BLUNDER ELIMINATION TECHNIQUES IN ADAPTIVE AUTOMATIC TERRAIN EXTRACTION

Fengliang Xu*, Neil Woodhouse, Zhizhong Xu, David Marr, Xinghe Yang, Younian Wang

ERDAS, 5051 Peachtree Corners Cir, Suite 100, Norcross, GA 30092 - (fengliang.xu@erdas.com)

COMMISSION III, ThS-3

KEY WORDS: Automatic Terrain Extraction, Digital Surface Model, Digital Terrain Model, Image Registration, Blunder Removal, Object Filtering, ERDAS, LPS

ABSTRACT:

This paper introduced a new automatic terrain extraction (ATE) module inside ERDAS's photogrammetric software LPS. This method uses a global DEM to initialize a surface model and iteratively refines it with image registration on different pyramid levels. Search range used in image registration is adaptively controlled by elevation range of matched feature points. However, mismatches and elevation blunders may cause search range to be out-of-scope and fail ATE process. We used three blunder elimination techniques to ensure the convergence of search range: positional cross-correlation, PCA-based blunder elimination, and object filtering. The method is tested with images from various sensors including frame cameras, satellites, and Leica's ADS40.

1. INTRODUCTION

1.1 Overview

Automatic terrain extraction (ATE) is used to extract digital surface model (DSM) from triangulated stereo images (Zhang, 2006; Zebedin, 2006). DSM can be either vector or raster format. Its density ranges from 1/3 to 1/10 of density of original image pixels, e.g., 10-meter resolution for satellite sensors such as Quickbird, 1-meter resolution for ADS40 imagery, etc. DSM can be used to generate digital terrain model (DTM), contour map, 3D building model, orthophoto, and true-ortho.

As lidar technology becomes popular and affordable, ATE begins to lose market on airborne platform. Lidar is superior in providing DSM faster, denser, more accurate (Hodgson, 2004; Ma, 2005): centimetre-level vertical accuracy from lidar vs. meter-level accuracy from ATE, 1~7 points per meter^2 from lidar vs. 1 point per meter^2 from ATE, 2-hours of filtering for a 1000-mile^2 by lidar vs. much longer time in matching, filtering, and manual post-editing by ATE. However, lidar alone cannot provide orthophotos. To generate quality orthophotos, images from other platform will need to go through automatic point measurement and triangulation before being integrated with lidar point cloud, and this adds to the cost and complexity of map generation, so ATE is still active in low-resolution and low-cost map generation. Furthermore, lidar can not generate dense and accurate enough points on satellite platform, which still relies on ATE for map generation.

The future of ATE is still ambiguous right now: it may fade out in next 5~10 years, or it will further develop in new directions, such as: 1) points from ATE (high horizontal accuracy) and lidar (high vertical accuracy) may be triangulated together to achieve centimetre level accuracy in both horizontal and vertical direction; 2) feature-extraction-based ATE and lidar point cloud may be integrated to provide buildings structures and true-orthos at real-time;

3) ATE may become more popular in map resynchronization and vector-to-raster registration for updating road map or detecting changes, etc.

In order to compete against lidar in both point density and accuracy, ERDAS's next generation ATE moves toward pixel-by-pixel and feature-based matching. Adaptive ATE is an intermediate step towards this goal.

1.2 Adaptive ATE

In ERDAS's traditional ATE, customer needs to set up search range for image registration manually by identifying terrain types such as high-mountain or rolling hills and then selecting corresponding strategies. To use a wider search range on mountains and a smaller one on hills within the same image pair, customer needs to manually digitize area-of-interest (AOI) for mountain regions and hill regions separately and assign different strategies to them. ATE will go from high-pyramid levels to low pyramid levels, at each level the search range used stays the same on image space but is actually reduced to half on object space, which means search range goes smaller and smaller by brute-force. This method is proven to be a reliable solution, but it requires much human operation and is slow on production line.

The performance of traditional ATE relies on customer-defined search ranges. To free customer from such kind of overhead, we developed an adaptive ATE module to define search range by terrain variation. The basic idea is: at the beginning of a matching circle, highest and lowest points (from terrain range) along an image ray from first image are projected to second image as starting and ending points, which defines a search range; and then, features are matched along epipolar-line within this range; finally, elevation range of matched points are used to update terrain range, ending this matching circle. Matching starts from a high image pyramid and terrain range is initialized with a global DEM generated with 3-second SRTM DEM (Slater, 2006); at each pyramid level, terrain range updated from matches on higher pyramid is used to limit search range at

current pyramid, and matching results from current pyramid will be used to update terrain range at next lower pyramid. By this way, the ambiguity of terrain variation is reduced at each pyramid level and search range should be reduced as well and converge to a small value, which is a function of terrain slope, accuracy, and pixel size.

