**Simulated Imagery Rendering Workflow for UAS-Based Photogrammetric 3D Reconstruction Accuracy Assessments**

**Abstract**

Structure from Motion (SfM) and MultiView Stereo (MVS) algorithms are increasingly being applied to imagery from Unmanned Aircraft System (UAS) to generate point cloud data for various surveying and mapping applications. To date, the options for assessing the spatial accuracy of the point clouds have primarily been limited to empirical accuracy assessments, which involve comparisons against reference data sets, which are both independent and of higher accuracy than the data they are being used to test. The acquisition of these reference data sets can be expensive, time consuming, and logistically challenging. Furthermore, these experiments are also almost always unable to be perfectly replicated, due to the numerous uncontrollable independent variables, such as solar radiation, sun angle, cloud cover, wind, objects in the scene moving, exterior orientation of cameras, and camera noise, to name a few. The combination of these factors leads to a situation in which robust, repeatable experiments are cost prohibitive, and the experiment results are frequently site-specific. Here, we present a workflow to render computer generated imagery using a virtual environment which can mimic all the independent variables that would be experienced in a real-world data acquisition scenario. The resultant modular workflow utilizes Blender, an open source computer graphics software, for the generation of photogrammetrically accurate imagery suitable for SfM processing, with explicit control on camera interior orientation, exterior orientation, texture of objects in the scene, placement of objects in the scene, and Ground Control Point (GCP) accuracy. The challenges and steps required to validate the photogrammetric accuracy of computer generated imagery are discussed, and an example experiment assessing accuracy of an SfM derived pointcloud from imagery rendered using a computer graphics workflow is presented.

# Introduction

Three dimensional geospatial point cloud data high spatial accuracy and point density has traditionally been acquired using either terrestrial or airborne lidar, but recent advances in Structure from Motion(SfM) and MultiView Stereo(MVS) algorithms have enabled the generation of image based pointcloud products comparable in density and accuracy to lidar data. SfM and MVS algorithms began development 30+ years ago, but have only relatively recently begun to be utilized for commercial surveying applications, leveraging advances in camera hardware, Unmanned Aerial Systems (UAS), computer processing power, and commercial SfM-MVS software. Research into SfM and MVS in the geomatics community is currently focused on both the accuracy and potential applications of the commercial SfM and MVS software packages such as Agisoft Photoscan Pro and Pix4d (Eltner et al., 2016). The accuracy of SfM-MVS can vary greatly depending on a variety of factors (Dandois, Olano, & Ellis, 2015; Smith & Vericat, 2015) which vary across different experiments (Clapuyt, Vanacker, & Van Oost, 2015), and the geomatics community is currently focused on determining potential use cases where SfM-MVS derived point clouds can begin to replace lidar as an alternative surveying tool, without sacrificing accuracy (Micheletti, Chandler, & Lane, 2015; Naumann, Geist, Bill, Niemeyer, & Grenzdörffer, 2013). The most common methodology for assessing the use cases and accuracy of SfM-MVS algorithms is to collect data in the field using a UAS and compare the data to lidar data. It is difficult to constrain many of the independent variables using this methodology, and the data acquisition can be expensive and time consuming. We propose an open source computer graphics based workflow to simulate various scenes and maintain full control over the ground-truth and the camera parameters. This workflow will allow researchers to perform more robust experiments to assess the feasibility and accuracy of SfM-MVS in various applications.

The 3D reconstruction methods used in most commercial software consist of an SfM algorithm first to solve for camera exterior and interior orientations, followed by an MVS algorithm to increase the density of the pointcloud. Unordered photographs are input into the software, and a keypoint detection algorithm such as Scale Invariant Feature Transform (SIFT) (Lowe, 2004) is used to detect keypoints and correspondences between images. A bundle adjustment is performed to minimize the errors in the correspondences. In addition to solving for camera interior and exterior orientation, the SfM algorithm also generates a sparse pointcloud. A bundle adjustment is performed to optimize the matches. Without any additional information, the coordinate system is arbitrary in translation and rotation with an inaccurate scale. To further constrain the problem and develop a georectified pointcloud, Ground Control Points (GCPs) and/or Initial Camera Positions are used to constrain the solution. The number of parameters that are solved for can also be reduced by inputting a camera calibration, however only inputting a camera calibration file with no camera positions or GCP coordinates will only help resolve the scale of the pointcloud coordinate system, not the absolute translation and rotation. These input data can be used to generate an absolute coordinate system either with a Helmert transformation after the pointcloud is generated (Clapuyt et al., 2015), or using an optimization within the bundle adjustment. The optimization of the bundle adjustment will generate more accurate results. The interior orientation, exterior orientation, and image correspondences for each image are used as the input to the MVS algorithm, which generates a denser pointcloud. Some of the common MVS algorithms generate more correspondences by utilizing a search along the epipolar line between corresponding images due to the known interior and exterior orientations. For this reason, the accuracy of the MVS algorithm is highly dependent on the accuracy of the parameters calculated with the SfM algorithm. A detailed explanation of the various MVS algorithms can be found in the book “Multi-view stereo: A Tutorial” (Furukawa & Hernández, 2015). Each of these algorithms also assumes that the scene is rigid with constant Lambertian surfaces, and deviations from these assumptions will affect the accuracy.

Numerous studies have been performed to quantify the accuracy of the SfM-MVS algorithms in a variety of environments (Naumann et al., 2013). A few datasets have been acquired in a lab environment, using a robotic arm to accurately move a camera and a light structure camera to collect reference data for a variety of objects of varying textures (Jensen, Dahl, Vogiatzis, Tola, & Aanaes, 2014; Seitz, Curless, Diebel, Scharstein, & Szeliski, 2006). While this dataset works well for testing the underlying algorithms--especially MVS--more application based experiments performed by the surveying community have shown to evaluate how error propagates in different environments. The most common and robust method has been to compare the SfM-MVS derived pointcloud to a groundtruth terrestrial lidar survey (Espositoa, Fallavollitaa, Wahbehb, Nardinocchic, & Balsia, 2014; Hugenholtz et al., 2013). This comparison can exhibit errors in vegetated areas due to vegetation moving by blowing in the wind, and increased uncertainty in the lidar pointcloud due to mixed pixel effects. Independent GPS control points have also been used, but result in fewer measurements for statistical analysis(Harwin, Lucieer, & Osborn, 2015). The use of independent control points may also exhibit a bias when the points used are easily photo identifiable targets (eg. checkerboards). These targets are traditionally very accurate keypoints and using them as independent control points will demonstrate a much better accuracy than a homogeneous surface. Therefore, the error reported from independent GCPs are not necessarily representative of the accuracy of the entire scene. The quality and uniqueness of detected keypoints in an image and on an object is called “texture.” The lack of texture of a scene has been shown to have one of the largest impact on the accuracy of SfM-MVS pointcloud (Harwin & Lucieer, 2012; Jensen et al., 2014; Micheletti et al., 2015; Naumann et al., 2013). The accuracy of SfM is also adversely affected by: poor image overlapping, lens distortion not fully modeled by the nonlinear lens distortion equation, poor GCP distribution, inaccurate GCP or camera positions, poor image resolution, blurry imagery, noisy imagery, varying sun shadows, moving objects in the scene, user error in manually selecting image coordinates of GCPs, low number of images, or a low number of GCPs (Smith & Vericat, 2015). With a computer graphics workflow, the user manually introduces all the uncertainty. The pointcloud produced through SfM processing is compared to the true surface, and any errors can be attributed to the user introduced uncertainty.

