AN IMPROVED CLASSIFICATION APPROACH FOR LIDAR POINT CLOUDS ON TEXAS COASTAL AREAS

L. Su, **J. Gibeaut**
Harte Research Institute for Gulf of Mexico Studies, Texas A&M University - Corpus Christi, Corpus Christi, TX 78412, USA - (Li.Su, James.Gibeaut)@tamucc.edu

KEYWORDS: Lidar, Bare-earth DEM, Classification, Slope

ABSTRACT:

There are two challenges in classifying lidar points into ground and non-ground points on Texas coastal areas, which usually has a low-lying landform consisting of morphological features including dunes, tidal and river channels with levees, barren flats, buildings, and trees with varying cover density. The first is to remove buildings and trees meanwhile keeping seawall, dunes, levees and channels. The second is to remove bushes and grasses. In this paper, a novel classification approach based on slope and neighbor properties is designed to meet these challenges. The innovation of this approach is to first determine the most suitable post-spacing for a given lidar point dataset and then to generate a raster with the post-spacing. Slope thresholds for landscape objects, such as buildings and trees, are derived from their own characteristic size. The classification has three main steps. Step 1 – identifying potential areas by removing steep slope cells. The slope calculation and removal are repeated several times. This step may incorrectly create holes. Step 2– restoring holes: the lidar points falling in a potential area are identified into two classes: the correctly removed or not. The latter are restored. Step 3– identifying bushes and grasses based on slope. Classifications have been carried out with a lidar point dataset of Mustang Island, Texas (a 40-km long barrier island) with promising results.

1. INTRODUCTION

Texas coastal areas typically have a low-lying landform consisting of dunes, tidal and river channels with levees, barren flats, buildings, and trees with varying cover density (Lehman, et al. 2009), To produce a detailed and accurate topographic model for this area, lidar points must be classified into ground and non-ground points. This classification process presents two challenges: 1) the removal of buildings and trees from the data while keeping data that represents seawalls, dunes, levees and channels, and 2) the removal of data representing bushes and grasses.

The data acquired by aerial lidar surveys consists of the horizontal (x, y) and elevation (z) coordinates of reflective points on surface morphological features. These points form a three-dimensional cloud of points with irregular spacing (Wehr and Lohr, 1999). Elevation changes on the terrain surface produced by the point cloud are one of the most important pieces of information used in lidar points classification methods (Vosselman, 2000; Shan and Sampath, 2005; Zhang, et al., 2003). Elevation changes are measured as slope, which is the rate of rise or fall of a quantity against horizontal distance and is a measure of how steeply a surface inclines (Gallant and Wilson, 2000). Computation of slope is a prerequisite step for lidar points classification algorithms that evaluate sudden changes in the terrain surface.

This paper proposes a slope-based approach for classification of lidar points into ground and non-ground points in Texas coastal areas. An area's slope is usually calculated from a digital terrain model produced by rasterizing lidar points; however, rasterization may lead to a loss of precision (Axelsson, 1999). In order to reduce this loss, the proposed approach first determines the most suitable post-spacing for a given lidar point dataset and then uses this information to generate an accurate raster representation of the data. From this raster, slope thresholds for the morphological features, such as buildings and trees, are derived and based on characteric size of the features.

2. WORKING AREA

Mustang Island, a 40-km long by 4-km wide barrier island located between Corpus Christi and the Gulf of Mexico was selected as the study area. The island is centered at at 27°44' N, 97°08' W and it is oriented generally northeast-southwest, with the Gulf of Mexico to the east and south, and Corpus Christi Bay to the north and west. Mustang Island is a coastal barrier island with sand dunes anchored by sparse mats of vegetation. On the gulf side of the island there is a continuous, well-defined foredune ridge consisting of several rows of dunes. The height of these dunes may reach 11 meters, though 5 - 6 meters above sea level is average. On the bay side of the island, topographic change occurs in complicated patterns across relict geomorphic features, such as storm-surge channels, dunes, deflation flats, and washover and flood-tidal delta deposits. Sandflats and areas of low coppice mounds are also characteristic of this region. Upland and wetland areas on Mustang Island are mostly void of trees and bushes taller than 2 m, but scattered clumps do occur with low mesquite trees and salt-marsh grasses in the upland areas. Wetland vegetation on Mustang Island, other than mangrove areas, is dominated by low density plants that are less than 0.5 m tall.

