NEXT GENERATION AUTOMATIC TERRAIN EXTRACTION USING MICROSOFT ULTRACAM IMAGERY

Dr. Bingcai Zhang, R&D Manager
Scott Miller, Director
Dr. Stewart Walker, Director
Kurt Devenecia, Product Manager
BAE Systems National Security Solutions
10920 Technology Place
San Diego, CA 92127-1874
bingcai.zhang@baesystems.com
scott.miller@baesystems.com
stewart.walker2@baesystems.com
kurt.devenecia@baesystems.com

ABSTRACT

With MSN Virtual Earth and Google Earth, 3D geospatial data are finding its way to average people's daily life. Digital terrain model (DTM) is one of the most important 3D geospatial data types. One of the key automation technologies in softcopy photogrammetry is to automatically generate DTM. The most reliable and widely used algorithm for DTM generation is the normalized image correlation. However, this algorithm has limitations when dealing with elevation discontinuities such as building edges because it is based on the assumption that elevation within a window is similar. Next generation automatic terrain extraction (NGATE) uses both image correlation as well as edge matching. The edge matching algorithm can deal with building edges or elevation discontinuities well. The results from image correlation are used to constrain and guide the edge matching process. At the same time, the results from edge matching are used to assist image correlation.

We applied NGATE on 66 images (GSD = 0.14 feet) from Microsoft 3DI Ultracam in an urban area. A DTM with 21 million 3D points from NGATE has the following characteristics:

- On natural terrain, NGATE DTM has an RMS (root mean square) error of 0.4 feet.
- On streets and parking lots, NGATE DTM has an RMS error of 0.3 feet.
- On center points of flat roof buildings, NGATE DTM has an RMS error of 0.5 feet.
- On corner points of flat roof buildings, NGATE can capture 94% of corners with an RMS error of 0.9 feet.
- On center points, edge points, corner points, and ground points of complex buildings, NGATE DTM has an RMS error of 0.9 feet. 90% of these points have an RMS error of 0.4 feet.
- Building edges are well preserved.
- Streets are precisely modeled.
- Positions and shapes of residential houses are accurately depicted.

DTM from NGATE are very dense and accurate similar to LIDAR data. LIDAR data has been successfully used to extract 3D buildings and we expect that we may achieve similar success with DTM from NGATE as a part of our ongoing research.

INTRODUCTION

Softcopy photogrammetry system can model human two eyes and recreate a 3D view of the real earth. Based on a 3D view, a human operator can extract 3D geospatial data (3D buildings, digital elevation model (DTM), etc.). In the last couple of decades, great progress has been made to automatically extract DTM from a 3D view (Zhang and Miller, 1997; Zhang, *et al*, 1998; Zhang, *et al*, 2006). However, it is far more difficult to automatically extract 3D buildings from a 3D view. Recently, there are great successes to automatically extract 3D buildings from LIDAR data (Rao, 2006; Yates and Rahmes, 2006; Zhang, 2006). LIDAR data are very dense and accurate DTM. If we could automatically extract DTM with LIDAR DTM density and accuracy from a 3D view, we would successfully extract 3D buildings from a 3D view using the same algorithms as LIDAR data.

Based on image intensities, the normalized cross correlation (Vosselman, *et al*, 2004) is the most reliable method to automatically generate DTM from a 3D view. This method matches the conjugates of a point along conjugate epipolar lines. Such a strategy assumes that scene points have the same or similar intensity in each image and there is significant image intensity variation (uniqueness) over both images. This method also assumes that X parallax or elevation within a window (e.g. 13 X 13 pixels) is similar. However, on building edges and corners, elevation changes abruptly. On man-made surfaces such as streets, building roofs, and parking lots, there are no significant image intensity variation. As a result, the normalized cross correlation fails on these areas.

Edge-based methods establish correspondence between image points by matching image-intensity patterns along conjugate epipolar lines. They detect edges and then seek matches between these edges' intersections with conjugate epipolar lines. This approach will not work well if the image regions have no edges or if it is difficult to detect edges precisely. Edge-based methods are not reliable as image correlations.

Next Generation Automatic Terrain Extraction (NGATE) uses both image correlation and edge matching to generate DTM. The edge matching algorithm can deal with building edges or elevation discontinuities well. The results from image correlation are used to constrain and guide the edge matching process. At the same time, the results from edge matching are used to assist image correlation. NGATE is so fast that every pixel (not post or point) is matched many times. Because of data redundancy by computing the same pixel elevation many times, NGATE has much better chance to detect and remove elevation blunders. To certain degree, DTM generated by NGATE is as dense and accurate as LIDAR data as reported in the following sections using high quality Microsoft 3DI Ultracam images in an urban area.

PROJECT DESCRIPTION

We applied NGATE on a project with high quality Microsoft 3DI Ultracam images. There are six strips and 66 images. Each image has 7500 lines X 11500 samples of three bands (red, green, and blue). The side overlap between strips is 60% and the end overlap between images in the same strip is 55%. The ground sample distance (GSD) is 0.14 feet. In other words, every pixel covers 0.14 feet X 0.14 feet area. Some area may be covered by up to six pairs of stereo images. Figure 1 shows the footprints of the 66 images.

