Processing LiDAR Datasets

By Robert J. Gammon, Executive Consultant, SG&I Functional Design, Intergraph Corp.

Light Detection And Ranging (LiDAR) is a technology that has great potential for a wide variety of applications in the realms of Defence and Intelligence, Government, Transportation and Utilities. However, managing and processing these data in a meaningful and timely fashion is proving to be a challenge for the geographic information system (GIS) and Remote Sensing community. Specifically, LiDAR platforms generate extremely large datasets, making the management and processing aspects of using these data a key element in leveraging their contents.

LiDAR platforms can generate 50-100 thousand points per second, and this level of data acquisition can result in tens of billions of points spread across thousands of files. To provide context, the Ohio Department of Transportation (DOT) in the U.S. recently acquired a set of LAS[1] formatted LiDAR files for the state (covering approximately 10,676,560 hectares). These datasets consisted of approximately 23,600 individual 1.5 KM by 1.5 KM LAS formatted LiDAR files, each containing approximately 750,000 plus records, or in total, in excess of 17 billion points.

As can be seen by the Ohio example, simply managing this number of files across an enterprise can prove trying, and as a result GIS/Remote Sensing vendors such as Intergraph have enhanced or are in the process of enhancing their technologies to easily allow for the integration/administration of LiDAR files into GIS/Remote Sensing workflows. Intergraph's TerraShare product, for example, contains modules specifically designed for managing and accessing large collections of LAS formatted LiDAR files.

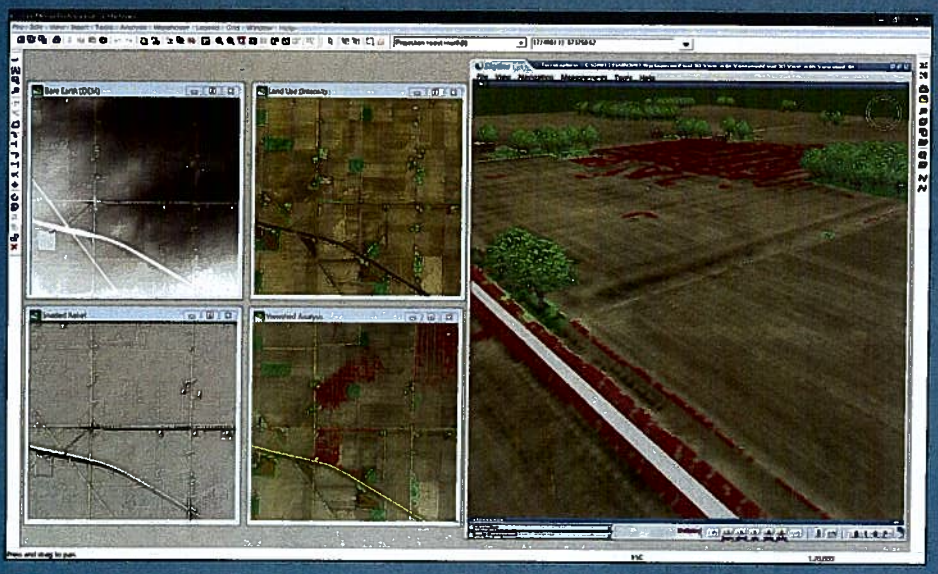
Managing LiDAR files is only one aspect of leveraging LiDAR datasets. Once a collection of files has been catalogued for easy retrieval, processing becomes the next impediment. Processing tens of millions (possibly billions) of points to produce meaningful results typically requires smaller areas of interest, data thinning or a combination of both approaches.

Smaller areas of interest may allow a user to include all of the points in a LiDAR dataset, but users are limited by the area that can be covered and hence processed. Assuming the data volume noted above and an area of interest covering 15 KM by 15 KM, the user would need a system that can collectively process 100 individual LAS formatted LiDAR files or approximately 75 million points. Processing 75 million points is possible; however, the number of points grows dramatically as the area of interest increases incrementally. With this in mind, some users are looking to grid-based systems for processing LiDAR datasets. Grid-based systems such as Intergraph's GeoMedia have point collision detection options that allow the user to thin data during the conversion process (i.e. from vector point to raster grid).

Point collision detection options allow users to rasterize point-based datasets effectively. For example, when multiple LiDAR points fall spatially into a shared/common cell, the contents (or value) of these points (e.g. Z value, Intensity or Color) can be processed using a statistical measure such as Mean, Median, Mode, Maximum or Minimum.

Ideally, one does not want to apply such measures to their data on every occasion; however, when larger areas of interest need to be processed, this approach can still produce accurate and detailed Digital Elevation Models (DEM). It should be noted that LAS formatted files include point attribution. These attributions allow users to create bare earth terrain models, terrain models with vegetation and/or buildings and "land-use" like maps and ultimately 3D models (as illustrated in the example provided below).

Leveraging LiDAR datasets to their fullest requires thought to create balance between data volume and accuracy. With systems of today, even the most robust technologies cannot always effectively process these datasets in their native form. Users invariably need to look at their data needs and determine how best to meet these needs (i.e. choose an area of interest and resolution that best fits the need of the project). With time, as technologies improve, users will be able to process more and more data in a timely and efficient manner; until this time, users will have to process these data in smaller chunks.



Note 1 - The LAS file format is a public file format that is commonly used for the interchange of LiDAR data. This file format was developed in concert with the American Society for Photogrammetry and Remote Sensing (ASPRS) and a committee of photogrammetric and remote sensing professionals, academic institutions, government agencies and private sector organizations.