Evaluation of image collection requirements for 3D reconstruction using phototourism techniques on sparse overhead data

Erin Ontiveros*, Carl Salvaggio, Dave Nilosek, Nina Raqueño and Jason Faulring
[a]Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY USA 14623

ABSTRACT

Phototourism is a burgeoning field that uses collections of ground-based photographs to construct a three-dimensional model of a tourist site, using computer vision techniques. These techniques capitalize on the extensive overlap generated by the various visitor-acquired images from which a three-dimensional point cloud can be generated. From there, a facetized version of the structure can be created. Remotely sensed data tends to focus on nadir or near nadir imagery while trying to minimize overlap in order to achieve the greatest ground coverage possible during a data collection. A workflow is being developed at Digital Imaging and Remote Sensing (DIRS) Group at the Rochester Institute of Technology (RIT) that utilizes these phototourism techniques, which typically use dense coverage of a small object or region, and applies them to remotely sensed imagery, which involves sparse data coverage of a large area. In addition to this, RIT has planned and executed a high-overlap image collection, using the RIT WASP system, to study the requirements needed for such three-dimensional reconstruction efforts. While the collection was extensive, the intention was to find the minimum number of images and frame overlap needed to generate quality point clouds. This paper will discuss the image data collection effort and what it means to generate and evaluate a quality point cloud for reconstruction purposes.

Keywords: point clouds, 3D reconstruction, phototourism, image collection, computer vision, structure from motion, WASP

1. INTRODUCTION

Phototourism is a method for utilizing and viewing image collections that has emerged from the application of computer vision techniques. It has allowed for the generation of point clouds and ultimately three-dimensional geometric reconstructions without the use of LIDAR data. Typical application of this technique is done using image collections generated from an internet search[1,2]. The application used in this paper however is the result of the work done here at Rochester Institute of Technology by Nilosek et al. (2012).[3] His implementation utilizes aerial imagery collected with our in house Wildfire Airborne Sensor Program (WASP) instrument. The ultimate goal is to be able to create a three-dimensional geometric model from the resulting three-dimensional point cloud.

Here in the DIRS group at RIT we have been considering how to optimize the generation of a three-dimensional point cloud from this workflow for three-dimensional geometry extraction. We then developed a test plan, which utilizes the WASP instrument and focuses on a complex urban area. From the resulting dataset we can determine what is the optimum imaging scenario for such extraction efforts.

This paper will discuss the development of the collection experiment for three-dimensional geometry extraction, photo tourism techniques applied to the data for point cloud generation, a study of how flight lines should be planned to optimize the resulting point cloud and finally, how the reader can get access to the imagery obtained in the collection experiment.

*ontiveros@cis.rit.edu; phone 1 585 475-4465; dirs.cis.rit.edu

Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII, edited by Sylvia S. Shen, Paul E. Lewis, Proc. of SPIE Vol. 8390, 83900K · © 2012 SPIE
CCC code: 0277-786X/12/$18 · doi: 10.1117/12.919319

2. COLLECTION EXPERIMENT FOR THREE-DIMENSIONAL GEOMETRY EXTRACTION

The ability to extract geometry from a three-dimensional point cloud is driven by the quality of the data set available. In order to have a chance at extracting three-dimensional information from a dataset there needs to be at least two images, preferably more, looking at each point in the scene. Otherwise holes or occlusions will occur. The researchers planning this collection had to plan for the highest overlap possible of imagery from multiple directions in order to prevent or minimize occlusions. To achieve this, overlap was planned into the flight lines as well as the acquisition rate, which is aircraft and sensor dependent.

2.1 Flight Planning

Typical large area coverage aerial flight plans follow a racetrack or raster pattern. These patterns are typically implemented to minimize overlap and maximize area coverage. For this work, more overlap means more points of reference and better three-dimensional extraction. While the traditional flight patterns with increased overlap may be sufficient, a third pattern was explored to see if it provided improved extraction and reconstruction. This alternative flight path, shown in Figure 1, will be referred to as the flower pattern.