## Computer Graphics for Remote Sensing Analysis

The field of computer graphics was first developed in the 19XXs and the advancement of the field has been predominantly driven by the desire to create more realistic video game and movie special effects. The software that turns a simulated scene with various geometries, material properties, and lighting into an image or sequence of images is called a render engine. While there are many different render engines available using many different algorithms, they all follow a basic workflow, or computer graphics pipeline.

First, a 3D scene is generated using vertices, faces, and edges. For most photo-realistic rendering, meshes are generated using an array of either triangular surfaces or quadrilateral surfaces to create objects. Material properties are applied to each of the individual surfaces to determine the color of the object. Most software allows for the user to set diffuse, specular, and ambient light coefficients as well as their associated colors to specify how light will interact with the surface. The coefficient specifies how much diffuse, specular, and ambient light is reflected off of the surface of the object, while the color specifies the amount of red, green, and blue light that is reflected from the surface. The material color properties are only associated with each plane in the mesh, so for highly detailed coloring of objects a large amount of faces are required. The more efficient way of creating detailed colors on an object without increasing the complexity of the surface of the object is to add a "texture" to the object. A texture is an image which is overlaid on the mesh in a process called u-v mapping. This is where each vertex is assigned a coordinate in image space in units of texels, which are synonymous to pixels but renamed to distinguish the fact that it is a texture and not a rendered image. The render engine interpolates the texel coordinates across the surface when the scene is rendered. For interpolated subpixel coordinates, the color value is either interpolated linearly or the nearest pixel value is used. The computer graphics definition for a "texture" object, is not to be confused with the SfM and photogrammetry definition of texture, which is the level of detail and unique photo identifiable features in an image.

Once a scene is populated with objects and their associated material and texture properties, light sources and shading algorithms must be applied to the scene. The simplest method is to set an object material as "shadeless", which eliminates any interaction with light sources and will render each surface based on the material property and texture with the exact RGB values that were input. The more complex and photorealistic method is to place light sources in the scene. Each light source can be set to simulate different patterns and angles of light rays with various levels of intensity and range based intensity falloff. Most render engines also contain shadow algorithms which enable the calculation of occlusions from various light sources. Once a scene is created with light sources and shading parameters set, simulated cameras are placed to create the origin for renders of the scene. The camera translation, rotation, sensor size, focal length, and principal point are input and a pinhole camera model is used. The rendering algorithm generates a 2D image of the scene using the camera position and all of the material properties of the objects. The method, accuracy (especially lighting), and performance of generating this 2D depiction of the scene is where most render engines differ.

There are many different rendering methodologies, but the one chosen for this research is Blender Internal Render Engine which is a rasterization based engine. The algorithm determines which parts of the scene are visible to the camera, and perform basic light interactions to assign a color to the pixel samples. This algorithm is fast, but is unable to perform some of the more advanced rendering features such as global illumination and true motion blur. A more detailed description of common rendering engines and algorithms can be found .

The use of synthetic remote sensing datasets to test and validated remote sensing algorithms is not a new concept. A simulated imagery dataset using Terragen 3 was used validate an optimized flight plan methodology for UAV 3D reconstructions (Martin, Rojas, Franke, & Hedengren, 2015). Numerous studies have been performed using the Rochester Institute of Technology’s Digital Imaging and Remote Sensing Image Generation (DIRSIG) using for various active and passive sensors. DIRSIG has been used to generate an image dataset for SfM-MVS processing to test an algorithm which automate identification of voids in three-dimensional pointclouds (Salvaggio & Salvaggio, 2013) and assess SfM accuracy using long range imagery (Nilosek, Walvoord, & Salvaggio, 2014). While DIRSIG generates radiometrically- and geometrically-accurate imagery, it is currently not available to the public.

# Materials and Methods

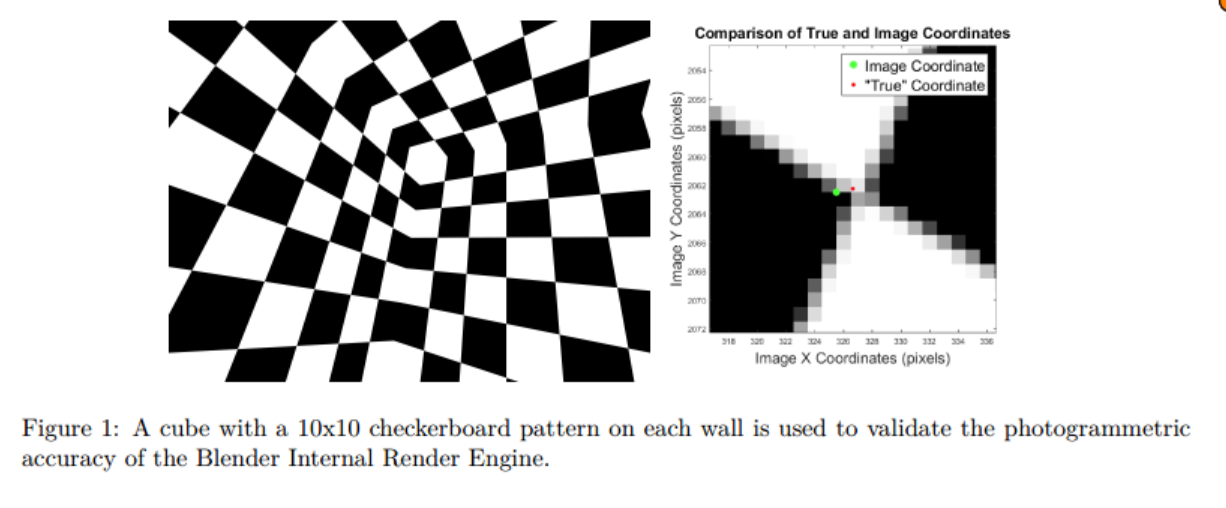
The use and validation of a computer graphics based methodology to render imagery for SfM analysis is presented in this paper. First, a series of tests are presented that should be performed to ensure that a render engine is generating photogrammetrically accurate imagery. The results of these tests for the Blender Internal Render Engine are presented and provide validation that the render engine is accurate. An example use case experiment is exploring the effect of Agisoft Photoscan Dense Reconstruction Quality on point cloud accuracy is presented utilizing the Blender Internal Render Engine. A few results from the example experiment are presented to demonstrate the potential of the methodology to perform sensitivity analyses. The results suggest that higher dense reconstruction quality settings result in a pointcloud which is more accurate and contains more points. Interestingly, results are presented which demonstrate that a lower dense reconstruction quality setting will sometimes generate points in a region where there is a data gap in a pointcloud generated with a higher reconstruction quality setting.

