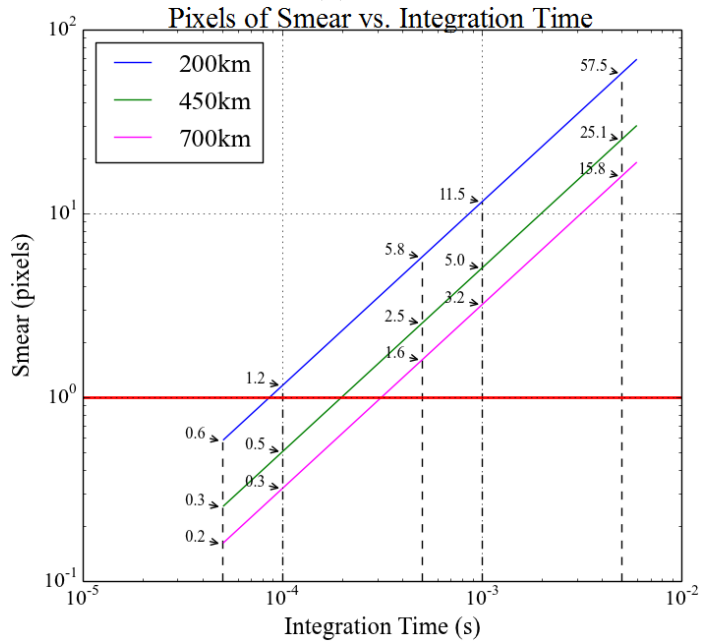
This chapter will begin with the presentation of the cubesat system characterization and discussion of the anticipated repercussions on image quality. Sources of error for the point clouds will be outlined before point cloud results and comparisons are examined in the range of spatial resolutions and signal-to-noise ratios expected for V4. The first dataset examined is the urban scene which contains height estimates of a building and normal estimates to its surfaces. The quarry dataset is analyzed through a local quarry depth estimate and volume estimates of a pile of rocks within the quarry.

3.1 System Evalution

Smear calculations for both V3 and V4 are shown in Fig. 3.1. Both designs were evaluated in order to illustrate the tradeoffs between the V4 design with a Q approaching 2 and the V3 design with a Q closer to 1. The calculations illustrated in Fig. 3.1 indicate that smear will be a problem even at short integration times (below 1 ms) for V4. For an altitude of 200 km, smear limits the V4 integration time to less than 0.1 ms.



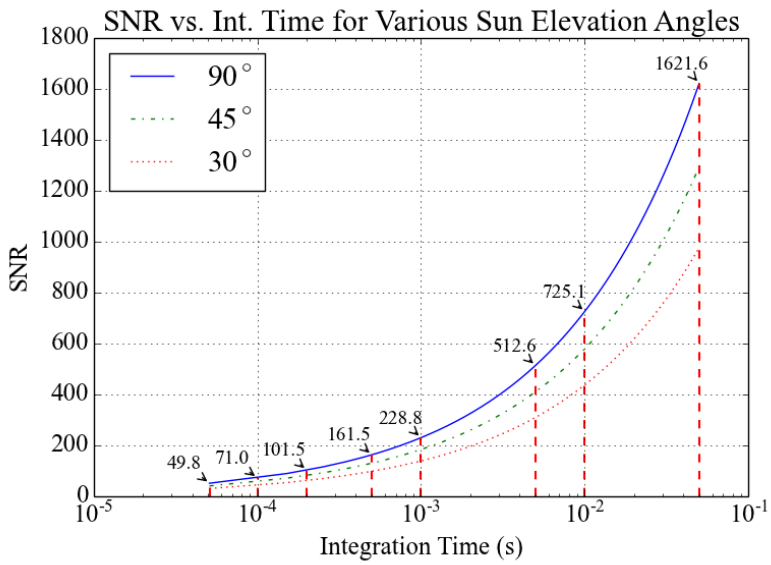
(a) V3



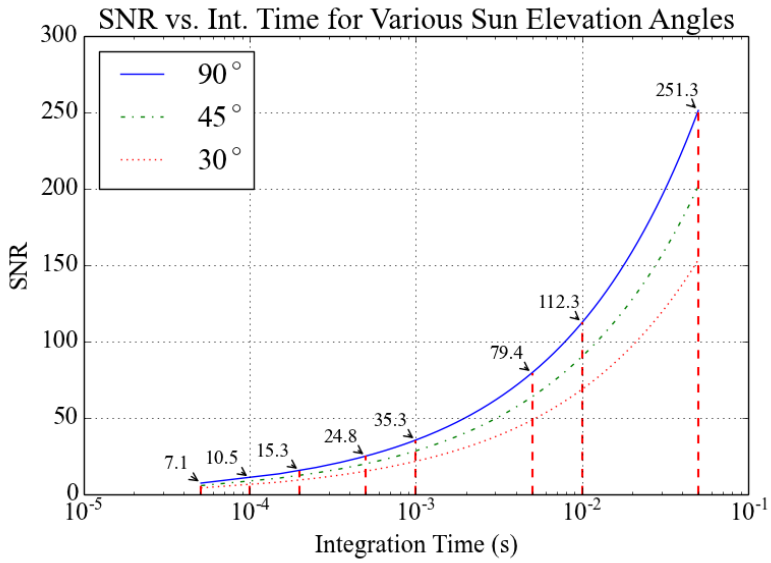
(b) V4

Fig. 3.1: Smear dependent on platform altitude and integration time are illustrated, with a red line denoting the noticeable 1 pixel of smear boundary.

An increase in altitude to sustainable orbits such as 450 km and 700 km allows a slight increase in allowable integration times to approximately 0.2 and 0.3 ms. In contrast, the larger detector elements of the V3 design places the one-pixel boundary of smear at higher integration times. To put the constraints on these integration times into perspective, SNR calculations for both designs are shown in Fig. 3.2.



(a) V3



(b) V4

Fig. 3.2: SNR estimates based on integration time, platform altitude, and system design specifications indicate that SNR. All calculations are for nadir viewing of a 10% reflective target. Various sun elevation angles are indicated by the legend.

The higher integration times for V3 and the larger detector elements for V3 allow the high SNRs shown in Fig. 3.2 (a). However, Fig. 3.2 (b) shows that the integration limits previously discussed for V4 put the SNR expectations for a 10% reflector from imagery collected at 200, 450, and 700 km at 10.5, 15.3, and 20.2, respectively. Note that these SNR calculations are for the full bit depth of the sensor, however, the 3D workflow requires high-contrast, 8-bit images as input. Stretching the contrast of the imagery will amplify the noise and decrease the SNR. Therefore, the actual SNR of the imagery input to the 3D workflow will be lower than the values previously specified. Such low SNR values will certainly present a problem for the feature extraction and matching process, which is pivotal to the remainder of the reconstruction workflow.

3.2 Point Cloud Characterization

3.2.1 Sources of Error

Due to the multiple stages present in the 3D workflow and measurement process, there are many sources of error that affect the results. These include:

1. Image Correspondence Error: occurs when features are improperly matched between images, which occurs more as spatial resolution is lost. This error contributes to the remaining types of error. There are two stages in the 3D workflow at which this error can occur: in the initial SiftGPU and matching stage, and within PMVS itself. Fig. 3.3 shows an example from SiftGPU matching.

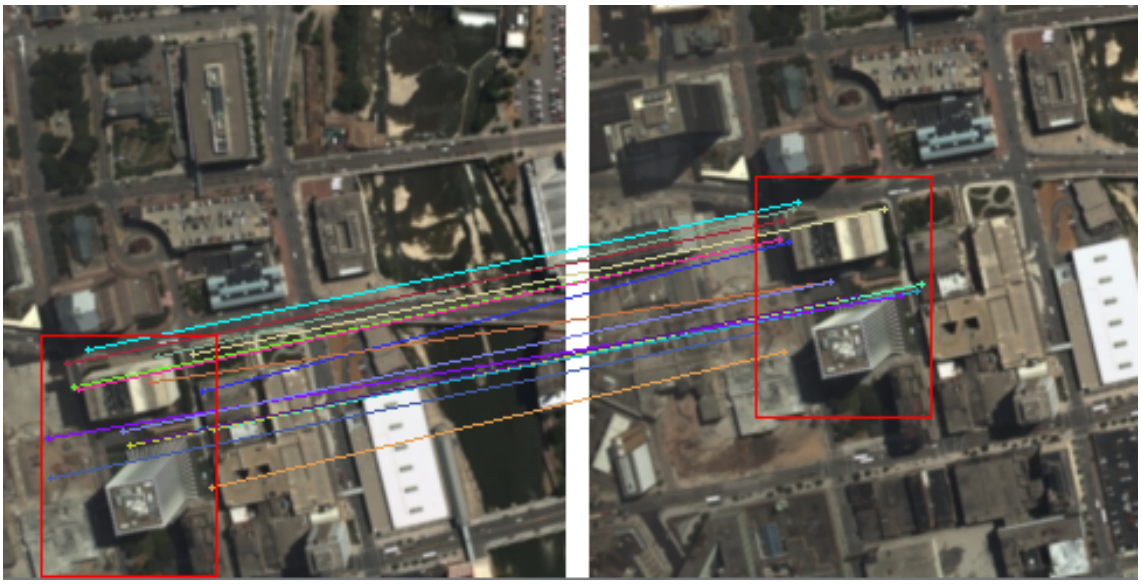


Fig. 3.3: Matches identified from the SiftGPU features are illustrated for subsets of an image pair. The red boxes have been placed to guide the eye to the same region of each image, they do not represent precisely the same locations. Only matches that fall within the red boxes are displayed. Poor image correspondences such as the lowermost orange line are identified.

2. Camera Pose Error: inaccurate calculation of the relative camera pose is affected by poor image correspondences. Error in the camera parameters also contribute to this error. Although bundle adjustment is used to attempt to eliminate the effects of poor image correspondence, a large number of incorrect matches will still result in error.

3. Triangulation Error: Error that occurs on a point-by-point basis caused by mismatches within PMVS and/or by camera pose error. Since PMVS uses the SfM-estimated camera centers to restrict the search space for feature matches, this error is often caused by error in the camera centers. Manifested in noisy points and misshapen surfaces.

4. Georegistration Error: a global error exacerbated by error in camera pose and manifested as an improper scaling, translation, and/or rotation of the entire point cloud. This error makes it difficult to automate isolating the same

region from each point cloud, which is why it was done manually for each case. Examples of this error based on GSD are shown in Fig. 3.4.

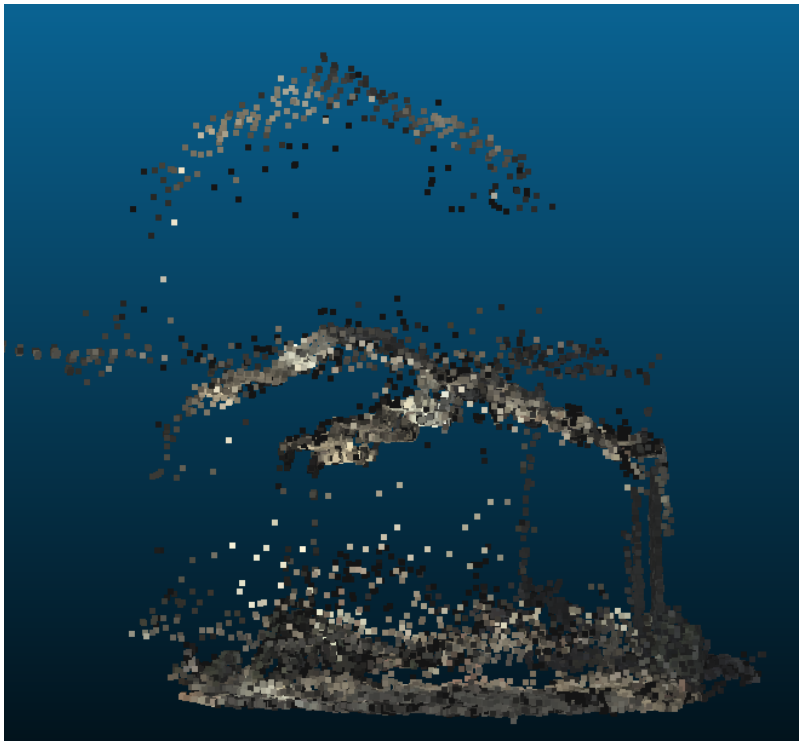


Fig. 3.4: Shifted versions of the Clinton Square building from each point cloud level illustrate error in the georegistration transform. The full resolution case is in the bottom right, while the 2m GSD result is in the upper left. The 2m GSD result has tens of meters of error from the full resolution case in all dimensions.

5. Alignment Error: error induced by the measurement process when rotating the point cloud with the Cartesian axes to enable calculations. Because all of the measurement regions were individually aligned by visual inspection, differences in the measurement can be attributed to the alignment process itself.

The propagation of these errors throughout the workflow are illustrated in Fig. 3.5.

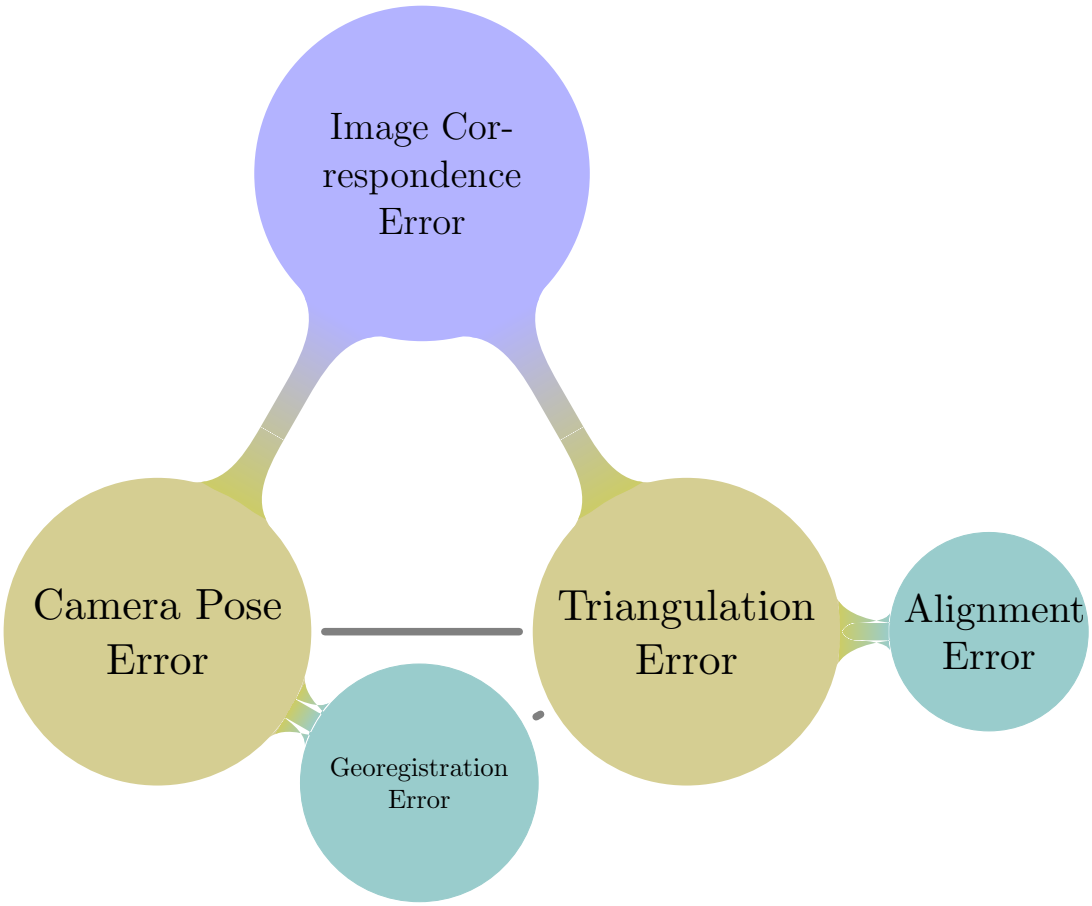
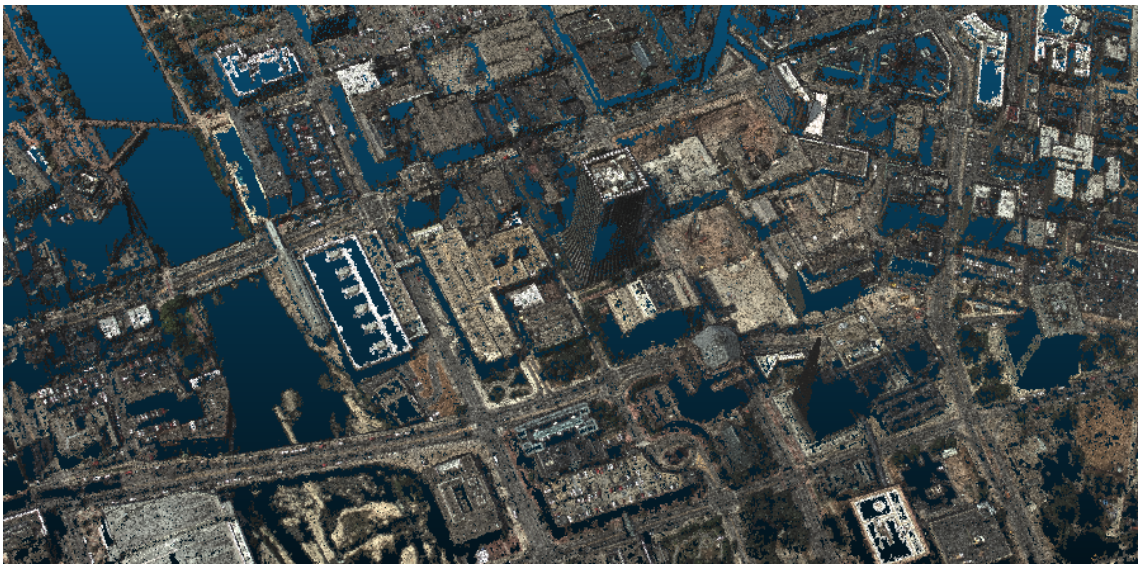