If there are no mismatches and blunders, both terrain variation and search range will converge through iterations. However, mismatch is inevitable in stereo image registration and blunder does exist, so search range won't reduce effectively and may ends up too big at low pyramid levels. That is why blunder elimination turns to be a very important part of adaptive ATE. We developed three blunder elimination techniques: positional cross-correlation, PCA-based blunder elimination, and object filtering. The first two are applied at the end of matching at each pyramid level to suppress mismatches, and object filtering is used at final pyramid to eliminate spikes, buildings, and trees to produce bare-earth.

2. BLUNDER ELIMINATION TECHNIQUES

2.1 Positional Cross-Correlation

Positional cross-correlation (PCC) measures the consistency of relative point locations between two sets of points on image space. It is calculated using the following equations:

$$p_x = \frac{E((x_1 - E(x_1))(x_2 - E(x_2)))}{\sqrt{E((x_1 - E(x_1))^2 \cdot E((x_2 - E(x_2))^2}}$$
$$p_y = \frac{E((y_1 - E(y_1))(y_2 - E(y_2)))}{\sqrt{E((y_1 - E(y_1))^2 \cdot E((y_2 - E(y_2))^2}} \quad (1)$$

where, p_x is PCC of x coordinates
p_y is PCC of y coordinates
(x_1, y_1) are image coordinates from set 1
(x_2, y_2) are image coordinates from set 2
$E()$ is an operation to calculate mean value

Fig. 1 shows an example: left shows points without blunders where PCC is high (1.00, 0.96); right shows points with blunders where PCC_y is low (0.85). A low PCC normally indicates mismatches, which can be identified by iteratively eliminating the most-inconsistent pair and re-calculating PCC until PCC is big enough.

This method works for terrain with slopes and also adapts well to various natural terrain such as mountains, hills, and flat planes. It can normally remove 30% of matches as blunders and make terrain estimation reliable for matching at next pyramid level. However, this method alone is not suitable for metropolitan area where high-rising buildings cause too much discontinuity.

The threshold for PCC is currently practised with empirical values. This method can eliminate approximate 5% of matched points that are normally big mismatches.
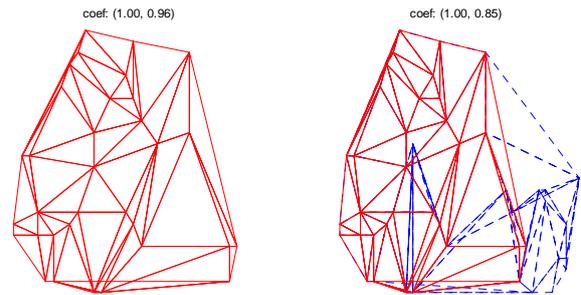
Figure 1. Positional correlation coefficients without blunders (left, vertex linked by solid lines) and with blunders (right, points linked only by dashed lines). Points are triangulated in objected space and linked in image space to show the displacement

2.2 PCA-based Blunder Elimination

This method is based on piecewise smoothness constraint on object space. Point cloud in neighbourhood is fitted to a principal plane using PCA decomposition (Rao, 1972) and points with big distance to this plane are eliminated. Distance threshold is dynamically changed with variation of distances. Fig. 2 shows an example.
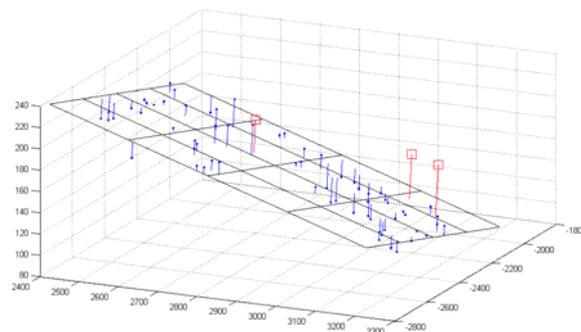
Figure 2. Principal plane, blunder points (in red, square-shaped) and in-range points (in blue)

2.3 Object Filtering

Blunders that survive above two methods can be further eliminated by object filtering. Buildings and trees can be filtered as well to generate bare-earth.

