May 19, 2015

Textured digital elevation model formation from low-cost UAV ladar/ digital image data

Taylor C Bybee, *Utah State University*
Scott Budge

Textured digital elevation model formation from low-cost UAV ladar/digital image data

Taylor C. Bybee, Scott E. Budge

Center for Advanced Imaging Ladar, Utah State University,
Logan, UT 84322-4120, (435) 797-3433

ABSTRACT

Textured digital elevation models (TDEMs) have valuable use in precision agriculture, situational awareness, and disaster response. However, scientific-quality models are expensive to obtain using conventional aircraft-based methods. The cost of creating an accurate textured terrain model can be reduced by using a low-cost (<$20k) UAV system fitted with ladar and electro-optical (EO) sensors.

A texel camera fuses calibrated ladar and EO data upon simultaneous capture, creating a texel image. This eliminates the problem of fusing the data in a post-processing step and enables both 2D- and 3D-image registration techniques to be used. This paper describes formation of TDEMs using simulated data from a small UAV gathering swaths of texel images of the terrain below. Being a low-cost UAV, only a coarse knowledge of position and attitude is known, and thus both 2D- and 3D-image registration techniques must be used to register adjacent swaths of texel imagery to create a TDEM.

The process of creating an aggregate texel image (a TDEM) from many smaller texel image swaths is described. The algorithm is seeded with the rough estimate of position and attitude of each capture. Details such as the required amount of texel image overlap, registration models, simulated flight patterns (level and turbulent), and texture image formation are presented. In addition, examples of such TDEMs are shown and analyzed for accuracy.

Keywords: lidar, ladar, 3D image creation, 2D/3D fusion, texel image, image registration, triangulation, textured digital elevation model

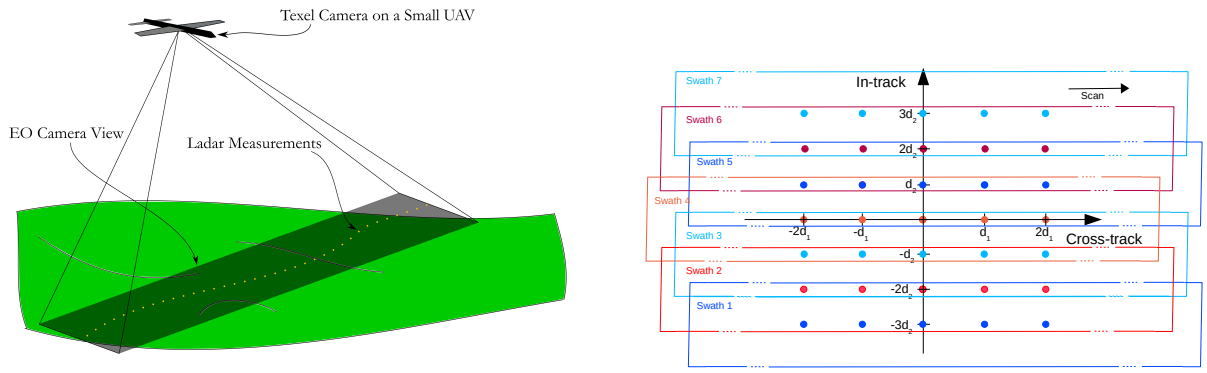
1. INTRODUCTION

Triangulation of aerial images has been used for decades to produce both orthophotos and digital elevation models.[1,2] One of the more straightforward methods involves finding matching points in calibrated images corrected for lens distortion, then solving for the relative 3D points.[3]

Previous work in 3D terrain modeling has included relying solely upon aerial imagery and image processing techniques. This method works well, but ambiguities exist in the resulting 3D model as there are no direct measurements made.[4] This can be offset to some degree by using ground control points, GPS,[5] or human intervention.[6] Another drawback is the computation time involved, but this can be mitigated using sparsity or sub-map concepts.[7] The alternative to using aerial imagery is to use expensive ladar, GPS, and inertial measurement unit (IMU) equipment that can pinpoint locations precisely while in a large unmanned aerial vehicle (UAV) or manned aircraft. The major drawback of this method is cost. This paper demonstrates a compromise between the two methods by incorporating ladar measurements, aerial imagery, and an approximate location and attitude of each capture in such a way to reduce cost while maintaining accuracy.

A texel camera is an instrument which fuses ladar and imagery upon capture. It combines both a ladar system and an electro-optical (EO) camera such that the measurements from both sensors are registered in both time and space. The registration is accomplished through a calibration process which involves mapping each ladar point to a point on the EO image, and does not change from capture to capture.[8]

This paper describes a triangulation method that incorporates both 2D EO image data and 3D ladar measurements from a low-cost, small UAV to create TDEMs. As the small UAV flies in a straight and level flight profile, the ladar scans a series of shots perpendicular (cross-track) to the flight path, as well as a narrow strip or swath of EO imagery, as illustrated in Figure 1(a). The simultaneous collection of the ladar points and the EO



(a) (b)

Figure 1: Acquisition of texel swaths. (a) Capture of a single texel swath from a small UAV. (b) Pattern of multiple adjacent texel swaths on the ground.

image are called texel swaths. As the small UAV flies forward, overlapping texel swaths are acquired, as shown in Figure 1(b). In addition to the texel swaths, the small UAV records the position and attitude of the camera for each swath with a low-cost micro electro-mechanical system (MEMS) GPS/IMU. The ladar measurements, however, are only as accurate in the world coordinate system as the position and attitude obtained from the GPS/IMU . Our goal is to use knowledge present in the aerial imagery to reduce the error in the camera position and attitude measurements using a unique cost function designed for texel images. This enables the ladar points to be accurate in the world coordinate system. From these accurate points, a TDEM can be formed.