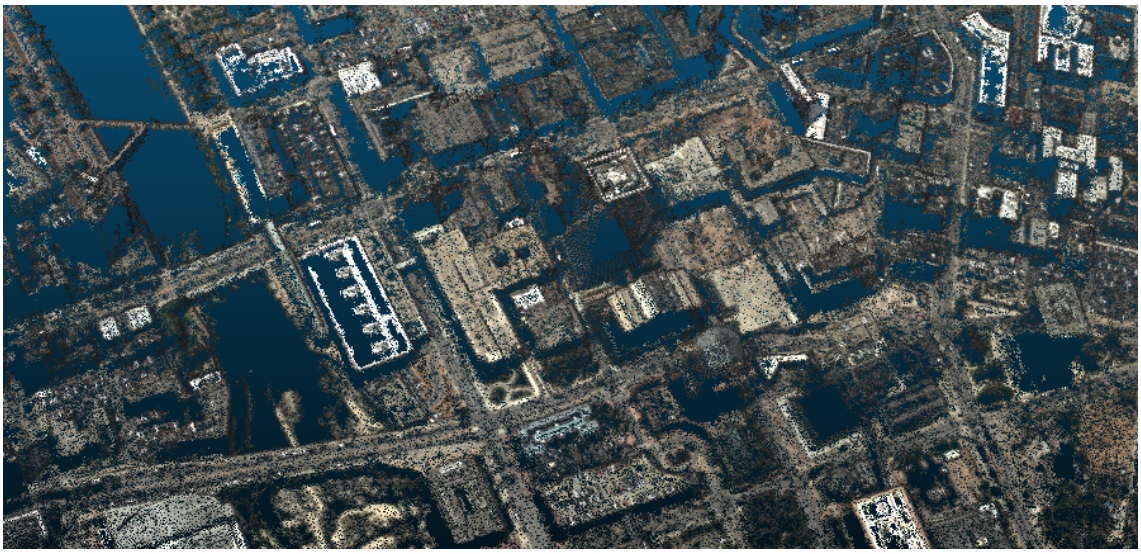
Fig. 3.5: Original image correspondence error propagates through the entire workflow and measurement process.

3.2.2 Building Metrics - Spatial Resolution

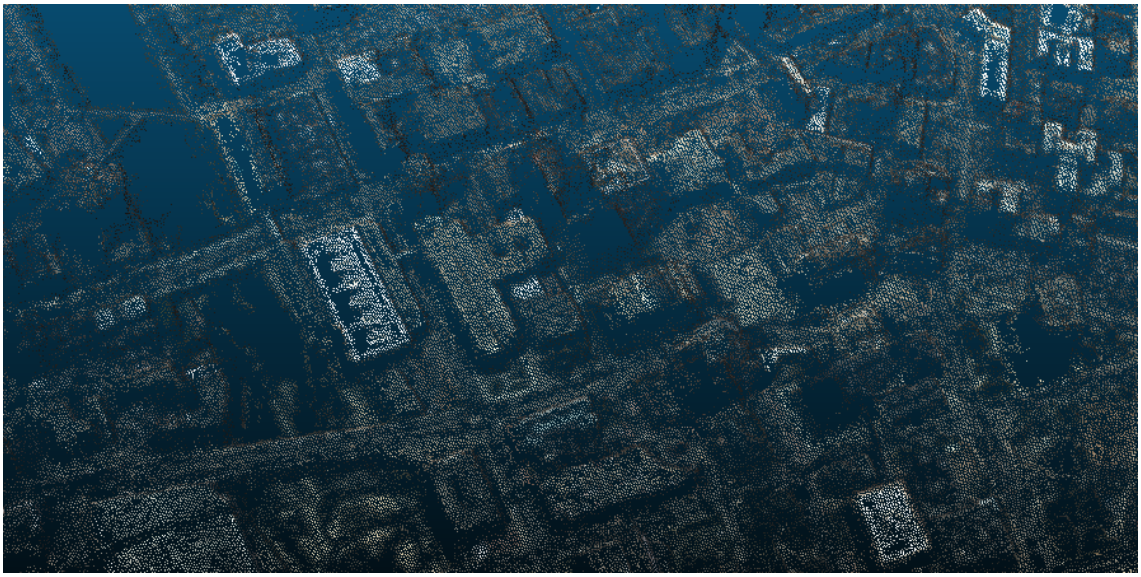
The point clouds produced at each level of image resolution are shown in Fig. 3.6.



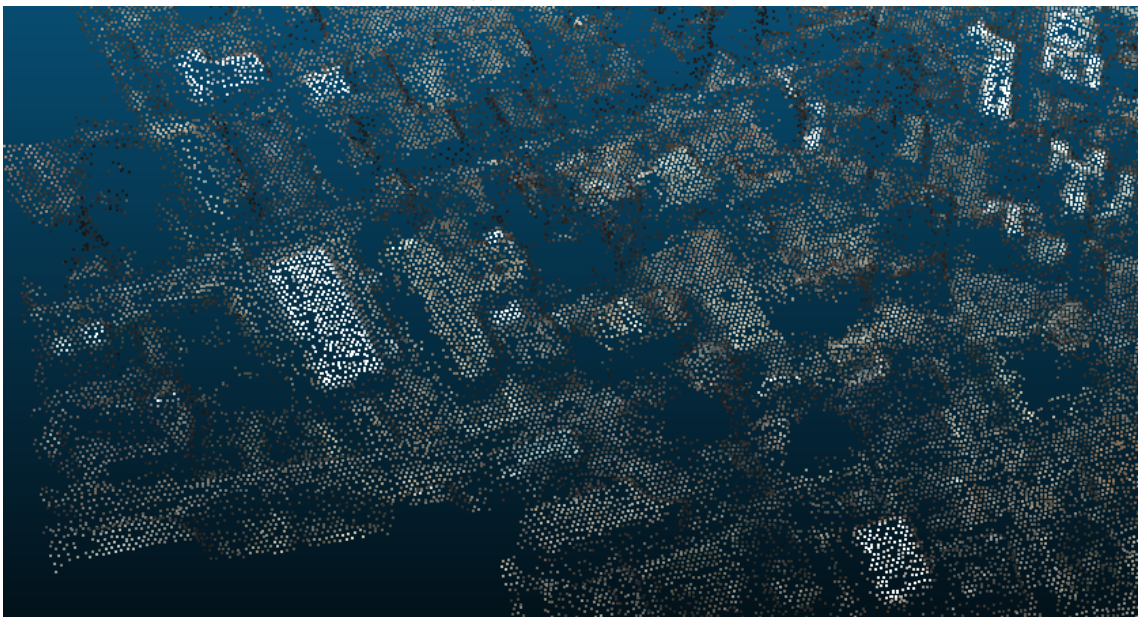
(a) GSD = 0.25m



(b) GSD = 0.5m



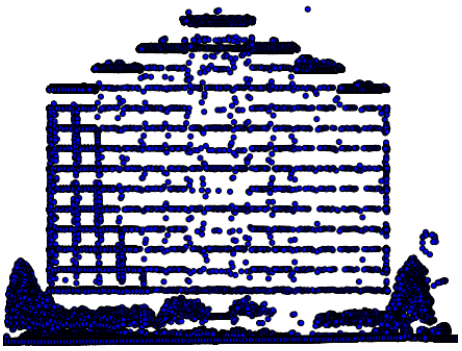
(c) GSD = 1.0m



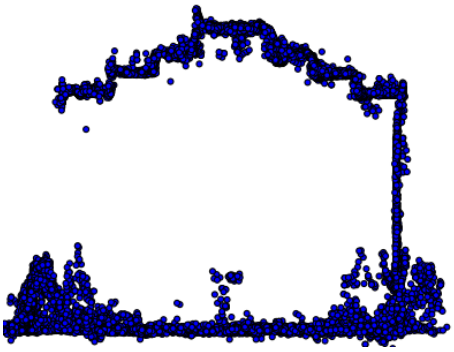
(d) GSD = 2.0m

Fig. 3.6: As the GSD of the input imagery increases from (a) to (d), the number of reconstructed points decreases. The point cloud shown in (d) has points that are larger in size in order to improve visibility.

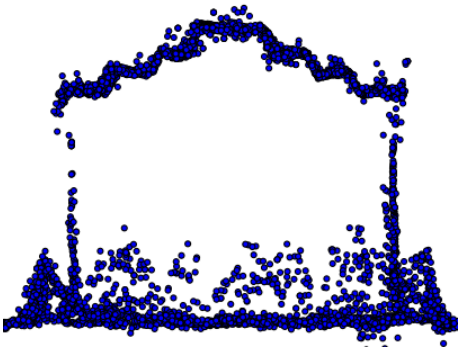
It can be seen in Fig. 3.6 that detail in the point clouds is lost with the loss of spatial resolution, as would be expected. With the simulated increase in GSD, the spacing between points increases making it more difficult to discern the geometry of structures within the scene. To further explore this loss of fidelity in the point clouds, the region of the point cloud containing the Clinton Square building in Fig. 2.5 was isolated. The projections of the 3D points of this building on to the xz plane from the LIDAR data and from the 3D workflow at each level of image resolution are illustrated in Fig. 3.7. The first thing to note from Fig. 3.7 is that the LIDAR building is noticeably more detailed and less noisy than the point cloud produced from the full resolution imagery. The LIDAR result captures points on all surfaces of the building, whereas the result in Fig. 3.7(b) omits points on the front, left, and back faces. This omission is due to the nadir collection of the imagery. The omitted areas of the building are not captured in the imagery, therefore they cannot be reconstructed. Despite these gaps, the image-based point cloud does capture some detail that is not in the LIDAR result. The vertical surfaces between the multi-level rooftop are reconstructed in Fig. 3.7 (b) and (c) but are not captured in the LIDAR result. The remaining image-based point clouds demonstrate noticeable distortion in the geometry of the building, culminating in the blob of points depicted at the 4m GSD level (Fig. 3.7 (f)) which hardly resembles the actual building. The lowest level of image resolution was omitted from further analysis for this reason.



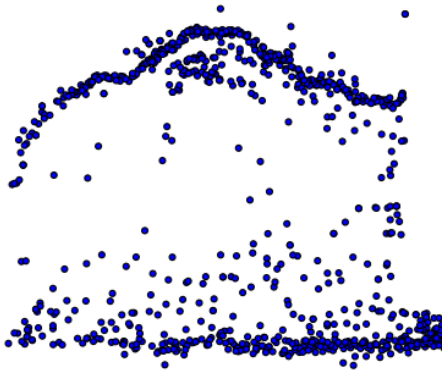
(a) LIDAR



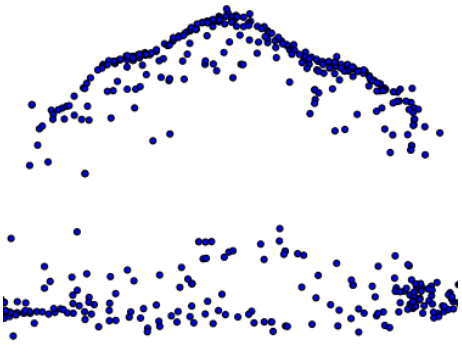
(b) GSD = 0.25 m



(c) GSD = 0.5 m



(d) GSD = 1 m



(e) GSD = 2 m

(f) GSD = 4 m

Fig. 3.7: Projections of the Clinton Square building onto the xz plane. Note that these projections only appear to be profiles because the front and back planes of the building were not reconstructed due to the nadir imagery collection.

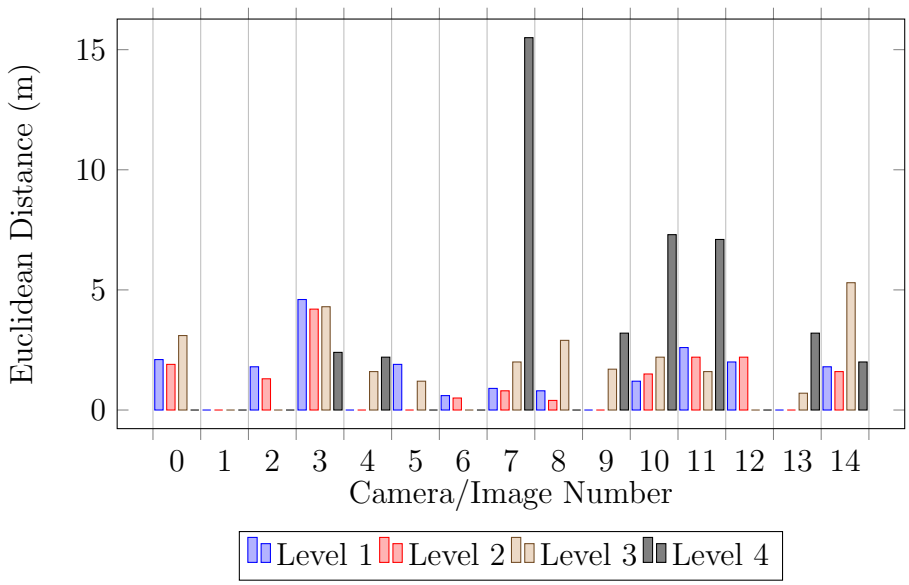
3.2.2.1 Building Height Estimation

The calculated heights for the first four levels are reported in Table 3.1. The spatial

Tab. 3.1: Building Height Estimates Based on GSD

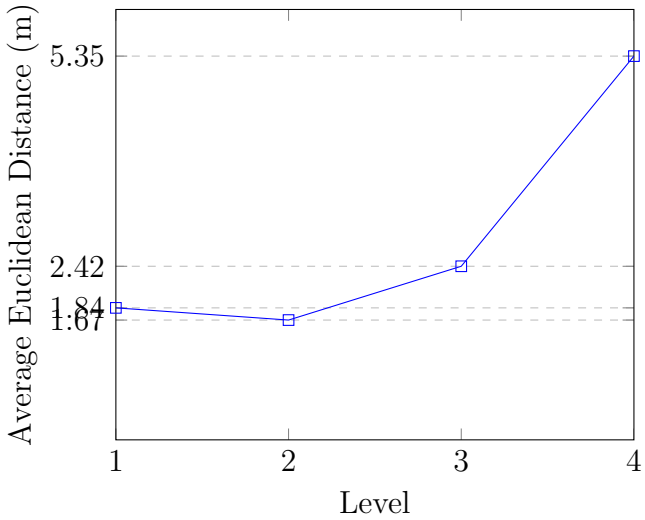
Level	GSD (m)	Height (m)
LIDAR	N/A	59.8
1	0.25	59.7
2	0.5	59.5
3	1.0	58.5
4	2.0	58.0
5	4.0	N/A

degradation correlates to increased error in the height estimates. The error between the first two levels was minor and was attributed to error introduced by aligning the building with the axes. The larger error of the second and third levels was expected, as features were lost due to the downsampling process. While the top surface of the building appears to be reconstructed appropriately, note that the definition of the ground worsened noticeably at Level 3 (Fig. 3.7(d)), affecting the height estimate. Furthermore, the sparse reconstruction and warped scene geometry are indicative of a larger problem that increases error in the height estimate: the loss of spatial resolution rendered the SfM process more susceptible to error in image correspondence. The image correspondence error translated to camera pose error. Figure 3.8 shows the increased error in each of the SfM estimated camera centers with respect to the GPS camera centers, which directly affects the model reconstructed by PMVS.