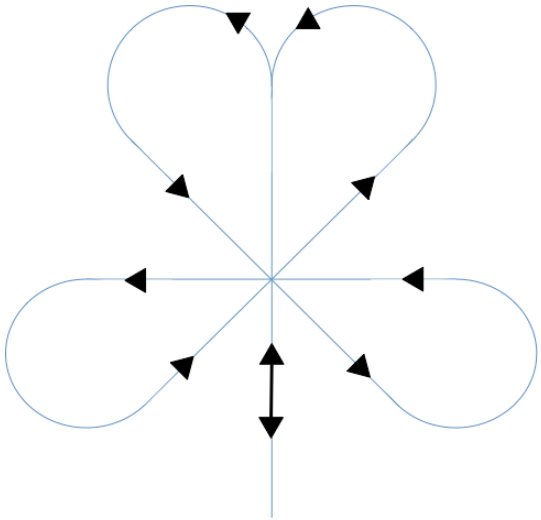

Figure 1. Alternative flower flight pattern to insure coverage of as many target facets as possible.

Our experiment looked at a complex urban area (high detail) using the above described traditional flight plans and the flower pattern in order to explore the limits of collection strategy and establish guidelines to guarantee quality three-dimensional extraction with minimal occlusions. The experiment was planned between the hours of 10 AM and 2PM in order to minimize shadows and activity at a nearby airport. The proposed parameters for the study are shown in Table 1.

Flying Height	4,600ft
Flying Speed	120kt
Resulting Forward Lap	75%
Side Lap	90%
Study Area Dimensions	3,800ft x 2,900ft (East-West x North-South)
Flight Lines	13 lines in the N-S direction 10 lines in the E-W direction 2 lines in the NE-SW direction 2 lines in the SE-NW direction (~300ft offset between lines)
Project Coverage Per Frame	3,400ft x 1,700ft
GSD	9.9 in

Table 1. Flight coverage study parameters for the field collection.

The collection was focused on the Rochester, NY downtown area, shown in Figure 2, centering on the buildings shown in Figure 3.

Figure 2. Overview of downtown Rochester, NY where the red line denotes the extent of the image set.

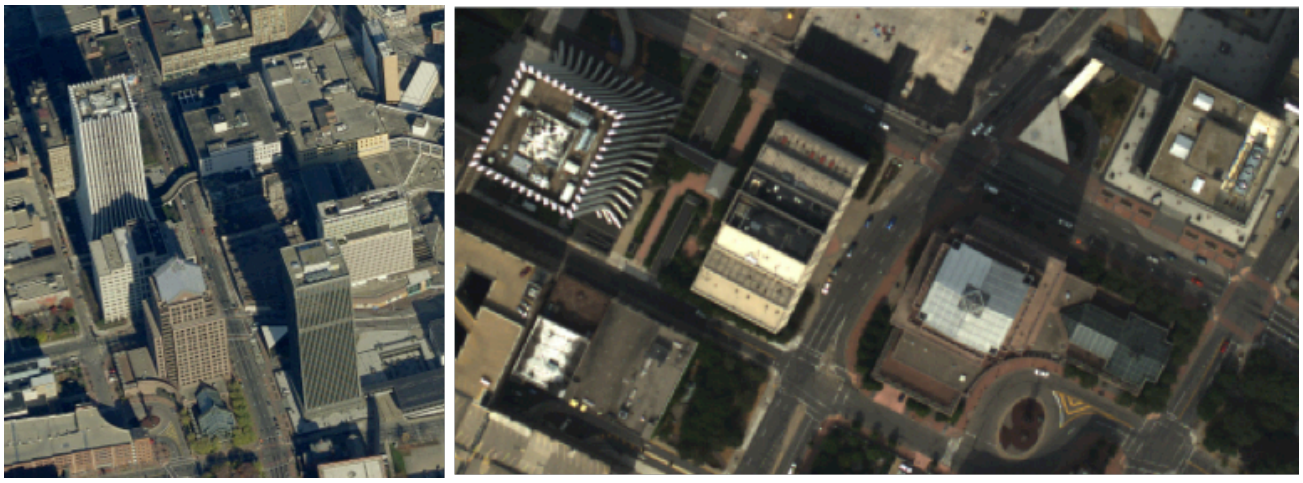

Figure 3. These images depict where the image collection was centered. The image on the right is the courtesy of Pictometry, the image on the left is a portion of a WASP image.

Kucera International, in addition to flying our WASP sensor for this effort, has also collected LIDAR data using their own Lecia AL60 LIDAR system on 1m postings over the study site. This paper will not discuss this data set other than to say it allows for direct comparison of the extracted data from these distinct modalities.