During the summer of 2005, the Bureau of Economic Geology of the University of Texas at Austin acquired topographic lidar data of the study area. The lidar survey was flown at a speed of about 100 kts and at a height of 500 to 800 m above the ground with at least a 60 percent overlap between flight lines. The Optech model ALTM 1225 lidar instrument was set at a 25 kHz pulse repetition rate. The aircraft stayed within 30 km of a GPS base station, and a ground calibration target was surveyed during each flight. These survey parameters provided lidar data points with a vertical accuracy of about 10 cm on non-vegetated surfaces.

The algorithms presented in this paper were applied to this lidar dataset to filter buildings, trees, and relatively tall vegetation, (particularly mangrove) from the data.

3. METHOD

2.1. Workflow for Generating a DEM

Figure 2 describes the approach used to generate a bare-earth DEM for the study area from the lidar point cloud data:

3.2 Estimation of the Best Post-Spacing

The D8 algorithm is considered the best method for calculating slope in geospatial raster environments (Jones, 1998) so this method was used for this study. This algorithm uses two factors to calculate slope: elevation differences between neighboring cells and raster cell size. To obtain calculated slopes that are a good approximation of the natural land contours, the raster surface produced by interpolating lidar points must be as close to the actual land surface as possible. Therefore, when converting the data to a raster format from a given lidar point dataset, the raster cell size must be based on the post-spacing of the lidar points, and thus the method used to determine the post-spacing is critical.

To better understand the relationship between lidar post-spacing and slope accuracy, consider that for a lidar dataset, surface elevation at lidar points is known within the context of the quality of the dataset. However, the elevation is unknown for surfaces located between the lidar points; therefore, denser lidar point clouds give a more accurate portrayal of the characterized surface. Since lidar points are generally randomly distributed over surveyed areas, when the data is converted to a raster, some cells will not contain any lidar points and others may contain multiple points, depending on the raster cell size. Since each cell must contain a single elevation value, the elevation at each cell can be calculated by using interpolation methods such as inverse distance weight (IDW) on nearby lidar points, e.g. using the three closest lidar points within a search range of a certain radius.

