A Novel Method for TLS Data Reduction Using Error-Modeling
Michael Gleason, John Pitlick, Barbara P. Buttenfield
University of Colorado, Boulder

Motivation and Objectives

Despite the proven advantages of Terrestrial Laser Scanning (TLS) for detailed geomorphologic research, raw TLS data demonstrate multiple traits that make them poorly suited for immediate use in geomorphologic applications. For example, the high density of survey points acquired by TLS results in characteristics that are largely undevelopable and computationally dense. These datasets, which stem to mean software programs used for terrain analysis (Buskley et al. 2008; Hergarten 2007), contain points uniformly spaced in three dimensions, which can be converted to point clouds, which are sparsely distributed across the scene (Petos et al., 2008). A general model of raw TLS data relates to their accuracy and errors in positional coordinates, scale accuracy and precision (RIEGL Laser Measurement Systems 2009; Hodge et al. 2008). The standard deviation of most error distributions is very small, or large magnitude (Hergarten and Hershfield 2007) or, alternatively, error contributions to the entire range are rare, errors of 10 cm or more may be common (Heritage and Hetherington 2007). Despite the reduction of errors: points with elevation values higher or lower than three standard deviations from the mean are identified as "extreme outliers" and removed. From the remaining points, the median elevation value is used to calculate the "central" elevation measurement. This median value is then used to build the cylinder: points within the cylinder are defined as those that have an elevation measurement within 2 cm of the median, while points beyond this range are outside the cylinder. The script then calculates a mean 3-D point from the points within the cylinder and moves on to the next search location. After repetition of this process for each data point, all of the mean 3-D points are output to a new shapefile.

Methods IMPLEMENTATION

The data reduction method was implemented as a Python script using Gdal/ogr. The method operates directly on Esri shapefiles, however, efficiency is vastly improved by using a quickshift spatial indexing, an ogr data structure (C). Figure 7 illustrates the workflow implemented by the script. For efficiency and simplicity, the script uses a conservative approximation of the modeled "error torus": a cylinder of height 4 cm and radius 8 cm. The script begins by populating a regularly-spaced set of search locations throughout the extent of the model surface as an optimized spacing of 8 cm. For each search location, a circle of radius 2 cm is calculated using a spatial filter and the mean and standard deviation of elevation values within the circle are calculated. These values are used to conduct an initial removal of gross errors with elevation values higher or lower than three standard deviations from the mean are identified as "extreme outliers" and removed. From the remaining points, the median elevation value is calculated to identify the "central" elevation measurement. This median value is then used to build the cylinder: points within the cylinder are defined as those that have an elevation measurement within 2 cm of the median, while points beyond this range are outside the cylinder. The script then calculates a mean 3-D point from the points within the cylinder and moves on to the next search location. After repetition of this process for each data point, all of the mean 3-D points are output to a new shapefile.

Preliminary Results

Preliminary results suggest that the developed method is highly effective at reducing TLS data volume, maintaining heterogeneity in point distribution, and enabling researchers to achieve greater efficiencies and accuracy in geomorphologic research, raw TLS data demonstrate multiple traits that make them poorly suited for immediate use in geomorphologic applications. For example, the high density of survey points acquired by TLS results in characteristics that are largely undevelopable and computationally dense. These datasets, which stem to mean software programs used for terrain analysis (Buskley et al. 2008; Hergarten 2007), contain points uniformly spaced in three dimensions, which can be converted to point clouds, which are sparsely distributed across the scene (Petos et al., 2008). A general model of raw TLS data relates to their accuracy and errors in positional coordinates, scale accuracy and precision (RIEGL Laser Measurement Systems 2009; Hodge et al. 2008). The standard deviation of most error distributions is very small, or large magnitude (Hergarten and Hershfield 2007) or, alternatively, error contributions to the entire range are rare, errors of 10 cm or more may be common (Heritage and Hetherington 2007). Despite the reduction of errors: points with elevation values higher or lower than three standard deviations from the mean are identified as "extreme outliers" and removed. From the remaining points, the median elevation value is calculated to identify the "central" elevation measurement. This median value is then used to build the cylinder: points within the cylinder are defined as those that have an elevation measurement within 2 cm of the median, while points beyond this range are outside the cylinder. The script then calculates a mean 3-D point from the points within the cylinder and moves on to the next search location. After repetition of this process for each data point, all of the mean 3-D points are output to a new shapefile.

Future Work

A comprehensive evaluation of the developed data reduction method is currently underway and expected to be completed by May 2012. This evaluation seeks to address the research questions listed below:

- How effective is the developed filter at addressing each of the three primary problems identified in raw TLS data (e.g., errors of 10 cm or more may be common, errors of 10 cm or more may be common, errors of 10 cm or more may be common)?
- How does the filtering process not only degrade the terrain-modeling capabilities of the data, particularly at resolutions greater than a decimeter?
- How can the filtering process be incorporated into the existing workflow for data processing and analysis in TLS applications?

The results presented on this poster represent an initial investigation into Question 1 and 2. Future investigations will address statistical methods (e.g., correlation, regression, etc.), geostatistics (e.g., experimental semivariograms), and geomorphometric analysis (e.g., slope, aspect) in a manner that addresses the three research questions described above.