## Render Accuracy Validation

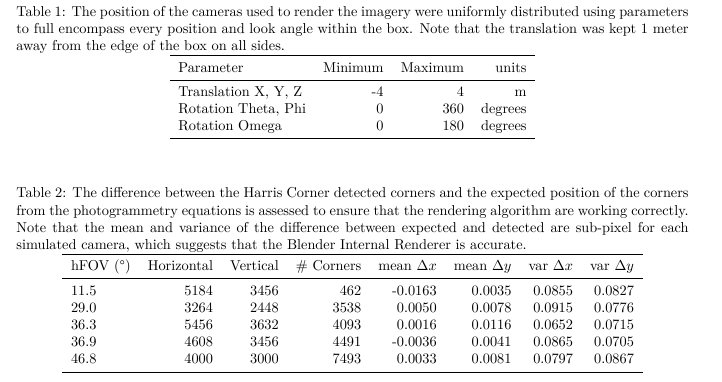
There are many different open source and commercial render engines available to generate imagery of simulated scenes, but before using a render engine to analyze surface reconstructions a series of validation experiments should be performed to ensure that the render engine is generating imagery as expected. Validation experiments are performed to ensure accurate rendering, and that any resultant error in a data product is due to SfM algorithm. In this paper, we present an assessment and validation using Blender Internal Render Engine, but this validation methodology could be applied to any render engine. (It should be noted that our focus in this study is on geometric accuracy, so procedures to validate the radiometric accuracy are beyond the current scope.) SfM keypoints are detected based on image texture gradients, and are not influenced by radiometric accuracy of the scene. For this experimentation methodology, it is more important for the object diffuse texture and colors to remain constant from various viewing angles. The authors recognize the render engine could also be validated by rigorously analyzing or developing the rendering source code, but believe these experiments provide a less time-consuming methodology which could also be applied to a closed source commercial render engine.

## Photogrammetric Projection Accuracy

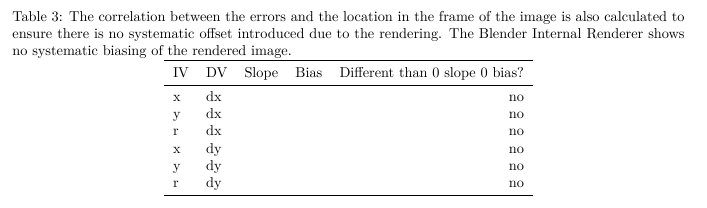
The first validation experiment is designed to ensure that the camera interior and exterior orientation are set accurately using a pinhole camera model. The pinhole camera model represents an ideal test case and is commonly the output from render engines. While Vertex Shaders can be programmed and implemented into a Computer Graphics workflow to accurately simulate lens distortion, the programming and implementation of this method is time consuming and can be confusing for someone not familiar with computer graphics. For this experiment, lens effects are simulated in postprocessing by distorting the pixels and reinterpolating onto a grid. Therefore, as long as the initial pinhole camera model assumption is photogrammetrically valid, the distortion will also be valid. This experiment is performed by creating a simple scene consisting of a 1000m3 cube with a 10x10 black and white checkerboard pattern on each wall, as depicted in Figure 1. The black and white corner of each checkerboard corner is at a known 3D world coordinates. A series of images are rendered using various camera rotations, translations, focal lengths, sensor sizes, and principal point coordinates. To ensure that the images are rendered correctly, the coordinates of the checkerboard corners are calculated from the rendered imagery using a corner feature detector and compared to the expected coordinates of the targets using photogrammetric equations. The difference between the image-derived coordinates and the photogrammetric equation derived coordinates should have a mean of 0 in both dimensions, and a subpixel variance on the order of the accuracy of the image corner feature detector. There should also be no correlation between the accuracy of the coordinate and the location of the coordinate in the image

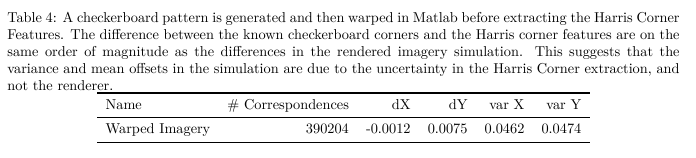


To validate the photogrammetric projection accuracy of the Blender Internal Render Engine using this experiment, a 1000 m3 cube was placed with the centroid at the origin. Five hundred images were rendered using five different interior orientations and random exterior orientations throughout the inside of the cube. These parameters were input using the Blender Python API, and the distribution of each input parameter are shown in Table 1. The accuracy of the imagery was observed qualitatively by plotting the photogrammetric equation calculated points on the imagery in MATLAB to ensure a rough accuracy. Once the rough accuracy is confirmed, a nearest neighbor is used to develop correspondences between the Harris corner coordinates and the photogrammetric equation derived coordinates. The mean and variance of the differences between the correspondences in each experiment are shown in Table 2. The correlation between the difference in x, y, and radius versus several parameters shows no statistically significant correlation. The correlation results are summarized in Table 3.



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Resolution | | |  | |  | |  | |  | |  | |  | |
| Focal Length  (mm) | Sensor Width  (mm) | Horizontal  (pixels) | Vertical  (pixels) | # Corners | | delta X  (pixels) | | delta Y  (pixels) | | var X  (pixels) | | var Y  (pixels) | | RMSE r  (pixels) | |
| 55 | 22.3 | 5184 | 3456 | 462 | | -0.0163 | | 0.0035 | | 0.0855 | | 0.0827 | | 0.4100 | |
| 4.1 | 4.54 | 3264 | 2448 | 3538 | | 0.0050 | | 0.0078 | | 0.0915 | | 0.0776 | | 0.4113 | |
| 16 | 23.5 | 5456 | 3632 | 4093 | | 0.0016 | | 0.0116 | | 0.0652 | | 0.0715 | | 0.3699 | |
| 4.11 | 6.17 | 4608 | 3456 | 4491 | | -0.0036 | | 0.0041 | | 0.0865 | | 0.0705 | | 0.3962 | |
| 2.9 | 6.17 | 4000 | 3000 | 7493 | | 0.0033 | | 0.0081 | | 0.0797 | | 0.0867 | | 0.4080 | |
| simulated checkerboard | |  |  |  | | -0.0012 | | 0.0075 | | 0.0462 | | 0.0474 | | 0.3059 | |

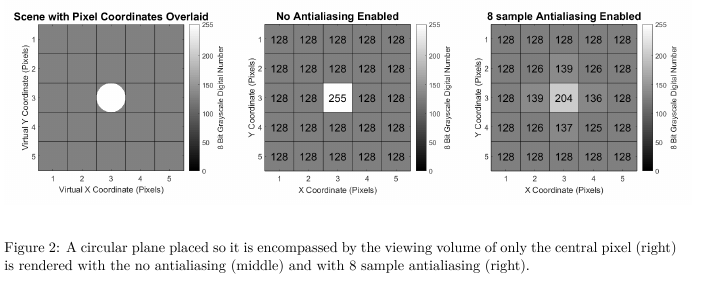




To ensure that the mean and variance are not an artifact of the rendering, an experiment was performed to determine the expected accuracy of the Harris Corner detector. 1000 Simulated checkerboard patterns were generated with random rotations, translations, and skew to create a synthetic image dataset. The known coordinates of the corners were compared to the coordinates calculated with the Harris Corner feature detector, and the results are shown in in Table 2. From these results, the hypothesis that the variance of the rendered image coordinate error is statistically different than the variance of the simulated image coordinate error is rejected. Therefore, all the variance can be statistically attributed to the Harris Feature Corner detection algorithm, rather than the render engine.

## Point Spread Function

The second validation experiment ensures that there is no unintended blurring applied to the rendered image. Specifically, this test determines that the point spread function of the rendered imagery is a unit impulse. This test is performed by simulating a white circular plane placed at a distance and size such that it exists in only one pixel. The rendered image should therefore only contain white in the one pixel and not be blurred into any other background pixels. This test is particularly important when antialiasing is performed, as the supersampling pattern and filter used to combine the samples can sometimes create a blurring effect. For example, the default antialiasing in blender uses a "distributed jitter" pattern and the Mitchel-Netravali filter, which uses supersampled values from neighboring pixels to calculate a pixel value. This effect can be seen in Figure 2, where the intensity of the white plane is influenced all eight of the neighboring pixels, even though the plane should only be visible in one pixel. While this photogrammetric inaccuracy is minimal, the error could propagate into the resultant SfM derived pointcloud especially when fine scale textures with high gradients are used.