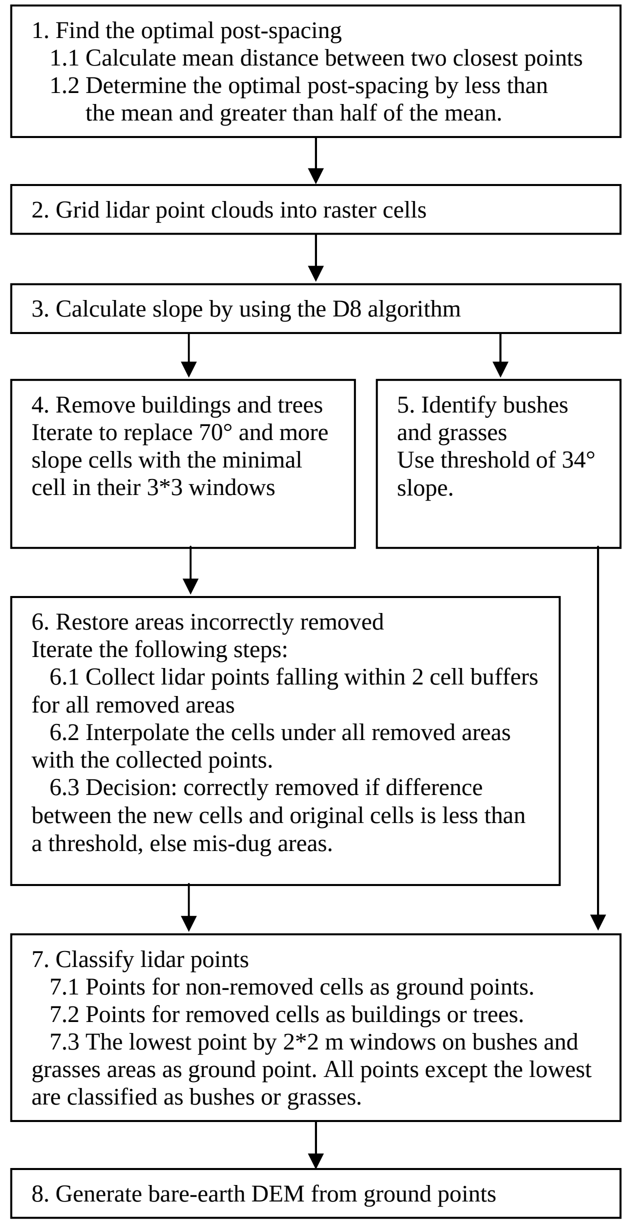

Figure 2. Workflow of the proposed approach

Therefore, accurately determining the post-spacing will result in better selection of a raster cell size, which in turn provides more accurate interpolation results and slope measurements.

Intuitively, the most suitable post-spacing should be related to the average distance between the two closest lidar points. This is confirmed in the following example that illustrates how slope accuracy depends on spacing for a given set of points. For the sake of simplicity, 2-dimensional points are used. In this example, There are six points with (x, y) coordinates (0.17, 3.0), (0.67, 3.0), (1.17, 3.0), (1.67, 0.0), (2.17, 0.0), and (2.67, 0.0), respectively, and the mean distance between the 2 closest lidar points is 0.50 m. Based on this information, IDW interpolation is used to generate elevations for rasters with cell sizes of 1.0, 0.5, 0.25 meter, respectively. The calculated slope measurements based on each of these rasters is illustrated in Table 1 and Figure 1.

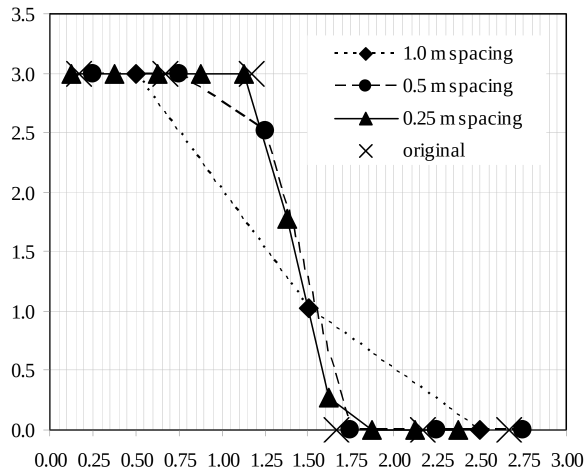

Figure 1. Original curve and interpolated curves

From Table 1 and Figure 1, it can be seen that smaller spacing gives a result closer to the slope of the line connecting the original points. The slope between the original points (1.17, 3.0) and (1.67, 0.0) is 81°. The curve is smoothed by 1 m spacing because a 1 m cell has a high probability of covering two points so the height of that cell will be an average of the covered points. The curve generated with 0.25 m spacing gives the closest result to the line connecting the original points. The reason is that 0.25 m spacing generates cells that contain none of the original points, so the height of these interim cells is interpolated from the values of nearby points. Using IDW interpolation, these interim cell values will fall on the line connecting the original points, for example, the sixth and seventh points, (1.375, 1.77) and (1.625, 0.27). The slope between these two points is same as the original slope.

The preceding example shows that any spacing less than the average distance between the two closest points is the most suitable spacing for a given point dataset. However, it should be noted that finer spacing leads to a larger computation load. The computation for finding the distance between the two closest points is $O(N^2)$ both space and time complexity for a dataset of *N* points.