2.2 WASP Camera System

The WASP sensor, which was initially built for wildfire detection and mapping, has since evolved into a multipurpose sensor. It is a multispectral mapping system mounted on a Piper Navajo twin-engine airplane flown at 5000 ft. for this experiment. It has coverage in the visible (red, green, and blue), long-wave infrared (LWIR), mid-wave infrared (MWIR), and short-wave infrared (SWIR). Each infrared (IR) camera is an Indigo Phoenix infrared imager that has a 25 micron pixel pitch and a 25mm focal length, resulting in a 640x512 image. The visible camera is a Geospatial Systems KCM-11 and has a 9 micron pixel pitch and a 50 mm focal length, resulting in a 4000x2672 image. The sensor takes advantage of the Applanix POS/AV-310 navigation system that utilizes GPS and the Litton LN-200 inertial measurement unit to create ortho-rectified images in real time.[4,5]

It is worth mentioning however, that for this research ortho-rectified images were not used; only the pre-ortho-rectified visible images were used.

3. THREE-DIMENSIONAL DENSE POINT CLOUD GENERATION WORKFLOW

The workflow used for generating a three-dimensional dense point cloud is a cascade of various computer vision techniques, as shown in Figure 4, that take in a collection of images and output a three-dimensional point cloud. Nilosek et al. (2012)[3] gives a thorough discussion of this overall process, so I will just give a brief summary of each technique will be provided here.

Figure 4. The computer vision techniques used in this work, the output of which is a 3-D point cloud.

The image collection of interest is the input to SiftGPU[6], this finds invariant features, or keys, in the images that would have potential for matching with the same features in other images. These keys are matched within SiftGPU and then

the keys and the matching points are passed to bundler along with the input imagery. Bundler[7] uses this information to trace back the points to find the camera position geometry. This output from bundler goes into CMVS[8], where it gets optimized so that points are not overlooked and reorganized so that it can be processed in parallel using PMVS. PMVS[9] is the algorithm that ultimately generates the three-dimensional point cloud.

4. IMAGE COLLECTION REQUIREMENTS STUDY

Although we had very high overlap planned into our data collection effort, the ultimate intention of this work is to determine what is the minimum imagery needed to generate a quality point cloud.

4.1 Test Plan

This study started with using all of the imagery and all of the flight lines as input to the workflow described above. Then data was iteratively downsized to find the optimal image input scenario by systematically removing the input images and flight lines. Then once the optimal set of input images were found, the focus would switch to the input parameters to the workflow.

It was expected that the point clouds would be easily compared and contrasted and that the number of input images and output points in the point clouds would be proportional to the quality.

4.2 Results

Upon looking at the results, it became clear that it would be difficult to generate a quality metric based on these point clouds. Here few examples will be presented to demonstrate the complexity of evaluating these point clouds.

Interacting with the point cloud is the best way to see what is portrayed in it. This very difficult to do that in a two-dimensional format but that is the constraint of this medium. Also, a full point cloud of the entire downtown Rochester area, as shown in Figure 5, is too large to glean any information from. Only cropped portions of the point cloud that depicts the region shown in Figure 3, will be discussed here.

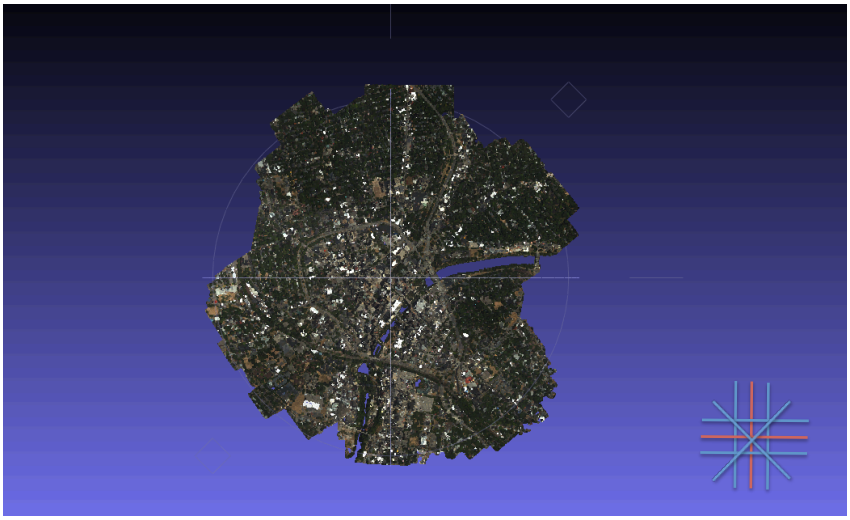

Figure 5. This image shows downtown Rochester as described by a point cloud output. The symbol in the corner shows that all images in all flight lines were used to generate this point cloud.