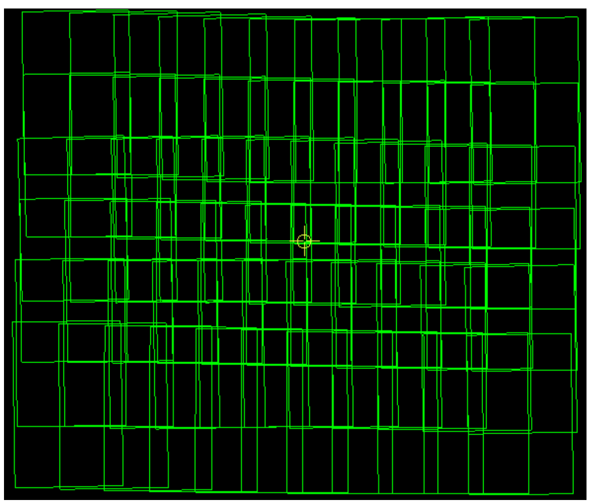

Figure 1. Footprints of 66 images. The area overlapped by at least two images is of 24,402,500 square feet. It is an urban area with industry buildings, residential houses, streets, and parking lots.

The project area is an urban area with residential houses, industrial buildings, streets, parking lots, etc. as shown in Figure 12. The images were acquired with leave off. The side overlap is greater than the end overlay, which is quite unique. In most photogrammetric projects, the end overlap is greater than the side overlap. Because of time difference between different strips, some shadows move significantly from one strip to the next strip. As a result, those shadows may float when they are matched using images from different strips. To compensate this problem, NGATE matches image pairs from the same strip with higher priority or weight. NGATE takes all 66 images and generates a huge DTM. For each pixel, NGATE uses up to six stereo image pairs to compute elevation. With each stereo pair images, NGATE uses both image correlation and edge matching to compute elevations. Both image correlation and edge matching use back matching algorithm to compute the same pixel elevation two times. In some cases, NGATE may use an additional window size to perform image correlation. Therefore, for some pixels, their elevations may be computed 6 x (2 + 2 + 2) = 36 times. Because of so much redundancy, NGATE has more than one chance to perform blunder detection and removal.

TEST RESULTS

A DTM of TIN format with about 21 million 3D points was generated using NGATE. To evaluate the accuracy of the DTM, we selected 5 different types of terrain: (1) center points of flat roof buildings; (2) corner points of flat roof buildings; (3) center points, corner points, edge points, and ground points of complex buildings; (4) natural terrain; and (5) streets and parking lots. The first three types of terrain are useful to extract 3D buildings. The last two types of terrain are useful to generate digital elevation model or bare earth model.

We generated 5 tables with root mean square (RMS) errors, standard deviations, biases, and percent removed points. The RMS error is computed by the following equation:

$$RMS = \sqrt{\sum \delta^2 / n} \qquad (1)$$

Where is the difference between manually measured elevation and computed elevation from NGATE or the residual and ***n*** is the number of manually measured elevation points used in each iteration.

The standard deviation is computed by:

$$std = \sqrt{\sum \delta^2 / (n-1)} \qquad (2)$$

The bias is computed as:

$$bias = \sum \delta / n \qquad (3)$$

Each table contains statistics from more than one iteration. The first iteration uses all the measured ground truth points. The second iteration removes the top few points with the largest residuals or elevation differences between measured elevation and computed elevation from NGATE. The “Percent Removed” measures how many points are removed from the calculations in each iteration.

The manually measured ground truth points are not without errors. They are measured using the same stereo image pairs. It is estimated that they have an RMS error of 0.2 feet. According to random error propagation law, the RMS error has two components: (1) RMS error from NGATE; and (2) RMS error from manually measured points. Therefore, the RMS errors in Table 1 to Table 5 are computed as:

$$RMS = \sqrt{RMS^2_{NGATE} + RMS^2_{manual}} \qquad (4)$$

Center Points of Flat Roof Buildings

Center point elevations are useful for applications such as estimating building heights. As shown in Figure 2, it is an area with lots of flat roof apartment houses. We manually measured 128 points on the centers of the apartment houses. We used these 128 points as ground truth and computed the elevation differences or residuals between the DTM generated from NGATE and these 128 points. The results are in Table 1. Figure 3 shows the locations of some of the 128 ground truth points and the 3D points generated from NGATE.

Table 1 indicates that elevations of center points on houses and buildings can be computed with an RMS error of 0.51 feet. After removing two points with largest residuals, the RMS error drops to 0.35 feet. In other words, NGATE can capture 98% of center points with an RMS error of 0.35 feet on those flat roof houses. As explained by Equation 4, the RMS error contains 0.2 feet RMS error from the manually measured points.

Figure 2. An area with flat roof apartment houses. This area is used for both center points and corner points RMS error computation in Table 1 and Table 2.