1 m spacing		**0.5 m spacing**		**0.25 m spacing**	
(*x*, *y*)	slope	(*x*, *y*)	slope	(*x*, *y*)	slope
(0.50, 3.00)		(0.25, 3.00)		(0.125,3.00)	
(1.50, 1.02)	63º	(0.75, 3.00)	0º	(0.375, 3.00)	0º
(2.50, 0.00)	46º	(1.25, 2.52)	44º	(0.625, 3.00)	0º
		(1.75, 0.00)	79º	(0.875, 3.00)	0º
		(2.25, 0.00)	0º	(1.125, 3.00)	0º
		(2.75, 0.00)	0º	(1.375, 1.77)	79º
				(1.625, 0.27)	81º
				(1.875, 0.00)	50º
				(2.125, 0.00)	0º
				(2.375, 0.00)	0º

Table 1. Point spacing and slopes

3.3. Calculation of Characteristic Slopes

Each morphological feature has its own characteristic size and shape. Buildings typically have 2 m or taller walls. Trees usually are 1.5 m or more high. Bushes, such as mangroves, generally are 1 m or less in the study area. These heights can be converted to slopes for a given lidar dataset based on the average distance between the closest two points. Given the average distance is 0.5 m, the slopes are 76º, 72º and 64º for buildings (2 m), trees (1.5 m) and bushes such as mangroves (1 m), respectively.

On Texas coastal areas, dunes are the main morphological feature. Dunes are formed when sand is deposited by wind. Eventually the deposit becomes so steep that it collapses under its own weight. The collapsing sand comes to rest when it is reduced to a steepness that keeps the dune stable. This angle, usually about 30-34°, is called the angle of repose (McKee, 1979). In other words, sandy coastal areas typically have 34° or less topographic fluctuation. 34° is 0.34 m rise on a 0.5 m run, or 0.20 m rise on a 0.3 m run. If vegetation in the area is 0.34 m or less in height, there is no way to differentiate low vegetation and topographic changes base only on heights.

3.4. Removing Buildings and Tall Vegetation such as Trees

To correct for buildings and tall vegetation, the data removal algorithm replaces each cell having a slope 70º or greater with the minimum cell height from within the 3*3 neighborhood window. In order to remove large buildings such as big hotels, the 70º removal operation may be repeated many times. Thirty five repetitions will cover an area with a 10.5 m radius at a cell size of 0.3 m, which is enough to remove all buildings in the study area. However, the removal algorithm also has the potential to misinterpret ground areas with high slope as buildings or trees, resulting in large areas that appear to be large "dug out" holes in the processed data.

3.5. Restoring "Dug Out" Areas

Areas that were misinterpreted by the removal algorithm must be restored to the data. The main reason for the misinterpretation is that the algorithm always replaces the steep cells (slope 70º or more) with the minimum elevation from its 3*3 neighbor. Since the tops of buildings and trees are typically considerably higher than the surrounding ground, and the buildings and trees usually have steep slopes on all sides, the removal will be carried out correctly. However, removal misinterpretation can occur for morphological features that have one steep side such as channels. Channels, which are narrow waterways between two close landmasses, typically have steep sides facing toward the water, which is at a lower elevation than the landmass surfaces. Thus the removal algorithm may excavate areas from the steep sides that extend into the landmasses.

The restoration routine is based on the assumption that buildings and trees have steep slopes on all sides, while other misinterpreted features do not. The routine consists of three components: 1) collecting nearby lidar points around the removed areas; 2) interpolating surfaces from the collected points; and 3) restoring the original points if the difference between the new surfaces and the original surfaces is greater than a certain threshold.

3.6. Removing Medium-Height Vegetation

For non-penetrable dense vegetation regions, lidar points will probably look like ground. In less dense areas, the laser pulses may penetrate the vegetation canopy resulting in. rugged appearing surfaces. Slopes of 34° or greater may be obtained from these surfaces because laser reflections may be produced from points at different height layers within the canopy. Therefore, Medium-height vegetation, such as bushes, is indicated by slopes of 35º or more. In order to identify this vegetation, a replacing operation is repeated two times with a 35° slope threshold. This operation replaces cells having 35º or greater slopes with the minimum cell from within its 3*3 neighborhood window, then- re-calculates new slopes.