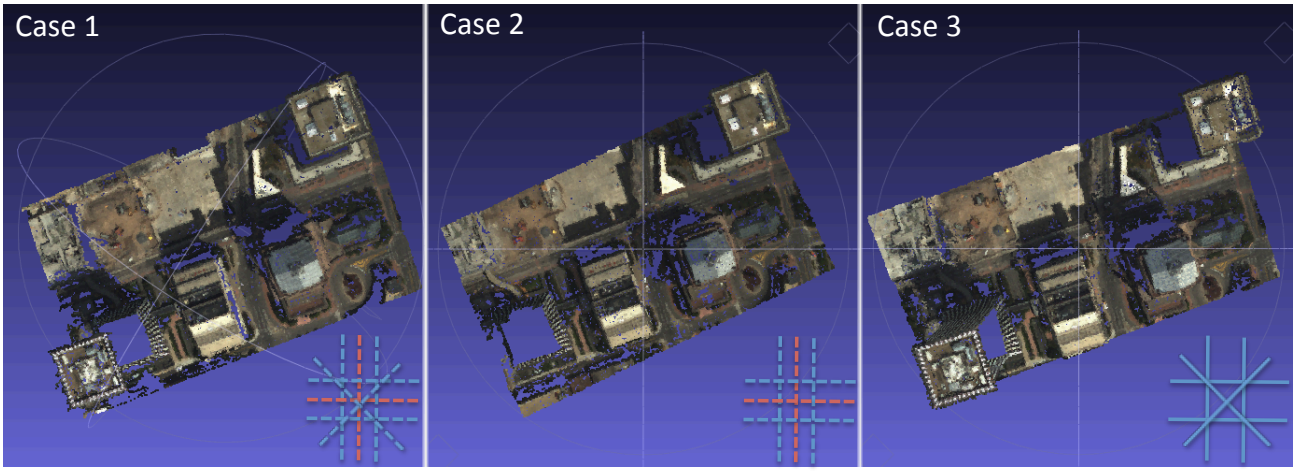

Figure 6. Three portions of point cloud case studies generated using different input images. The input images are notated by the symbols in the lower right corner. Case 1 was generated using every other image in all of the flight lines; case 2, every other image in the EW and NS flight lines, and case 3 all images in every other flight line.

Figure 6 shows three case studies of output point clouds, each having different image combinations as input. All have small holes that show up in different areas. However, there is a large hole in case 2 that isn't obvious at first glance, if you look closely you will notice the roof of the building in the lower left corner is completely missing. Also, interesting to note is that case 3 is the only one able to reproduce the sides of the building in the lower left corner.

For an even closer look at what the point distribution looks like, we can zoom in on just one of the roofs, shown in Figure 7. Again we are seeing a varying distribution of holes. Case 2 is the only case to clearly show a linear division on the upper portion of the roof. Also, the aerial photo in Figure 3 clearly shows five rectangular objects on the roof; one is centered on the roof, two on the left hand side of the roof, one below the middle rectangle and finally one that runs parallel to the middle rectangle on the right. In all cases we can glean the center rectangle, although it is more difficult in case 3. The two rectangles on the right side of the roof show up in all cases but the one underneath and along the right hand side of the middle rectangle are less defined in cases 1 and 3.