(a)

Average Error of SfM Camera Centers vs. Level



(b)

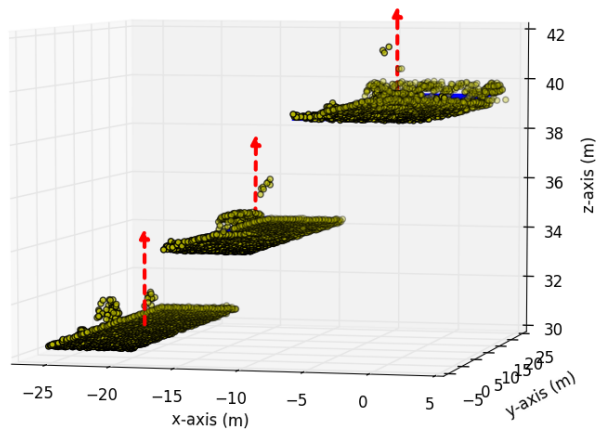
Fig. 3.8: The Euclidean distance between each SFM estimated camera center and its GPS center counterpart is shown for each of the four GSD levels. Note that the same cameras were not used in each level of reconstruction due to CMVS.

Figure 3.8 shows that the SfM-calculated camera pose have significantly larger error from the GPS camera centers at levels three and four. This error propagated

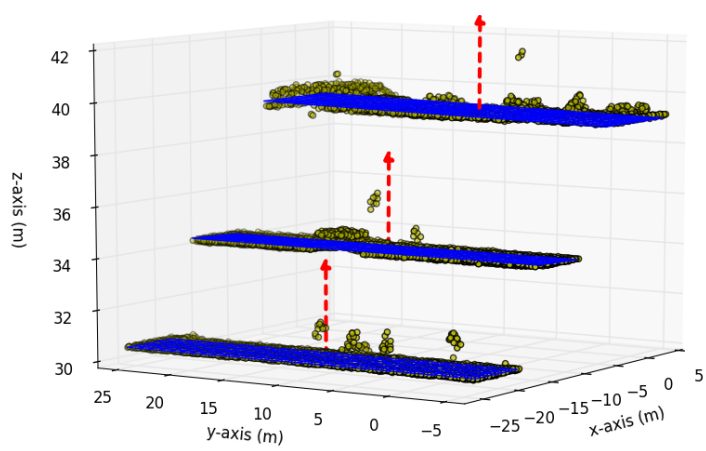
through the triangulation process of the PMVS algorithm, manifested by the false z-coordinates that fell between the actual z-coordinates of the multi-level surfaces. In the end, it is image correspondence error that increases with the loss of spatial resolution and initiates a chain reaction of error that affects the building height estimate.

3.2.2.2 Surface Normal Estimation

For reference, the surfaces that will be examined to assess the reconstructed geometry of the 3D workflow are first illustrated by the LIDAR data in Fig. 3.9.



(a) Surface 1



(b) Surface 2

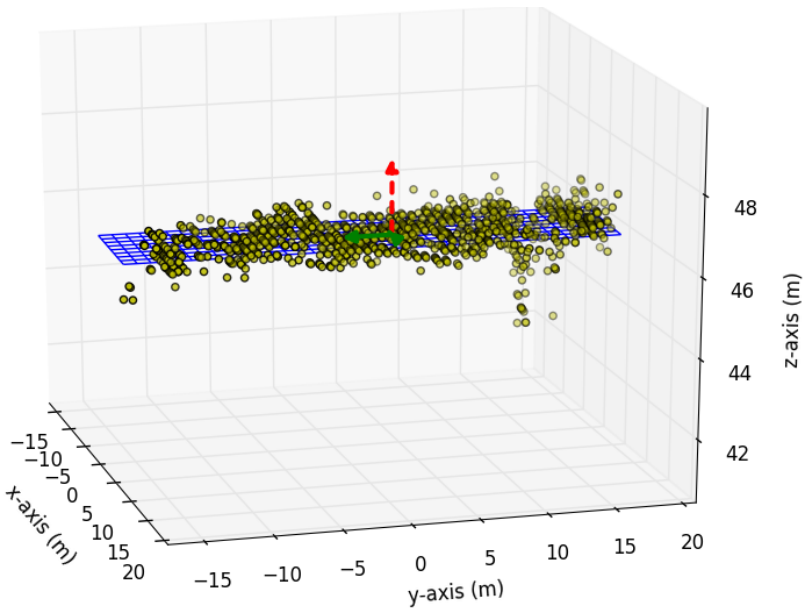
Fig. 3.9: Three surfaces of the Clinton Square building captured from the LIDAR platform provide a reference for the reconstruction of the same surfaces from imagery. Note that the z-axis scale is different than that of the x and y axes for viewing purposes; the variation of the points in the z-dimension is much smaller than the variation in the other two dimensions.

The LIDAR point cloud demonstrates densely populated surfaces, with few noisy points as shown in Fig. 3.9. Note that some small structures are present within the surfaces, though the surface normal estimation approach assumes the entire section is one flat surface. The results of surface normal estimation for all data are summarized in Table 3.2.

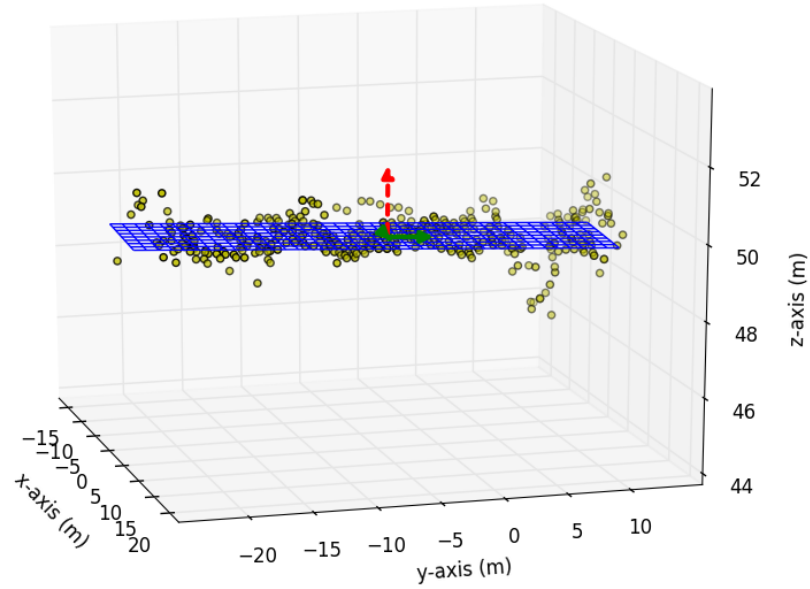
Tab. 3.2: Surface Normal Estimates Based on GSD

GSD(m)	Num. of Pts.	Normal	Variance (m^2)
		Surface 1	
LIDAR	4401	[-0.001, -0.006, 1.000]	0.029
0.25	1130	[0.011, 0.002, 1.000]	0.151
0.5	374	[0.020, 0.018, 1.000]	0.224
1.0	80	[-0.020, 0.024, 1.000]	0.407
2.0	40	[0.035, 0.072, 0.997]	0.688
		Surface 2	
LIDAR	3789	[0.015, 0.000, 1.000]	0.024
0.25	793	[0.037,-0.009, 0.999]	0.146
0.5	321	[-0.033, 0.009, 0.999]	0.251
1.0	87	[-0.507, 0.013, 0.862]	0.661
2.0	16	[-0.545, 0.016, 0.838]	1.743
		Surface 3	
LIDAR	3857	[0.003, 0.001, 1.000]	0.017
0.25	700	[0.001,-0.005, 1.000]	0.155
0.5	330	[0.010, 0.018, 1.000]	0.177
1.0	80	[-0.163, 0.001, 0.987]	0.712
2.0	23	[-0.262,-0.010, 0.965]	0.894

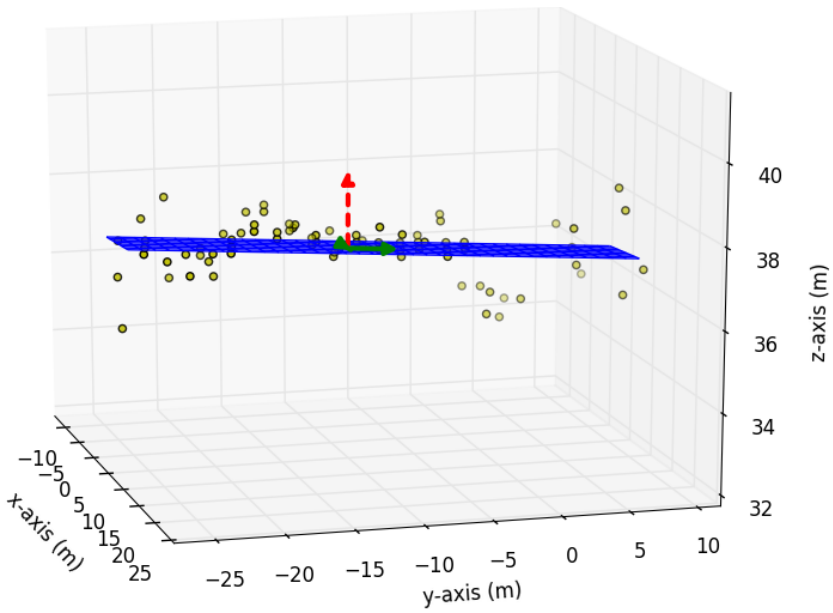
The variance of the surface points increased as the spatial resolution was degraded. This can be seen in close-ups of the best (Surface 1) and worst (Surface 2) case surfaces, shown in Figs. 3.10 and 3.11 respectively.



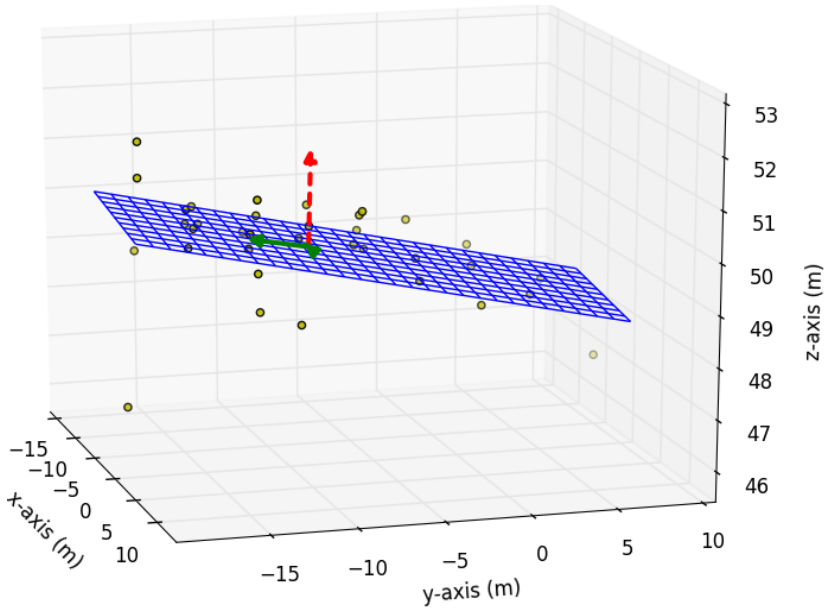
(a) GSD = 0.25m



(b) GSD = 0.5m

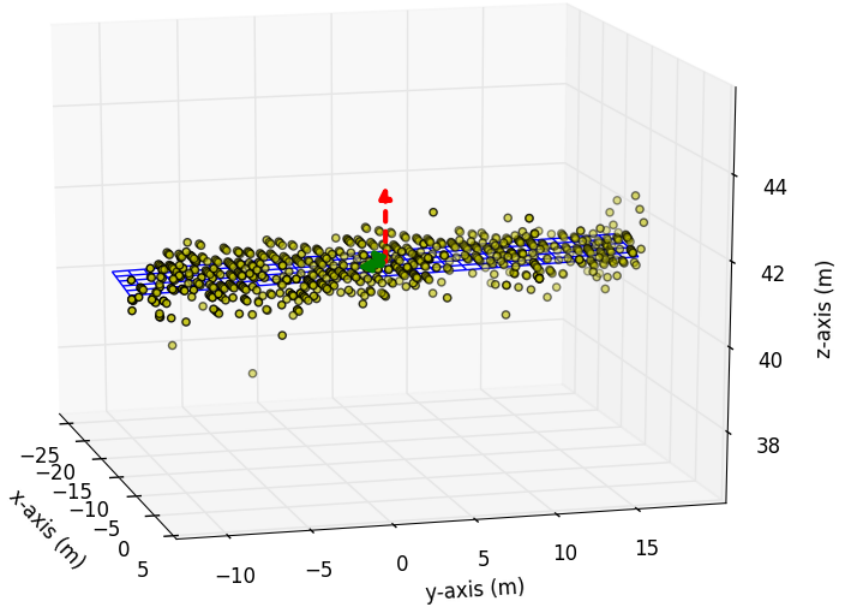


(c) GSD = 1.0m

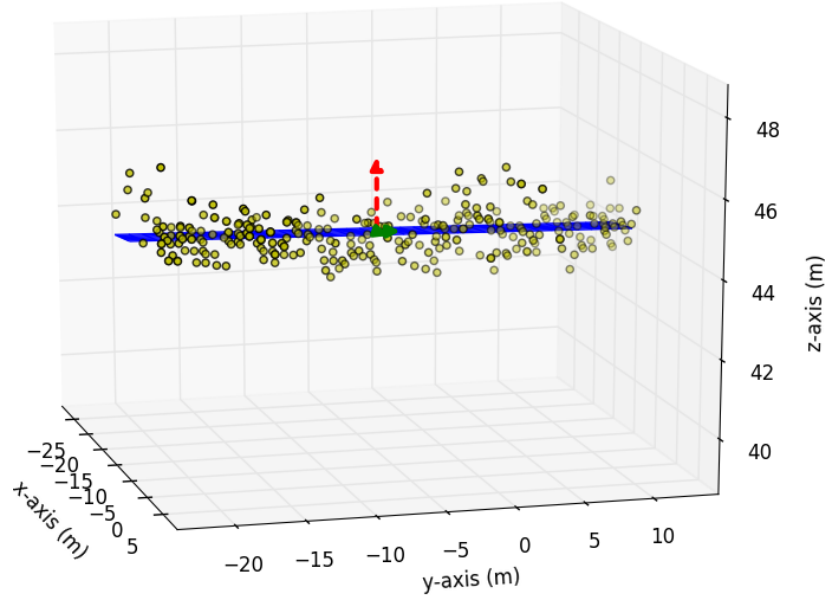


(d) GSD = 2.0m

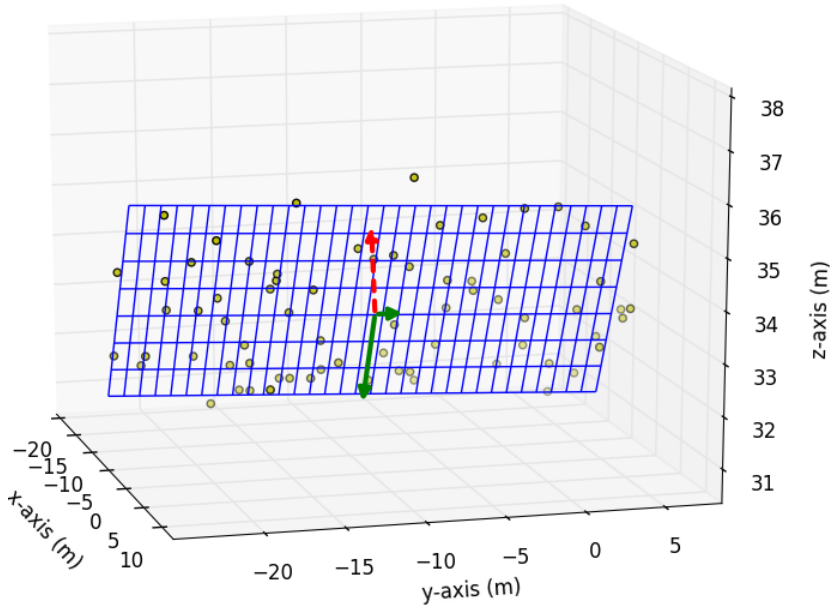
Fig. 3.10: The plane fit to the points by the space defined by the first two PCA vectors, and corresponding normal are shown. The z-axis has been scaled differently than the x and y axes in order to highlight variation in the z-dimension.



