

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

Targeted 3D modeling from UAV imagery

R. Abraham Martin, Benjamin K. Heiner, John D. Hedengren

R. Abraham Martin, Benjamin K. Heiner, John D. Hedengren, "Targeted 3D modeling from UAV imagery," Proc. SPIE 10645, Geospatial Informatics, Motion Imagery, and Network Analytics VIII, 106450D (27 April 2018); doi: 10.1117/12.2304436

SPIE.

Event: SPIE Defense + Security, 2018, Orlando, Florida, United States

Targeted 3D Modeling from UAV Imagery

R. Abraham Martin^a, Benjamin K. Heiner^b, and John D. Hedengren^a

^aBrigham Young University, 350 Clyde Building, Provo, UT, USA

^bAFRL RYAT, Wright Patterson AFB, OH, USA

DISTRIBUTION STATEMENT A: Approved for public release. APRS-RY-17-1031

ABSTRACT

Reconstruction of 3D objects from UAV EO imagery yields useful information, but can be time consuming and computationally expensive. View planning reduces processing time by selecting the optimal image set needed to reconstruct a scene. This paper demonstrates how view planning is used in a targeted manner to select a subset of images from a large existing image set to model specific vehicles or structures. Potential applications of the method include enabling 3D target classification algorithms and rapid geo-location. The method could also facilitate on-board reconstruction. The view planning algorithm is tested on five different targets, and is shown to reduce processing time for target models by up to a factor of 50 with little decrease in accuracy.

Keywords: ATR, view-planning, Structure-from-Motion, SfM, UAV

1. INTRODUCTION

The rise of Unmanned Aerial Vehicles (UAVs) allows the capture of large quantities of aerial visual data. UAVs are highly mobile and portable sensor platforms, and are increasingly popular in applications in agriculture, military, mining, emergency services and others. Maturing technologies in the area of computer vision allow increasing amounts of valuable information to be extracted from aerial imagery. In particular, techniques such as Structure from Motion (SfM), and Multi-View Stereo (MVS) make it possible to generate 3D models from 2D UAV imagery.¹ However, these techniques are computationally expensive, and scale poorly with increasing numbers of images. View planning helps to alleviate these problems by automatically selecting the images needed to reconstruct a target object, thus reducing the required processing time and resources. This paper demonstrates a targeted view planning approach in which the best images to reconstruct specific vehicles or structures are selected from a large set of previously collected images. This differs from previous work by the authors in which the view planning problem is solved prior to image collection.²

2. RELATED WORK

View planning is the problem of finding optimal sensor positions to cover an object or scene. It is a well-studied field, and formulations have been developed for a number of different sensors including EO,³ sonar⁴ and laser range sensors.⁵ Advances in 3D reconstruction from monocular cameras along with the high mobility of UAV systems have made UAV mapping, surveying, and surveillance a ripe area for application of view planning, and many researchers have addressed the topic in recent years.⁶⁻¹⁰ View planning is an NP-Complete problem,¹¹ meaning that no efficient exact solution method is known, but several approximation methods such as greedy search or linear relaxation have been used by researchers to find acceptable solutions.

Though much of the view planning literature has roots in the inspection of machine parts and other specific objects,^{12,13} most UAV related work in the area has focused on view planning for an entire scene.¹⁴ For example, Trujillo et al. present a method for UAV flight planning to map areas with multiple structures and areas of

Further author information: (Send correspondence to B.K.H.)

R.A.M.: E-mail: abe.martin@byu.edu, Telephone: 1 435 592 1826

B.K.H.: E-mail: benjamin.heiner@us.af.mil, Telephone: 1 937 713 4113

J.D.H.: E-mail: john.hedengren@byu.edu

vegetation,⁸ and Cheng et al. treat the case of 3D coverage of urban regions.⁷ Cases where single specific objects are targeted for reconstruction are generally limited to close up inspections, such as that proposed by Liu et al.,¹⁵ or that shown by Martin et al.,² in which view plans are generated for modeling infrastructure objects.