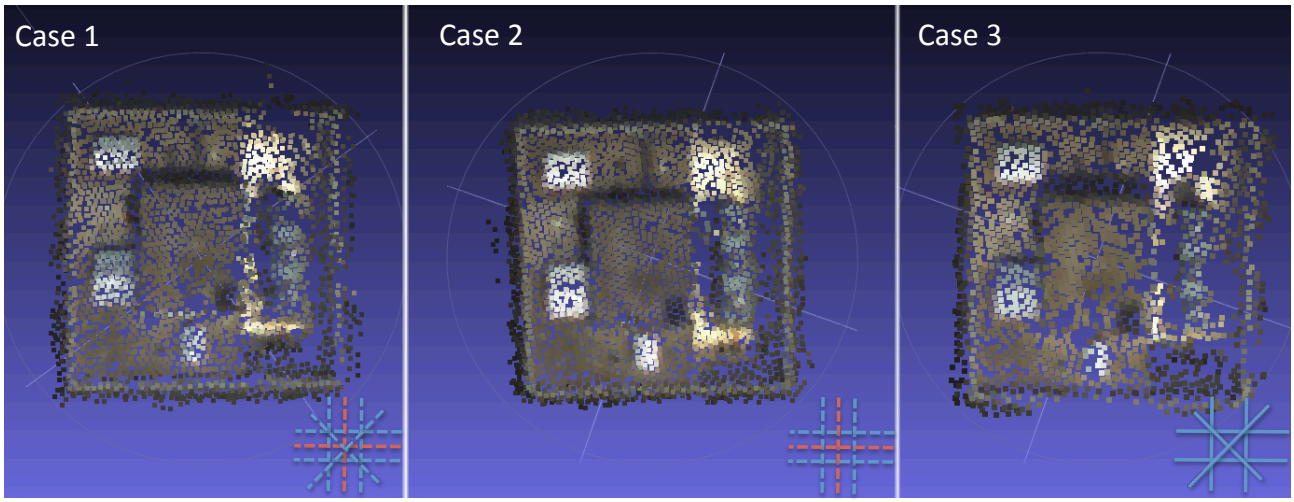

Figure 7. This shows a roof of one of the buildings to better show the output point distribution.

After looking at more results like this, we decided that I may need to scale back the study and rather than consider the whole dataset at once, would hand pick some images and see what the results look like. Again focusing on the same region shown in Figure 3, choosing first eight images, two from each directionality (NS, EW, NESW, NWSE), after

looking at those results adding one more image. The image boundaries for the handpicked eight and nine image cases are shown in Figure 8.

