Current capabilities and community needs for software tools and educational resources for use with LiDAR high resolution topography data

August 8, 2008

J R. Arrowsmith (Arizona State University; ramon.arrowsmith@asu.edu)
Nancy Glenn (Idaho State University)
Christopher J. Crosby (San Diego SuperComputer Center)
Eric Cowgill (UC Davis)

Introduction

LiDAR (Light Detection and Ranging) data are transforming diverse facets of earth surface studies, including geomorphology, hazard assessments, forestry, and fish ecology. The major increase in spatial resolution (1 to 2 orders of magnitude better than what is currently typically available—e.g., USGS National Elevation Dataset) LiDAR data enable users to identify, measure, and quantitatively characterize features in the landscape such as earthquake offsets, ungullied hillslopes, or tree canopy height/density at the scales necessary for analysis. Although the LiDAR data make unprecedented opportunities, efficient quantitative analysis of these data is challenging because of their size and heterogeneity. Current Airborne Laser Scanning (ALS) methods typically yield several point measurements per square meter over areas encompassing 10s to 1000s of km^2. Although Terrestrial Laser Scanning (TLS) typically covers smaller areas (0.1 to 1 km^2), current approaches sample the surface at 1 or 2 orders of magnitude higher spatial resolution than airborne LiDAR. Thus both ALS and TLS produce data sets containing 100s of millions to billions of individual measurements to be processed and analyzed. The use of repeat LiDAR scans in time-series analysis for change detection and the coupling of imagery with the scanning processes further magnifies these computational needs.

Users working with LiDAR data encounter two main data types: 1) the attributed point cloud measured by the laser scanner, and 2) high-resolution (0.1 to 1 m/pixel) DEMs (digital elevation models) derived from the point measurements. **Point clouds** comprise a set of measured positions in three-dimensional space of point locations on a scanned surface (e.g., bare ground, trees, roads, buildings, etc.). Points are typically attributed with return intensity, time, and number, in the case of multiple returns for a single shot. In addition, they can be classified as ground, structure, vegetation, etc. during post-processing. LiDAR point clouds are three-dimensional data because they can contain multiple height measurements for a given X, Y position. The main sources of uncertainty in the point locations include accuracy of the laser scanner, Inertial Measurement Unit (IMU; in the airborne case) and GPS system, registration of individual scans, and georeferencing of the resulting point cloud. Importantly, point cloud data are scattered (i.e., non-gridded), so that both point spacing and point density (number of points per m^2) vary within the scan. Current topographic analysis tools are limited in their ability to work with such scattered data, although some software tools (most commonly used with TLS data), perform 3D tessellation and texture mapping on the point cloud.

DEMs are gridded height fields in which elevations are represented on a 2 dimensional map grid with constant spacing and only one height value at each horizontal grid node. As such, they are "2.5D" representations of true 3D surfaces. High resolution DEMs are generated from LiDAR data by gridding and interpolating the measured point cloud. Thus, in addition to the uncertainties in point locations described above, they contain an additional level of uncertainty because points in the DEM do not necessarily lie on the true surface. Conversion of the point cloud to digital elevation model is a destructive step, in that the original cloud cannot be regenerated from the derived DEM. Many users work with DEMs because current implementations of a number of important topographic analysis tools, such as slope and flow direction, presume gridded data.

Software tools

The software needs of the community will vary due to the differences in data types and the diversity of user needs. The major LiDAR workflow steps of direct interest to most end users begin with a point cloud that has already been merged from individual scans, georeferenced, and attributed with point intensity and return number.

Software tools, tutorials, and test data sets are needed to support both general tasks common to many users (point cloud analysis, DEM generation, and QA/QC) and application-specific tasks (stream-profile analysis, neotectonic mapping, etc.). Depending upon the application, DEM generation may occur at the end of a processing workflow to preserve the accuracy of the point data for transects, height analysis, etc. Regardless of the workflow, software tools should be diverse enough to adapt to user needs.

When working with LiDAR data, most people use multiple pieces of software because each performs certain tasks well. All GIS or remote sensing-type software packages with raster support will allow some amount of LiDAR DEM analysis and visualization. The US Army Corps of Engineers prepared a comprehensive survey of Terrain Visualization Software: http://www.tec.army.mil/TD/tvd/survey/survey_toc.html. As Table 1 indicates, current commonly used tools can vary from expensive commercial packages (e.g., Polyworks, TerraScan), to less costly commercial resources that are often site licensed in academic environments (e.g., ArcGIS, ENVI/IDL, MATLAB) to free extensions or codes for commercial packages (e.g., LiDAR Tools, GeomorphTools), to free open source software (e.g., GRASS, Points2Grid, LViz; LidarViewer, Real-time interactive Mapping System).