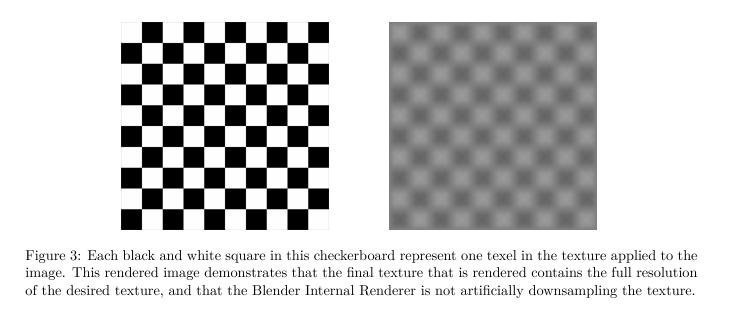


To validate the point spread function of the Blender Internal Render Engine, a sensor and scene are set up such that the geometry of the circular plane is only captured with one pixel in the render of a 5x5 pixel image. This experiment ensures that any other pixels that contain values different than the background digital number of 128 are an artifact of the rendering. Rendered imagery is shown with and without antialiasing in Figure 2. The antialiasing used is the default settings for the Blender Internal Render Engine (8 Samples, Mitchell-Netravali filter). The rendered image with no antialiasing contains no blurring of the image, while the antialiased image contains a slight amount of blurring. The antialiased imagery supersamples the scene and renders a smoother, more photorealistic imagery, and is deemed to be suitable for experimentation.

## Texture Resolution

The final validation experiment ensures that any textures applied to the objects in the scene are applied in a manner which maintains the resolution of the imagery without and compression or subsampling. This validation experiment is performed by applying a texture on a flat plane and rendering an image that contains a small number of the texture pixels. By qualitatively looking at the image, it should be clear that the desired number of pixels are in the frame, and no smoothing is being applied. When rendering textures in computer graphics there is an option to perform interpolation of the texture, which yields a smoother texture. An example of a texture with and without interpolation is shown in Figure 3.

To validate the texture resolution of the Blender Internal Render Engine a black and white checkerboard pattern where each checkerboard square is 1x1 texel is applied to a flat plane such that each texel represents a 10cm x 10cm square. An image is rendered using a focal length and sensor size such that each texel is captured by 100 x 100 pixels, as shown in Figure 3 with and without interpolation. The rendered images in Figure 3 are qualitatively observed and it was determined that the rendering has not subsampled or compressed the texture image.



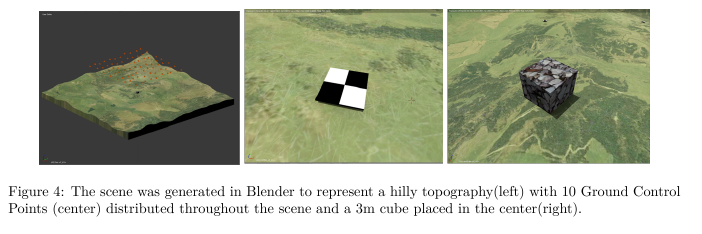
## Use Case Demonstration

An example experiment was generated as a proof of concept to demonstrate a potential workflow for testing the effect of various independent variables on SfM accuracy. This experiment specifically is generated to observe how the dense reconstruction quality setting in Agisoft Photoscan Pro effects the dense pointcloud accuracy. The dense reconstruction quality setting in Photoscan is used to downsample the image prior to MVS processing. The percentage of image downsampling is shown in Table 6. The results from this experiment suggest that a higher quality dense reconstruction setting will generate more accurate results, which agrees with the stated expectation in the Agisoft Photoscan Pro manual. The computer used to render and process the data for this experiment is a Windows 7 Desktop PC with an Intel Xeon CPU E5-1603 @ 2.80GHz, GeForce GTX 980 Graphics card (4Gb), and 32Gb of RAM.

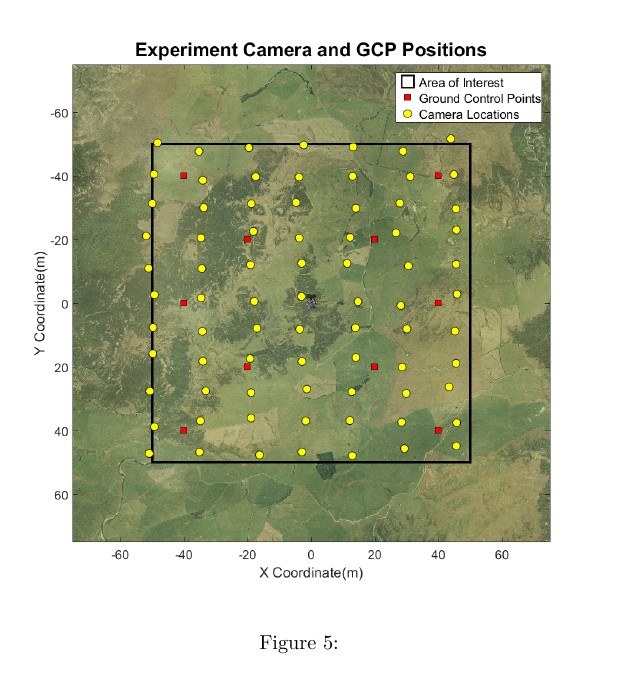
**Use Case Experiment Design**

A 200m x 200 m square mesh was manually generated to simulate a topography with rolling hills using a 1 meter grid. A large 27 m3 meter cube was placed in the center of the scene to test surface reconstruction accuracy on regions with sharp corners and edges. Ten 1m x 1m x 0.05m square, checkerboard pattern Ground Control Points (GCPs) are distributed evenly throughout the scene 0.25 meters above the ground surface. The material of all the objects in the scene are modeled as perfect Lambertian surfaces. The topographic surface was textured using a combination of two textures. The first texture is a 7200x7200 pixel aerial image for an effective texel footprint of 2.78 cm square. The second texture is a 3456x3456 pixel image of grass was tiled ten times in both the x and the y dimension for an effective repeating image pattern 34560 x 34560 pixels, and a texel footprint of 0.58 cm square on the topography. The image of grass was taken with a DSLR and manually edited to create a seamless texture for tiling with no edge effects between tiles. The aerial image and grass texture were merged together by setting the grass texture with an alpha of 0.15 and the aerial image layered beneath it with an alpha value of 1. The cube was textured using a 3456x3456 pixel seamless image of rocks that was derived from a DSLR image taken by the authors. This resulted in an effective texel footprint of 0.35 cm on the cube. Each of the textures is set so that the coloring on the scene is interpolated between texels so that there are no abnormal effects on the edge of pixels. Oblique images of each object in the scene are shown in Figure 4.