Table 1. Center points of flat roof buildings. After removing two points with largest residuals, the RMS error drops to 0.35 feet and becomes consistent afterwards.

Number of Points	RMS Error (feet)	Standard Deviation (feet)	Bias (feet)	Percent Removed
128	0.5056	0.4935	-0.1182	0.0000
126	0.3478	0.3414	-0.0732	1.5873
125	0.3354	0.3414	-0.0825	2.4000
124	0.3225	0.3151	-0.0744	3.2258

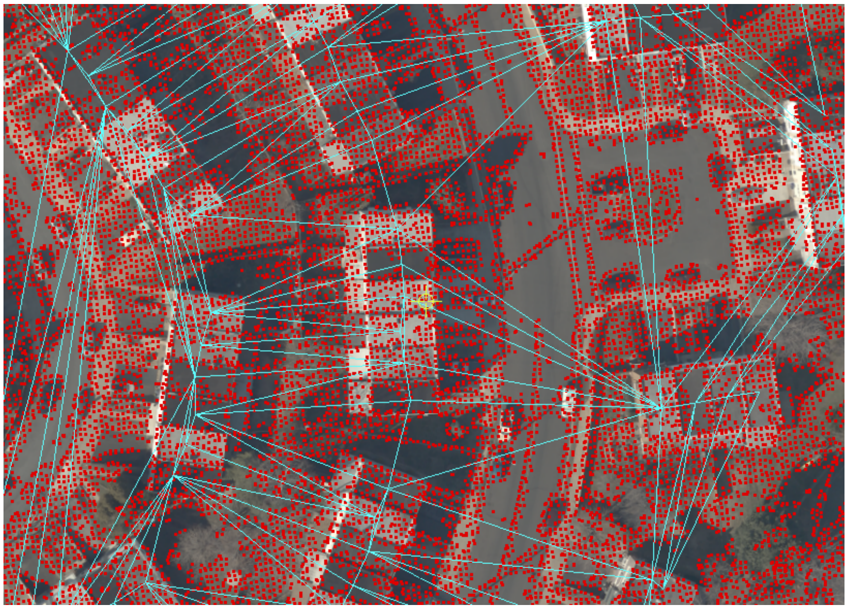

Figure 3. Red dots are 3D points generated by NGATE. Vertices of cyan triangles are ground truth points which are on the centers of apartment houses. Only one in every sixteen 3D points is shown here.

Corner Points of Flat Roof Buildings

Accurate building corners are critical for some applications such as automatic building extraction and targeting. In targeting application, we need to measure accurate XYZ coordinates of buildings from a single image and a DTM. The accuracy of XY coordinates at building corners depends on the accuracy of DTM at building corners. It is an extremely difficult problem for three reasons: (1) image correlation may fail due to elevation discontinuity at building edges; (2) interpolating elevations at corners may result in erroneous numbers; (3) some corners may not have strong edges. As illustrated in Figure 4, the elevation at a corner abruptly changes 23 feet. A small error in XY dimensions may result in a large error in elevation. We manually measured 112 points on the corners of the apartment houses. We used these 112 points as ground truth and computed the elevation differences between the DTM generated from NGATE and these 112 points. The results are in Table 2. Figure 5 shows the locations of some of the 112 ground truth points and 3D points generated from NGATE.

Figure 4. The corner of a building has an elevation difference of 23 (197 – 174) feet. Image correlation fails on corners due to elevation discontinuity. Edge matching depends on available edges on building corners.

Table 2 indicates that NGATE can capture 94% of flat roof building corners with an RMS error of 0.9 feet. NGATE has missed 6% of the corners due to week image edges or no image edges. It should be noted that these are not tall buildings. The success rate may decrease with very tall buildings. NGATE accuracy on building corners varies significantly depending on the quality of image edges on those corners. As shown in Figure 4, the image edge quality on a corner may depend on its background. In Figure 4, the background changes from shadow to grass, and from grass to concrete. In summary, the numbers in Table 2 is not very conclusive. They may vary significantly from different types of buildings and from different areas.

Table 2. Corner points of flat roof buildings. After removing 3 largest residuals, the RMS error drops to 1.5 feet. After removing 6 largest residuals, the RMS error drops to 0.9 feet. In other words, there are 6 corners that NGATE has missed. For the rest of the 106 corners, NGATE has an RMS error of 0.9 feet.