The remainder of the paper is as follows. Section 2 describes the registration model and the associated image processing techniques. Section 3 describes the optimization algorithm and the TDEM formation. Section 4 describes the experimental results, and, finally, Section 5 concludes the paper with a summary and ideas for further work.

2. SELECTION OF PROJECTION POINTS

Texel swaths are potentially gathered several times per second, along with a coarse measurement of the camera position and attitude made with a low-cost GPS/IMU. These texel swaths are assumed to have some amount of overlap from one to the next. Because each measured 3D point can be viewed from more than one camera, triangulation can be used to improve the estimate of the camera locations and 3D ladar points. Once these are known, the 3D points and the corresponding EO imagery from each swath can be combined with other swaths to form a TDEM.

Each texel swath j acquires n_j ladar points, a set denoted as I_j. By mapping all the ladar points in M swaths into a common coordinate system $\mathcal{O}$ using the GPS/IMU measurements, the entire set of 3D points can be indexed using i, and each point is represented as a 3D vector of Cartesian coordinates, $\mathbf{b}_i = [b_{i_x}, b_{i_y}, b_{i_z}]$. The total number of 3D points in the system is $N = \Sigma_{j=0}^{M-1} n_j$. A ladar point $\mathbf{b}_i$ belongs to the texel swath j if $i \in I_j$. Calibrated 2D projection points are defined as the projection of points $\mathbf{b}_i$ into the normalized image plane of texel swath j when $i \in I_j$. These points are determined by the texel camera calibration, and are the intersection of the ladar ray with the normalized image plane. These calibration values remain constant for a given instrument. Because the EO camera is calibrated, the pixel locations corresponding to the 2D projection points are also known.

In order to fuse texel swaths, common points must be identified in several EO images. Each ladar point has a measured 3D location and a calibrated 2D projection point in one texel swath, as well as projections of the 3D point into several adjacent texel swaths. The proposed algorithm uses the assumption that a small patch of EO

imagery surrounding the projection of a ladar point can be matched to corresponding patches on several other texel swaths using simple image processing techniques.

Because the texel swaths are captured very quickly, the scene in the EO image does not change significantly from one swath to the next. This allows the effective use of normalized cross-correlation (NCC) to find corresponding projected image points. A pair of image patches are considered putative correspondences (image patches potentially representing the same image point) if the NCC score is above a threshold.

In order to reduce the NCC search size to find the matching ladar point image patches, a homography relationship from swath to swath is found using feature points. This also allows perspective correction of the EO images for better correlation on ladar point projections. The homography transform is given by

$$\begin{bmatrix} c_1 \\ r_1 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{bmatrix} c_2 \\ r_2 \\ 1 \end{bmatrix}, \tag{1}$$

where r_m and c_m represent the row and column of matching points in the images (in pixel coordinates), and the h_{ij} represent the transform parameters.

A simple Harris corner detector,[9] with a response given by Noble,[10] was chosen for this application. Once the Harris feature points are found in each pair of images, a small patch surround each feature point in the first EO image is compared to a small patch around each feature point in the second EO image. The feature points with the largest NCC score greater than a threshold are considered putative correspondences. These putative correspondences are then filtered again by the RANSAC algorithm[11] to remove outliers, and a homography is found using a linear least-squares method.

Once this homography is found, then the calibrated 2D projected image point in a texel swath can be mapped into other texel swaths, and the image patch around these points can be warped and resampled using the homography so that the image patches have the same perspective. This serves as a starting point to search the area using NCC to find the small patch with the highest NCC score which is then considered a "match" for the projected image point. A homography assumes the scene in which the image points are located are on a plane. This is not usually the case, which is why an area surrounding the expected location of the ladar point must be searched. This reduces the effects of parallax.

Normalized cross-correlation fails when there is a rotation between swaths or when significant lighting differences (such as changing shadows) are present. However, because adjacent swaths are captured nearly simultaneously, it can be safely assumed there are no significant lighting changes. If there were significant rotation from swath to swath, such as from turbulence, then knowledge of the position and attitude between the camera locations can help remove the rotation between the images. This is done by creating a simple rotational transform based on the IMU attitude measurements, and rotating the image patches surrounding the Harris points in the second swath into the space of the first swath, effectively removing the rotation, before the NCC is performed.

This process is summarized in Algorithm 1. Note that the algorithm initially assumes there is no rotation from image to image, as it attempts to find a homography. If this fails, then the attitude knowledge is used to remove any rotation, and the process starts again. In practice, Step 4 is not needed often.

Once a homography is found between sets of adjacent swaths, it does not need to be explicitly recalculated using the method in Steps 1 - 3 for Steps 5 and 6 to find matching ladar projections in swaths not adjacent to one another. To find these matching ladar image patches, the starting point for the search can be determined by merely cascading the homographies of the pairs of images between the swaths. The images are warped and resampled using this cascaded homography. Cascading the transforms significantly reduces computation time.

At the completion of Algorithm 1 for each pair of texel swaths, a set of 2D image points (x_{ij}, y_{ij}) has been determined. These constitute the projection of the i^{th} ladar point into the normalized image plane of the j^{th} texel swath. An example of a few of the (x_{ij}, y_{ij}) is shown in Figure 2. In the figure, each color corresponds to projected points from one swath. The line of points down the center of each swath are projected from the ladar points acquired with that swath. Missing points result when the NCC process does not produce a maximum above the threshold.