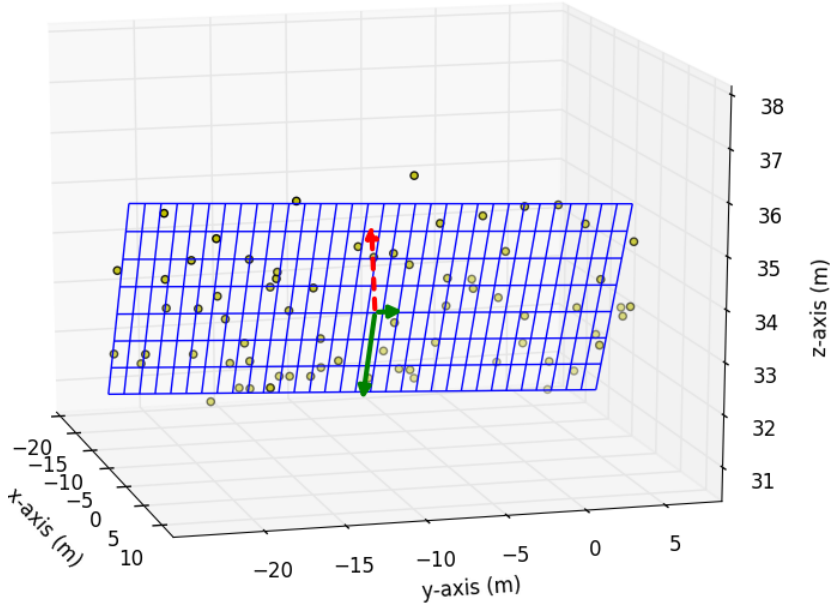
(a) GSD = 0.25m



(b) GSD = 0.5m



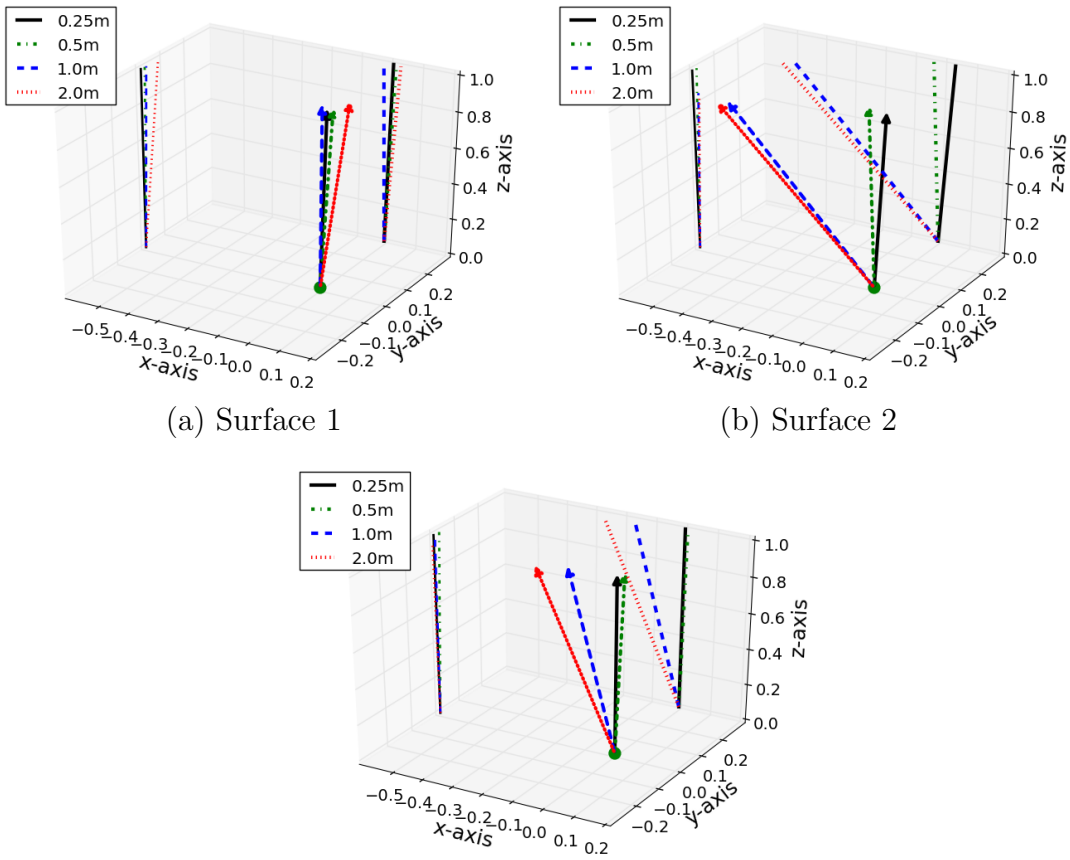
(c) GSD = 1.0m



(d) GSD = 2.0m

Fig. 3.11: The plane fit to the points by the space defined by the first two PCA vectors, and corresponding normal are shown. The z-axis has been scaled differently than the x and y axes in order to highlight variation in the z-dimension.

The image correspondence error previously discussed in the building height estimation was also responsible for this trend. The variation in the direction of the estimated surface normals is illustrated in Fig. 3.12. Using the estimated normal



(a) Surface 1

(b) Surface 2

(c) Surface 3

Fig. 3.12: Estimated surface normals are shown for each surface in Fig. 2.5 along with their projections onto the xz and yz planes for comparison.

from the full resolution data as truth, the maximum angle between the remaining estimated normals and truth for the first surface was 4.3 degrees. The second surface had a maximum angle, or error, of 35.2 degrees and the third had an error of

15.3 degrees. All errors are summarized in Fig. 3.13. The variation in these errors

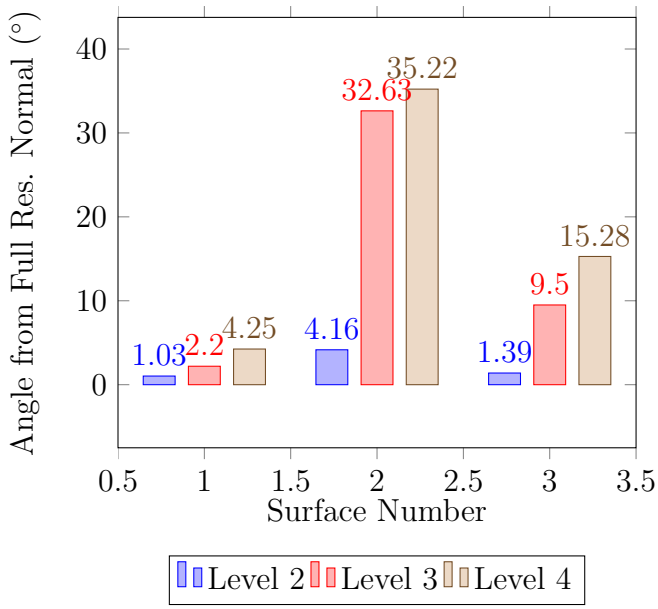


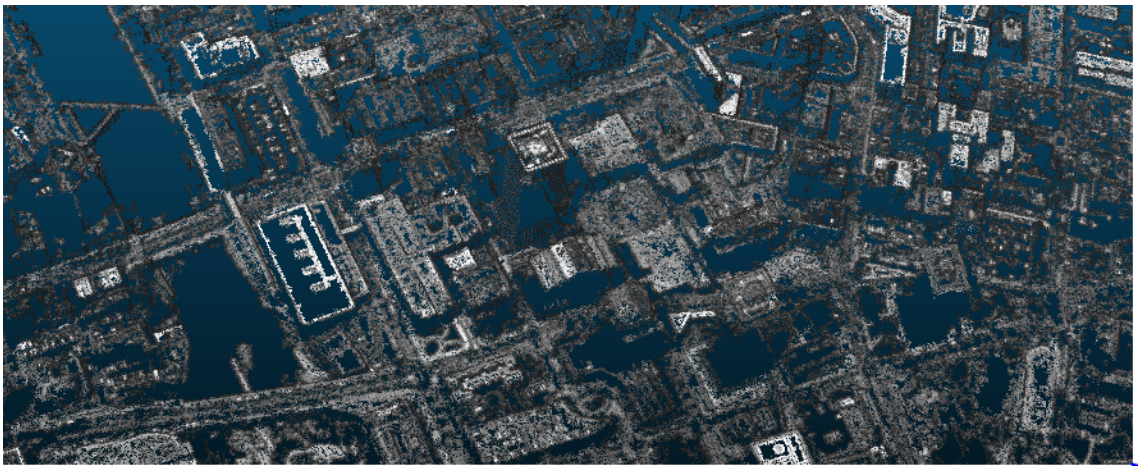
Fig. 3.13: The surface normal errors are shown with respect to the highest resolution point cloud.

is explained by the features on each surface, which are shown in Fig. 2.5. Because the second surface is so barren, its points were prone to larger error in image correspondence than the points on the other two surfaces. In contrast, the first surface has larger objects on it that are still visible at lower resolution, thus the accuracy of the reconstruction of this surface was more robust to the loss of spatial resolution. These cases of surface normal estimation illustrate how the scene content affects the reconstruction accuracy.

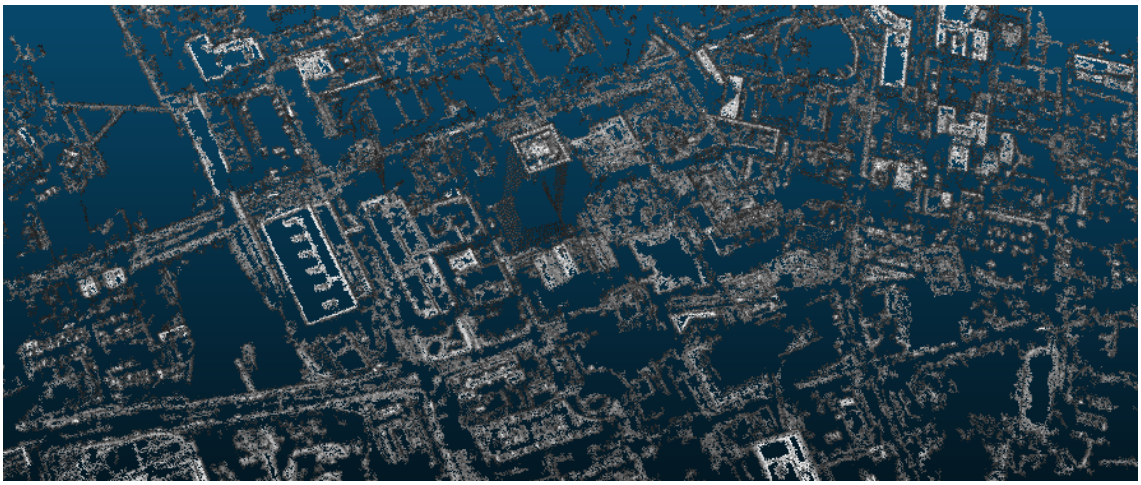
3.2.3 Building Metrics - SNR

The low SNR chosen from the one-pixel boundary of smear drastically reduced the number of reconstructed points. Therefore, intermediate signal-to-noise ratios were

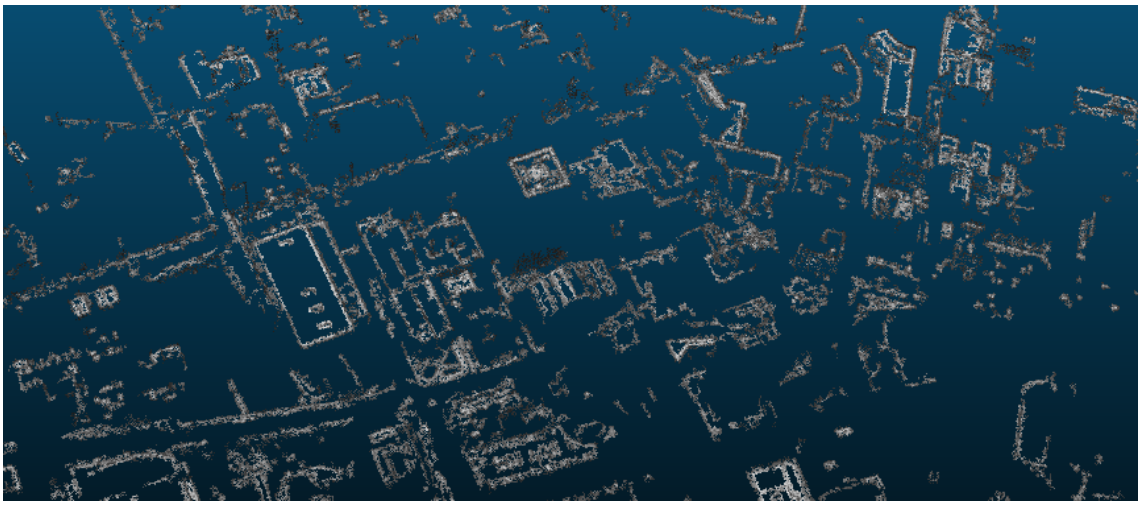
simulated. Views of the point clouds produced from this range of SNR values are shown in Fig. 3.14.



(a) SNR = 73.4



(b) SNR = 35.3

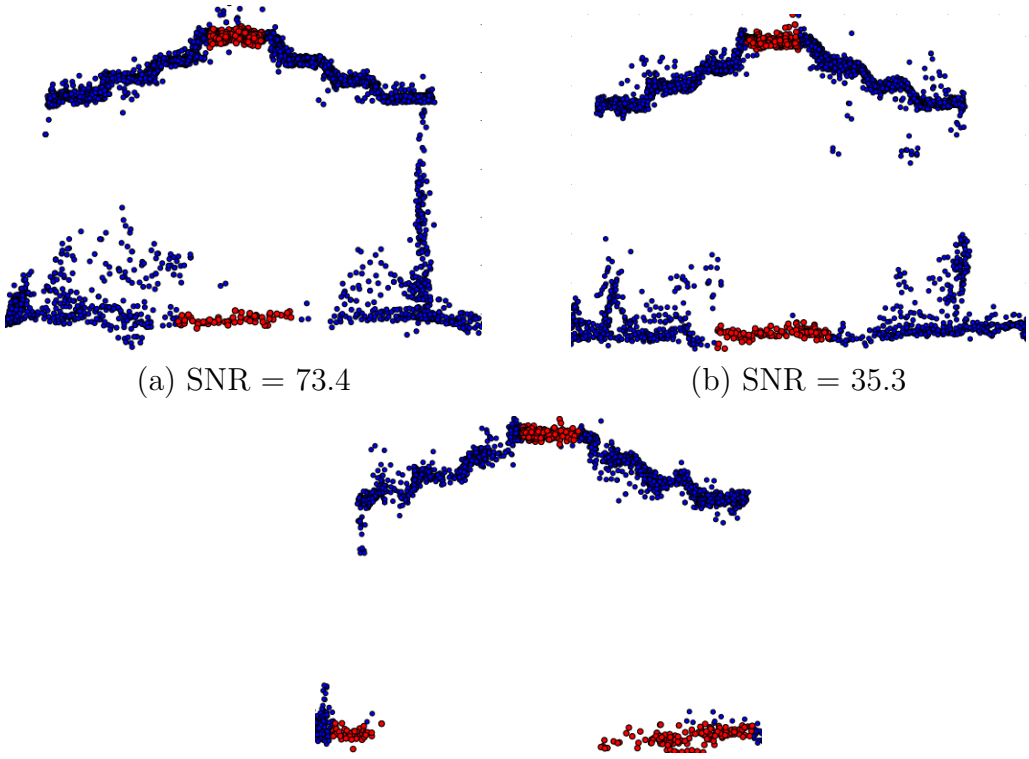


(c) SNR = 15.3

Fig. 3.14: Shown from left to right, a wide area of the point clouds are shown as the SNR is decreased.

Recall from Fig. 1.6 that images with lower SNR cause false extrema to be detected by the DoG filter. As the DoG filter is also used in PMVS to detect patch centers that will be reconstructed, this high SNR causes unreliable noise artifacts to be picked out as strong features. These are not reliable patch centers as they obscure actual matching features between the images, thus only the strongest edges that can be detected over the noise remain in the reconstruction at low SNR values.

Projections of the Clinton Square building from input imagery of various SNR (based on 10% target reflectance) are shown in Fig. 3.15.



(a) SNR = 73.4

(b) SNR = 35.3

(c) SNR = 15.3

Fig. 3.15: From left to right, the projection of the Clinton Square building on to the xz plane is shown for decreasing SNR values. Red points indicate values that were used to produce height estimates.