Number of Points	RMS Error (feet)	Standard Deviation (feet)	Bias (feet)	Percent Removed
112	2.5596	2.5086	0.5609	0.0000
109	1.4609	1.4465	0.2469	2.7523
106	0.9111	0.9135	0.0585	5.6604
103	0.6994	0.7016	-0.0409	8.7379
100	0.5763	0.5678	-0.1139	12.0000
97	0.4807	0.4636	-0.1355	15.4639
96	0.4591	0.4354	-0.1523	16.6667

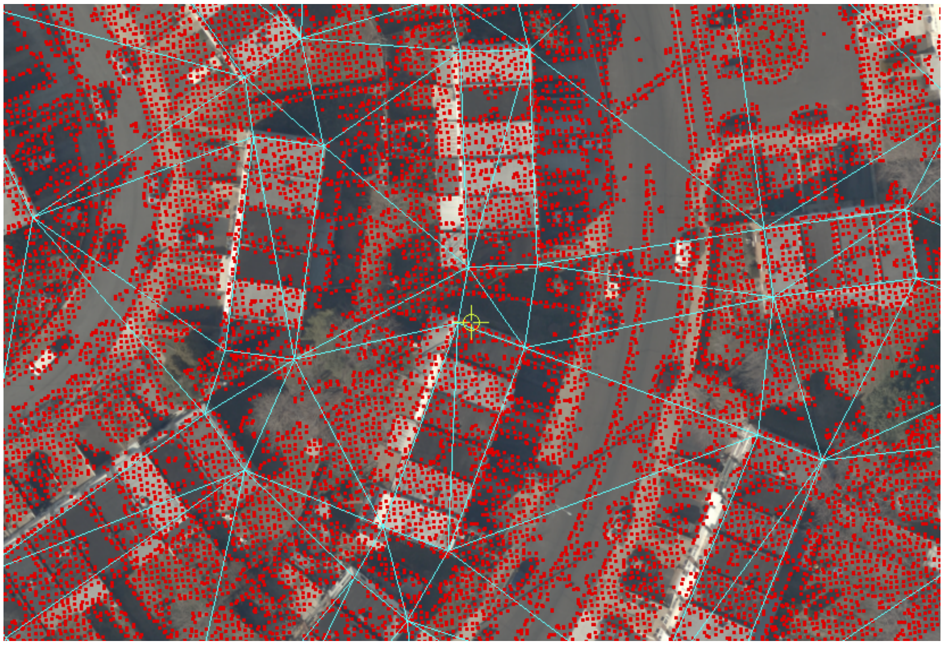

Figure 5. Red dots are 3D points generated by NGATE. Vertices of the cyan triangles are ground truth points which are on the corners of apartment houses. Only one in every sixteen 3D points from NGATE is shown here. Only a small subset of the 112 ground truth corner points is shown in this picture.

Complex Buildings

Complex buildings are a very common type of buildings. Complex buildings have complex roofs as illustrated in Figure 6. NGATE may not generate any 3D points on the interior of a man-made planar surface because there are no image edges and intensity variations as shown in Figure 6 and Figure 7. Fortunately, NGATE may generate enough 3D points on building roof ridges, corners, and edges because it is likely there are image edges in these locations and NGATE uses edge matching. With enough 3D points on roof ridges, edges, and corners, we can interpolate the interior elevations of a planar surface. To accurately model complex buildings, we need accurate 3D points on corners, edges, roof ridges, as well as 3D points on near-by ground. Most complex buildings have planar surfaces. Figure 7 shows the manually measured ground truth points which form cyan triangles. These manually measured points are on the corners, edges, roof ridges of the building, and the near-by ground. There are 139 manually measured ground truth points. A subset of the 139 points are shown in Figure 7. The red dots are 3D points generated by NGATE. For applications like automatic building extraction and targeting, 3D points must be dense enough and there must be enough 3D points located at breaklines where terrain slope changes significantly. Figure 7 indicates that there are enough automatically generated 3D points on the corners, edges, roof ridges, and the near-by ground to accurately model the complex building.

Table 3 indicates that NGATE can achieve an RMS error of 0.9 feet on the corners, edges, roof ridges, and near by ground of complex buildings. There are void areas on roofs and ground where there are no 3D points from NGATE. For void areas, we can interpolate its elevations using adjacent 3D points as long as there are enough 3D points on breaklines. For example, in Figure 6 and Figure 7, we can interpolate elevations in the void roof areas using 3D points from building edges and roof ridges. There are void areas on the ground close to buildings as shown in Figure 7. For a digital elevation model (DEM), we can use adjacent non-roof 3D points to interpolate its elevations. In this case, all 3D points on building roofs are removed. For automatic building extraction, the algorithm to search a near-by ground point needs to search farther from a building edge.