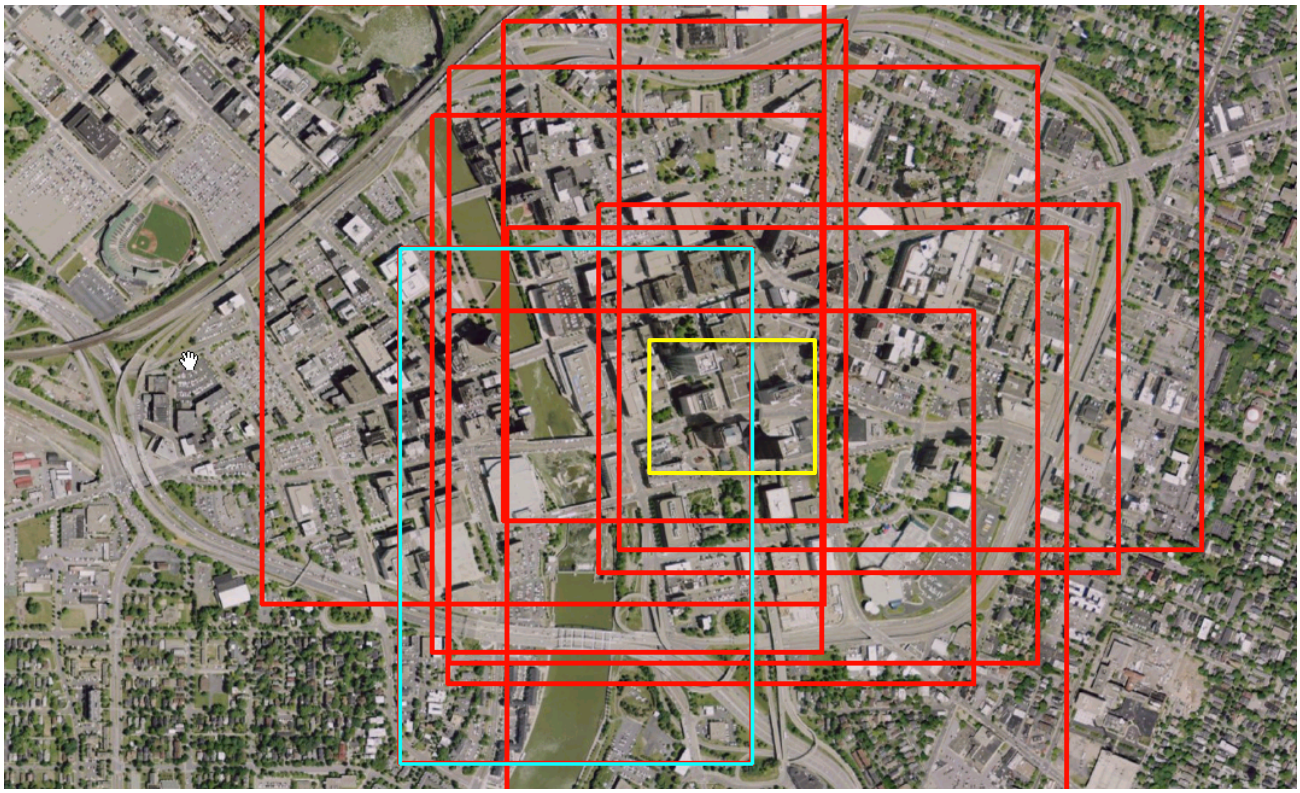

Figure 8. Shown here are the image boundaries for the handpicked eight-image case shown in red, and the addition of the one in blue for the handpicked nine-image case. The yellow boundary is the region shown in Figure 3.

The full output point clouds for the eight and nine image cases are shown in Figure 9. The extent of these point clouds look similar, however, there are significantly more holes in the eight-image case. If we take a close look at these point clouds, as shown in Figure 10, we see that the image on the left is missing buildings, specifically the ones that have been discussed in this paper.

If you consider the image boundaries, shown in Figure 8, you will see that for the eight-image case all of the images cover the region of interest shown in yellow. However, for the nine-image case the additional image does not capture the full region of interest. So, why would the addition of an image that does not fully cover the region of interest make such a significant difference in the output point cloud? Also, comparing the output from the handpicked nine-image case to the first three cases they are quite similar. This leads us to question how these images are being utilized by the point cloud generation workflow.

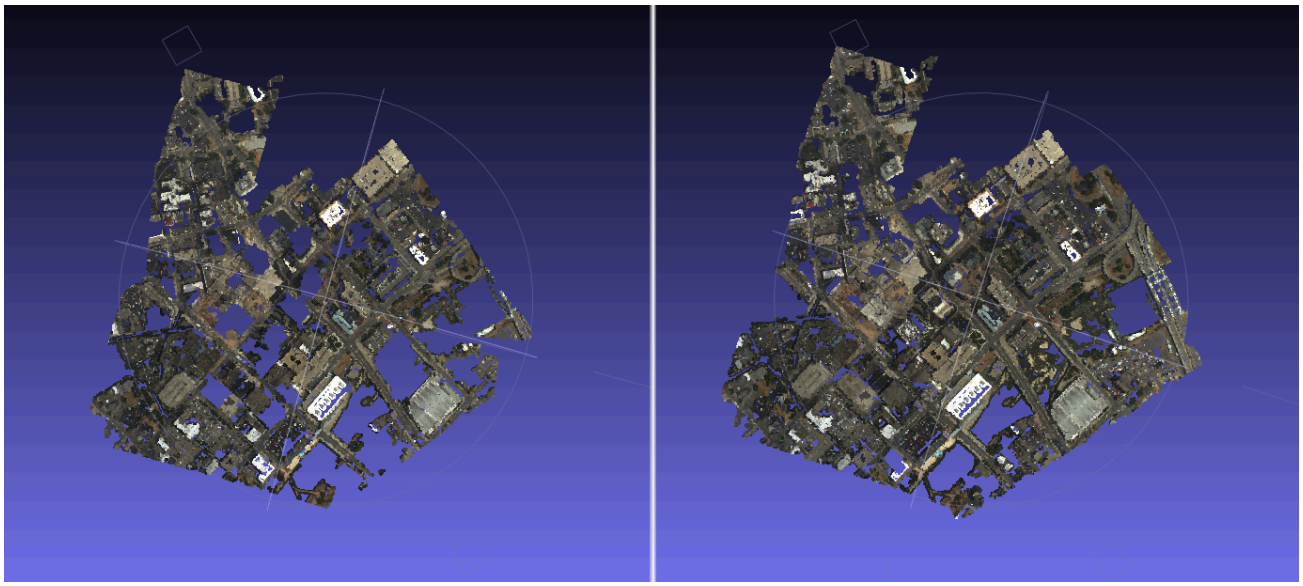

Figure 9. This shows the full output point clouds for the eight handpicked image case (right) and nine handpicked image case (left).