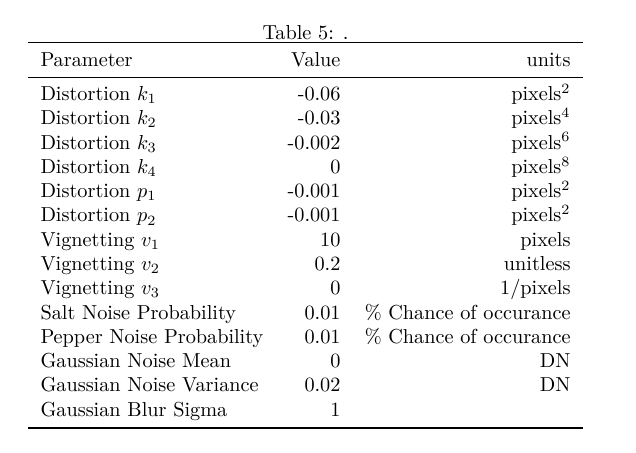
The scene was illuminated using a "Sun" style of lamp in Blender, where all the light rays are parallel to each other. The light is initially directed at nadir and the angle is linearly interpolated to a 30-degree rotation about the x axis for the final image. The sun has a constant intensity of 1, and emits white light with values of 1 for R, G, and B. To improve scene texture in shadowed regions, an ambient light source was added with an intensity value of 0.25 and values of 1 for R, G, and B.

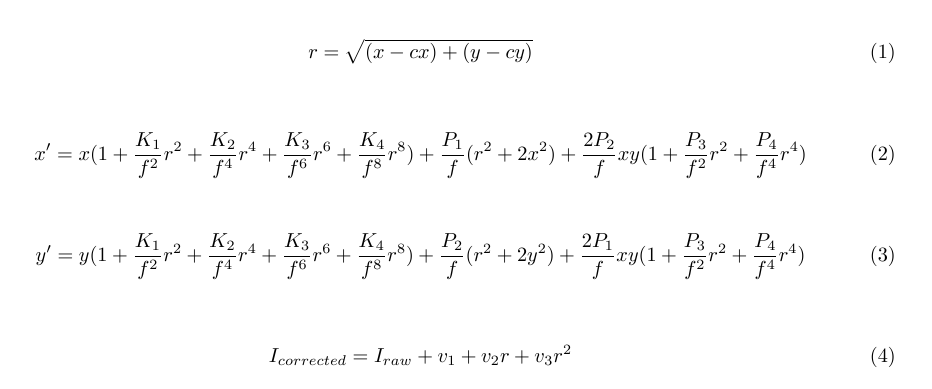


A camera was created in Blender with parameters meant to emulate a Sony A5000 camera with a 16mm lens and 5456x3632 (20 Mp) pixel sensor. An array of simulated cameras are placed on a flight path to create a Ground Sampling Distance (GSD) of 1.00 cm and an overlap and sidelap of 75% each. To remove imaging on the edge of the simulated topographic surface, the inner 100m x 100m of the topography were selected as the Area of Interest (AOI). The trajectory consists of 77 simulated cameras distributed across 7 flight lines with nadir looking imagery, as shown in Figure 5. To generate imagery that is more representative of a real-world scenario with a UAS, 1 sigma Gaussian noise in meters is added to the camera translation in each of the three dimensions and 2 sigma Gaussian noise in degrees is added to the camera rotation about each of the three axes. Imagery was then rendered using Blender Internal Render Engine with the default 8-sample antialiasing enabled. The processing to render the imagery took 2 hours and 50 minutes.



The imagery that is output from Blender is postprocessed in MATLAB to simulate nonlinear brown distortion (Equation 1-3) (), vignetting (Equation 4), salt/pepper noise, Gaussian noise, and Gaussian blur. To accurately apply fisheye distortion and Gaussian blur, the imagery was rendered at a larger resolution than the sensor and then cropped after the filtering was applied. The constants used in this postprocessing are shown in Table 5. The post-processing of imagery in MATLAB took 50 minutes.



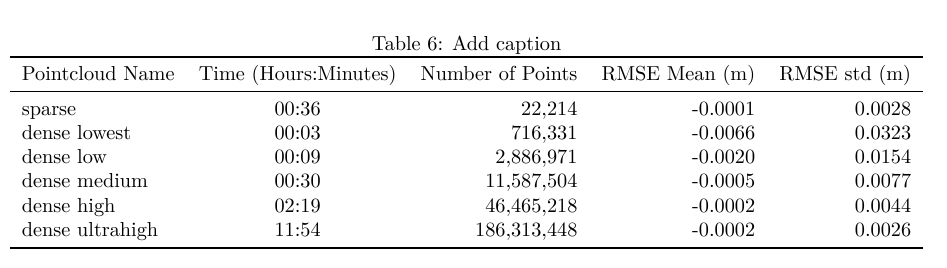


**Use Case Processing Methodology**

The resultant imagery was processed using the commercial software Agisoft Photoscan Pro using the settings shown in Table 6. The dataset was processed by inputting the position of the cameras, the position of the GCPs, and the camera calibration file. Additionally, the pixel coordinates of the GCPs, which are traditionally clicked by the user by varying degrees of accuracy are calculated using photogrammetric equations and input into the program. All the values were input into Agisoft Photoscan Pro with no uncertainty. A nonlinear adjustment is performed using the "optimize" button, and the reported total RMSE for the GCPs is 0.38 mm. This represents an idealized scenario which is currently unrealistic for a real world scenario as traditionally there will be error in the GCP surveyed markers, the UAS position, the user clicking of pixel coordinates for GCPs, and in the calculation of the camera calibration.

|  |  |  |
| --- | --- | --- |
| Processing Parameter | Value | units |
| Align Photos | high | N/A |
| Max tiepoints | 40000 | N/A |
| Max keypoints | 4000 | N/A |
| Pair Preselection | Disabled | N/A |
| Input Camera Calibration | yes | N/A |
| Lock Camera Calibration | yes | N/A |
| Input GCP targets | yes | N/A |
| Input GCP pixel coordinates | yes | N/A |
| Input Image Positions | yes | N/A |
| Camera Accuracy | 0.005 | m |
| Camera Accuracy (degrees) | 2 (not used) | degrees |
| Marker Accuracy | 0.005 | m |
| Scale Bar Accuracy | 0.001 (not used) | m |
| Marker Accuracy | 0.01 | pixel |
| Tie Point Accuracy | 1 | pixel |

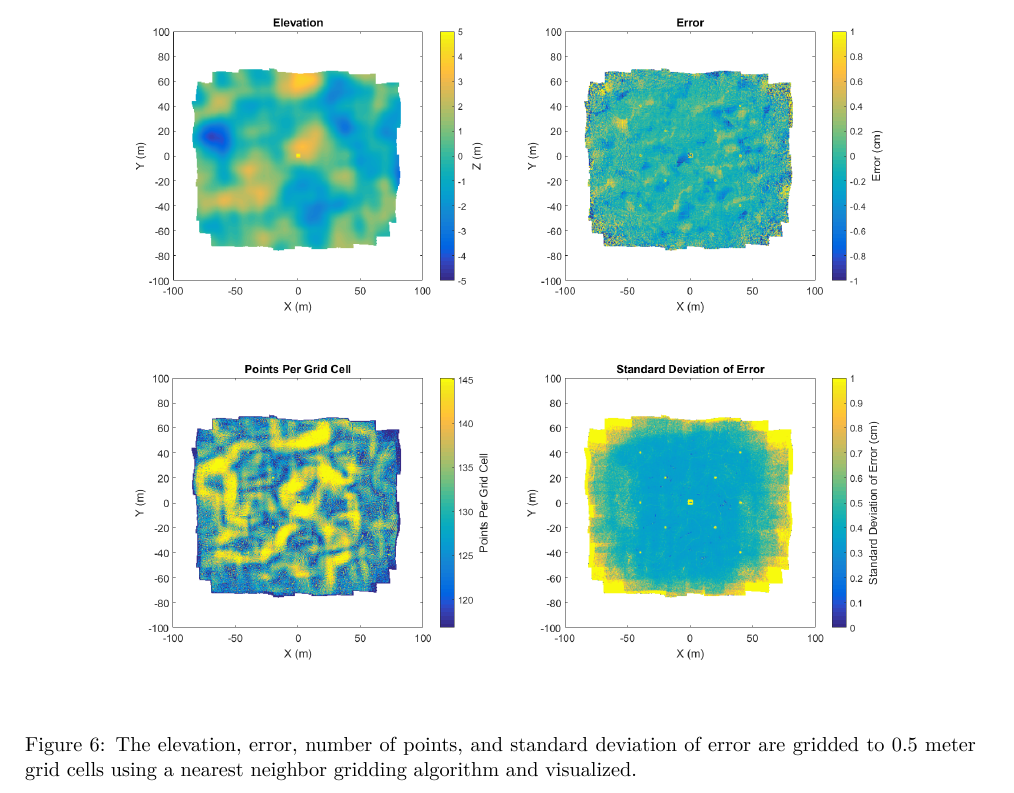
A dense reconstruction is performed using the "aggressive" filtering and each of the quality settings available in Photoscan (lowest, low, medium, high, highest) to generate five different pointclouds. The higher the quality setting, the more “detailed and accurate geometry” is generated. The limiting factor is the time and CPU processing power required to process large datasets. Ultrahigh becomes quickly unattainable to users without purpose built CPUs with a large amount of RAM. The processing time and number of points for each pointcloud is shown in Table 6.