The view planning approach used in this paper belongs to a family of sampling-based methods, where a large number of candidate image positions are generated and a minimal subset is selected.¹⁶ This is similar to the method employed by Schmid et al. They likewise take the approach of planning a desired image set for an area by referencing a rough digital surface model (DSM).¹⁷

Bircher et al. present research similar to the current results.¹⁸ They calculate inspection paths for a structure or object based on CAD models. They also demonstrate re-planning from a previously created model, and extend their algorithm to both fixed wings and quad-rotors. Their approach focuses on minimizing the length of the path taken in the inspection, and does not attempt to minimize the number of images required.

This paper makes the following contributions to the state of the art:

- Targeted view planning for modeling specific vehicles and structures
- Minimal image subset selection from existing large image sets

A description of the workflow, data, and algorithms used follows, along with a presentation of the results obtained.

3. METHODS

Selecting a set of images for a given location is done in several steps. The general workflow is as follows:

1. Load rough terrain data and input images
2. Calculate image/terrain visibility matrix
3. Greedily select a minimum image subset
4. Generate a 3D model from the image subset

These steps are described in more detail in the following subsections.

3.1 Data Input and Conditioning

The first step in the workflow is to load and condition the required UAV imagery and terrain data for the location of interest.

3.1.1 Load Input Data

As with all model based view planning approaches, a reference model of the target area is required. All tests performed in this paper use Level 2 (30m) DTED elevation maps from the publicly available SRTM data sets as a rough reference for the area of interest. As DTED data is generally available in the MSL elevation system, coordinates must be transformed to WGS84 to match image GPS data. Imagery is also loaded, along with image meta-data. The algorithm requires that image latitude, longitude, elevation, roll, pitch, yaw, size and focal length be known.



(a) Overview of flight area containing 1637 images of the Muscatatuck Urban Training Center.



(b) KML selection of targeted reconstruction area highlighted in bright green.

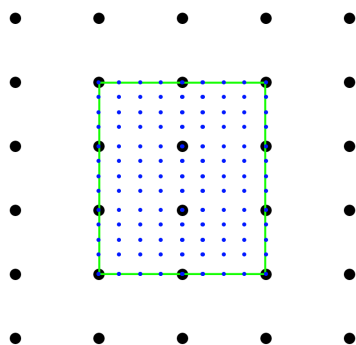
Figure 1. Viewing angles are sorted into histogram bins according to their azimuth and elevation.

3.1.2 Pre-process Terrain

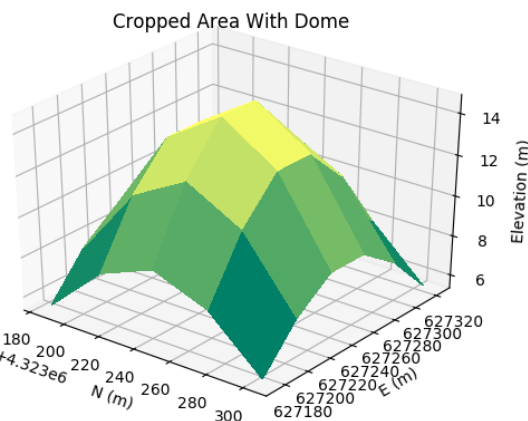
A key feature of this implementation is the ability to plan image sets to reconstruct a target vehicle or structure rather than an entire scene, such as shown in Figure 1a. To facilitate this, the input terrain data is cropped to using either a user defined KML (Figure 1b), or with four corners information from one of the input images.

It is often the case that the terrain data is sparse enough and the target area small enough that the target area contains very few or no terrain points. In this case, the density of the terrain data is increased by first cropping to an area slightly larger than the target, interpolating with a 2D cubic spline, and then cropping the denser terrain to the desired area (Figure 2a). This preserves useful elevation and slope information for small areas.

For a small target such as a vehicle or structure, the sparsity of the available terrain data often leads to insufficient oblique views being selected. In the absence of detailed geometry information about the target, augmenting the DTED data with a dome formed from a 2D Gaussian centered at the target area helps to promote side views, as shown in Figure 2b.²



(a) Illustration of DTED elevation map interpolation to improve terrain density.



(b) Example of 2D Gaussian dome used to promote oblique views of the target area.