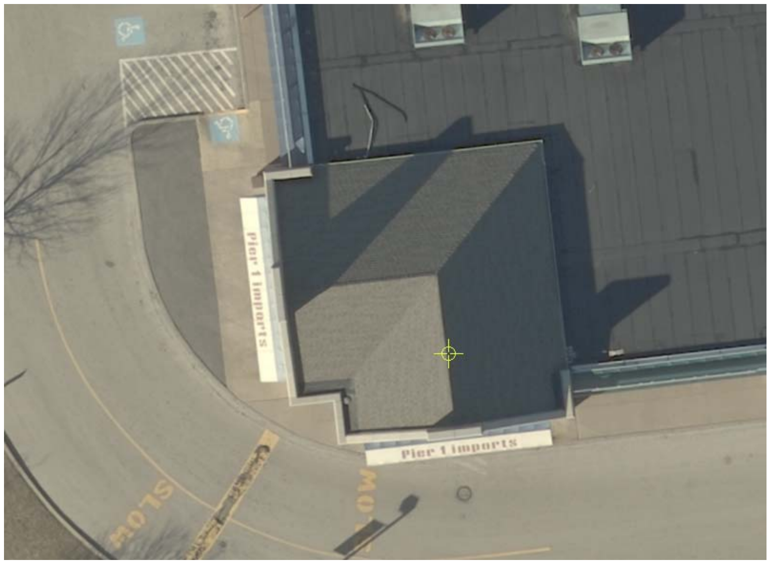

Figure 6. A complex building with complex roofs. Man-made surfaces such as roofs and streets are normally planar. In the interior of a plane, NGATE may not generate any 3D points because there are no edges and no intensity variations. However, NGATE may generate 3D points on the edge of a plane or the roof ridges as shown in Figure 7.

Table 3. NGATE has an RMS error of 0.9 feet on the corners, edges, roof ridges, and near-by ground of a complex building. After removing 10% points with largest residuals, the RMS error drops to 0.4 feet.

Number of Points	RMS Error (feet)	Standard Deviation (feet)	Bias (feet)	Percent Removed
139	0.8815	0.8721	-0.1484	0.0000
134	0.6283	0.6266	-0.0713	3.7313
129	0.4622	0.4634	-0.0244	7.7519
127	0.4198	0.4214	0.0005	9.4488

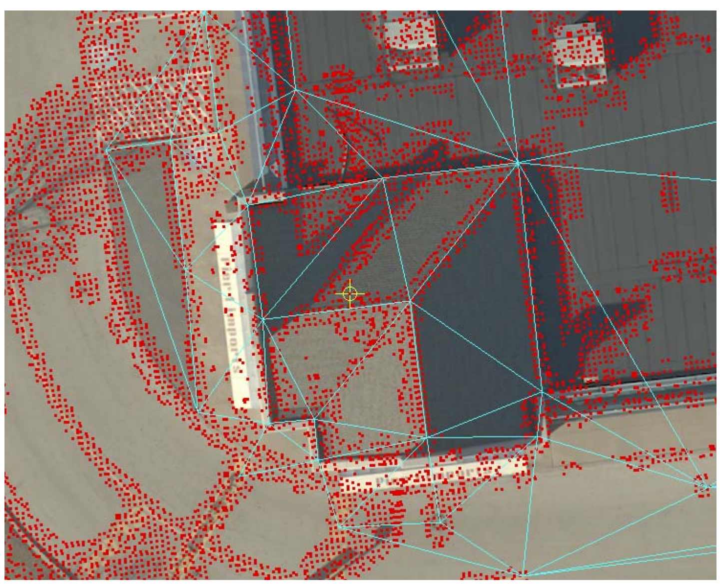

Figure 7. Red dots are 3D points generated by NGATE. Vertices of the cyan triangles are ground truth points which are on the corners, edges, roof ridges of a complex building, as well as on the near-by ground. There are many 3D points on the roof ridges, corners, and edges of the building. These points can be used to interpolate elevations for roof void areas.

Natural Terrain

Natural terrain is an important category because most of terrain is natural. However, it is pretty hard to find a natural terrain in this project area. As illustrated in Figure 8, there is some natural terrain between a parting lot and a road, and between a road and a highway. Natural terrain is the easiest area to automatically generate 3D points. As illustrated in Figure 9, NGATE generated lots of 3D points on natural terrain. However, there are almost no 3D points on the road. This is because there are no edges on the road and all the pixels on the road are almost identical to each other. This is not really a problem because the elevation on the road can be interpolated from the elevations on both sides of the road.

Table 4 indicates that the total RMS error is about 0.66 feet including trees and bushes. After removing 3 points on trees/bushes, the RMS error drops to 0.4 feet. In other words, the RMS error on open natural terrain excluding tress and bushes is 0.4 feet. We manually measured 259 points using the same stereo pair images in a 20 feet post spacing grid as shown in cyan mesh in Figure 9. Some of the 259 points are on the road. A few of them are on trees/bushes where we cannot determine their true elevations. As shown in Figure 9, NGATE did not generate any 3D points on the road. The elevations on the road can be accurately interpolated from the 3D points on both sides of the road.

The RMS error from NGATE on open nature terrain is square root (0.4 x 0.4 – 0.2 x 0.2) = 0.35 feet as explained in Equation 4. The 0.2 feet RMS error is from the manually measured ground truth points. When evaluating quality and accuracy of DTM, we should use only well defined elevation points on solid ground as ground truth. Therefore, the 3 points on trees and bushes should be excluded because their true elevations cannot be determined.