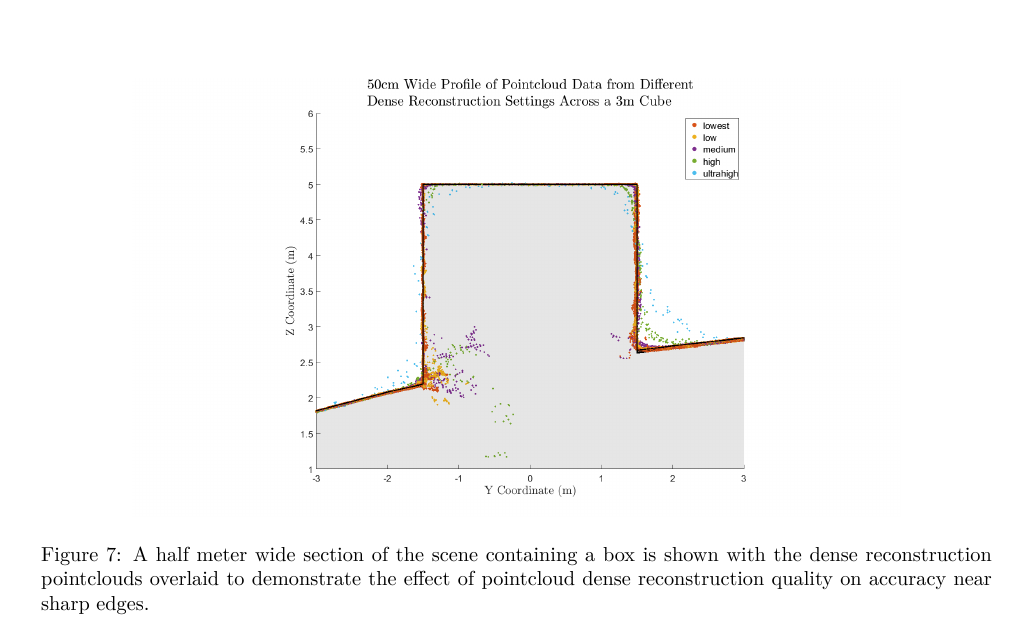
Each of the dense pointclouds is processed using CloudCompare (“CloudCompare (version 2.8),” 2017) and compared to the groundtruth blender mesh using the "point to plane" tool. This tool calculates the signed distance of every point in the pointcloud to the nearest surface on the mesh, using the surface normal to determine the sign of the error. Each pointcloud was then exported and analyzed in MATLAB to determine how the dense reconstruction quality setting effects the pointcloud error.

# Use Case Results

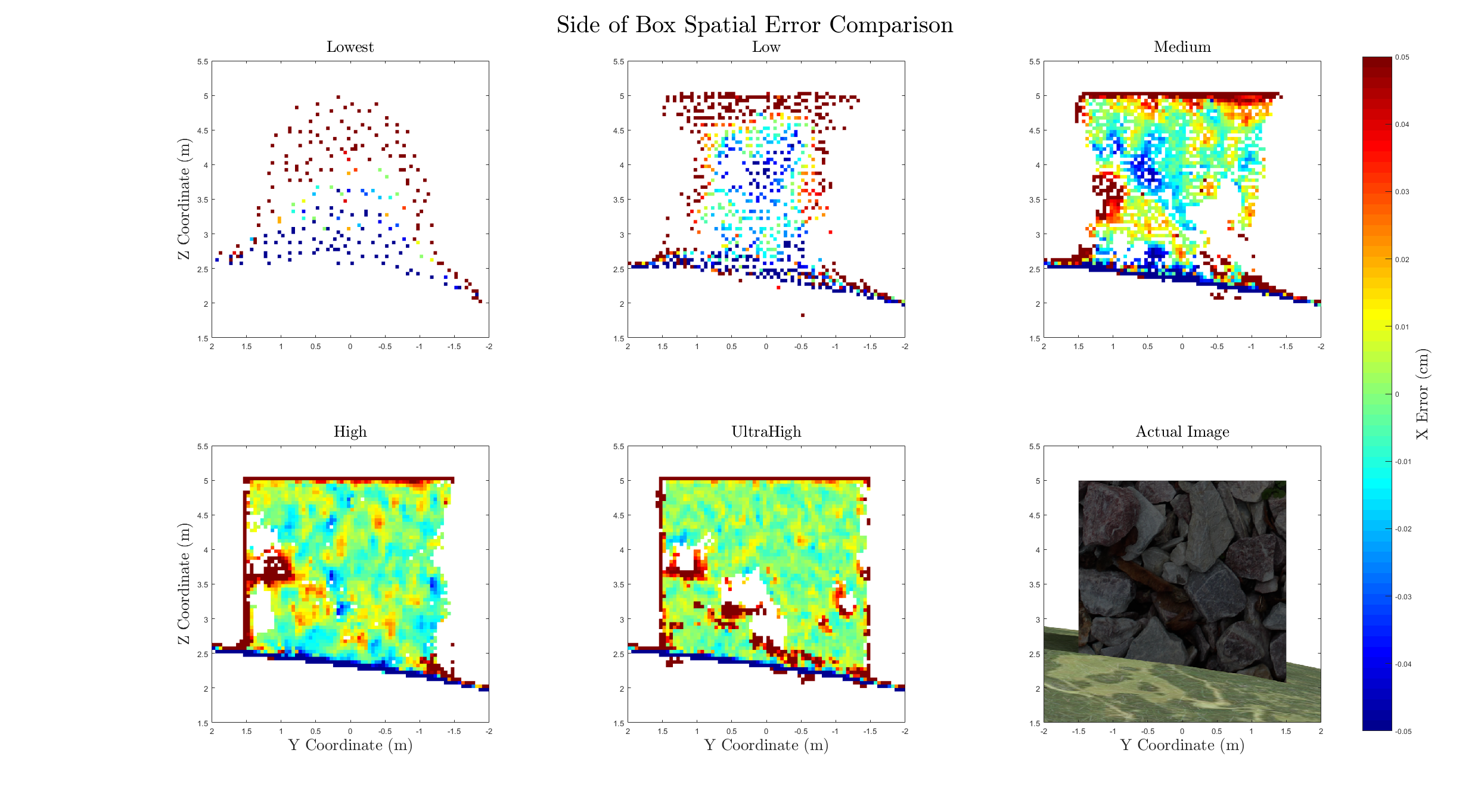
The error was first visualized spatially for each reconstruction by gridding the pointcloud elevation and error using a nearest neighbor gridding technique. The number of points and standard deviation of points in each grid cell are also visualized. The results for the medium quality dense reconstruction are shown in Figure 6. These plots are useful to begin to explore the spatial variability in both the density and the errors in the data. One initial observation for this dataset is that there is a larger standard deviation of error at the edges of the pointcloud outside the extents of the AOI. This is due to the poor viewing geometry at the edges of the scene, and suggests that in practice these data points outside of the AOI should be either discarded or used cautiously.