It can be seen that the projection of the rooftop loses the definition shown in Fig. 3.15 (a) as SNR decreases. This is likely due to an increased error in correspondences between the images due to increased noise. Not only do the shapes of the surfaces start to warp, but they also appear to be more noisy in terms of their triangulated locations. The results of the height estimation process are summarized in Table 3.3.

Tab. 3.3: Building Height Estimates Based on SNR

SNR	Height (m)
LIDAR	59.8
73.4	59.5
35.3	59.1
15.3	57.2

The error between the LIDAR measurement and those corresponding to the highest two SNR levels are within the error of the alignment and measurement process. The height measurement for the third and lowest SNR suffered a larger error. This SNR was identified as the highest SNR that can be achieved by the V4 imager due to a one pixel limit on smear at an orbit altitude of 450 km. Again, this error was attributed to image correspondence error caused by the decreased SNR of the input imagery. While the completeness and the geometry of the point cloud suffer noticeably at the lowest SNR value, the height estimate only deviated from the LIDAR estimate by less than three meters.

3.2.3.1 Surface Normal Estimation

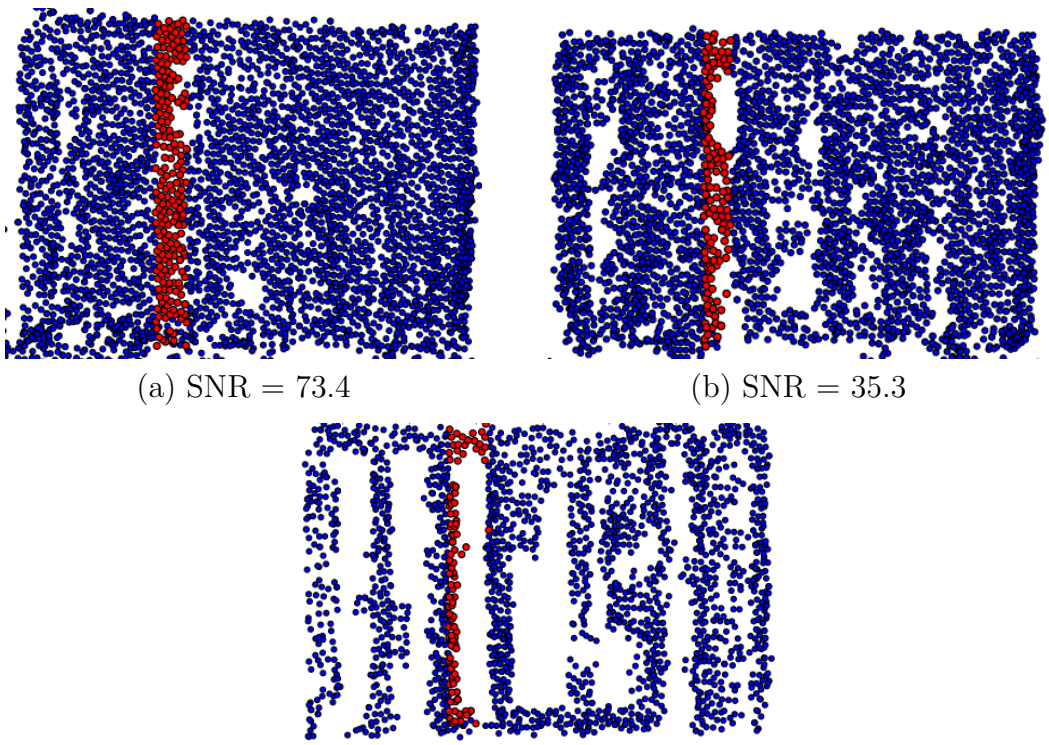
The surface normals were estimated for the best case (Surface 1) and worst case (Surface 2) surfaces, as identified from the spatially degraded data in Section 3.2.2. The results are summarized in Table 3.4 The estimated surface normals dependent

Tab. 3.4: Surface Normals Estimates Based on SNR

SNR	Num. of Pts.	Normal	Variance (m^2)
		Surface 1	
73	334	[0.017, 0.004, 1.000]	0.311
35	271	[0.003, 0.024, 1.000]	0.476
15	106	[0.010, -0.007, 1.000]	0.558
		Surface 2	
73	233	[-0.031, -0.014, 0.999]	0.271
35	153	[-0.029, 0.014, 0.999]	0.717
15	127	[-0.297, -0.007, 0.995]	0.572

on SNR reflect a similar trend as those dependent on spatial resolution. As the SNR

worsens, the variance of the points comprising the surface in the normal direction increases. Again, this reflects the effects of image correspondence error on the final point cloud. The exception is the larger variance for the SNR = 35 case for Surface 2, relative to the SNR=15 case. The reason for this exception is illustrated in Fig. 3.16. The red points used for the surface normal calculation in Fig. 3.16 (a) and (b) show



(a) SNR = 73.4 (b) SNR = 35.3

(c) SNR = 15.3

Fig. 3.16: For the three SNR cases, projections onto the xy plane of the Clinton Square building are shown. Red points indicate those used for the normal calculations of Surface 2.

that points throughout the surface are included in the calculation, whereas Fig. 3.16 (c) shows that only points on the edges of the surface are reconstructed and used for the calculation. The lack of reconstructed points on the interior of the surface is

due to overwhelming noise that obscures features available in that area. At this low SNR, only points on the edges of the surface can still be detected and matched in the 3D workflow. These are stronger features, so it is expected that their variance in the normal direction will be lower. This effect is also responsible for the fairly consistent accuracy of the surface normal direction dependent on SNR. Because lowering the SNR aggressively filters out weaker features from the reconstruction, it has less of an effect on the surface normal accuracy than the spatial degradation presented in Section 3.2.2.

3.2.4 Quarry Metrics

The measurement regions that will be discussed in this section are illustrated in the point cloud produced from the full resolution imagery, shown in Fig. 3.17.

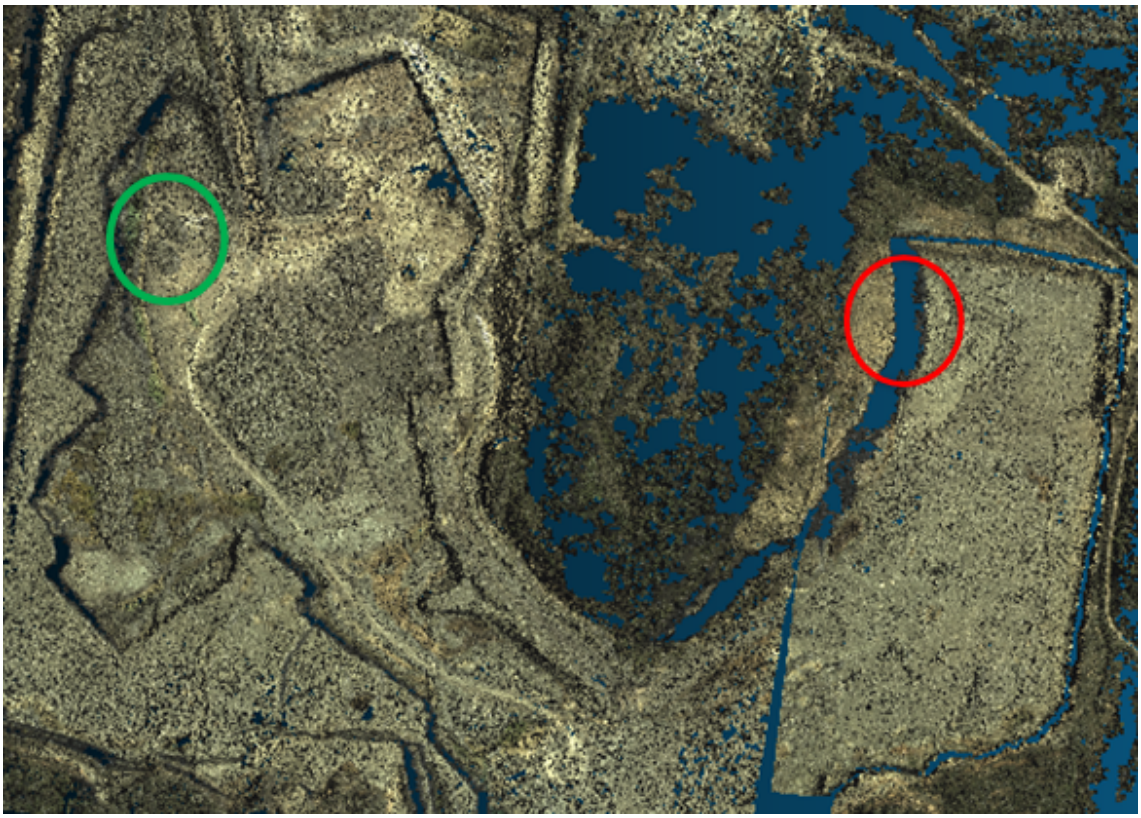
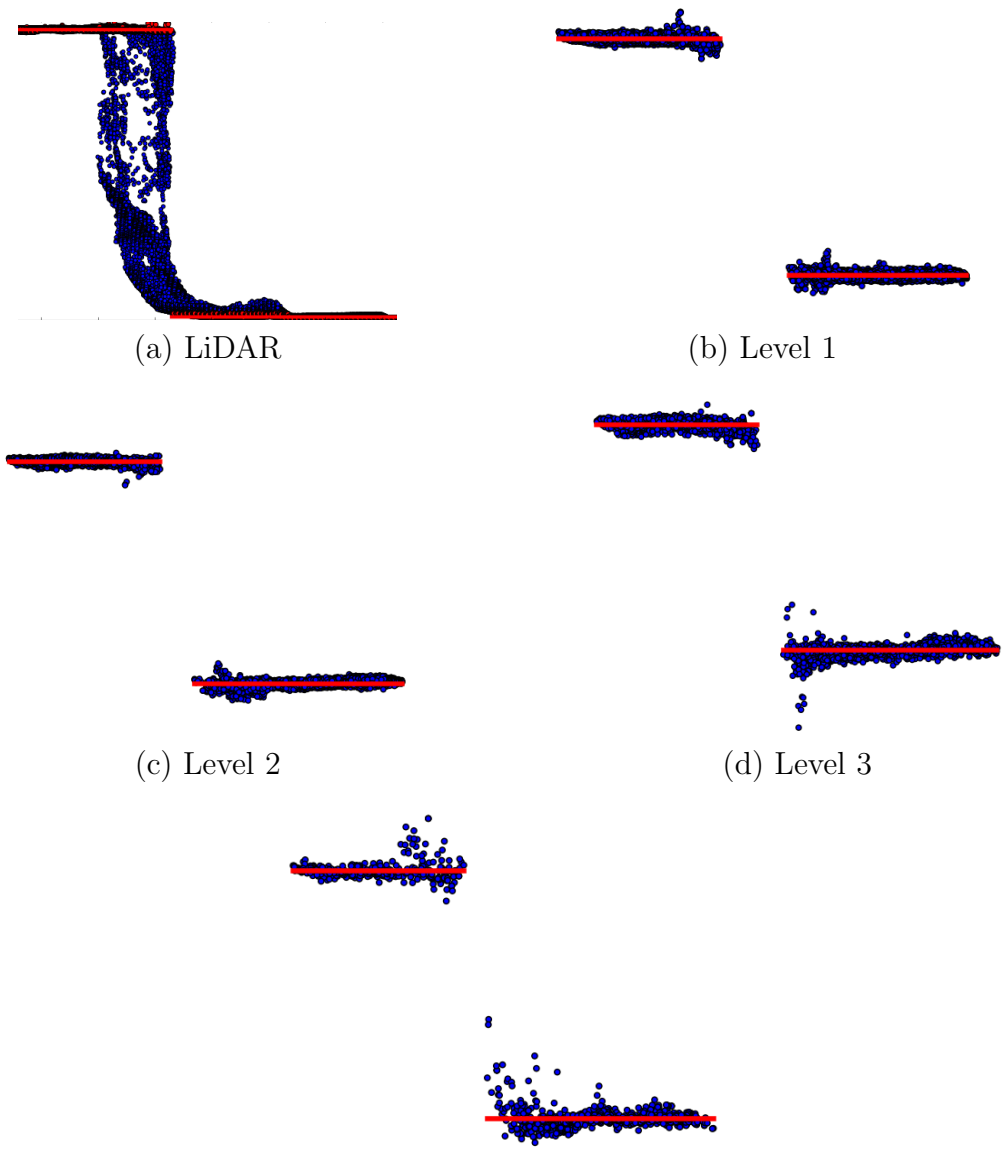


Fig. 3.17: The measurement region where the depth of the quarry was calculated is shown in red while the rock piles are indicated by the green circle. Refer to Fig. 2.7 to see these regions from actual WASP imagery.

3.2.4.1 Wall Depth Estimation

The projections of the 3D points of the quarry wall onto the yz plane at each level of image resolution are illustrated in Fig. 3.18.



(a) LiDAR (b) Level 1

(c) Level 2 (d) Level 3

(e) Level 3

Fig. 3.18: Note that the apparent gap between the two surfaces was manifested because of shadow cast in that region of the imagery (see Fig. 2.7 (a)). Red lines indicate the mean z-coordinates of the surfaces used for the height estimate.

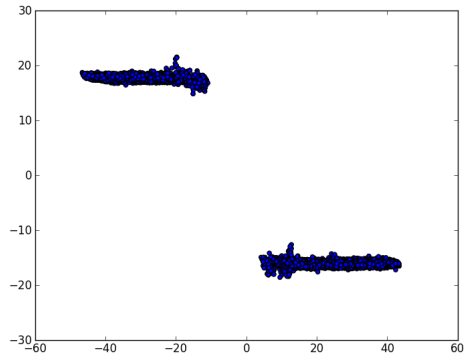
The LIDAR data in Fig. 3.18 captures the actual downward slope that is obscured by shadow in the WASP imagery, and therefore omitted from the reconstructions.

Note that the top and bottom surfaces in the measurement region have much better definition in the LIDAR data, whereas the reconstructions show more uncertainty in the z-dimension. The depth estimates for the first four levels are reported in Table 3.5.

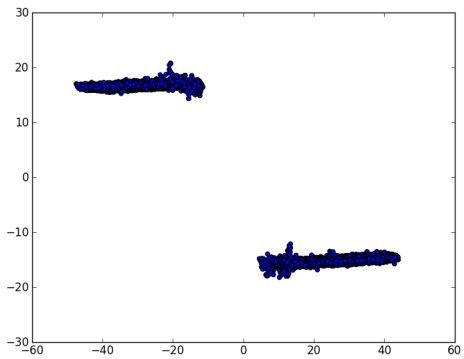
Tab. 3.5: Quarry Depth Estimates Based on GSD

Level	GSD (m)	Height (m)
LIDAR	N/A	33.99
1	0.15	33.73
2	0.30	33.81
3	0.60	33.59
4	1.2	34.15

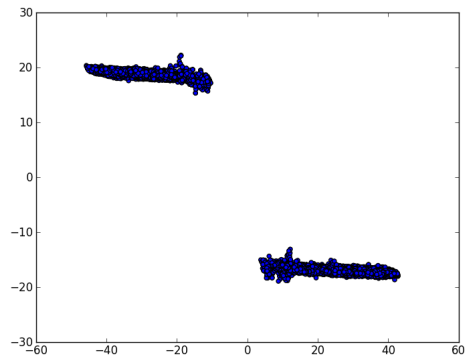
Although the reconstructions are noticeably different, the mean values used for the depth calculation and indicated in Fig. 3.18 by red lines, were not significantly affected. This is reflected in the reported depth measurements by no more than one meter of error at any level. These results demonstrate that even at the 1.2 meter GSD, the depth measurement is accurate. Due to such small error, the measurement is most limited by the alignment error, as approximately one meter of variation can occur due to uncertainty in aligning the surfaces with the ground plane. This alignment error is illustrated by the slight variations in rotation shown in Fig. 3.19.



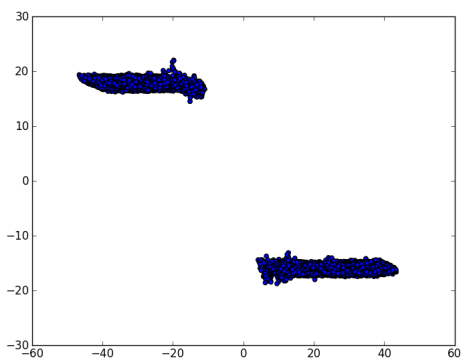
(a) Reference



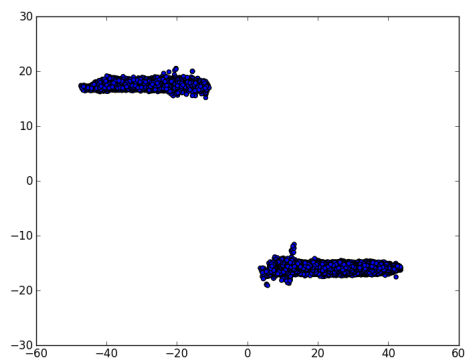
(b) -2° about x-axis



(c) +2° about x-axis



(d) -2° about y-axis



(e) +2° about y-axis

Fig. 3.19: The aligned cloud that was used for the calculation reported in Table 3.5 is shown in (a), while (b) and (c) show the cloud rotated by a difference of 2° from (a) about the x-axis in either direction and (d) and (e) show rotation about the y-axis.