Figure 2. Raw DTED elevation data is preprocessed to improve view planning results.

3.2 View Planning Algorithm Description

Once data input is complete, the actual process of view planning begins. The algorithm used contains three sections, visibility calculations, angle sorting, and image selection.

3.2.1 Visibility Calculations

To determine the relationship between the images and the terrain points, the algorithm employs a visibility matrix similar to the 3M matrix introduced by Scott.¹¹ For each image, 3D frustum culling is performed to calculate which terrain points are visible in the image. Hidden point removal¹⁹ is used to eliminate points obscured by occlusions. This produces an $m \times n$ binary visibility matrix V , where m is the number of images, and n is the number of terrain points.

High levels of image overlap are desirable for 3D reconstruction. To this end, in addition to checking for a binary point visibility, the visible points are also scored according to a 2D Gaussian distribution (Equation (1)) centered on the midpoint of the image. The closer to the center of the image each point is, the higher the score it receives for that image. This encourages images with the target area closer to the center of the image. Since more images are centered on the relatively small target area, this naturally improves overlap between images. An illustration of the concept is shown in Figure 3.

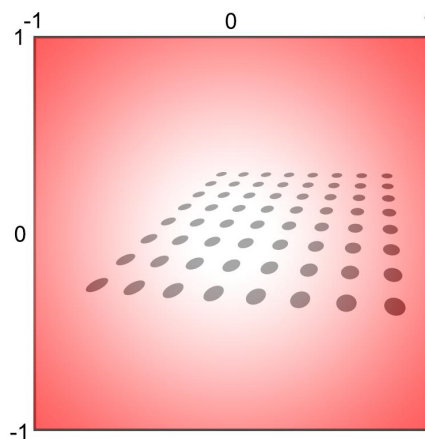


Figure 3. A Gaussian scoring function is used to increase image overlap by promoting well centered images.

$$G_i = \exp \left[-\frac{x_i^2}{2(0.4)^2} + \frac{y_i^2}{2(0.4)^2} \right] \quad (1)$$

Another issue encountered is that the algorithm always chooses images that are zoomed out as far as possible, because these images cover the most terrain points. However, this does not account for the additional information and detail that can be gained from high zoom imagery. At the same time, simply choosing high zoom imagery is not enough because the imagery may not cover the target area, or cover very little of the target area. To correct for this, each image was evaluated to determine the spatial distribution of visible points within the image. The image was divided into 25 zones. Each zone was checked to determine whether it contained a visible point. The image was then scored according to the percentage of zones that included a point. This 'point spread' ratio W was then used as a weighting factor when selecting the final image set. The rational is that in zoomed out imagery or off target imagery, the target area will cover little of the total image, while in high zoom on target imagery, the target area will cover much of the image. This is illustrated in Figure 4, which would receive a point spread score of 12/25, or 0.48:

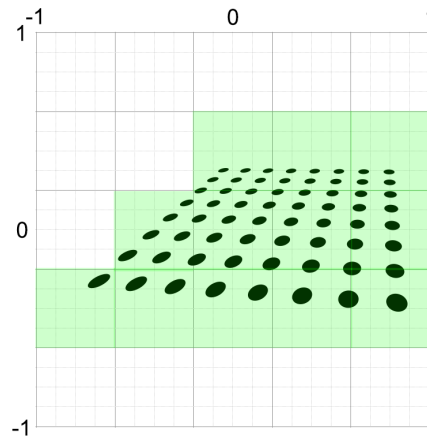


Figure 4. Images are prioritized for focal length by weighting by the spatial distribution of target points in the image plane.

3.2.2 Angle Sorting

Following visibility calculations the visibility matrix is expanded by including viewing angles. For each image, the angle between the image pointing vector and the surface normal at each visible point is calculated, as well as the heading angle. These are termed the elevation and azimuth viewing angles respectively. The calculated angles are then sorted into bins, three elevation and five azimuth, as shown in Figure 5a and Figure 5b. The visibility matrix is filtered and expanded based on which points are visible from each angle bin, going from an $m \times n$ matrix V to an $m \times (n \times b)$ matrix $\bar{V}$, where b is the total number of bins. The expanded matrix $\bar{V}$ is termed the augmented visibility matrix. Each row of the augmented visibility matrix now represents the terrain points that are visible from the viewpoint of a specific image, and which angle range they are visible from.