Algorithm 1 Selection of image points in adjacent texel swaths.

1. Find Harris feature points in adjacent swaths.
2. Find corresponding Harris feature points using NCC.
3. Determine the homography using linear least-squares and RANSAC. This step assumes there is no rotation between the swaths.
4. If RANSAC fails, perform the following.
   (a) Use the position and attitude knowledge from the GPS/IMU project the 3D points from each swath into image of the other swath.
   (b) Using these image points, and the corresponding image points known from the calibration process, compute an initial homography between swaths.
   (c) Using the homography, map the second image into the first, and repeat Step 3 with the mapped second image and Harris points. This enables the Harris feature finding and NCC to work effectively.
   (d) The final homography used in Step 5 is the cascade of the homographies in Steps 4b and 4c.
   (e) If RANSAC fails again, then assume an identity transform – this may cause the algorithm to fail.
5. Determine the projection of each ladar point in the first image, then map each point and surrounding pixels to the other image using the homography found in Step 3.
6. Search an area centered around the expected ladar point location using NCC to find each of the matching locations. Save these locations for optimization.

3. BUNDLE ADJUSTMENT AND TEXTURE CREATION

A given 3D point i can be mapped into the coordinate system of camera j using

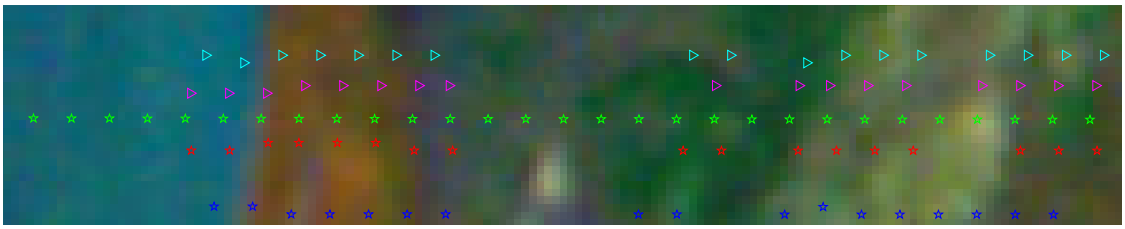
$$\hat{\boldsymbol{\chi}}_{ij} = \begin{bmatrix} \hat{\chi}_{ij_x} \\ \hat{\chi}_{ij_y} \\ \hat{\chi}_{ij_z} \end{bmatrix} = \begin{bmatrix} 1 - \frac{2(q_{j2}^2+q_{j3}^2)}{q_{j0}^2+q_{j1}^2+q_{j2}^2+q_{j3}^2} & \frac{2(q_{j1}q_{j2}-q_{j0}q_{j3})}{q_{j0}^2+q_{j1}^2+q_{j2}^2+q_{j3}^2} & \frac{2(q_{j1}q_{j3}+q_{j0}q_{j2})}{q_{j0}^2+q_{j1}^2+q_{j2}^2+q_{j3}^2} \\ \frac{2(q_{j1}q_{j2}+q_{j0}q_{j3})}{q_{j0}^2+q_{j1}^2+q_{j2}^2+q_{j3}^2} & 1 - \frac{2(q_{j1}^2+q_{j3}^2)}{q_{j0}^2+q_{j1}^2+q_{j2}^2+q_{j3}^2} & \frac{2(q_{j2}q_{j3}-q_{j0}q_{j1})}{q_{j0}^2+q_{j1}^2+q_{j2}^2+q_{j3}^2} \\ \frac{2(q_{j1}q_{j3}-q_{j0}q_{j2})}{q_{j0}^2+q_{j1}^2+q_{j2}^2+q_{j3}^2} & \frac{2(q_{j2}q_{j3}+q_{j0}q_{j1})}{q_{j0}^2+q_{j1}^2+q_{j2}^2+q_{j3}^2} & 1 - \frac{2(q_{j2}^2+q_{j1}^2)}{q_{j0}^2+q_{j1}^2+q_{j2}^2+q_{j3}^2} \end{bmatrix}^T \begin{bmatrix} b_{i_x} - t_{jx} \\ b_{i_y} - t_{jy} \\ b_{i_z} - t_{jz} \end{bmatrix}, \tag{2}$$

where the location and attitude of each camera j can be represented using seven parameters: a quaternion rotation and a 3D translation vector, given by $\mathbf{a}_j = \{q_{j0}, q_{j1}, q_{j2}, q_{j3}, t_{jx}, t_{jy}, t_{jz}\}$.[12] These values are referenced to a world coordinate system $\mathcal{O}$. A quaternion is used because it allows for straightforward optimization resulting in an orthonormal rotation. Other methods have been developed using Euler angles,[13] but employ only projections, and not measured ranges.