ArcGIS is the principal environment for 2.5D-based cartography and data integration for many earth scientists, although its 3D rendering capability is limited with the large files that characterize LiDAR DEMs. Nancy Glenn and colleagues from Idaho State University have developed a free set of LiDAR Tools which is an extension to ENVI (http://geology.isu.edu/BCAL/tools/EnviTools/index.html) (Glenn et al., 2006; Streutker and Glenn, 2006). Kelin Whipple (Arizona State University) and colleagues have developed free extensions to MATLAB and ArcGIS to extract stream profiles from DEMs and analyze their steepness index and concavity (http://www.geomorphtools.org/Tools/StPro/StPro.htm). George Hilley (Stanford) has created and released a large number of MATLAB functions that also enable analysis of DEMs in slope-area space. Members of the UC Davis W.M. Keck Center for Active Visualization in the Earth Sciences (KeckCAVES) (http://www.keckcaves.org/) have created tools to allow real-time interactive visual analysis of massive point cloud (LidarViewer) and DEM (RIMS) data (Bernardin et al., 2006; Bernardin et al., 2008; Kellogg et al., 2008) as the first steps in developing a comprehensive point cloud analysis tool (Kreylos et al., 2007; Gold et al., 2007).

Despite these many tools, there remains a considerable need for expansion of software resources that can handle the challenges posed by LiDAR point cloud and DEM data. For example, an open-source toolkit for various platforms (Matlab, C, ENVI/IDL, etc.) for basic operations is a valuable target that was identified at the June 2008 Studying Earth Surface Processes with High-Resolution Topographic Data Workshop. Furthermore, software tools that are well linked with on-line data sources or archives can take advantage of significant computational resources beyond the user's desktops. A comprehensive software scheme for high resolution topography data should include a field computing (e.g., mobile, PDA, tablet) to desktop to grid- or "cloud"-based architecture.

Many LiDAR point clouds are initially acquired as community datasets, and all are valuable in many ways beyond the original motivation for their collection. Some are valuable as iconic datasets on which important research has been performed (e.g., Roering and others Oregon Coast Range LiDAR DEMs, etc.). Others are valuable because they serve the needs of another scientific discipline (e.g. data being valued by the ecology community for its representation of vegetative canopy may be useful for earth scientists). Data collected and preprocessed by commercial vendors and NCALM (National Center for Airborne Laser Mapping) are typically provided to the purchaser (individual PI, state or federal agency, UNAVCO) on DVD or portable hard drive. The degree to which these data are then made available to the general community and the format in which they are provided is currently quite variable. Although no single data clearinghouse for community data has been established, there are several sites where such data may be downloaded and processed. The USGS CLICK effort (http://lidar.cr.usgs.gov/) provides data primarily in raw form [unclassified LAS] as provided by the dataset owner. The CLICK site provides no data processing and minimal QA/QC. Alternatively, Web-based LiDAR data access, data management, and data processing has been pioneered by the GEON (GEON LiDAR Workflow; Crosby, et al., 2006; Jaeger-Frank, 2006) and NOAA's LDART tool (http://www.csc.noaa.gov/crs/tcm/about_ldart.html). The GEON LiDAR system begins with user-defined selection of a subset of point data and ends with download (including dynamically generated metadata) and visualization of DEMs and derived products. Users perform point cloud data selection, interactive DEM generation and analysis, and visualization all from an internet-based portal. Users may experiment with DEM resolution and DEM generation algorithms so as to optimize terrain models for their application. By using cyberinfrastructure resources, this approach allows users to carry out computationally intensive LiDAR data processing without having appropriate resources locally. This system gives users access to datasets of interest and basic tools to process and interact with the data. But, at some point, the user will need to process the data independently for their specific science objectives.

Although reprocessing the point cloud to improve positional accuracy is generally not a high priority for the majority of users, the legacy of these datasets can be greatly extended if all raw LiDAR measurements are archived. Archiving allows the opportunity to reprocess the data in the future as algorithms develop. Establishment of such archives is critical to support change detection studies.