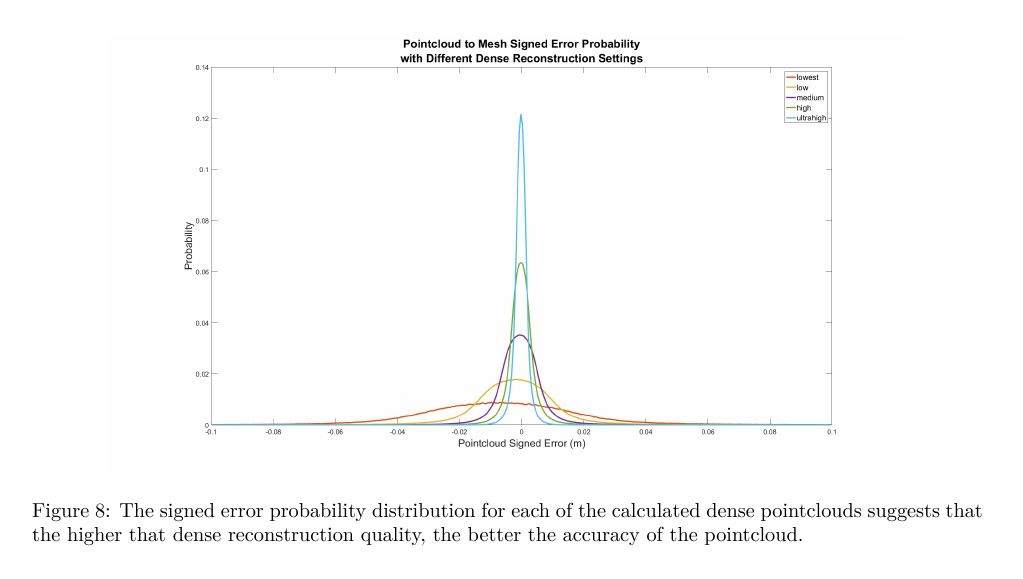
To qualitatively observe the effect of different quality dense reconstructions, a plot showing the true surface and the points from each construction in a 0.5-meter-wide section of the 27 m3 box is shown in Figure 7. Notice that the accuracy of each pointcloud at the sharp corners of the box improves as the quality of the reconstruction increases, which is consistent with the Agisoft Photoscan Pro manual. This observation suggests that higher quality dense reconstruction settings will increase accuracy in regions with sharp corners.



A visualization of the horizontal error of points along one side of the box is shown in Figure 8. All points within 0.25 meters horizontally of the face of the box are compared to the true x coordinate of the box. This 1D error calculation along the X dimension shows how well the face of the box is calculated, and errors along the edge of the box and along the ground surface should be ignored. The regions that are white demonstrate where there are no data points. The size and location of these data gaps varies between each pointcloud. For example, the High pointcloud contains points in the lower center of the cube, while the ultrahigh does not. While the data gap in the ultra-high appears to be correlated to a region of low texture on the actual image, further research is required to determine the cause.



A more quantitative, statistical assessment is performed to assess the error throughout the entire scene by calculating a histogram for the distribution of error in each pointcloud, as shown in Figure 8. These distributions bolster the conclusion derived from the box profile plot, which is that higher quality dense reconstruction settings yield more accurate results than a lower quality reconstruction. While the accuracy of the GCPs, as provided in Agisoft Photoscan, averaged 0.38 mm (RMSE), the standard deviations of the points from the dense reconstruction ranged from 2.6 mm to 32.3 mm, as shown in Table 6. This observation indicates that the GCP accuracy table is insufficient as a metric to depict the accuracy of the resultant dense pointcloud. While these conclusions suggest general trends, further experimentation is required in order for the accuracy and magnitudes of the distribution to be generalized.



# Discussion

The use case demonstration provides just one example of the type of rigorous analysis that can be obtained by utilizing computer generated imagery, rather than field data. It is important to note that the data and results associated in this experiment are closely coupled to the texture and topography of the scene. Future work will vary these independent variables to assess their effect on pointcloud accuracy.

The first conclusion from this experiment is that the error and standard deviation of error are larger for points outside of the area of interest, which in this experiment was -50 m to 50 m in both the x and y directions. This is shown in the spatial error plot in Figure 6. The cause of this error is the poor viewing geometry of these points, where they are only seen by a few cameras at oblique angles. In practice, these points should be included in the final data product with caution, as it is shown here that the errors can be significantly greater than those within the AOI.

The second conclusion from this experiment is that a higher quality dense pointcloud results in a more accurate pointcloud, as shown qualitatively in Figure 7 and quantitatively in Figure 8. The quality settings in Photoscan determines the amount of downsampling of the imagery that should occur before performing the reconstruction algorithm. The downsampling of the imagery removes some of the finer texture details in the imagery, and therefor reduces the quality of the keypoint matching. The authors recommend using the highest quality dense reconstruction setting that the computer processing the dataset can handle. However, if there are noticeable data gaps in the pointcloud, consider processing the pointcloud on a lower dense reconstruction setting and merging the pointclouds. For this experiment, a relatively small number of 20Mp images (77) were used to create the dense pointcloud, which took almost 12 hours for the highest pointcloud setting. The resultant pointcloud for this setting also contained 186 Million points, which caused some pointcloud data viewers and processing to fail due to memory issues. For this reason, ultra-high may not be a possible solution for all experiments.

The third conclusion is that the RMSE of the GCP control network as shown in Agisoft Photoscan Pro is insufficient to characterize the accuracy of the resultant dense pointcloud. In this extremely idealized experiment, where the GCP positions, pixel coordinates of GCPs, camera positions, and camera calibration were all input precisely, the GCP control network 3D RMSE was 0.38 mm. The smallest standard deviation of the pointcloud error, using the "ultra-high" quality setting, is 2.6mm and the largest standard deviation, using the "lowest" setting, is 32.3 mm, as shown in Table 6. Further experimentation is needed to determine the relationship between the GCP total RMSE and the RMSE of the dense pointcloud. This workflow is well suited to perform the experimentation required to make accurate generalized conclusions with confidence.

## Methodology Implications

This methodology generates photogrammetrically accurate imagery rendered using a pinhole camera model of a scene with various textures and lighting, which is then processed to assess SfM pointcloud accuracy. The rendered imagery can be postprocessed to add noise, blur, nonlinear distortion, and other effects to generate imagery more representative of that from a real-world scenario. The accuracy of the camera trajectory, GCP position, camera calibration, and GCP pixel coordinates in each image can also be systematically adjusted to simulate uncertainty in a real world scenario. The ability to adjust these parameters enables a user to perform a sensitivity analysis with numerous combinations of independent variables.