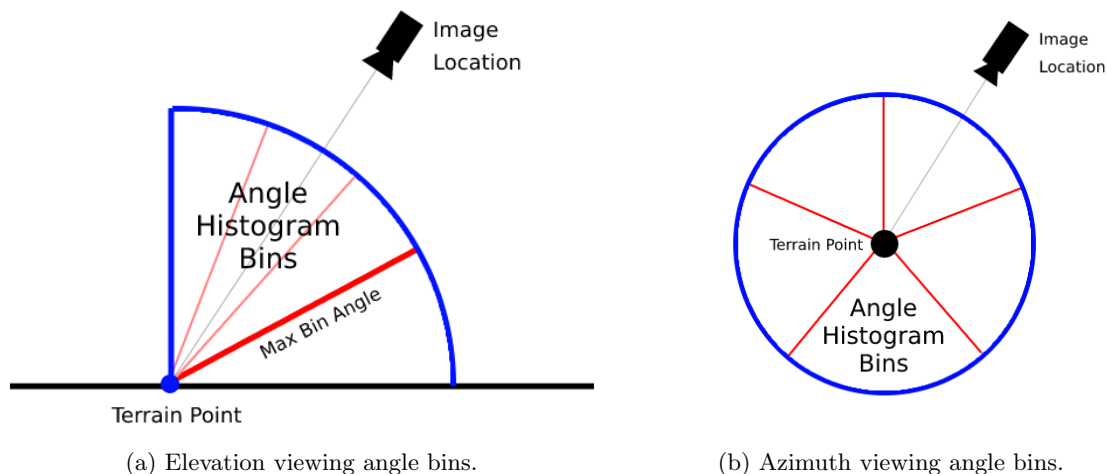


Figure 5. Viewing angles are sorted into histogram bins according to their azimuth and elevation.

3.2.3 Image Selection

Following angle sorting, a minimum subset of images is selected. First the visibility of the original image set is evaluated. Then each image in the original image set is scored according the formula given in Equation (2).

$$S_j = W_j \sum_{i=1}^{n \times b} v_{j,i} G_{j,i} \quad (2)$$

Where $j = 1 \dots m$, $i = 1 \dots n \times b$, $\bar{V}$ is the augmented visibility matrix, G is the Gaussian overlap score, and W is the spatial distribution weighting.

The highest scoring image S_{max} is added to the selected subset. The corresponding row, and the columns covered by that row are eliminated from the visibility matrix. The images are then re-scored, and the new highest scoring image is selected. This is repeated until the selected subset covers the same number of terrain points from the same number of angles as the original set, minus some user defined tolerance.

4. EXPERIMENTAL SETUP

4.1 Test Data

Experiments are performed using data collected from the Muscatatuck Urban Training Center as part of the AFRL SUSEX program. The data consists of 16 MP EO imagery, together with metadata for each image, including latitude, longitude, elevation, roll, pitch, yaw, focal length, and pose bias for sensor/platform alignment. The data are publicly available and can be downloaded at https://s3.amazonaws.com/sdms-susex-mutc-2015/24_Sept_2015_WAMI_Flight_1.tar.gz. The full image set is processed using Pix4D to produce a baseline model for the experiments.

Five target areas are selected for testing. These include several vehicles and structures of varying sizes. Each target area is defined using a KML polygon in Google Earth. Figure 6 shows the fourth target area, several buses in a parking lot.



Figure 6. Fourth target area, several buses in a parking lot, highlighted in white.

4.2 Procedure

Each target area is processed using the view planning algorithm, and a subset of the original imagery is selected for each. Each subset of imagery is processed in Pix4D to produce a point cloud model of the target. All models are created using the settings detailed in Table 1.