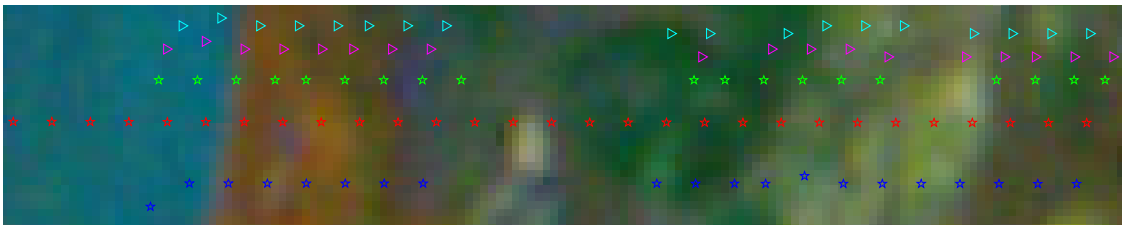
The point can then be projected onto the j^{th} normalized image plane by dividing by the z-coordinate, giving the 2D projection $\hat{x}_{ij}$ and $\hat{y}_{ij}$. The range to the ladar point $\hat{\lambda}_{ij}$ is the Euclidean distance between the camera center and the ladar point. These can be combined into a vector given by

$$\hat{\boldsymbol{X}}_{ij} = \begin{bmatrix} \hat{x}_{ij} \\ \hat{y}_{ij} \\ \hat{\lambda}_{ij} \end{bmatrix} = \begin{bmatrix} \frac{\hat{\chi}_{ij_x}}{\hat{\chi}_{ij_z}} \\ \frac{\hat{\chi}_{ij_y}}{\hat{\chi}_{ij_z}} \\ \sqrt{\hat{\chi}_{ij_x}^2 + \hat{\chi}_{ij_y}^2 + \hat{\chi}_{ij_z}^2} \end{bmatrix}, \tag{3}$$

which is merely another way to represent a 3D point in space. It represents a point with normalized projection coordinates and a range value rather than with Cartesian coordinates.



(a)



(b)

Figure 2: Image points projected from ladar points found using correlation in adjacent swaths. Colors correspond to the points projected from ladar points in the current swath and swaths before and after the current swath. Each color corresponds to points from one swath. The line of points down the center of each swath are calibrated 2D points projected from the ladar points acquired with that swath. (a) First swath (zoomed to show detail). (b) Second swath (zoomed to show detail).

There are two types of points being projected into a given image plane: those with a measured range value from the camera j ($i \in I_j$) and otherwise. First, if $i \in I_j$, then there is a range measurement and the projection error is defined in terms of distance from the calibration point in the image plane to the projection of the point into the image plane. Secondly, if $i \notin I_j$, then there is no range measurement, and the projection error is defined in terms of the distance from the correct location as determined by NCC to the projection of the point into the image plane.

The above errors can be combined into a single cost function given by

$$\begin{aligned}\mathcal{E}^2 = &\sum_{j=0}^{M-1}\sum_{i\in I_j}\frac{1}{\sigma_I^2}[(x_{ij}-\hat{x}_{ij})^2+(y_{ij}-\hat{y}_{ij})^2] + \sum_{j=0}^{M-1}\sum_{i\notin I_j}\frac{1}{\sigma_{\bar{I}}^2}[(x_{ij}-\hat{x}_{ij})^2+(y_{ij}-\hat{y}_{ij})^2] \\ &+ \sum_{j=0}^{M-1}\sum_{i\in I_j}\frac{1}{\sigma_\lambda^2}(\lambda_{ij}-\hat{\lambda}_{ij})^2,\end{aligned} \tag{4}$$

where σ_I^2, $\sigma_{\bar{I}}^2$, and σ_λ^2 are the variances on the 2D calibrated projected points, the 2D points found using Algorithm 1, and the range measurements, respectively. If the measurements are assumed to be from a Gaussian distribution, $\mathcal{E}^2$ is the log-likelihood of the estimated parameters.

3.1 Nonlinear Optimization using the Levenberg-Marquardt Algorithm

The error for the system is minimized to create a maximum-likelihood estimate of the true camera positions needed to create an accurate TDEM. We desire to move all cameras $\mathbf{a}_j$ and points $\mathbf{b}_i$ such that the total squared

error $\mathcal{E}^2$ in the system is minimized. Both camera locations $\mathbf{a}_j$ and coordinates of each ladar measurement $\mathbf{b}_i$ must be adjusted to minimize the error. This is a nonlinear problem, due to the quaternion rotation and image plane projection, that can be solved using an iterative technique. The Sparse Levenberg-Marquardt Algorithm (LMA) as described by Hartley and Zisserman[3] can be used to do this, as it incorporates the fact that many relationships between $\mathbf{a}_j$ and $\mathbf{b}_i$ are nonexistent and do not contribute to the error in the system. In practice, the associated terms are merely left out of the sum.

This is an extremely large optimization problem given there are many 3D points and cameras in the system. The variance values must be selected carefully to ensure that points and cameras move freely, yet stay within the camera calibration and triangulation constraints. The computation of the solution can be reduced by imposing a window on the number of swaths before and after the current swath to include in the optimization.

The typical application of LMA uses an initial estimate of the parameters as a seed to the algorithm. This is provided by the coarse measurements of position and attitude from the GPS/IMU sensor the small UAV.

3.2 TDEM Creation

Once the system error is minimized to the desired degree, the data in each texel swath can be combined in such a way to form a single point cloud with an overlaid texture. The method in doing this depends on the desired application of the data. One common motive is to create an orthorectified TDEM.

For an orthorectified image, the camera location is considered to be at infinity, and the texture is merely the 3D colored surface projected onto a plane by using only $(\hat{\chi}_{i\mathcal{O}_x}, \hat{\chi}_{i\mathcal{O}_y})$. These are then converted into pixel coordinates using the intrinsic camera calibration parameters.