Lots of no data areas will be produced if medium-height vegetation such as bushes is simply removed. If the vegetation is not too dense, laser pulses have chances to hit ground through the canopy. The improvement is to select minimal elevation point by 2*2 m or 3*3 m windows. The relatively tall vegetation such as bushes usually has smaller individual size relative to tall vegetation such as trees. For sparse area, ground points can be definitely found. For very dense areas, some points lower than the bush/grass tops will be found. Obviously these points will be closer to ground.

4. RESULTS AND DISCUSSIONS

The working area, Mustang Island, involves 18 USGS 1:24,000 quadrangles. They are Cinw (NE, NW, SE, SW), Cisw (NE, NW, SE, SW), Osoc (NE-SE), Pita (NE, SE), Porta (NE, NW, SE, SW), Porting(NE, SE) S-Cinw (NW) and Sbi (NE). Based on statistics on distance between the two closest lidar points (minimum 0.01 m, mean 0.56 m, median 0.55 m, and mode 0.54 m) the optimal post-spacing is set to 0.3 m. Following the proposed workflow, a bare-earth DEM was produced after removing tall features such as buildings and trees, and medium-height vegetation such as bushes and mangroves, while keeping seawall, dunes, levees and channels.

To show detailed results of the proposed algorithm, Port Aransas quadrangle (Porta - NW) is examined here. Figure 2 shows the slope results for a raster with 0.3 m cells located in the lower part of the quadrangle map. Figure 3 shows two enlarged areas of Figure 2 (upper left corner and lower right corner). The pictures demonstrate that beaches and upland areas have a slope of 10° or less and dunes are generally less than 34°. Tall features such as buildings typically have a slope of 70° or more on their edges, and these edges enclose the areas inside. Other tall features with a 70° or greater slope are relatively small features such as trees or light poles. When 34° is set as a threshold to identify medium-height vegetation or dunes, misclassification may occur. Vegetated dunes may have steeper slopes since vegetation holds the sand in place.

Figure 4 shows results of the removal and restoration algorithms. In this example, the removal algorithm has correctly removed buildings and tall vegetation, while misinterpretation by the algorithm produced several incorrect large and small holes. The restoring routine filled the large holes and also some thin regions surrounding buildings and tall vegetation giving a reasonable final result for the removal process

Figure 2. Slope with 0.3 m cell on lower Port Aransas quadrangle.

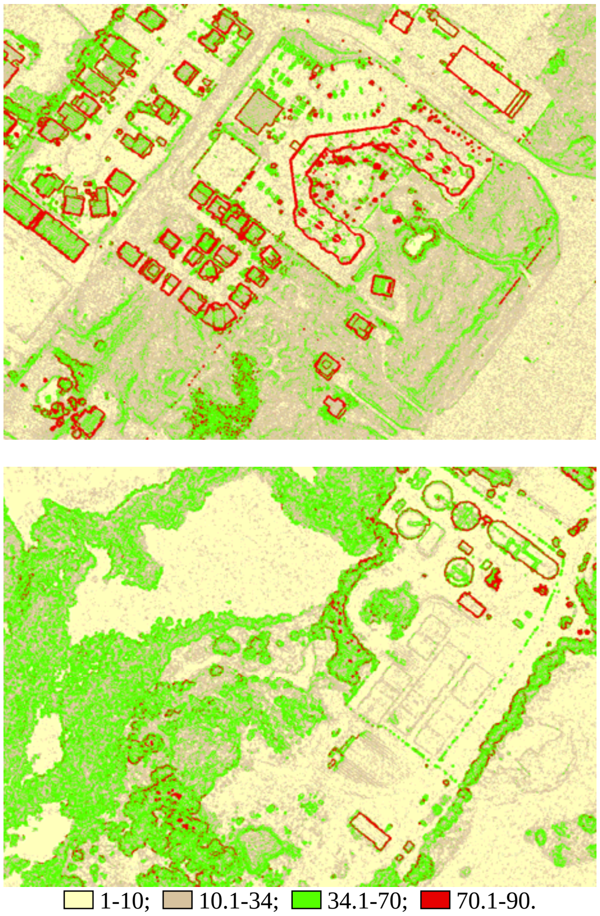

Figure 3. Two enlarged areas of Port Aransas quadrangle.