The time required for processing each subset model is recorded, and the accuracy of each subset model is evaluated using CloudCompare,²⁰ a commonly used point cloud analysis tool, and then the two point clouds are finely aligned using the iterative closest point algorithm. The distance between each point in the subset model

Table 1. Pix4D Program Settings

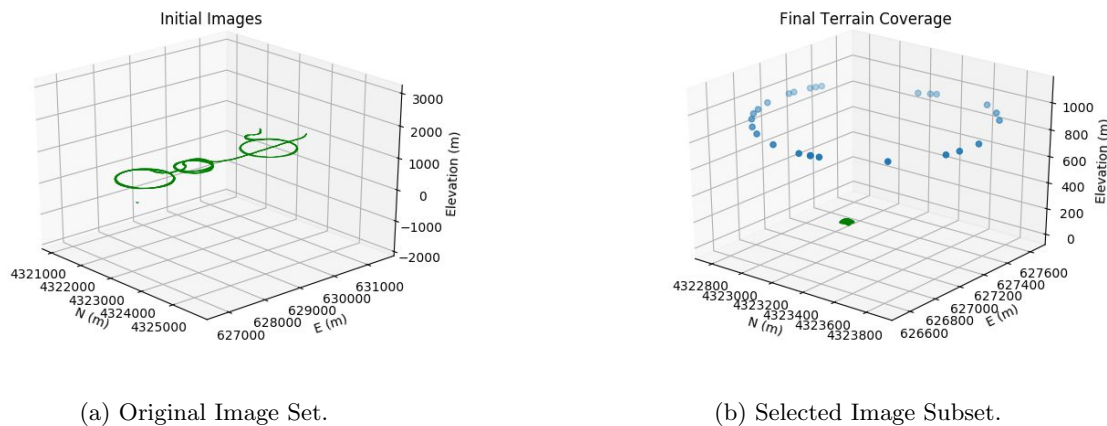
Parameter	Value
Keypoints Image Scale	Full
Matching Image Pairs	Free Flight
Image Scale	1/2
Point Density	Optimal
Minimum Matches	3

and its nearest neighbor in the baseline model is computed, and this is used to find the mean error of the subset model relative to the baseline.

5. RESULTS AND DISCUSSION

This section presents the results of the experiments performed in the previous section. Graphical results of model quality and tabular results of processing time and model accuracy are included for all five target areas. The processing times computed are on a laptop with a quadcore 2.50 GHz i5 processor and 8 GB of RAM.

Figure 7a shows the locations of the 1637 images in the original Muscatatuck image set. The full image set covers the entire flight area, which is shown in Figure 1a. These images were collected by a fixed wing UAV platform sampling at a high rate, resulting in a high density of images along the flight path. Processing the entire image set is possible, but computationally expensive, as will be shown.



(a) Original Image Set.

(b) Selected Image Subset.

Figure 7. The initial image set from Muscatatuck flight contains 1637 images and covers the entire flight area, while the selected image subset for target area 4 contains only the 29 images needed to reconstruct the target.

Figure 7b shows the locations of the 29 images in the image subset chosen to reconstruct target area 4. Notice that the algorithm selected images from the loop on the left side of Figure 7a, which circles this particular target. It can also be seen that the algorithm selected images which view the target area from a variety of angles, and that only a small fraction of the original images are needed to reconstruct the target. Image overlap is also preserved.

Figure 8 shows the point clouds of the five target areas produced from the original images and those produced from the chosen image subsets. Note that the subset models have been cropped to show only the target areas. In most of the cases the target point clouds are visually very similar to the originals, and in some cases nearly identical. In each case however, far fewer images and far less time are required to produce the subset model. At this point, the produced point clouds could be cropped further and used for ATR applications.