There are potentially several texel swaths with EO pixels that correspond to each final texture pixel. To simplify the process of creating the texture, it is assumed that the best pixel value for pixel p is interpolated from texel swath j, where swath j is determined by the projection of 3D point i such that the pixel p is closest to the projection of $i \in I_j$. The appropriate interpolated pixel value is determined by a homography relationship between the final texture and the texel swath j using their projected points. The homography between the final texture and each texel swath j is computed once at the start of the texturing process. Because a homography relationship assumes a planar scene, this has the potential to introduce parallax.

Once the TDEM is formed, then the entire system of points, cameras, and textures can be placed anywhere in space by applying a rigid rotation and translation to each point and camera. One suggestion would be to identify ground control points in the TDEM and reference those to a global coordinate system.

4. RESULTS

4.1 Test Data

The test data was gathered using a second-generation short-range handheld texel camera as described in Sec. 1. This texel camera consists of a PMD CamCube 2.0 depth camera (204x204 range pixels) as well as a Micron/Aptina EO camera (1024x1024 EO pixels) fitted with a VectorNav (VN-100) AHRS. The field of view (FOV) of the camera is 41°x41°. The range error of the camera is about 1 cm, and the attitude error of the VectorNav is 2° in heading and 0.5° in pitch and roll. The entire array of range measurements were removed except a strip near the middle of the EO image to mimic a single scan of ladar data. The EO image was trimmed around the range measurements, creating a visual "swath". The AHRS was mounted on the camera to record the attitude of each capture and a meterstick was used to determine the translation from a reference point. The camera was placed about a meter from a model landscape, and moved slightly from one capture to the next, simulating flight of a small UAV. This setup is shown in Figure 3(a). An example EO image of the landscape captured from the texel camera is shown in Figure 3(b).

(a) (b)

Figure 3: Data Set Acquisition. (a) Texel camera and landscape. (b) Example EO image captured using the texel camera.

4.2 Metric for Analyzing the TDEMs

The quality of a TDEM is both subjective and objective in nature. Fisher and Tate[14] describe sources of error in DEMs. In order to make data meaningful, it must be presented in a way appealing to the human eye. In other words, the TDEM must both "look good" and resemble the real world. This is very subjective. However, a more quantifiable method of determining the quality of a TDEM is to measure the distance between two points on the model and compare with actual surveyed points. The error is the difference between the two distance measurements. This then becomes a question of which points are chosen to evaluate the model. To estimate the average quality, many points can be chosen in throughout the model.

For this paper, to acquire "surveyed" points for the landscape model, a single-frame texel image is captured and evaluated which has a FOV covering the entire "flight" of the simulated small UAV. The desire, then, is to achieve registration to within the calibration of the single-frame texel image. This assumes the single-frame texel image is "ground truth" for measurements on the landscape.

In the following experiments, the error is analyzed by manually choosing ten 3D points on each of two single-frame texel images (one for each data set) and corresponding points on each of the registered TDEMs. They are chosen such that they are spread across the scene. The points are combined in $\binom{10}{2} = 45$ ways, and values for the error mean, the error standard deviation, and the root-mean-squared error between corresponding points on the ground truth and the TDEM are determined for each registered TDEM.

4.3 Registration Results and Discussion

4.3.1 Level Flight

This data set consists of moving the camera a few millimeters for each capture approximately one meter away from the model landscape, and contains 150 texel swaths. The camera maintained a fairly constant attitude, but a significant translation offset was introduced part-way through the data set.

The parameters used for this optimization include $\sigma_I = 0.001$ (1.36 pixels), $\sigma_{\bar{I}} = 0.002$ (2.72 pixels), and $\sigma_\lambda = 0.001$ m. No windowing was enforced, apart from that caused by the EO trimming. Convergence for the LMA was determined to be a 0.5% change in successive error, with a maximum number of iterations set at 100. The EO image width with the ladar strip centered in the EO image was adjusted to determine the accuracy of the registration if fewer adjacent swaths are seen. The results of these experiments are shown in Table 1. The errors in the TDEM created from using only the measured translation and IMU data confirm the need for a registration algorithm to create an accurate TDEM. It is interesting that the errors in the 160 pixel swaths do

Trim Size (pixels)	**Error Mean** (mm)	**Error Std Dev** (mm)	**RMS Error** (mm)	**Comments**
Before Registration	5.24	40.1	40	Includes a significant translation offset
200	0.690	4.08	4.10	Good registration, LMA converged relatively quickly
180	0.493	4.13	4.11	Good registration
160	1.02	4.40	4.47	Noticeable errors, but good registration overall
140	0.214	4.47	4.42	Noticeable errors, but good registration overall
120	8.10	16.0	17.8	Significant errors

Table 1: Error for Various EO Widths for Level Flight Data Set

not follow the trend for error mean and RMS error. This may be due to the random inlier selection inherent in RANSAC.

Figure 4 shows TDEMs both before the optimization and after the optimization using the 200 pixel trim size. The algorithm does very well for correcting the translation offset, especially when the EO image width is quite wide. Visually, the large translation offset is gone, and there are several minor adjustments to the other swaths making it visually appealing. As the quantities in Table 1 suggest, these resultant TDEMs are accurate to within the accuracy of the texel camera, and more accurate as the number of adjacent swaths seen increases.