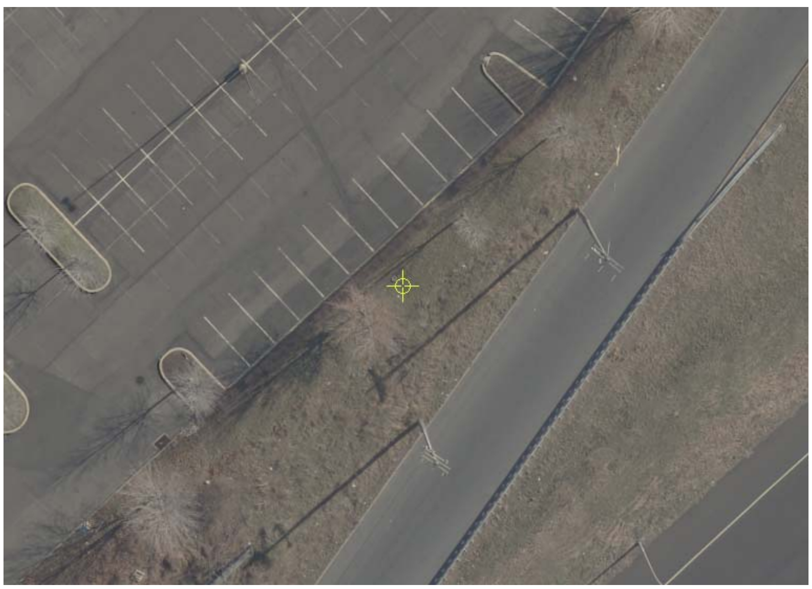

Figure 8. Some natural terrain between a parking lot and a road, and some natural terrain between a road and a highway. There are a few small trees and bushes. Elevations on trees and bushes are not well defined. Therefore, they should not used to evaluate accuracy of NGATE DTM. The elevation variation is about 25 feet in this area.

Table 4. 259 manually measured ground truth points are a regular grid with a post spacing of 20 feet. They were manually measured using the same stereo image pairs and have an RMS error of 0.2 feet. After removing 3 points on trees/bushes, the RMS error drops to 0.4 feet. After considering the ground truth point error, the RMS error from NGATE DTM is about 0.35 feet.

Number of Points	RMS Error (feet)	Standard Deviation (feet)	Bias (feet)	Percent Removed
259	0.6590	0.6556	0.0784	0.0000
256	0.3941	0.3916	0.0506	1.1719
250	0.3276	0.3219	0.0641	3.6000
245	0.2929	0.2855	0.0681	5.7143
244	0.2879	0.2793	0.0721	6.1475

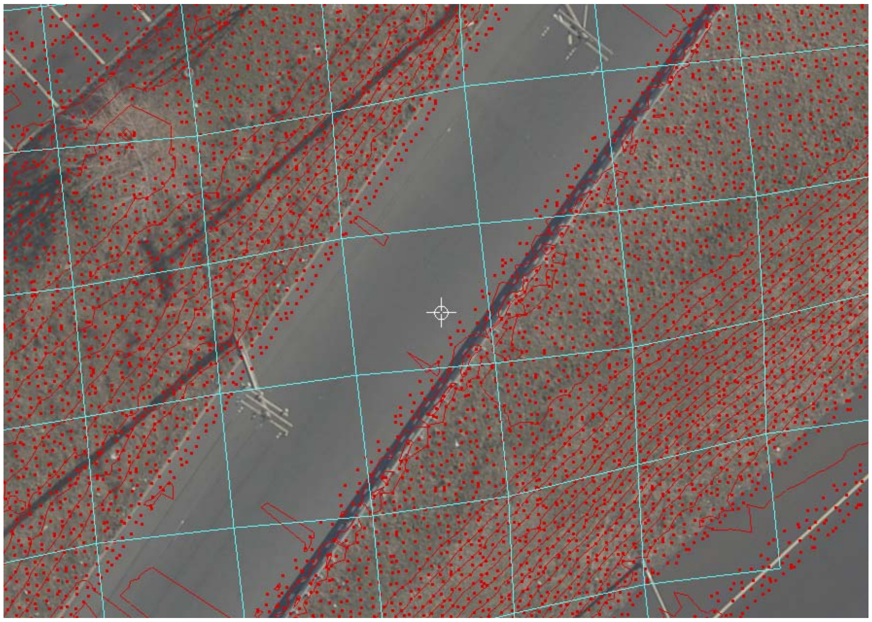

Figure 9. Red dots are 3D points generated by NGATE. Red lines are 1 foot contours. Intersections of the cyan meshes are ground truth points which were manually measured and have fixed locations. A few of them are on trees/bushes. Only one in every four 3D points is shown here. There are no 3D points on the road. But the elevations on the road are accurately interpolated using 3D points on both sides.

Streets and Parking Lots

Streets and parking lots are common terrain types in urban areas. As illustrated in Figure 10, there are parking stripes on the parking lots and center dividers and traffic signs on the streets. There is a small shadow cast by a light pole on the right side. There is a long shadow from the lower side to the right side cast by a light pole. These are excellent edges for NGATE to match because NGATE uses edge matching in addition to image correlation. As shown in Figure 11, NGATE generated many 3D points on these edges. In general, NGATE matches features/pixels which are distinguishable from its surroundings on a man-made surface. For areas without distinguishable pixels on a man-made surface, we can interpolate elevation from adjacent matched 3D points as shown in Figure 11 because the man-made surface normally consists of a number of planar 3D planes.