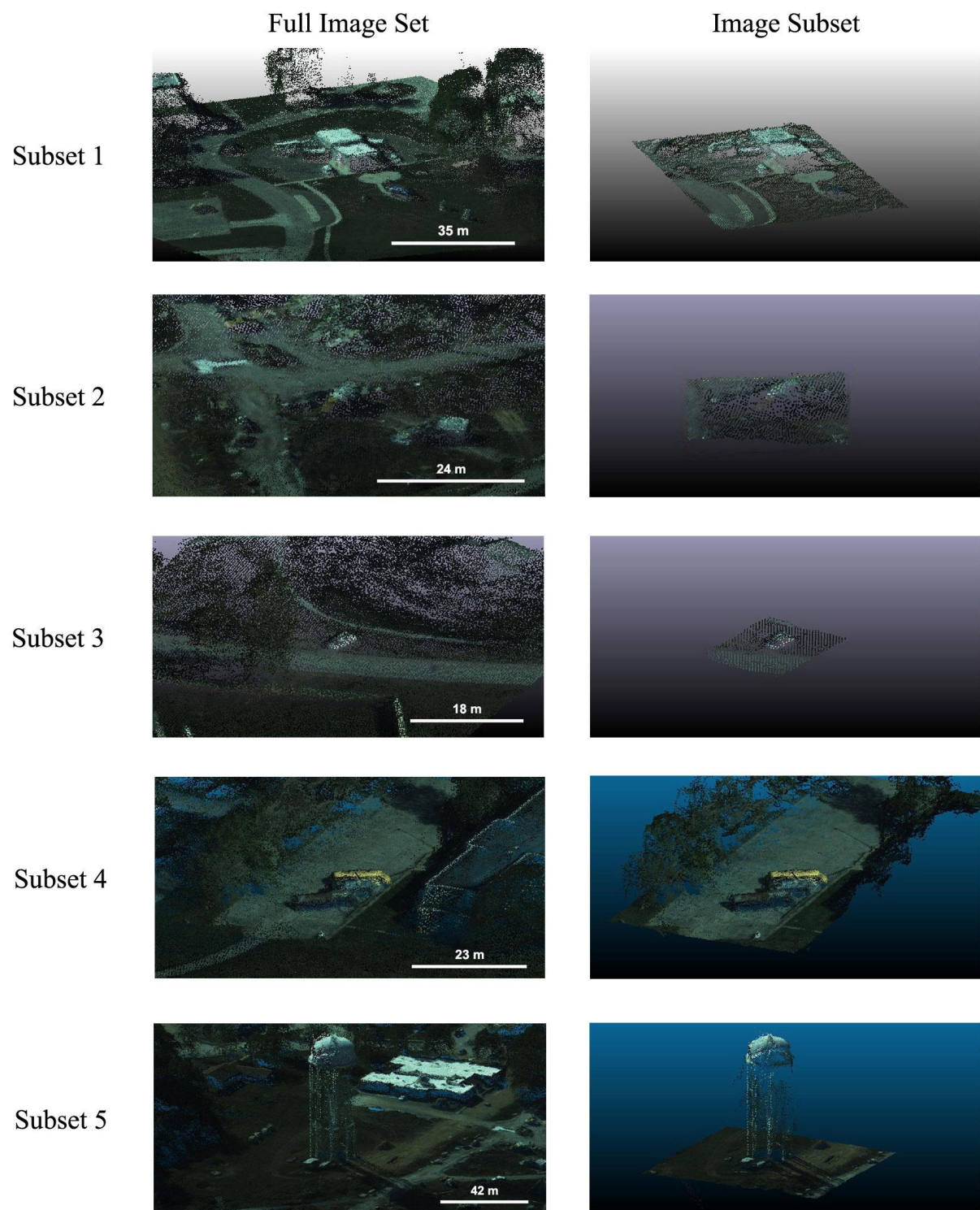


Figure 8. Point clouds for the five test areas reconstructed from the full image set (left) and the selected image subsets (right).

An example of the the cloud to cloud distance comparison between the original and subset model is shown in Figure 9 using the results from area 4. As can be seen, the majority of the error is found in the trees and vegetation at the edges of the scene. This is to be expected, as vegetation is generally difficult to model accurately using SfM techniques.