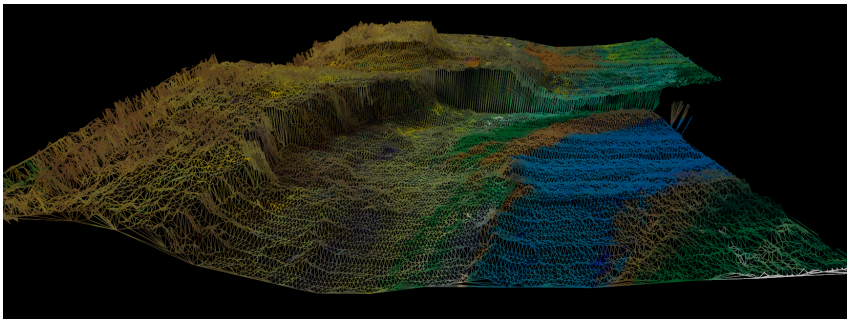
4.3.2 Turbulent Flight

This data set consists of moving the camera a few millimeters for each capture approximately one meter away from the model landscape and contains 177 texel swaths. The turbulence consists primarily of "roll" relative to the landscape. Figure 5 shows the approximate amount of roll present in the data set. Considering the camera field of view is approximately 40°, the nearly 30° peak-to-peak roll is very significant. The roll is not intended to exactly replicate the real-world roll of a small UAV; the roll is occurring at a fast rate compared to the forward motion of the camera and, thus, is an extreme case demonstrating the robustness of the algorithm.

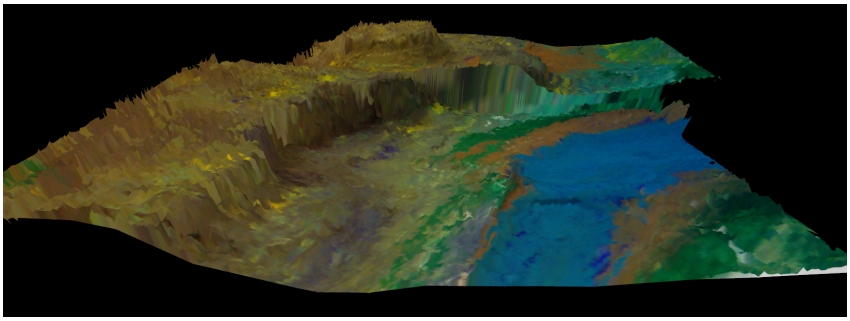
The parameters used for this optimization are the same as for Section 4.3.1. The results of these experiments are shown in Table 2. The errors in this experiment are greater than in the level flight experiment, however,

Trim Size (pixels)	**Error Mean** (mm)	**Error Std Dev** (mm)	**RMS Error** (mm)	**Comments**
Before Registration	-8.32	28.4	28.1	Contains significant roll
260	7.48	8.83	11.5	Good registration
240	7.86	8.67	11.6	Good registration
220	4.94	12.0	12.9	Good registration locally, but noticeable errors
200	4.71	12.6	13.4	Good registration locally, but noticeable errors
180	1.31	15.1	15.0	Good registration locally, but noticeable errors
160	1.05	13.1	13.0	Good registration locally, but noticeable errors

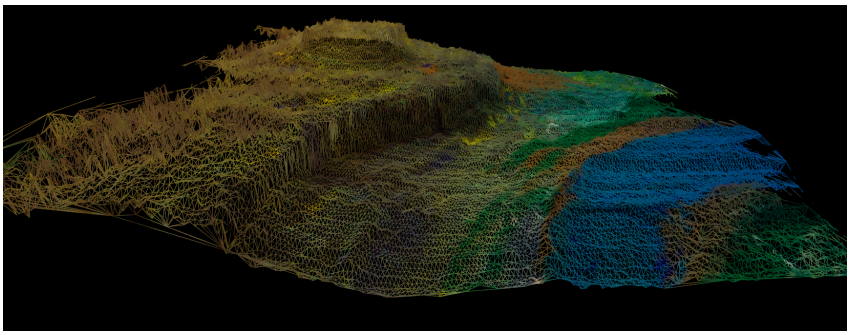
Table 2: Error for Various EO Widths for Turbulent Flight Data Set



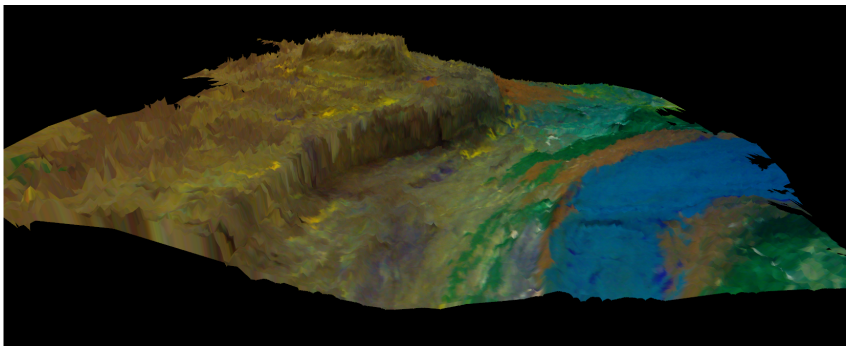
(a)



(b)



(c)



(d)

Figure 4: Smooth Flight Data Set Registration. (a) Wire-frame before optimization. (b) TDEM before optimization. (c) Wire-frame after optimization. (d) TDEM after optimization.