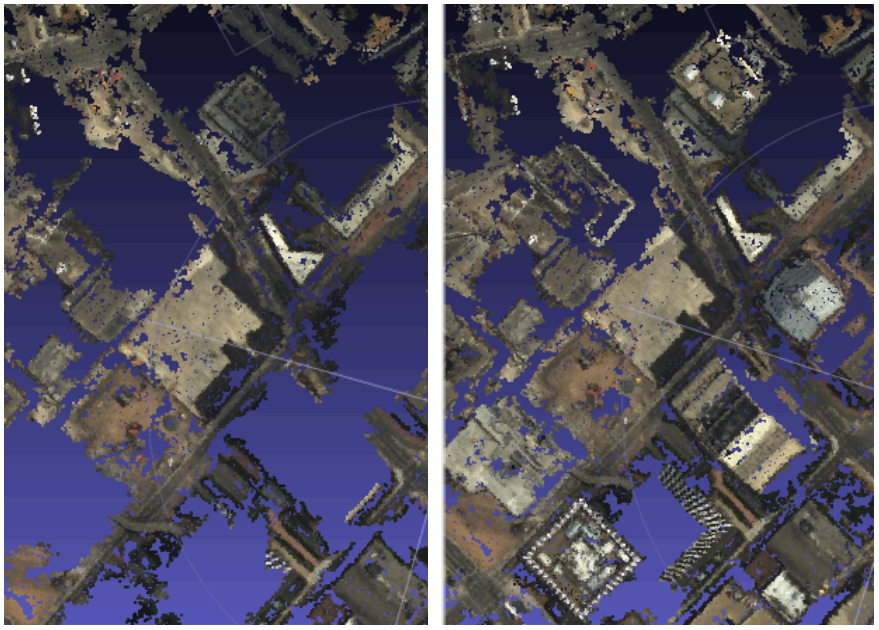

Figure 10. This shows a zoomed in portion of the output point clouds for the eight handpicked image case (right) and nine handpicked image case (left).

4.3 Conclusion

The DIRS group has planned and executed an extensive image collection for three-dimensional reconstruction. This will aid in the analysis of these phototourism techniques. The results shown are just the beginning of a large analysis undertaking. There is still so much to be understood about what these few point clouds show. Perhaps image selection cannot be decoupled from the three-dimensional dense point cloud workflow. It is clear we need to better understand the workflow. For example, we know that CMVS clusters images to optimize processing. How do these clusters get

defined and how do they drive matches and ultimately the output points? Is this the reason we loose the rooftop in case three?

Phototourism is a burgeoning field and we are just beginning to utilize and understand these techniques. There is much more research to be done. Moving forward we intend to focus on smaller datasets and more thoroughly investigate how these images navigate the workflow.

5. DATA

The DIRS group at RIT has a history of large-scale collections and this is just the latest effort. Since this collection can significantly aid the research community and this new research arena, we are sharing the data for free online. It will be available for download along with the LIDAR data, as well as the code used to make the three-dimensional point clouds[3]. To gain access to this data go to http://dirs.cis.rit.edu/resources you will be able to gain access to this data.

REFERENCES

[1] Snavely, N.; Seitz, S.M.; and Szeliski, R., "Photo tourism: exploring photo collections in 3D," in ACM SIGGRAPH 2006 Papers, 835–846. (2006)

[2] Snavely, N.; Seitz, S.M.; and Szeliski, R., "Modeling the world from Internet photo collections," International Journal of Computer Vision, 80(2):189-210, November 2008.

[3] Nilosek, D.R.; Sun, S.; Salvaggio, C., "Geo-Accurate Model Extraction from Three-Dimensional Image-Derived Point Clouds," Proceedings of SPIE, SPIE Defense and Security, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII (2012)

[4] Casterline, M.V.; Salvaggio, C.; Garrett, A.J.; Faulring, J.W.; Bartlett, B.D.; Salvaggio, P.S., "Improved temperature retrieval methods for the validation of a hydrodynamic simulation of a partially frozen power plant cooling lake," Proceedings of SPIE, SPIE Defense and Security, Thermosense XXXII, Utilities and Fluid Dynamics, 7661 (2010)

[5] http://lias.cis.rit.edu/projects/wasp

[6] http://cs.unc.edu/~ccwu/siftgpu/

[7] http://phototour.cs.washington.edu/bundler/

[8] http://grail.cs.washington.edu/software/cmvs/

[9] http://grail.cs.washington.edu/software/pmvs/