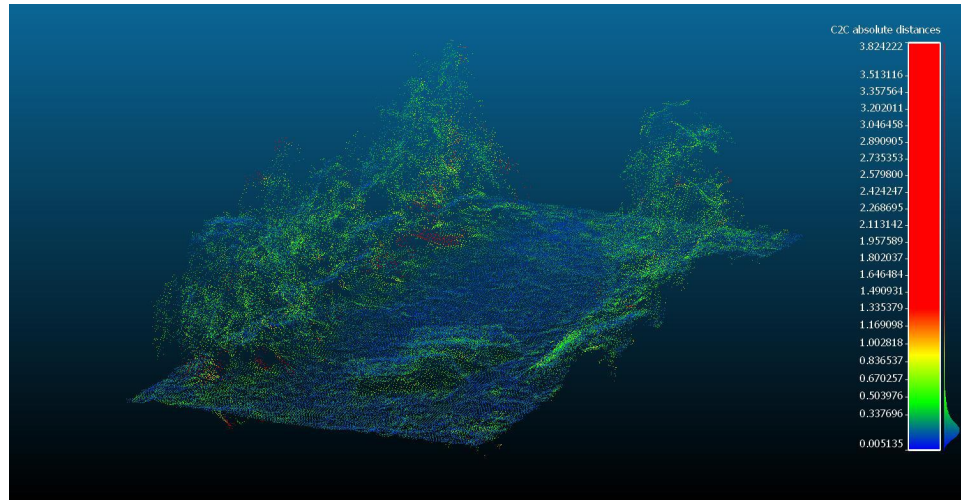


Figure 9. Model accuracy was determined by measuring the average cloud to cloud distance between original and subset models. Subset model 4 (three buses) is shown here.

Results for the processing time required to reconstruct each target area are summarized in Tables 2 and 3. In all cases, the time needed for reconstruction was reduced drastically from the original image set. This is an important result, as it leads to a much quicker turn-around time for 3D classification applications. These results also imply that if longer processing times are acceptable the subset models could be processed effectively on lower cost and quality hardware. This result is promising for applications such as 3D model reconstruction on-board a UAV using limited resources. This in turn could lead to further avenues of research such as producing real-time target models on-board, and using the produced models to re-plan a new flight path to capture improved imagery of the target.

Table 2. Processing Time Results

Image Set	Target	Images	Time (min)
Original	N/A	1637	3016
Subset 1	Building with Vehicles	29	72
Subset 2	Delivery Truck	11	10
Subset 3	White Van	20	17
Subset 4	Three Buses	29	58
Subset 5	Water Tower	28	35

Accuracy results consisting of the RMS error between the full image set point cloud and the subset image set point clouds for the test sites are reported in Table 3. Mean errors were generally on the order of 20 cm when compared to the original models, which was acceptable for the level of detail available in the original images. The authors anticipate that higher quality input imagery would produce an even lower level of error.

6. CONCLUSION AND FUTURE WORK

This paper presents a targeted view planning algorithm used to select minimal image subsets from large aerial image datasets to model specific vehicles and structures on the ground. DTED information for the area is loaded, and KML files are used to select a target area. Target visibility is evaluated for each image, and a subset of

Table 3. Model Accuracy Results

Image Set	RMS Error (m)	Standard Deviation (m)
Original	0	0
Subset 1	0.25	0.15
Subset 2	0.28	0.18
Subset 3	0.20	0.10
Subset 4	0.28	0.22
Subset 5	0.30	0.43

images is selected for 3D processing. The algorithm is tested on five separate target areas, and the resulting models are compared against those from the full image set. It is found that the number of images and the required processing time can be greatly reduced with only small losses in model accuracy.

Future work in this area includes several potential applications of this technology. The reduced processing times achieved make SfM modeling a more viable option for use in target recognition algorithms. Iterative model reconstruction is another promising area, in which 3D models of the target are rapidly produced on-board and used to plan new flight paths to collect additional target imagery.

ACKNOWLEDGMENT

The authors would like to acknowledge Clark N. Taylor (AFRL) for providing valuable insight and direction for the project. This work was supported by the Air Force Research Laboratory Automatic Target Recognition Center (AFRL ATRC) Summer Program.

Appendix

Source code for the algorithms presented in this paper is publicly available at <https://github.com/AFRL-RY/Explore-Coarse-3D-Reconstruction-Path-Planning-ATRC-Summer-2017>.