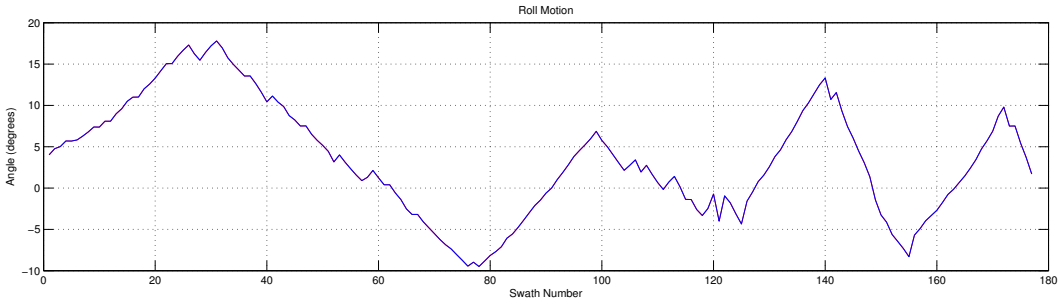


Figure 5: Roll profile for the turbulent flight data set.

the proposed registration method significantly improved the registration from the case using only the measured translation and IMU data.

Figure 6 shows the TDEMs both before and after the optimization, using the 240 pixels wide trim size. The quantities in Table 2 show that the registration significantly improves the distance error variance from the case using only the GPS/IMU data, despite the presence of the camera roll. There are, however, artifacts from the roll, especially when the trim width is small. These artifacts are the "vertical" movement of the swaths (which do not reflect reality), making the flat parts of the landscape look "bulged." However, the overlaid texture looks good, which suggests the projection points from NCC contained some errors.

5. CONCLUSION

The major contribution of this paper is the cost function presented in Section 3 for texel swaths, and the image processing method to select the points used in the cost function. This cost function enables the Sparse LMA technique to determine the best location of 3D points and camera locations in terms of minimum squared-error.

The quantitative and visual analysis of the method shows promising performance for the intended scenario. The method was able to create TDEMS by stitching together swaths of texel image gathered as the simulated small UAV "flew" across the model terrain. The accuracy of the resulting TDEM was very good, with less distance error in the level flight simulation than the native accuracy of the range sensor.

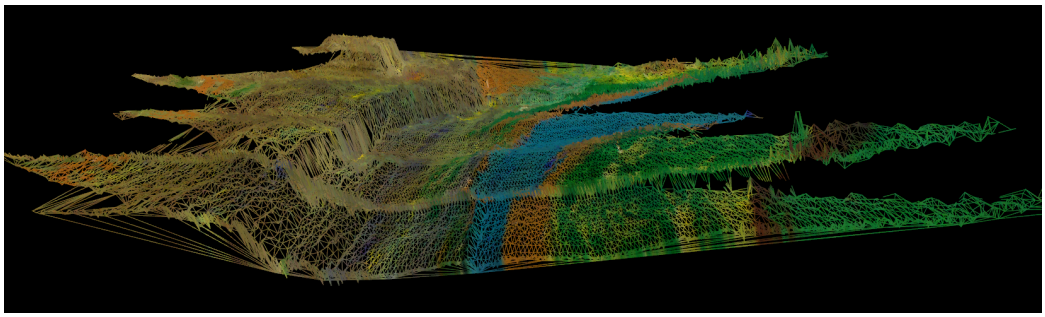
The test using an extreme turbulence model also showed promise, with significant improvement over the TDEM created from the raw IMU and translation measurements. Further refining of the proposed method should lead to better performance in the presence of motion from turbulence.

Limitations This method assumes there are very small changes from one image to the next, and simple image processing techniques can be employed. This method also assumes the texel swaths are approximately ordered by their position on the ground. It also assumes the scene is nearly planar, as a homography is used to determine both the approximate location of projected points, as well as in the stitching process for the final texture.

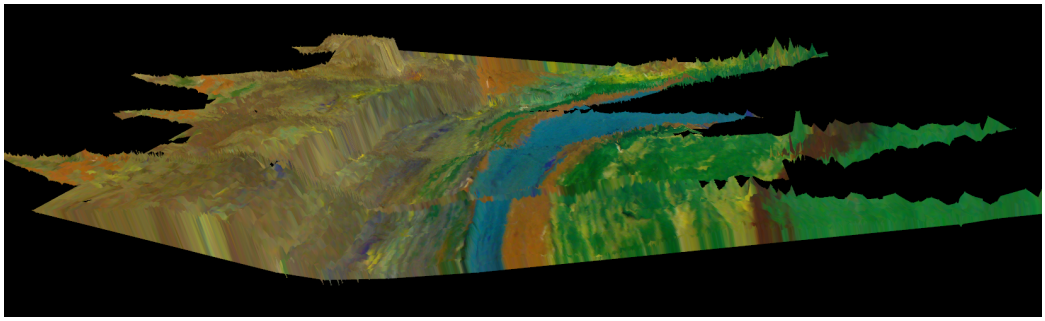
It also relies heavily on the accuracy of the points gathered using NCC. These points are susceptible to noise, perspective difference, and parallax. Techniques must be developed to ensure these points are accurate so outliers are not introduced into the optimization.

Future Work In this technique the original camera locations and point locations (relative to the associated camera) are not considered during or after the optimization process, apart from being used as a seed. This means additional knowledge could be incorporated into the optimization in addition to the methods presented in this paper.