The slightly rotated (2° about the x or y-axis in either direction) point clouds in Fig. 3.19 show that slight rotation error can cause noticeable alignment error. It is evident that the surfaces are no longer parallel with the axes in Fig. 3.19 (b) and (c). The error in Fig. 3.19 (d) and (e) is slightly less noticeable, but is manifested in the increased thickness of the surfaces in the vertical direction. Therefore, these limits were used to perform a statistical analysis of the effects of rotation on the depth estimate. Using the original alignment angles for Level 1 from Table 3.5 as the mean value of a normal distribution, 1000 randomly drawn values from the distribution were applied to the point cloud illustrated in Fig. 3.19 (a), for each of which the depth measurment was recalculated. The standard deviation of this normal distribution was set to one degree, so that approximately 95% of the values would be less than or equal to two degrees from the mean. The results of this analysis about both the x and y axes are summarized in Table 3.6.

Tab. 3.6: Effect of Alignment Error on Quarry Depth Estimate for Level 1

Axis of Rotation	Min (m)	Max (m)	Mean (m)	Std. Dev. (m)
x	30.59	36.51	33.70	0.94
y	33.35	34.04	33.73	0.093

The mean values of the depth estimate from this analysis are extremely close, if not equal to the 33.73 m estimate reported in Table 3.5, as expected. The minimum and maximum values, however, show up to 2 m of error induced by the previously mentioned limits. Note that the same limit of rotation about the y-axis results in a standard deviation that is an order of magnitude lower than for rotation about the x-axis. This difference occurs because the points on the surfaces span much more space along the y-dimension than along the x-dimension (refer to Fig. 2.8 (a)). Therefore, the same angle of rotation about either axis will cause the extrema along each axis to deviate by a different amount from the center of rotation. The analysis of the alignment error verifies that it can cause larger errors than observed in Table 3.5, verifying that it is the limiting error in this analysis. A more reliable

method for calculating the depth would be invariant to rotation, possibly a method which fits a plane to the two surfaces and calculates the distance in the direction nearest to perpendicular to both planes. However, the estimates produced with this approach are accurate to within 1.67% error from the LIDAR data.

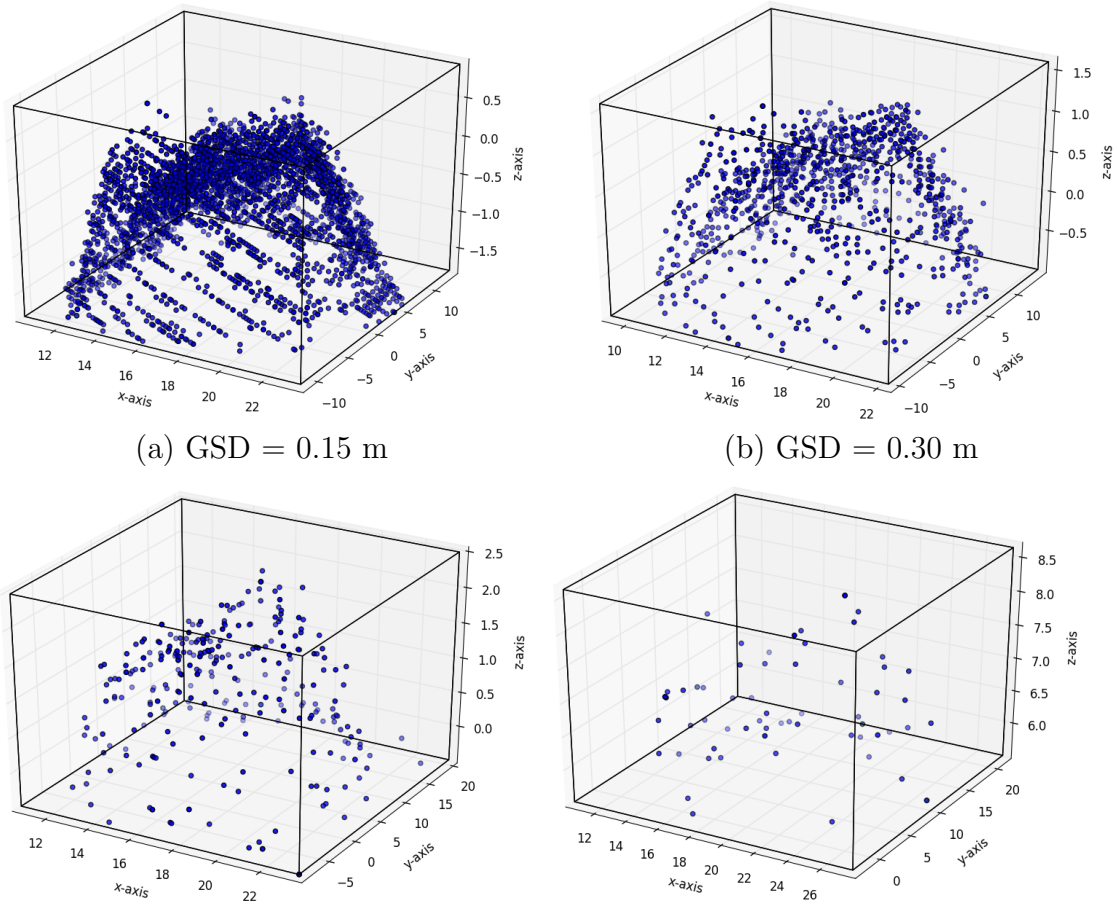
3.2.4.2 Volume Estimation

The bounding boxes for the rock piles are shown at each level of image resolution in Fig 3.20 and the results are summarized in Table 3.7. Recall that the purpose of the bounding box measurement was to capture the effect of image resolution on important dimensions of the asymmetrical rock pile. The volume of the box is not meant as an accurate measure of the volume of the rock pile, rather as a rough, zeroth-order estimate.

Tab. 3.7: Rock Pile Bounding Box Estimates

Level	GSD (m)	x (m)	y (m)	z (m)	Volume
LIDAR	N/A	17.1	22.0	3.0	1129
1	0.15	13.2	24.5	2.6	841
2	0.30	15.9	25.3	2.6	1046
3	0.60	13.0	30.0	3.0	1170
4	1.2	17.2	25.3	3.1	1349

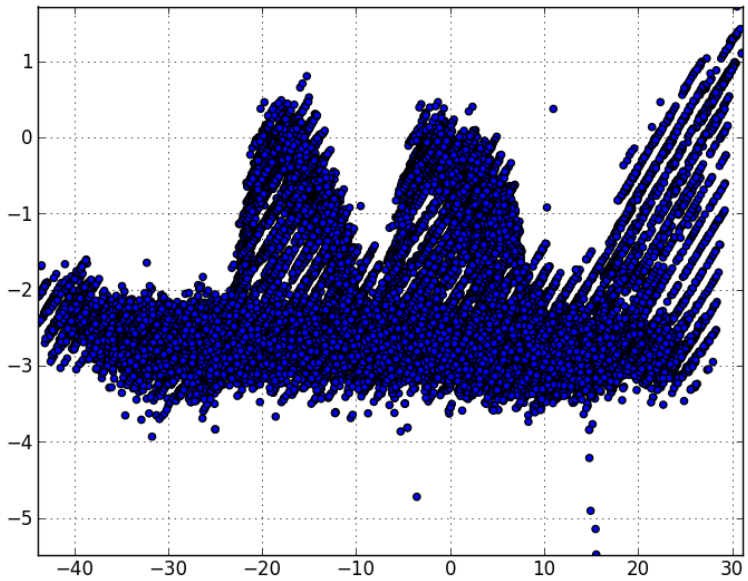
Note that the largest error with respect to the LIDAR estimate is observed at Level 1, or the smallest GSD. This occurrence was due to the uncertainty in the boundary between the ground plane and the pile of rocks, which is a large source of error. Even at Level 1, this boundary is prohibitively noisy, as seen in Fig. 3.21.



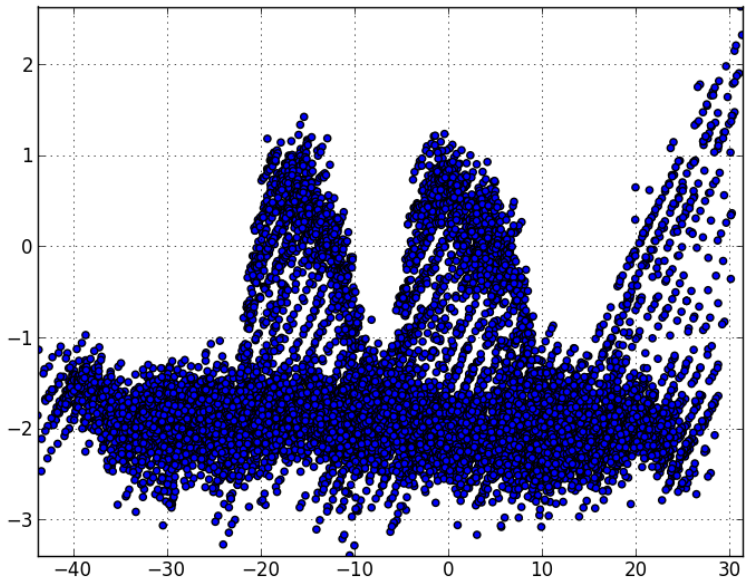
(a) GSD = 0.15 m (b) GSD = 0.30 m

(c) GSD = 0.60 m (d) GSD = 1.20 m

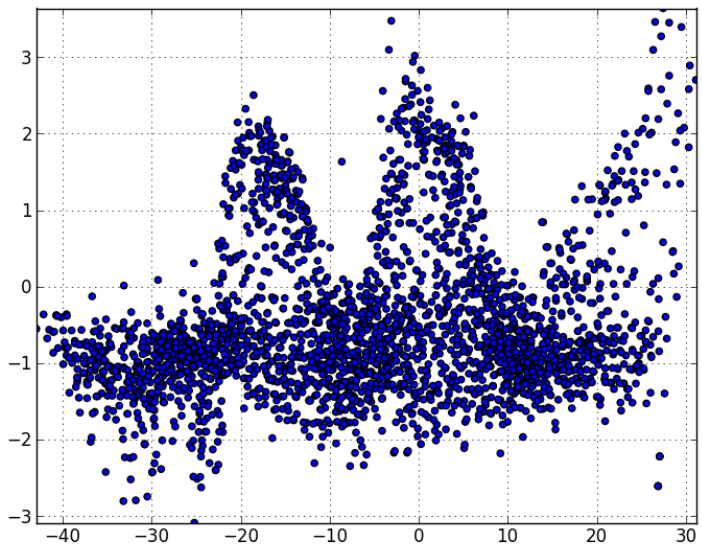
Fig. 3.20: The bounding boxes determined to calculate a volume estimate for the rock pile are shown for each GSD.



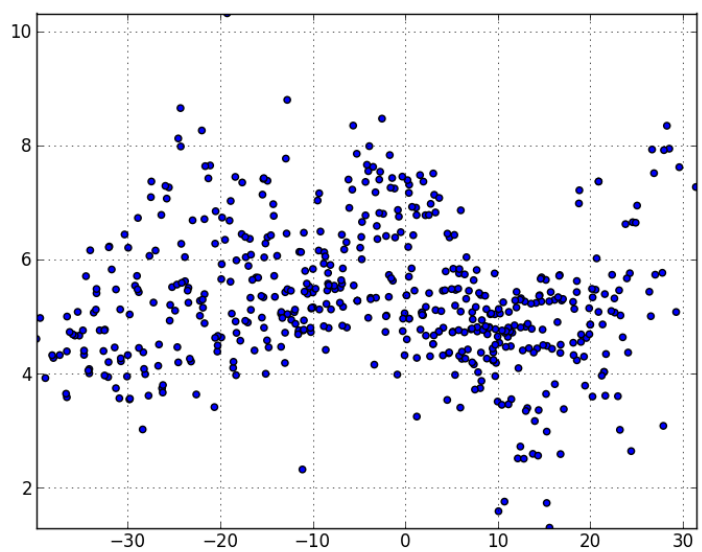
(a) GSD = 15 cm



(b) GSD = 30 cm



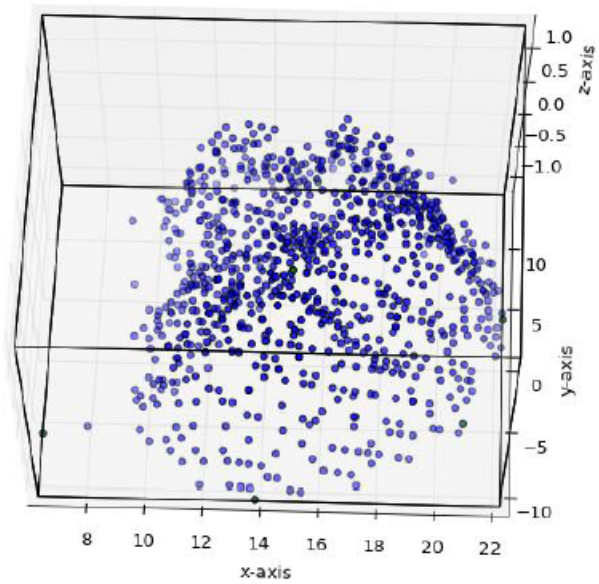
(c) GSD = 60 cm



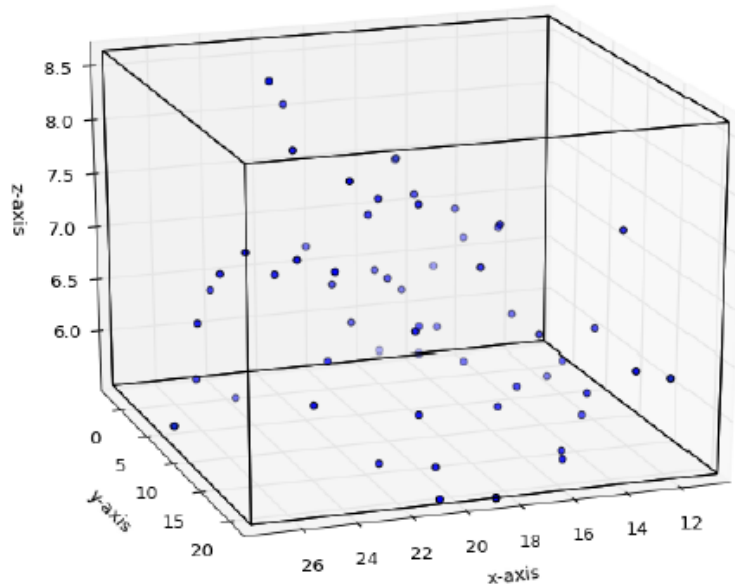
(d) GSD = 1.2 m

Fig. 3.21: The approximate two meters of uncertainty in the ground at the full resolution (a) makes it difficult to isolate the rock pile from the ground for the volume estimate. The uncertainty worsens as GSD gets larger.

Because there is such a large uncertainty in the ground plane for the full-resolution case, the edges of the rock pile that taper off are lost in the thresholding process. Therefore, the volume estimate for the full-resolution case is lower than the LIDAR volume estimate which has less uncertainty in the ground plane (refer to Fig. 2.9 for the LIDAR points). This uncertainty in the ground plane is a combination of triangulation error, alignment error, and the fact that the ground surrounding the rock pile is not completely flat. The slightly curved ground can be seen in the LIDAR point cloud in Fig. 3.23 (a). Lowering the ground threshold for the image-based point clouds causes several outliers to be included in the bounding volume, drastically affecting the volume calculation. For this reason, the rock points were thresholded near the top of the rock-ground interfaces illustrated in Fig. 3.21. Yet, outliers still remained as shown in Fig. 3.22. These outliers contribute to the large changes in the volume estimate between the image-based point clouds. Additionally, noisy points from poor image correspondences around the pile-to-air interface contribute further error that is also magnified by the volume calculation.