REFERENCES

- [1] Lucieer, A., Jong, S. M. d., and Turner, D., "Mapping landslide displacements using structure from motion (sfm) and image correlation of multi-temporal uav photography," *Progress in Physical Geography* **38**(1), 97–116 (2014).
- [2] Martin, R. A., Blackburn, L., Pulsipher, J., Franke, K., and Hedengren, J. D., "Potential benefits of combining anomaly detection with view planning for UAV infrastructure modeling," *Remote Sensing* **9**(5), 1–20 (2017).
- [3] Zhao, J. and Sen-ching, S. C., "Optimal visual sensor planning," in [*Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*], 165–168, IEEE (2009).
- [4] Englot, B. and Hover, F., "Planning complex inspection tasks using redundant roadmaps," *Proc. Int. Symp. Robotics Research*, 1–16 (2011).
- [5] Scott, W. R., Roth, G., and Rivest, J.-F., "View planning for automated three-dimensional object reconstruction and inspection," *ACM Computing Surveys* **35**(1), 64–96 (2003).
- [6] Nex, F. and Remondino, F., "UAV for 3D mapping applications: a review," *Applied Geomatics* **6**(1), 1–15 (2014).
- [7] Cheng, P., Keller, J., and Kumar, V., "Time-optimal UAV trajectory planning for 3D urban structure coverage," in [*2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*], 2750–2757 (2008).
- [8] Trujillo, M. M., Darrah, M., Speransky, K., Deroos, B., and Wathen, M., "Optimized flight path for 3D mapping of an area with structures using a multirotor," in [*2016 International Conference on Unmanned Aircraft Systems, ICUAS 2016*], 905–910 (2016).

- [9] Chen, S., Li, Y., and Kwok, N. M., [*Active vision in robotic systems: A survey of recent developments*], vol. 30 (2011).
- [10] Xiaowei, X., Zhengjun, L., and Zhiquan, Z., “Flight route Designing and mission planning Of power line inspecting system Based On multi-sensor UAV,” *IOP Conference Series: Earth and Environmental Science* **17**(1), 012192 (2014).
- [11] Scott, W. R., “Model-based view planning,” *Machine Vision and Applications* **20**, 47–69 (2007).
- [12] Tarbox, G. and Gottschlich, S., “Planning for Complete Sensor Coverage in Inspection,” *Computer Vision and Image Understanding* **61**(1), 84–111 (1995).
- [13] Tarabanis, K. a., Tsai, R. Y., and Allen, P. K., “MVP sensor planning system for robotic vision tasks,” *IEEE Transactions on Robotics and Automation* **11**(1), 72–85 (1995).
- [14] Martin, R., Rojas, I., Franke, K., and Hedengren, J., “Evolutionary View Planning for Optimized UAV Terrain Modeling in a Simulated Environment,” *Remote Sensing* **8**(1), 26 (2015).
- [15] Liu, C.-a., Dong, R.-f., and Wu, H., “Flying Robot Based Viewpoint Selection for the Electricity Transmission Equipment Inspection,” *Mathematical Problems in Engineering* (2014).
- [16] Jing, W., Polden, J., Lin, W., and Shimada, K., “Sampling-based view planning for 3D visual coverage task with unmanned aerial vehicle,” *IEEE International Conference on Intelligent Robots and Systems* **2016-Novem**, 1808–1815 (2016).
- [17] Schmid, K., Hirschmüller, H., Dömel, A., Grixia, I., Suppa, M., and Hirzinger, G., “View planning for multi-view stereo 3D Reconstruction using an autonomous multicopter,” *Journal of Intelligent and Robotic Systems: Theory and Applications* **65**, 309–323 (2012).
- [18] Bircher, A., Alexis, K., Burri, M., Oettershagen, P., Omari, S., Mantel, T., and Siegwart, R., “Structural Inspection Path Planning via Iterative Viewpoint Resampling with Application to Aerial Robotics,” *IEEE International Conference on Robotics & Automation* , 6423–6430 (2015).
- [19] Katz, S., Tal, A., and Basri, R., “Direct visibility of point sets,” *ACM Transactions on Graphics* **26**(3), 24 (2007).
- [20] CloudCompare, “Cloudcompare.” <http://cloudcompare.org/> (2017). Accessed: 2017-07-12.