Our object filter is called "ebb process". It is similar to ebb of water: suppose at beginning all buildings and trees are flooded with water; as water ebbs away, building/tree tops will first come out of water and appear as standalone regions that are relatively simple to analysis; and then water level will drop for several meters before ground appears. If a region is high and small enough before it connects to terrain, it will be classified as an object (building, tree, or spike); otherwise, it will be merged into terrain. Fig. 3 and 4 show an example. Fig. 3 is a DSM together with slope edges that indicate boundaries of objects. Fig. 4 shows ebb process while elevation drops from 278m to 250m.
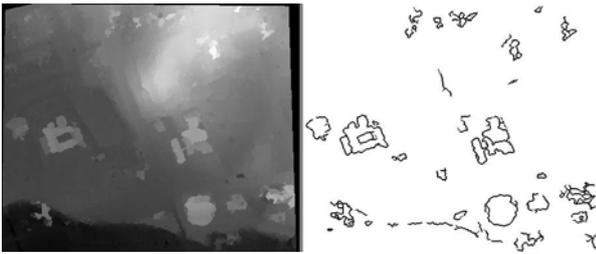
Figure 3. DSM (left) of a test region and corresponding slope edges (right)

DTM in ERDAS LTF format are generated and compared with manually digitized baseline data, as in Table 1.

Data Type	DTM		
	Point #	Check #	RMSE (meter)
Urban (Hamburg)	155014	170	1.51
Rural (Quasco)	551492	65	1.48
Mountain (Mexico)	195097	100	1.82

Table 1. ATE accuracy for frame sensors

Object filtering can effectively remove most spikes as well as points on buildings and trees. It can cut off approximately 60% of manual editing time. It can also improve bare-earth quality: for an urban dataset (Hamburg), RMSE drops from 4.84m before object filtering to 1.51m after object filtering.

Fig. 5 shows one example: the majority of buildings on a flat terrain are removed except for an over-size building and a high-rising road that are beyond threshold.

Figure 4. Status at elevation levels 277.4m, 272.4m, 265.4m, 252.2m, ordered from top to bottom. Legend: white (terrain), black (under water), yellow (slope edge), red/green (new isolated regions). Right: white (extracted terrain), red (filtered objects)

This object filter can also be used to filter lidar point cloud. Preliminary test shows a good performance.

3. EXPERIMENTAL RESULTS

3.1 Frame Sensors

We tested adaptive ATE on frame images with three typical scenes: urban, rural, and mountainous areas. 2m-resolution

Figure 5. DSM before (above) and after (below) object filter

Fig. 6 shows another example where buildings on a slope are removed, yet small terrain variation is still preserved.

Fig. 7 shows a DSM from rural area. Points are located on both ground and trees and contours reveal the coverage of trees. After object filtering points on trees are removed and contours become much smoother.

Fig. 8 shows a mountainous scene. The initial terrain range from global DEM is 50~1700m. Adaptive ATE can adapt to both ridge and valleys.
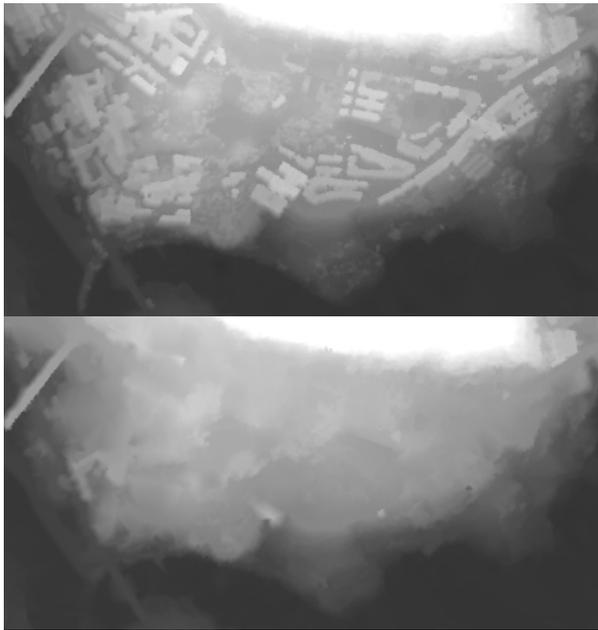
Figure 6. DSM before (above) and after (below) object filtering
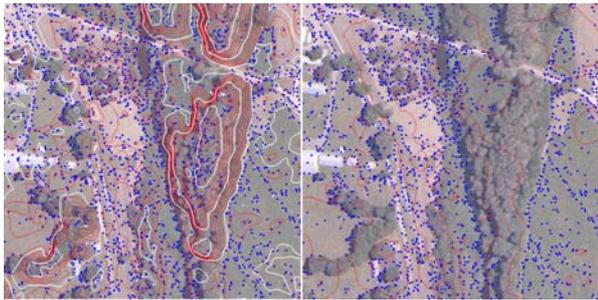
Figure 7. Points and contours before (left) and after (right) object filtering
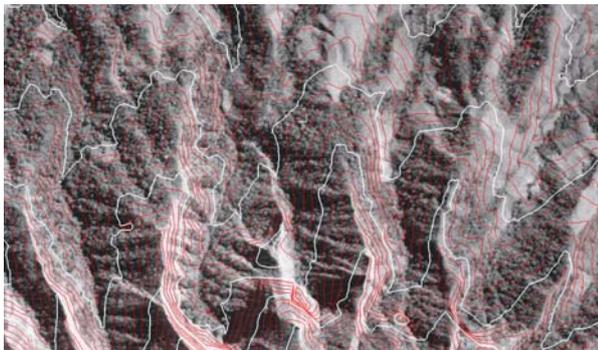
Figure 8. Contours of ridges and valleys from mountains