Table 5 indicates that the total RMS error is around 0.3 feet. We manually measured 310 points using the same stereo pair images in a 10 feet post spacing grid as shown in cyan mesh in Figure 11. All the 310 points are on solid ground. As explained in Equation 4, the 310 ground truth points have an RMS error too. Assuming that the ground truth points have similar RMS error, the real RMS error of the NGATE DTM is about 0.3/1.414 = 0.21 feet. It should be noted that many LIDAR hardware vendors use similar type of terrain to evaluate LIDAR accuracy. Table 5 also indicates that there are no large residuals because all ground truth points are on solid ground.

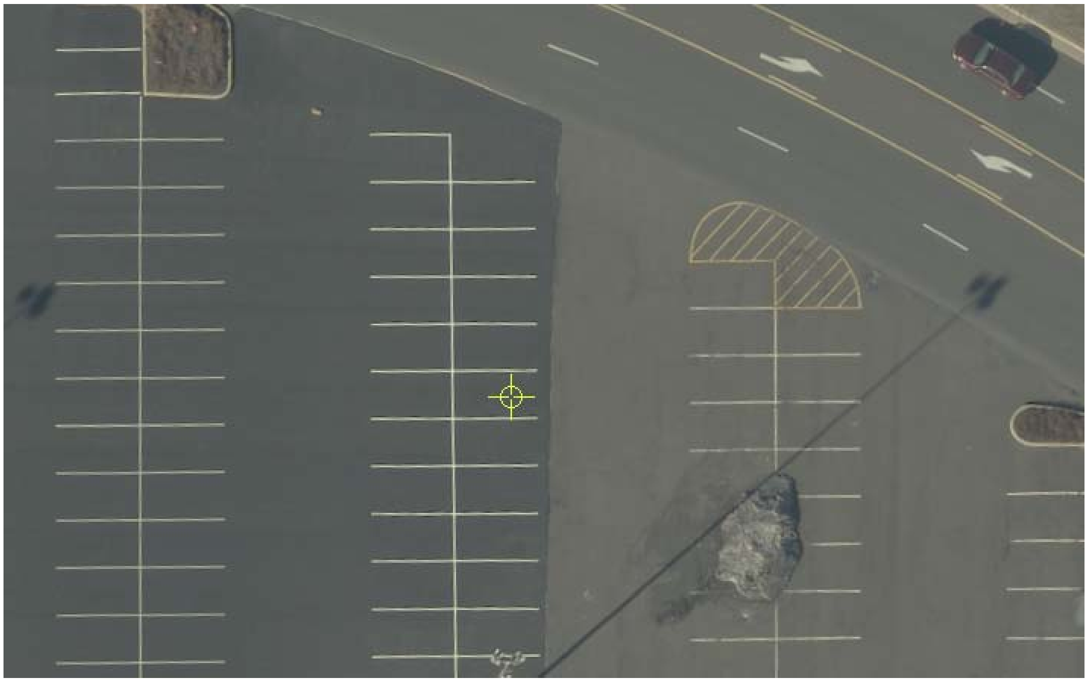

Figure 10. Parking lots and streets. There are parking stripes on the parking lots. There are center dividers and traffic marks on the streets. There are two shadows cast by light poles. On man-made surfaces, they are excellent edges for NGATE to match because NGATE uses edge matching in addition to image correlation. For areas without edges and intensity variations, their elevations can be accurately interpolated from adjacent 3D points as shown in Figure 11.

Table 5. 310 ground truth points are a regular grid with a post spacing of 10 feet. All the 310 points are on solid ground. They were manually measured using the same stereo image pairs. Therefore, they have RMS error similar to the RMS error from NGATE DTM. According to Equation 4, the RMS error from NGATE DTM is 0.3/1.4 = 0.21 feet. This number is the one that some LIDAR vendors use to evaluate their LIDAR accuracy.

Number of Points	RMS Error (feet)	Standard Deviation (feet)	Bias (feet)	Percent Removed
310	0.3057	0.2524	-0.1731	0.0000
309	0.3005	0.2480	-0.1703	0.3236