While this methodology enables the user to perform repeatable, accurate experiments without the need for time consuming field work, there are currently some limitations in the experiment methodology when utilizing the Blender Internal Render Engine. First, the internal render engine does not handle global illumination, and therefore light interactions between objects are not modeled. A second limitation of the lighting schema is that the radiometric accuracy has also not been independently validated. There are a few methods within the render engine which effect the exposure and gamma of the resultant imagery. For this experiment, these exposure and gamma values were set at the default and provided imagery that was not over or underexposed. While the lighting in the scene using the Blender Internal Render Engine does not perfectly replicate physics based lighting, the absolute color of each surface of an object is constant and perfectly Lambertian. The keypoint detection and SfM algorithms utilize gradients in colors and the absolute colors of the scene, and the accuracy should not be effected by the imperfect lighting.

Another broader source of inaccuracy in the Blender Internal Render Engine methodology is the methodology to convert the scene to pixel values relies in an integration over a finite number of subpixel super-sampling. This deviates from a real world camera where the pixel value is a result of an integration over all available light. The Blender Internal Render Engine uses the term "antialiasing" to describe a supersampling methodology for each pixel, which can supersample up to 16 samples per pixel. This small, finite number of samples per pixel can induce a small amount of inaccuracy when mixed pixels are present. These small inaccuracies, though, are small enough to be deemed negligible for this experiment.

A potential source of uncertainty induced into the system is the use of repeating textures to generate a scene. In the use case provided in Section X, the grass texture is repeated 10 times in both the x and y direction. This repeating pattern was overlaid onto another image, to create different image color gradients in an attempt to generate unique texture features. Despite this effort, it is possible that keypoint detection and matching algorithms could generate false positives which may bias the result if not removed or detected as outliers. This phenomenon could also occur in a real-world scenario, where manmade structures often exhibit a repeating pattern of similar shapes and colors. In this experiment, this effect was not observed, but if the scene is not generated carefully, these repeating textures could induce a significant amount of inaccuracy in the SfM processing step.

# Conclusion

The methodology presented utilizing simulated UAS based image acquisition with the Blender Internal Render Engine enables research into the performance of SfM and MVS algorithms without the need for costly field experiments. The accuracy of the simulated groundtruth data enhances the confidence in sensitivity analyses and allows for numerous repeat experiments. An example use case is presented to explore the effect of the Agisoft Photoscan Pro dense reconstruction setting on the accuracy of the output pointcloud. Results suggest that the errors decrease as the dense reconstruction setting is increased. Secondary results suggest that the data points outside of the AOI should be either discarded or used with caution, as the accuracy of those points is higher than that of the pointcloud within the AOI.

As the application of SfM-MVS algorithms continue to expand into new fields, it is important to first test the accuracy of the pointcloud when performing new experiments. This methodology could also be used from alternative sources of imagery, such as handheld or vehicle based imagery. Future work will focus on various sensitivity analyses to assess the accuracy of new algorithms, applications, and potential new sensors. The radiometric accuracy of the Blender algorithm and its implications with regard to SfM accuracy assessments should also be evaluated in future work.

Clapuyt, F., Vanacker, V., & Van Oost, K. (2015). Reproducibility of UAV-based earth topography reconstructions based on Structure-from-Motion algorithms. *Geomorphology*, *260*, 4–15. https://doi.org/10.1016/j.geomorph.2015.05.011

CloudCompare (version 2.8). (2017). Retrieved from http://www.cloudcompare.org/

Dandois, J. P., Olano, M., & Ellis, E. C. (2015). Optimal altitude, overlap, and weather conditions for computer vision uav estimates of forest structure. *Remote Sensing*, *7*(10), 13895–13920. https://doi.org/10.3390/rs71013895

Eltner, A., Kaiser, A., Castillo, C., Rock, G., Neugirg, F., & Abell??n, A. (2016). Image-based surface reconstruction in geomorphometry-merits, limits and developments. *Earth Surface Dynamics*, *4*(2), 359–389. https://doi.org/10.5194/esurf-4-359-2016

Espositoa, S., Fallavollitaa, P., Wahbehb, W., Nardinocchic, C., & Balsia, M. (2014). Performance evaluation of UAV photogrammetric 3D reconstruction. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 4788–4791). IEEE. https://doi.org/10.1109/IGARSS.2014.6947565

Furukawa, Y., & Hernández, C. (2015). Multi-View Stereo: A Tutorial. *Foundations and Trends® in Computer Graphics and Vision*, *9*(1–2), 1–148. https://doi.org/10.1561/0600000052

Harwin, S., & Lucieer, A. (2012). Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from Unmanned Aerial Vehicle (UAV) imagery. *Remote Sensing*, *4*(6), 1573–1599. https://doi.org/10.3390/rs4061573

Harwin, S., Lucieer, A., & Osborn, J. (2015). The impact of the calibration method on the accuracy of point clouds derived using unmanned aerial vehicle multi-view stereopsis. *Remote Sensing*, *7*(9), 11933–11953. https://doi.org/10.3390/rs70911933

Hugenholtz, C. H., Whitehead, K., Brown, O. W., Barchyn, T. E., Moorman, B. J., LeClair, A., … Hamilton, T. (2013). Geomorphological mapping with a small unmanned aircraft system (sUAS): Feature detection and accuracy assessment of a photogrammetrically-derived digital terrain model. *Geomorphology*, *194*, 16–24. https://doi.org/10.1016/j.geomorph.2013.03.023

Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., & Aanaes, H. (2014). Large scale multi-view stereopsis evaluation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 406–413. https://doi.org/10.1109/CVPR.2014.59

Lowe, D. G. (2004). Distinctive image features from scale invariant keypoints. *Int’l Journal of Computer Vision*, *60*, 91–11020042. https://doi.org/http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94

Martin, R., Rojas, I., Franke, K., & Hedengren, J. (2015). Evolutionary View Planning for Optimized UAV Terrain Modeling in a Simulated Environment. *Remote Sensing*, *8*(1), 26. https://doi.org/10.3390/rs8010026

Micheletti, N., Chandler, J. H., & Lane, S. N. (2015). Investigating the geomorphological potential of freely available and accessible structure-from-motion photogrammetry using a smartphone. *Earth Surface Processes and Landforms*, *40*(4), 473–486. https://doi.org/10.1002/esp.3648

Naumann, M., Geist, M., Bill, R., Niemeyer, F., & Grenzdörffer, G. J. (2013). Accuracy Comparison of Digital Surface Models Created By Unmanned Aerial Systems Imagery and Terrestrial Laser Scanner. *International Archives of Photogrammetry and Remote Sensing*, *XL*(September), 4–6. https://doi.org/10.5194/isprsarchives-XL-1-W2-281-2013

Nilosek, D., Walvoord, D. J., & Salvaggio, C. (2014). Assessing geoaccuracy of structure from motion point clouds from long-range image collections. *Optical Engineering*, *53*(11), 113112. https://doi.org/10.1117/1.OE.53.11.113112

Salvaggio, K. N., & Salvaggio, C. (2013). Automated identification of voids in three-dimensional point clouds. *Spie*, *8866*, 88660H--88660H. https://doi.org/Artn 88660h\rDoi 10.1117/12.2023740

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., & Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *1*, 519–528. https://doi.org/10.1109/CVPR.2006.19

Smith, M. W., & Vericat, D. (2015). From experimental plots to experimental landscapes: Topography, erosion and deposition in sub-humid badlands from Structure-from-Motion photogrammetry. *Earth Surface Processes and Landforms*, *40*(12), 1656–1671. https://doi.org/10.1002/esp.3747