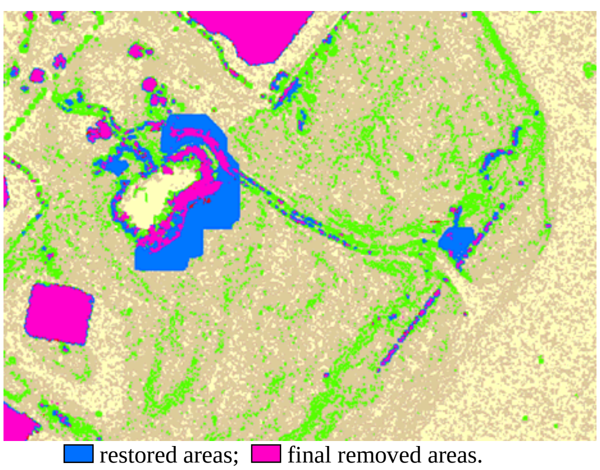

Figure 4. Detailed areas of removing and restoring.

5. CONCLUSIONS

A Mustang Island DEM produced with the proposed approach has been used to support projects for Coastal Bend Bays & Estuaries Program (Gibeaut, et al., 2010).

The case study of Mustang Island illustrates the capability of this algorithm. The results constitute a proof-of-concept for the proposed method.

REFERENCES

Jones, K. H., 1998. A comparison of algorithms used to compute hill slope as a property of the DTM. *Computer & Geosciences*, 24(4):315-323.

Chow, T. E., and Hodgson, E. M., 2009. Effects of lidar post-spacing and DEM resolution to mean slope estimation. *International Journal of Geographical Information Science*. 23(10):1277–1295.

Lehman, R. L., O'Brien, R., and White, T., 2009. Plants of the Texas Coastal Bend. Texas A&M University Press.

Baltsavias, E.P., 1999. Airborne laser scanning: basic relations and formulas, *ISPRS Journal of Photogrammetry & Remote Sensing*, 54:199–214.

Wehr, A. and U. Lohr, 1999, Airborne laser scanning - an introduction and overview, *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3):68-82.

Gallant, J. C., and Wilson, J.P., 2000. Primary topographic attributes, *Terrain Analysis: Principles and Applications*, Wilson J.P. and Gallant, J. C. (Eds.), New York: John Wiley & Sons, 479p.

Vosselman, G., 2000. Slope based filtering of laser altimetry data, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pt. B4, 33:958–964.

Shan, J., Sampath, A., 2005. Urban DEM Generation from Raw Lidar Data: A Labeling Algorithm and its Performance, *Photogrammetric Engineering and Remote Sensing*, 71(2): 217-226.

Zhang, K., Chen, S.C., Whitman, D., Shyu, M.L., Yan, J. and Zhang, C., 2003. A Progressive Morphological Filter for Removing Nonground Measurements From Airborne LIDAR Data, *IEEE Transactions on Geoscience and Remote Sensing*, 41(4):872-882.

Axelsson, P.E., 1999. Processing of laser scanner data – Algorithms and applications, *ISPRS Journal of Photogrammetry and Remote. Sensing*, 54(2–3):138–147.

McKee, E., 1979. An introduction to the study of global sand seas. In *A Study of Global Sand Seas*, E. McKee (eds.), pp.1- 20. Washington, U. S. Geological Survey Paper 1052.

Gibeaut, J.C., Barraza, E., Radosavljevic, B., 2010. Estuarine Wetland Habitat Transition Induced by Relative Sea-Level Rise on Mustang and North Padre Islands, Texas: Phase 1. *Publication CBBEP (Coastal Bend Bays & Estuaries Program) – 64,* Project Number – 0822.