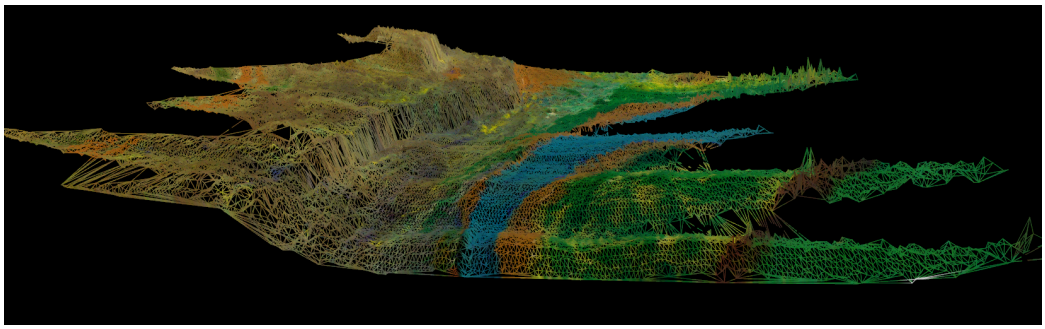
Further work may also include optimizing the 3D location of points that do not have distance measurements; in other words, a hybrid between the method described in this paper solely using imagery around ladar points



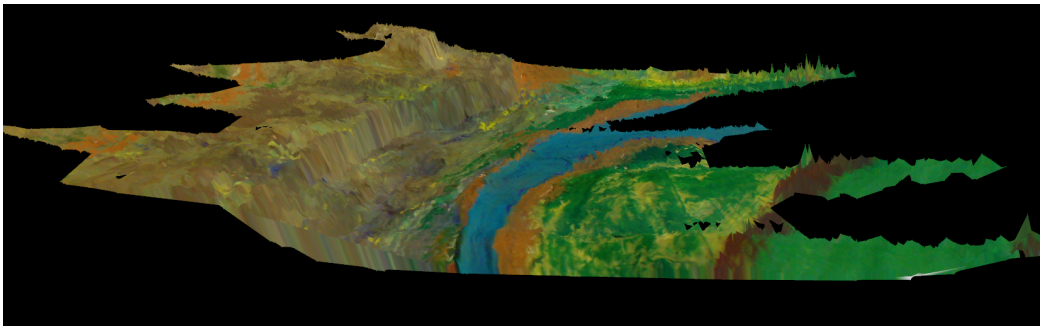
(a)



(b)



(c)



(d)

Figure 6: Turbulent Flight Data Set Registration. Notice the significant roll motion. (a) Wire-frame before optimization. (b) TDEM before optimization. (c) Wire-frame after optimization. (d) TDEM after optimization.

and existing methods solely using imagery. Texturing the point cloud needs to be further investigated, including incorporating super-resolution techniques in the texture creation. Nonlinear image mapping may also be employed to mitigate the planar-assumption effects using a homography.

In addition, more reliable image processing techniques can be used to ensure more accurate, or even sub-pixel accuracy, of matching points in the images. Other research areas may include using surveyed "ground control points" in the optimization to ensure real-world accuracy.

Acknowledgments The use of the Armadillo C++ Matrix Library is greatly appreciated as it enabled ease of development with high computational efficiency in the algorithm.[15] We acknowledge the Space Dynamics Laboratory at USU for funding this research. We acknowledge the Utah Water Research Laboratory at USU for providing insight to the end goals of the TDEM creation.

REFERENCES

[1] Schut, G., "An analysis of methods and results in analytical aerial triangulation," *Photogrammetria* **14**(0), 16–33 (1957–1958).

[2] Agouris, P. and Schenk, T., "Automated aerotriangulation using multiple image multipoint matching," *Photogrammetric Engineering and Remote Sensing* **62**(6), 703–710 (1996).

[3] Hartley, R. and Zisserman, A., [*Multiple view geometry in computer vision*], Cambridge university press (2003).

[4] Vallet, J., Panissod, F., Strecha, C., and Tracol, M., "Photogrammetric Performance of an Ultra Light Weight Swinglet UAV," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **3822**, C253 (Sept. 2011).

[5] Ahmad, A. and Karimian, H., "Performance of aerial triangulation based on airborne GPS data," *Faculty of Geoinformation & Real Estate, University Technology Malaysia* (2006).

[6] Schenk, T., "Digital aerial triangulation," *International Archives of Photogrammetry and Remote Sensing* **31**, 735–745 (1996).

[7] Maier, R., Sturm, J., and Cremers, D., "Submap-based bundle adjustment for 3D reconstruction from RGB-D data," in [*Pattern Recognition*], 54–65, Springer (2014).

[8] Budge, S. E. and Badamikar, N. S., "Calibration method for texel images created from fused lidar and digital camera images," *Opt. Eng.* **52**, 103101 (Oct. 2013).

[9] Harris, C. and Stephens, M., "A combined corner and edge detector.," in [*Proc. of Fourth Alvey Vision Conference*], 147–151 (1988).

[10] Noble, J. A., *Descriptions of image surfaces*, PhD thesis, University of Oxford (1989).

[11] Fischler, M. A. and Bolles, R. C., "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM* **24**(6), 381–395 (1981).

[12] Horn, B. K., "Closed-form solution of absolute orientation using unit quaternions," *JOSA A* **4**(4), 629–642 (1987).

[13] Konolige, K. and Agrawal, M., "Frameslam: From bundle adjustment to real-time visual mapping," *IEEE Trans. Robot.* **24**, 1066–1077 (Oct 2008).

[14] Fisher, P. F. and Tate, N. J., "Causes and consequences of error in digital elevation models," *Progress in Physical Geography* **30**(4), 467–489 (2006).

[15] Sanderson, C., "Armadillo: An open source C++ linear algebra library for fast prototyping and computationally intensive experiments," (2010).