(a) GSD = 0.30m

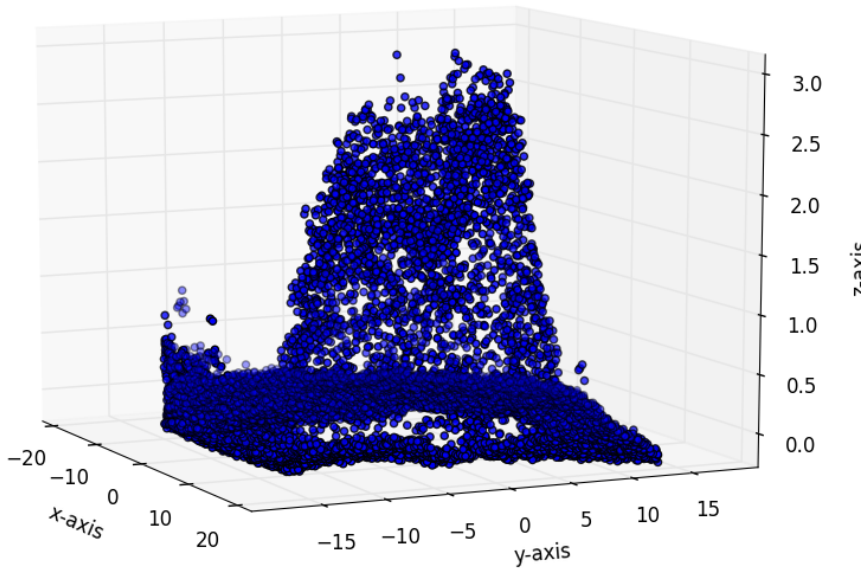


(b) GSD = 1.2m

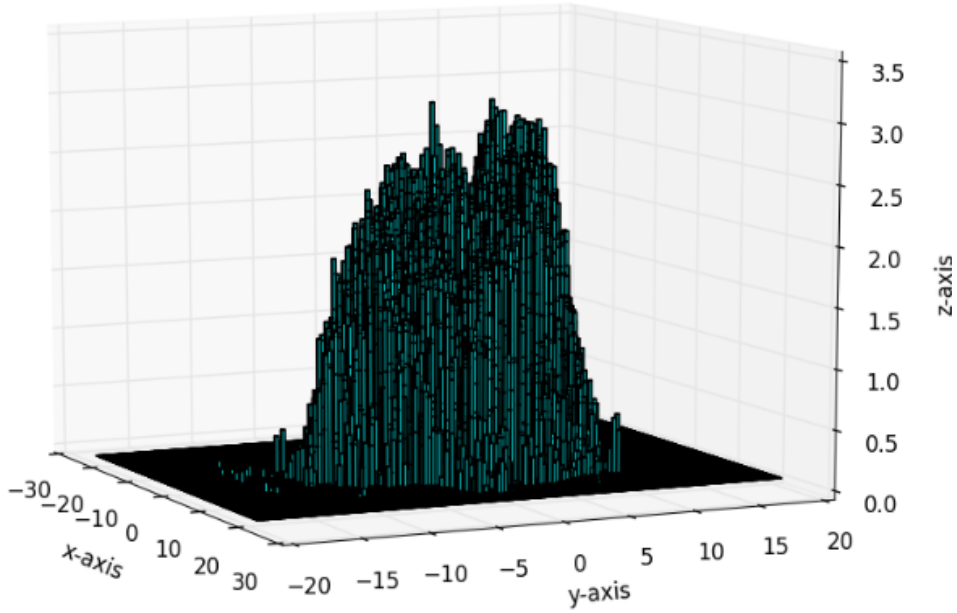
Fig. 3.22: Outliers in both (a) and (b) noticeably effect the limits of the bounding box, adding error to the volume estimate.

The pseudo-integration approach to the volume estimate was also very sensitive

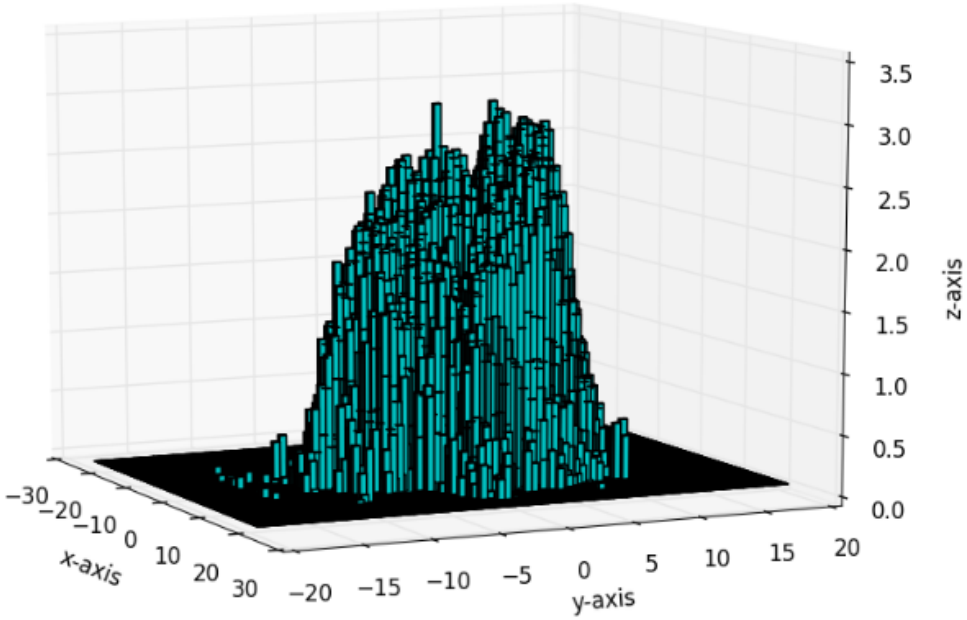
to the uncertainty in the ground plane. In addition, this approach required determination of the proper size of the grid elements, or bin size. The results of various bin sizes for the LIDAR estimate are illustrated in Fig. 3.23. As shown in Fig. 3.23, the volume of the pile largely depends on the bin size. If the bin size is too small, some bins will have a height of zero and the volume will be inaccurate. However, if the bin size is too large, the volume of the entire pile will be overestimated. The 0.5 m bin size was chosen to best represent the volume of this data because it resulted in no zero volume elements within the observable boundary of the rock pile. This volume estimate was compared to estimates from the image-based point clouds shown in Fig. 3.24, but note that the bin size must change for each data set to accomodate the variable point spacing.



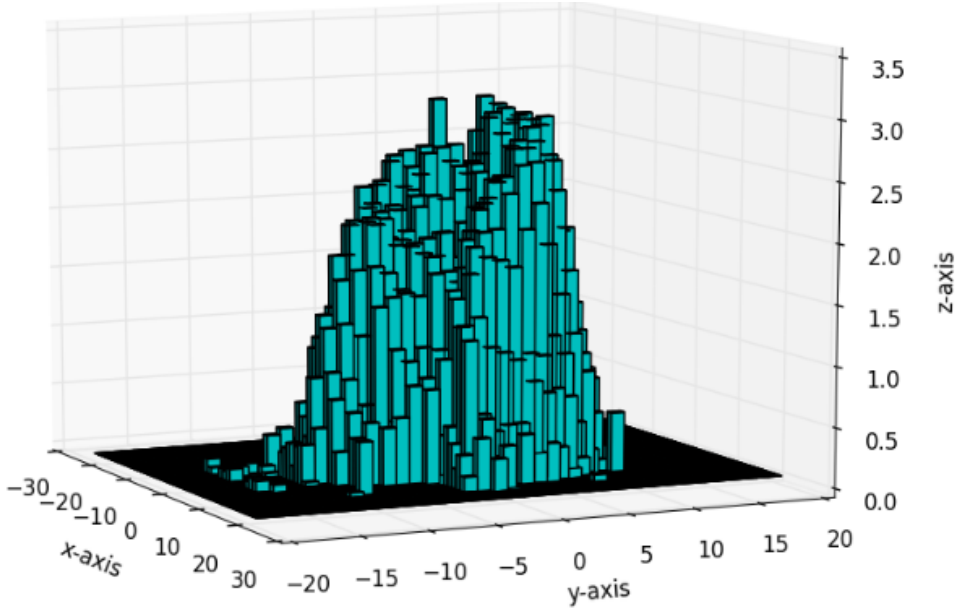
(a) LIDAR Point Cloud



(b) Bin = 0.25m
Volume = 215 m^3

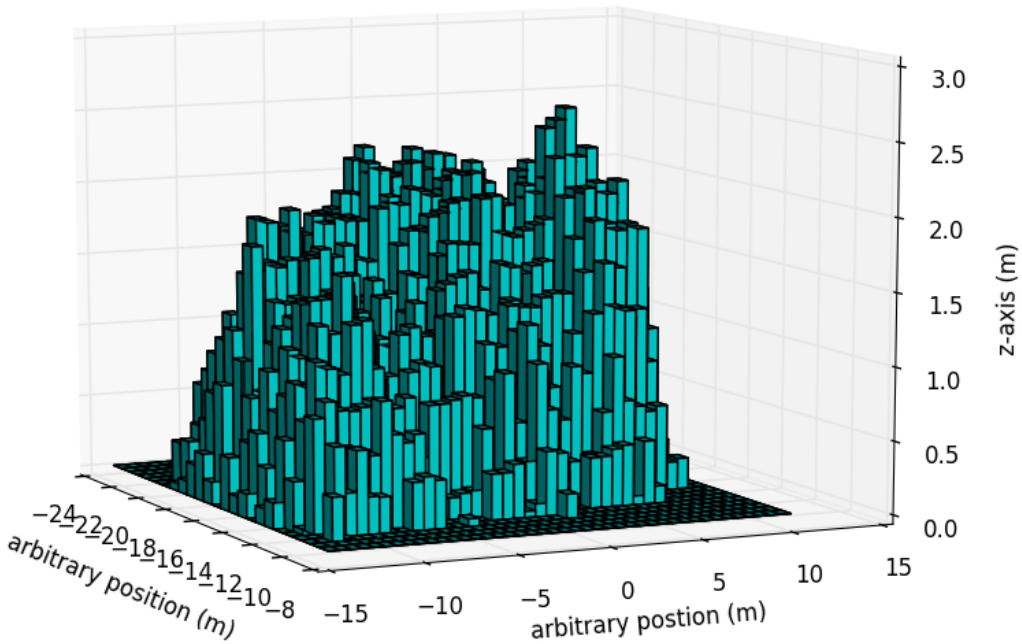


(c) Bin = 0.5m
Volume = 365 m^3

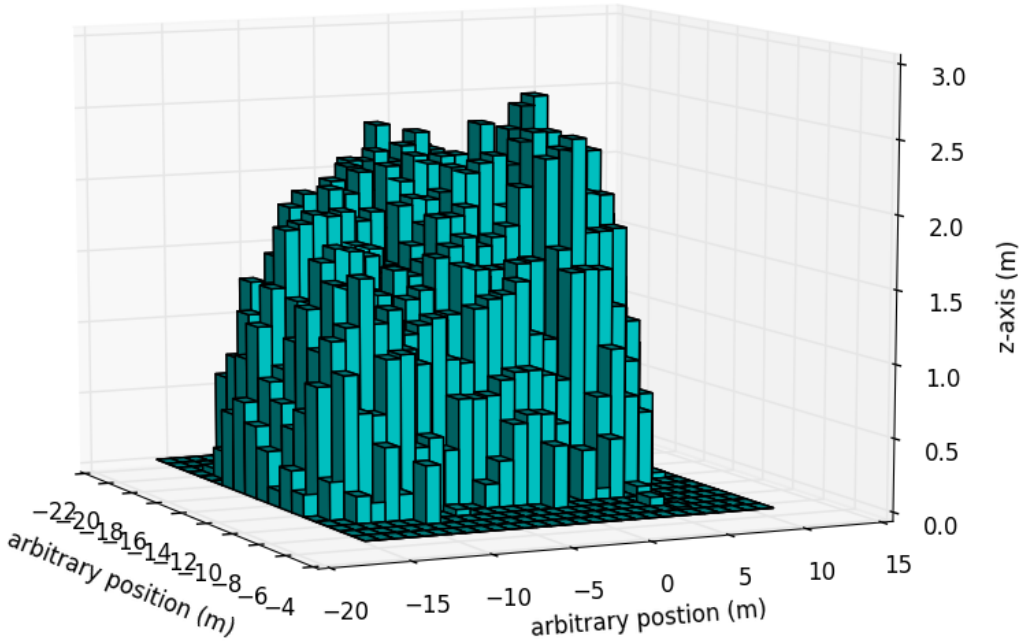


(d) Bin = 1m
Volume = 433 m^3

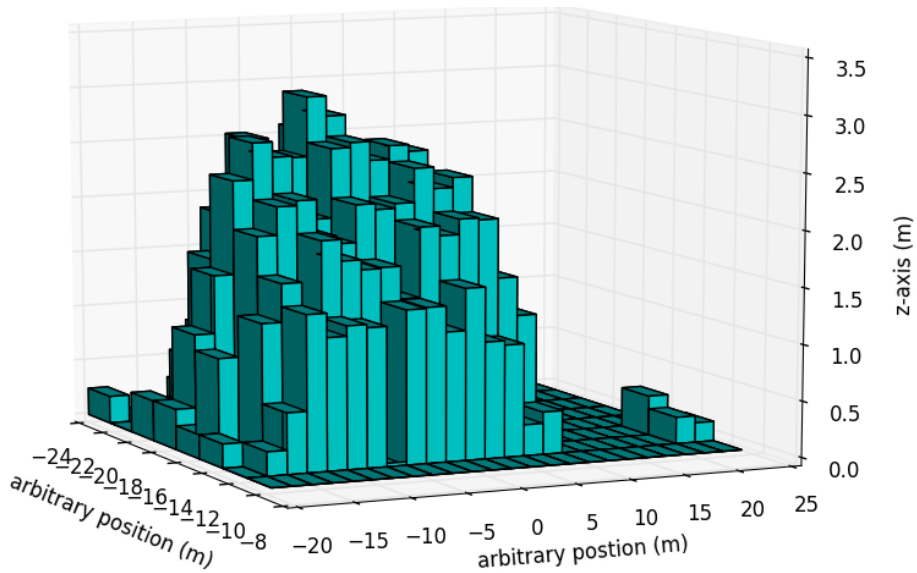
Fig. 3.23: The volumes estimated from the LIDAR data using the second approach are shown for various bin sizes. The volume changes significantly based on the bin size.



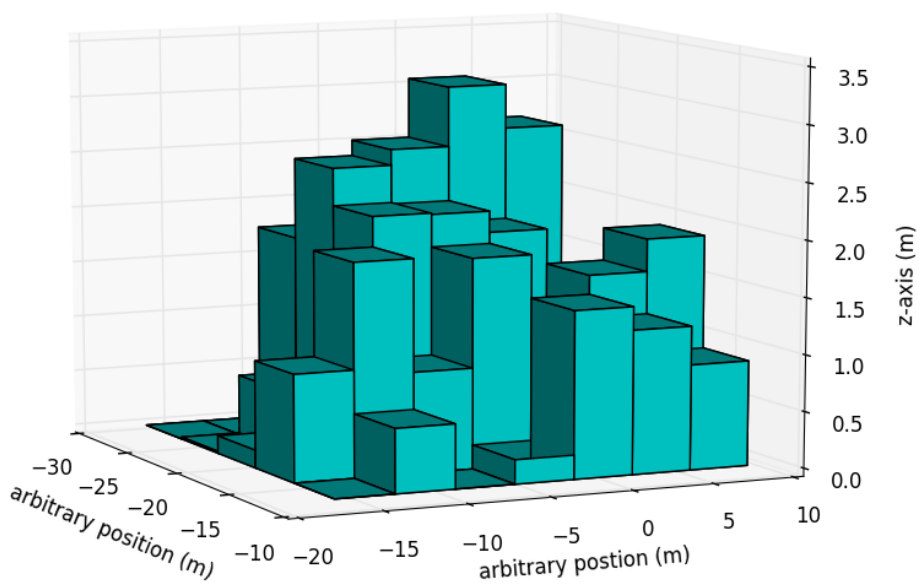
(a) Bin Size = 0.6 m
Volume = 290 m^3



(b) Bin Size = 0.9 m
Volume = 351 m^3



(c) Bin Size = 1.8 m
Volume = 480 m^3



(d) Bin Size = 3.6 m
Volume = 505 m^3

Fig. 3.24: The volumes estimated using the second approach are shown for the image-based point clouds. Note that the bin size increases accordingly due to the decrease in point density.