For frame images, adaptive ATE is slower than traditional ATE and ranges from 2 times to 16 times depending on terrain type. However, the quality from adaptive ATE is normally better in terms of point density, distribution, and accuracy. Adaptive ATE needs more computer time but less human time because it reduces overhead and post-editing work.

3.2 Satellite Sensors

Adaptive ATE also works for satellite images with good geometry (e.g, after triangulation with control points). Table 2 shows some results.

Sensor Type	Accuracy	
	Check #	RMSE (meter)
alos	565	11.87
cartosat	255	7.36
Eros	127	2.99
ikonos	147	3.8
quickbird	95	3.54
worldview	700	9.15

Table 2. ATE accuracy for satellite sensors

Fig. 9 shows an IKONOS image overlapped with extracted points and contours. Since satellite images have much lower resolution, point density from ATE is also very low, and the performance of object filtering is not as good as on frame images.
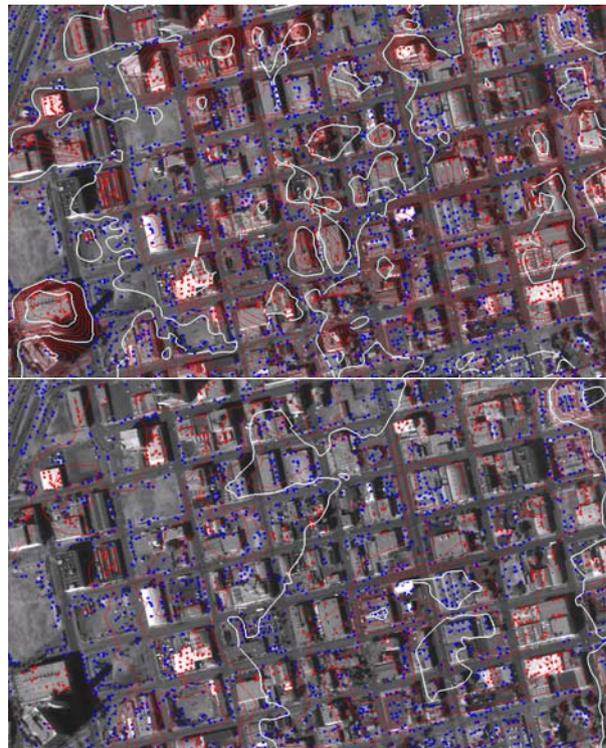
Figure 9. IKONOS image overlapped with extracted points (red: points on objects; blue: points on ground) and contours, before (above) and after (below) object filtering

The converge angles of some satellite stereo pairs are very small, so space intersection from sensor model can be very sensitive to small changes of parallax: at high pyramid levels, image coordinates of correctly-matched points are not accurate (at pyramid n, the uncertainty of a pixel is 2^n times bigger than a pixel at pyramid 0 or original resolution) , so calculated z values from space intersection may be far out of valid range and be subject to elimination as blunders. It is possible to see that

the z values of same feature points are in range A at pyramid one and range B at pyramid two, yet range A and B do not overlap. ATE needs to address such issues by relaxing criteria for z values at high pyramid levels.

The quality of adaptive ATE on satellite images is not significantly better than traditional ATE, because terrain variation is not significant either relative to flying height of satellites. The speed of adaptive ATE, however, can be faster than traditional ATE, because traditional ATE works best for epipolarimages where terrain variation mainly affects x-parallax, yet adaptive ATE does not assume epipolarimages, and most satellite images are not epipolarimages either.

3.3 ADS40 Sensors

Leica's airborne ADS40 sensor provides multi-spectral pushbroom images with three looking angles (-14°, 0, +28°), high resolution (up to 5cm), and 5 bands (pan, rgb, and infrared). Normally, forward/backward configuration gives best configuration and highest accuracy, but it also has biggest distortion and is most difficult for matching. Normally customer uses forward/nadir or backward/nadir pair for ATE. The current version of ATE is handling images pair by pair. In next version, we will process three looking angles at the same time for better reliability and consistency.