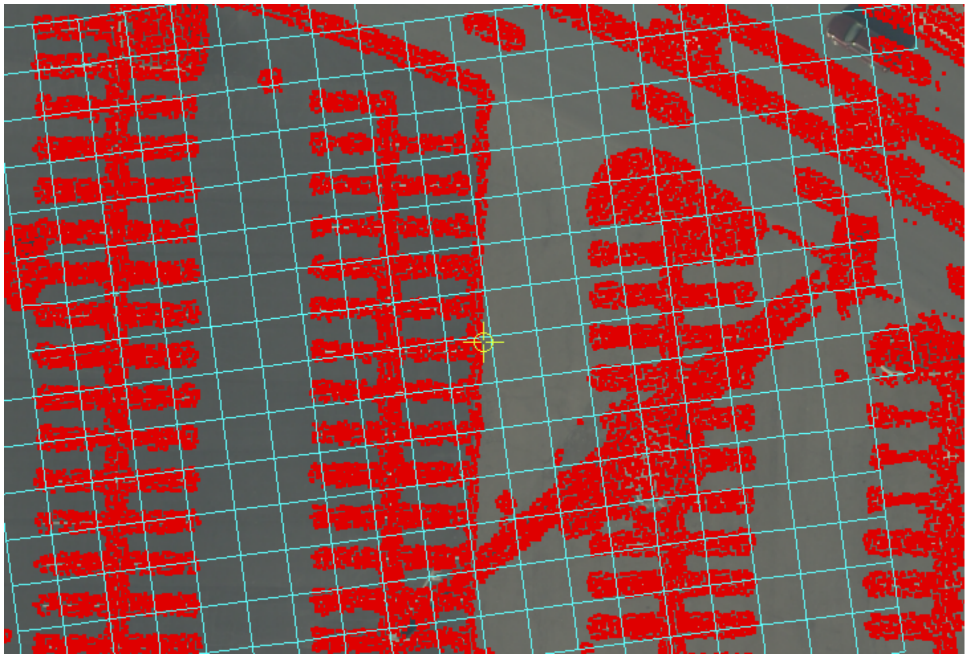

Figure 11. The intersections of the cyan meshes were manually measured ground truth points. NGATE generated many 3D points on all parking stripes, center dividers, and traffic signs. On the left side, there is small shadow cast by a light pole. From the lower side to the right side, there is a long shadow cast by a light pole. NGATE matched both shadows and generated many 3D points on them. For most man-made surfaces, elevations of void areas can be interpolated from adjacent 3D points. For example, elevations between stripes can be accurately interpolated from 3D points on the stripes. Elevations on streets can be accurately interpolated from 3D points on street center dividers and traffic marks.

SUMMARY

Test results using images of 0.14 feet GSD from Microsoft 3DI Ultracam aerial digital camera indicates that NGATE has the following capabilities:

1. On natural terrain, DTM generated by NGATE has an RMS error of about 0.4 feet.
2. On streets and parking lots, DTM generated by NGATE has an RMS error of about 0.3 feet.
3. On center points of flat roof buildings, DTM generated by NGATE has an RMS error of about 0.5 feet.
4. On corner points of flat roof buildings, NGATE can capture 94% of corners with an RMS error of about 0.9 feet.
5. On center points, edge points, corner points, and ground points of complex buildings, DTM generated by NGATE has an RMS error of about 0.9 feet. 90% of these points have an RMS error of 0.4 feet.

There are two important factors that contribute to the above results. One factor is the NGATE technology. It utilizes all information in images. In some cases, NGATE computes elevation of every pixel up to 36 times. The other factor is the quality of the images from Microsoft 3DI Ultracam. It is much better than traditional film-scanned images in terms of running NGATE.

Figure 12 is the mosaic of the 66 images ortho rectified by the DTM from NGATE. The entire processes from DTM generation, ortho rectification, and mosaic can be done in a batch process. In other words, no human operators are needed for any editing purposes. The DTM generation of 21 million 3D points took 24 hours on a DELL Precision 670 with two dual CPUs. NGATE actually generated 602 million 3D points in 24 hours, but we filtered out most of them to reduce data storage.

Figure 12. Mosaic of 66 images ortho-rectified using DTM generated by NGATE. The mosaic has zigzag edges because the edges are from the DTM boundary generated by NGATE. The DTM boundary is computed by overlapping the 66 image footprints and union the areas overlapped by at least two images.

With very dense and accurate 3D points, automatic building extraction technology from LIDAR data may be applied to DTM generated by NGATE. We have already started working on this next step in our next generation photogrammetric system.

REFERENCES

Rao R. (2006). Automatic Feature Extraction from Terrestrial and Airborne LIDAR. *International LIDAR Mapping Forum*, Denver, 13-14 February 2006.

Vosselman, G., M. Sester, and H. Mayer (2004). Basic computer vision techniques. *Manual of Photogrammetry, Fifth Edition, American Society for Photogrammetry and Remote Sensing* (edited by McGlone, J.C., Mikhail E.M., and J. Bethel). pp 455-504.

Yates J.H., and M. Rahmes (2006). Evaluation of LIDAR 3D Model Generation. *International LIDAR Mapping Forum*, Denver, 13-14 February 2006.

Zhang B. (2006). Processing LIDAR Data in SOCET SET – Exploring Its Full Potential. *International LIDAR Mapping Forum*, Denver, 13-14 February 2006.

Zhang B., S. Miller, K. Devenecia, and S. Walker (2006). Automatic Terrain Extraction Using Multiple Image Pairs and Back Matching. *Proceedings of ASPRS.*

Zhang, B., Olander, N., Paderes , F.C., Miller, S.B., and A.S. Walker. (1998). Automatic TIN Generation from Imagery. *Proceedings of ASPRS.*

Zhang B., and S. Miller (1997). Adaptive automatic terrain extraction, *Proceedings of SPIE, Volume 3072, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision* (edited by D. M. McKeown, J. C. McGlone and O. Jamet). pp. 27-36.