The second volume approach also underestimates the volume of the Level 1 point cloud with respect to the LIDAR estimate. As was also observed with the first volume approach, due to the drastic increase in point spacing of the remaining levels, the volume estimate increases as the GSD becomes larger. Again, this is because triangulation error produces points that fall outside the true extent of the pile. As both volume estimates are dependent on detecting the outer limits of the pile, calculations are very sensitive to errors generated from the 3D workflow. While both volume approaches yield significant error, it is important to note that the pile is only approximately 3 m high, a dimension which is extracted with a maximum 13.3% error according to Table 3.7. This small height causes the portion of the pile that is lost in the 2+ meters of ground plane error to be a significant contribution to the error in volume.

Chapter 4

Conclusions and Future Work

At sustainable orbits (roughly above 400 km which translates to a GSD of 1.4 m), the combination of a large GSD and low SNR will strongly affect the completeness and geometric accuracy of the cubesat-based point clouds. However, this work shows that some meaningful information can still be extracted with the chosen 3D workflow despite these limitations. An order of magnitude simulated change in GSD, from 0.25 meters to 2 meters, results in an error of approximately 2 meters for a building height estimate with a LIDAR truth of 60 meters. Over the same range of GSD, however, surface normal approximations suffer significantly, deviating a maximum of 35 degrees at the 1m GSD mark for a surface with few features. Signal-to-noise ratios evaluated between the best-case V4 estimate of 15 and a higher value of 73 demonstrated up to 3 meters of error in the height estimate of the same building. However, surface normals from imagery in this SNR range were much more consistent than those estimated at the largest GSDs. This result indicates that the low SNR of V4 is the limiting factor for the height estimate, but GSD is the limiting factor for the surface normal estimates.

From the quarry scene, it was demonstrated that estimating the local depth of a quarry over the GSD range of 15 cm to 1.2 meters was within 1 meter of error from the LIDAR data. This error is within the error induced by the measurement process

itself, therefore no discernible error is attributed to the loss of spatial resolution over this range. However, volume estimates of a rock pile in the quarry incur significant error and continue to increase as GSD becomes larger. The inherent nature of both volume estimate approaches depends on detecting the extrema of the boundary of the rock pile, therefore they are much more sensitive to error than the other metrics used.

As the results from both scenes demonstrate, the independent impacts of spatial resolution and SNR on the quality of resulting point clouds are significant depending on the application. Though the completeness and geometry of the reconstructions suffer noticeably beyond a GSD of 0.5 m, height and surface normal estimates are still informative with the aformentioned errors. Thus, the ultimate utility of the point cloud depends on the application and the error tolerance of the user.

It is important to note that the quality metrics reported here are dependent on the 3D workflow that was chosen. Image correspondence error due to the feature extraction and matching process of choice, subsequent camera pose error, and triangulation error due to PMVS may all be mitigated by using other stages in the workflow. This is particularly significant for the volume estimates, which have high error. Alternate stages may minimize the ultimate triangulation error to which the extrema-dependent volume estimates are sensitive. Furthermore, fine-tuning the parameters in the workflow used in this work may also achieve that end.

It remains to determine the impacts of smear, the optical transfer function, jitter, and other MTF effects on point cloud quality, not to mention the simultaneous impact of large GSD and low SNR. While smear was used to limit the integration time for SNR predictions, it was not simulated in the images, nor were the other previously mentioned effects. Due to the low SNR to which the high-Q design of V4 is limited, decreasing Q towards 1 may provide a more suitable cubesat imager for multi-view, image-based 3D reconstruction, if this can be done without too much loss of spatial resolution due to increased GSD. Other options include the investigation of lower noise detectors that can fit in the cubesat payload or using

a slightly larger satellite that would be able to house higher-quality detectors and optical configurations.

As the same sets of images were degraded in this analysis and used to represent imagery at different altitudes, DIRSIG simulation of images for input to 3D reconstruction workflows would be more illuminating. DIRSIG would be able to capture differences in parallax dependent on the imaging altitude, a factor which is not included in this work. Difference in parallax dependent on altitude would also contribute to point cloud quality. The difficulty with using DIRSIG images for point cloud generation lies in producing a scene with enough texture and variation in content to allow accurate feature matches.

Opportunities for additional work that would facilitate studies such as the one presented in this work are numerous due to the error-ridden workflow. While there are errors present in the PMVS reconstruction, georegistration transform, and measurement process, they are all linked to error in the initial detection and matching of features performed in this work with SiftGPU. Therefore, the comparison of the robustness of other existing feature detection algorithms to that of SiftGPU would be beneficial. A more robust feature extractor that produces more accurate image correspondences would reduce much error. Furthermore, sophisticated quality metrics that directly rate to the utility of the point clouds would also be beneficial. The metrics presented in this work were simply intended to provide examples of a qualitative assessment of information that can be extracted from point clouds.

APPENDIX

Appendix A

Point Cloud Generation Failure at 2.4 m GSD - Natural Scene

The 3D workflow outlined in Section 1.2.5.3 failed at the 2.4 m GSD case for the natural scene due to the use of CMVS. At this large GSD, a majority of the 31 input images were rejected. The number of images retained by CMVS as a function of GSD, or pixel size on the ground, is shown in Fig. A.1. Note that the number of images may change between running the same set of images through the workflow a second time, due to the random nature of the algorithm.

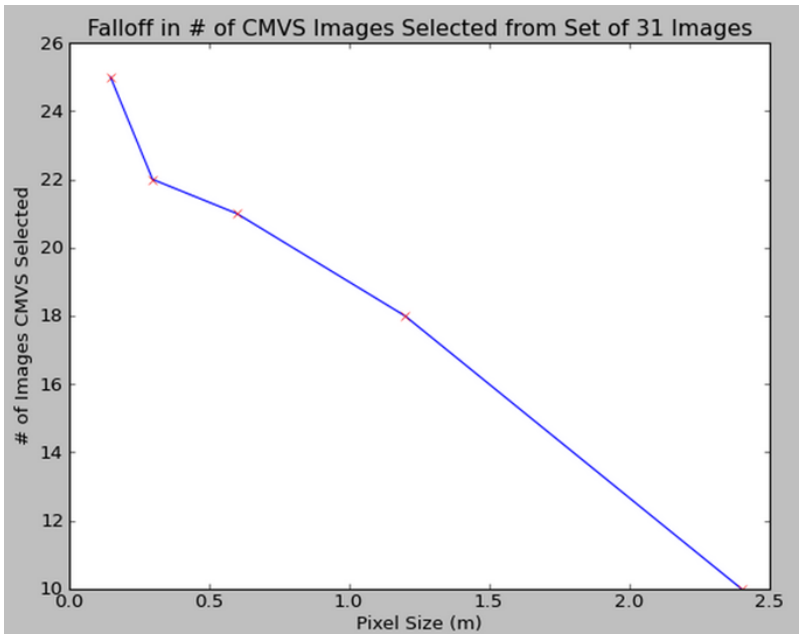


Fig. A.1: The number of images selected for reconstruction by CMVS at each GSD level decreases as GSD increases, resulting in only 10 of the 31 input images being used in the 2.4 m GSD case.

The resulting point cloud for the 2.4 m GSD case is shown in Fig. A.2. This failure example illustrates that CMVS was designed for large image collections with redudant images. Redundant images were not included in this input data set. Note that a successful reconstruction at this GSD can be achieved by removing CMVS from the 3D workflow with a different implementation[38].

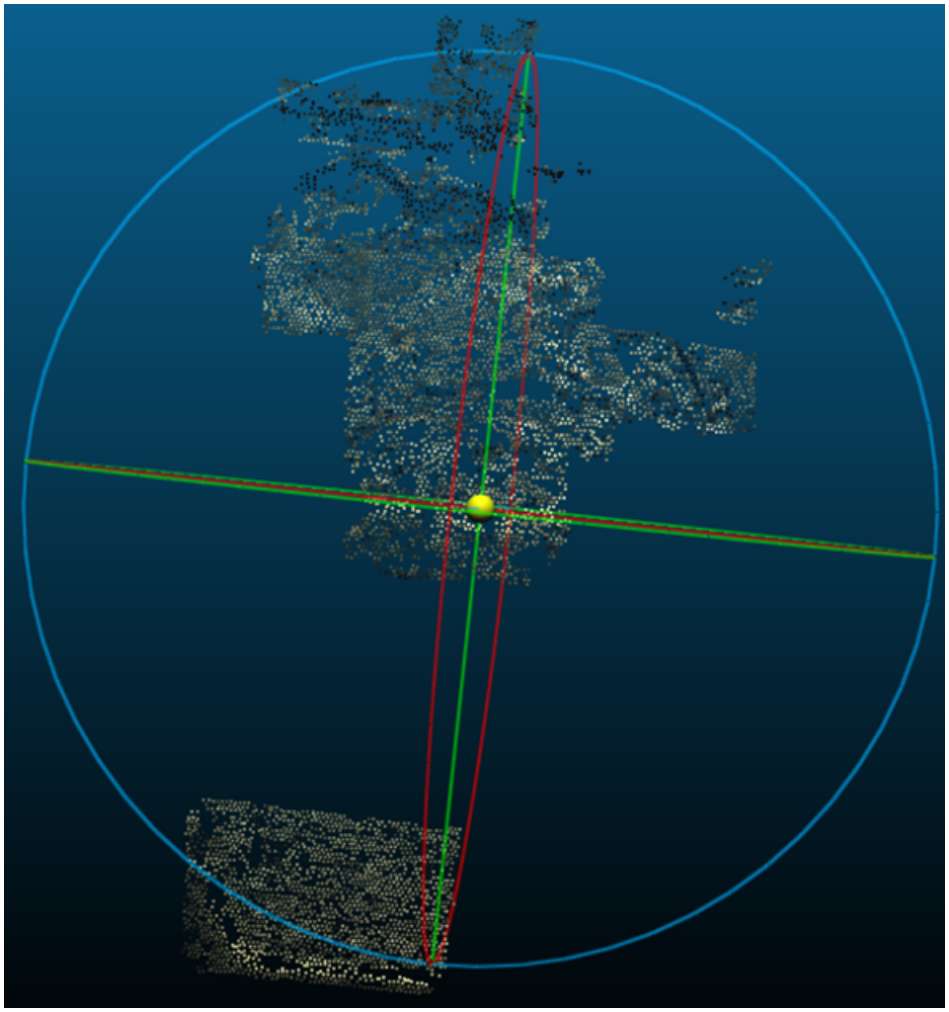


Fig. A.2: Failure at 2.4 m GSD

Bibliography

[1] D. Nilosek. (2014, May) Geo-accurate dense point cloud generation. [Online]. Available: dirsapps.cis.rit.edu/3d-workflow

[2] D. J. Walvoord, A. J. Rossi, B. D. Paul, B. B. Brower, and M. F. Pellechia, "Geoaccurate three-dimensional reconstruction via image-based geometry," in *Proc. SPIE Defense, Security, and Sensing*, vol. 8747, 2013.

[3] S. Brown, *DIRSIG*. [Online]. Available: http://dirsig.org/documentation

[4] J. Puig-Suari, C. Turner, and W. Ahlgren, "Development of the standard cubesat deployer and a cubesat class picosatellite," in *Aerospace Conference, 2001, IEEE Proceedings.*, vol. 1. IEEE, 2001, pp. 1–347.

[5] C. L. Gaux, "STARE cubesat communications testing, simulation and analysis," Master's thesis, Naval Postgraduate School, 2012.

[6] "Sporesat." [Online]. Available: http://www.nasa.gov/sporesat/#.U6txg_ldUuA

[7] "Kicksat." [Online]. Available: kicksat.net/

[8] M. Wall, "Record-breaking 33 'cubesats' to launch from space station this month," February 2014. [Online]. Available: http://www.space.com/24546-cubesat-record-space-station-launch.html

[9] "Flock 1." [Online]. Available: https://www.planet.com/flock1/

[10] "Skybox imaging." [Online]. Available: www.skyboximaging.com/technology

[11] L. M. Simms, W. Vries, V. Riot, S. S. Olivier, A. Pertica, B. J. Bauman, D. Phillion, and S. Nikolaev, "Space-based telescopes for actionable refinement of ephemeris pathfinder mission," *Optical Engineering*, vol. 51, no. 1, 2012.

[12] V. Riot, L. Simms, D. Carter, T. Decker, J. Newman, L. Magallanes, J. Horning, D. Rigmaiden, M. Hubbell, and D. Williamson, "Government-owned cubesat next generation bus reference architecture," 2014.

[13] J. Chu, "Inflatable antennae could give cubesats greater reach," September 2013. [Online]. Available: http://newsoffice.mit.edu/2013/inflatable-antennae-could-give-cubesats-greater-reach-0906

[14] A. Pertica, private communication, 2014.

[15] R. D. Fiete, "Image quality and λfn/p for remote sensing systems," *Optical Engineering*, vol. 38, no. 7, pp. 1229–1240, 1999.

[16] J. M. Irvine, "National imagery interpretability rating scales (NIIRS): overview and methodology," in *Optical Science, Engineering and Instrumentation'97*, vol. 3128. International Society for Optics and Photonics, 1997, pp. 93–103.

[17] J. C. Leachtenauer, W. Malila, J. M. Irvine, L. Colburn, and N. Salvaggio, "General image-quality equation: GIQE," *Appl. Opt.*, vol. 36, no. 32, Nov 1997.

[18] R. D. Fiete and T. Tantalo, "Comparison of SNR image quality metrics for remote sensing systems," *Optical Engineering*, vol. 40, no. 4, pp. 574–585, 2001.

[19] A. Berk, G. P. Anderson, P. K. Acharya, L. S. Bernstein, L. Muratov, J. Lee, M. Fox, S. M. Adler-Golden, J. H. Chetwynd Jr, M. L. Hoke *et al.*, "Modtran5:

2006 update," in *Defense and Security Symposium.* International Society for Optics and Photonics, 2006, pp. 62 331F–62 331F.

[20] J. R. Schott, *Remote sensing.* Oxford University Press, 2007.

[21] D. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proc. of the Seventh IEEE Int. Conf. on*, vol. 2, 1999.

[22] M. Flood, "Laser altimetry: From science to commerical lidar mapping," *Photogrammetric Engineering and Remote Sensing*, vol. 67, no. 11, 2001.

[23] E. P. Baltsavias, "Airborne laser scanning: existing systems and firms and other resources," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 54, no. 2, pp. 164–198, 1999.

[24] M. Lemmens, "Airborne lidar sensors," *GIM international*, vol. 21, no. 2, pp. 24–27, 2007.

[25] "Nasa: Icesat." [Online]. Available: http://icesat.gsfc.nasa.gov/icesat/index.php

[26] K. Harris, M. Ryle, N. Martini, P. Patel, A. Baba, A. Gonzalez, S. Gillespie, and T. Grande, "Application for rso automated proximity analysis and imaging (arapaima): Development of a nanosat-based space situational awareness mission," 2013.

[27] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362–1376, 2010.

[28] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 807–814.

[29] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, "Towards internet-scale multi-view stereo," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1434–1441.

[30] C. Wu, *SiftGPU: A GPU Implementation of Scale Invariant Feature Transform*. [Online]. Available: http://cs.unc.edu/~ccwu/siftgpu/

[31] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3d," *ACM transactions on graphics (TOG)*, vol. 25, no. 3, pp. 835–846, 2006.

[32] D. R. Nilosek, "Analysis and exploitation of automatically generated scene structure from aerial imagery," Ph.D. dissertation, Rochester Institute of Technology, 2014.

[33] W. Kabsch, "A discussion of the solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, vol. 34, no. 5, pp. 827–828, 1978.

[34] E. Ontiveros, C. Salvaggio, D. R. Nilosek, N. Raqueño, and J. Faulring, "Evaluation of image collection requirements for 3d reconstruction using phototourism techniques on sparse overhead data," in *SPIE Defense, Security, and Sensing*, vol. 8390, 2012, pp. 83 900K–83 900K–9.

[35] A. Giannandrea, N. Raqueno, D. W. Messinger, J. Faulring, J. P. Kerekes, J. van Aardt, K. Canham, S. Hagstrom, E. Ontiveros, A. Gerace *et al.*, "The share 2012 data campaign," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2013, pp. 87 430F–87 430F.

[36] G. Bradski, *Dr. Dobb's Journal of Software Tools*.

[37] *CloudCompare: 3D point cloud and mesh processing software*, 2013. [Online]. Available: http://www.danielgm.net/cc/

[38] A. Rossi, H. Rhody, C. Salvaggio, and D. Walvoord, "Abstracted workflow framework with a structure from motion application," in *Image Processing Workshop (WNYIPW), 2012 Western New York*, 2012, pp. 9–12.