Figure 10. Contours overlapped with ADS40 images, before (above) and after (below) object filtering

ADS40 normally has long strips with tremendous data flow and needs to be split into small blocks for efficient memory handling. Certain overlap between blocks is used to prevent block effect, which is discontinuity on block boundary and is very obvious on contour map.

Fig. 10 shows the contours from ATE, draped on an urban scene. The contours without object filtering reveal shapes of buildings, and contours after object filtering shows general trend of terrain.

Fig. 11 shows a DSM (without object filtering) and DTM (with object filtering) from ATE on ADS40 images. Most buildings and trees are removed and small variation of terrain is still preserved.
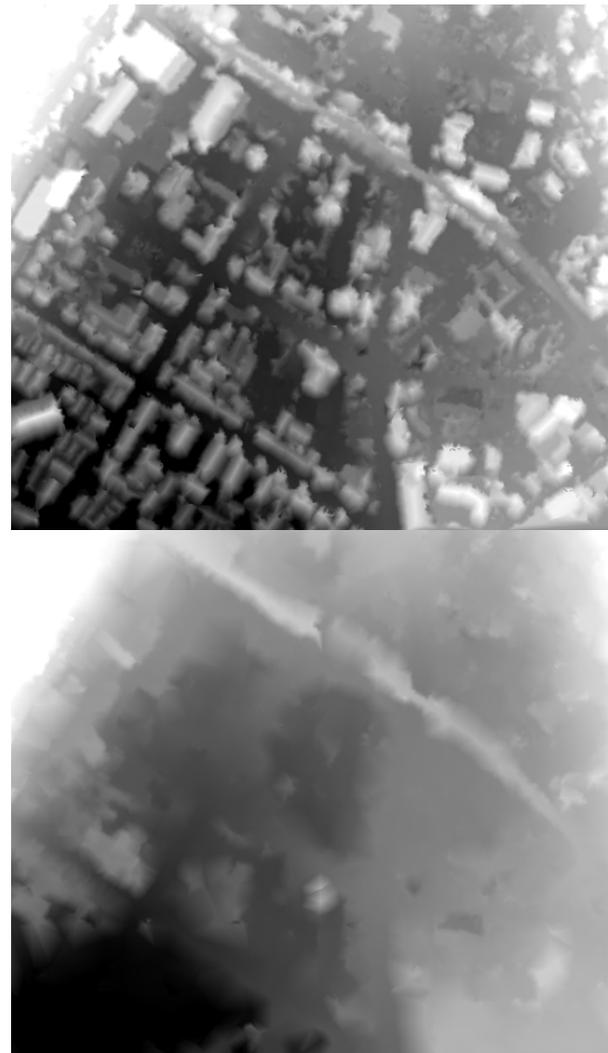
Figure 11. DSM (above) and DTM (below) from ADS40

4. SUMMARY

This paper introduced ERDAS's adaptive ATE and associated blunder detection techniques. The performance of ATE on frame, satellite, and ADS40 images are presented and discussed.

Generally, adaptive ATE assumes piecewise continuity. During matching with different image pyramid levels, ATE iteratively applies terrain range to limit search range and uses matched points to update terrain range. It is suitable for natural terrain. For urban scene, it needs object filtering for good performance.

The next generation of ERDAS ATE will address occlusion detection and pixelwise matching to improve matching on urban scenes.

REFERENCES

Hodgson, M., Bresnahan., P., 2004. Accuracy of Airborne Lidar-Derived Elevation: Empirical Assessment and Error Budget. *Photogrammetric Engineering and Remote Sensing*, 70(3), pp. 331-339.

James, S., Garey, G., Johnston, C., Haase, J., Heady, B., Kroenung, G., Little, J., 2006. The SRTM Data "Finishing" Process and Products. *Photogrammetric Engineering and Remote Sensing*, 72(3), pp. 237-247.
195–211.

Ma, R., 2005. DEM Generation and Building Detection from Lidar Data. *Photogrammetric Engineering and Remote Sensing*, 71(7), pp. 847–854.

Rao, J., 1972. Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.*, 67, pp. 112–115.

Zebedin, L., Klaus, A., Gruber, B., Karner, K., 2006. Towards 3D map generation from digital aerial images, *ISPRS Journal of Photogrammetry & Remote Sensing*, 60, pp 413–427.
Zhang, L., Gruen, A., 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry & Remote Sensing*, 60, pp