Additional needs at the community level include:

- A single-point internet-based clearing house for LiDAR point cloud and DEM data that makes it simple for dataset holders to make their data available to the user community and for users to discover data of scientific interest. All publically-funded LiDAR missions should be required to post data on this site within a specified timeframe (1-2 years). Ideally, this system should also provide tools for users to perform basic data processing, analysis and visualization tasks (e.g. the GEON LiDAR system). The site should provide comprehensive metadata characterizing each data set and all processing steps used to produce derived products such as an attributed, classified, merged and georeferenced point cloud or a bare-earth DEM. Standards for data delivery are included in this requirement.
- A data archive (possibly combined with above) for the preservation of all raw measurements collected from the laser scanner, inertial reference system, and GPS receivers during a LiDAR survey. LiDAR data are typically valuable in many ways beyond the original motivation for their collection. Thus, a data archive will ensure maximum utility and longevity for these data by making it possible to reprocess the data for different applications or as community standards and algorithms evolve. The archive is also an important resource should errors be discovered in current processing approaches.
- Support for development of algorithms for conducting topographic analyses directly on the scattered point cloud data. Common operations include calculations of slope, slope-aspect, stream profiles, catchment areas, and topographic roughness and curvature. Although such analyses are widely used, current implementations generally presume gridded data. Performing such operations directly on the point cloud is appealing for several reasons. It removes the processing steps required to generate the DEM. The operations should be more accurate because they are performed directly on the measured data, rather than a model of the surface. Analysis of the point cloud directly removes the need to discard or interpolate data in areas of high or low measurement density, respectively.
- Format conversion capability: no one software solution will be achieved for the entire community interested in these data. Therefore, delivery in and conversion between common file formats for both point data (LAS, ASCII) and DEM data (ASCII grid, binary grid, etc.) is necessary. Data from publicly supported data acquisitions should be released in such common non-proprietary formats.
- The community would most benefit from a Wiki or similar system where users could post tools, tutorials, scripts etc. that they have found useful in building LiDAR processing workflows to address their science goals. A community forum for idea and method exchange. Existing venues that could be adopted by the community include: the HydroVent (http://**pasternack**.ucdavis.edu/hydrovent.html), GEON Forums (http://www.geongrid.org), the USGS CLICK Bulletin Board

(http://lidar.cr.usgs.gov/), or email listserves (TLS listserv from U. New Mexico, lidar@asu.edu, GEOMORPH-L@listserv.boisestate.edu). The GEON LiDAR team (led by Chris Crosby) is building the OpenTopography Portal (http://www.opentopography.org/) would be a logical place to host such a Wiki and some of the other community-based functionality we have identified.

Development of community-oriented data systems and software libraries can be enhanced with external support for collaboration with computer scientists and employment of professional programmers to build a framework on top of which the community could develop specific tools and workflows. Support for such an effort could come from NASA or NSF collaborative geoscience initiatives. Such support will be particularly important for developing new algorithms to handle quantitative analysis of point data, because a number of these algorithmic challenges are on the frontier of scientific computing.

Educational resources

Training on technology, tools, scientific and management applications is an area in which significant impacts can be made. Enabling students, scientists, and managers to analyze their data independently and for science/management-specific needs will provide for improved application. Recent topography and LiDAR-oriented workshops were sold out (2007 Geological Society of America Meeting: *New Tools for Quantitative Geomorphology: Extraction and Interpretation of Stream Profiles from Digital Topographic Data* & *Processing* and *Analysis of GeoEarthscope and Other Community LiDAR Topography Datasets* http://www.geosociety.org/meetings/2007/cw_gsa.htm and UNAVCO: http://www.unavco.org/edu_outreach/uscs/2008/LiDAR_Course_2008.html). The demand for data and knowledge on how to handle and analyze high-resolution topography is very high. Such 1-2 day courses with 20-30 people are one of the most effective mechanisms for the engagement of the communities interested in the data and for the propagation of the scientific discoveries and enhanced management that come from their analysis.

In addition to training workshops, documentation via web-based tutorials and curricula needs to be created and/or improved upon. Documentation will not only provide users the knowledge about the tools but provide an opportunity for the tools to be enhanced by the community.

Finally, free, quick and easy tools for visualization of LiDAR data within a widely used system such as Google Earth (see here: http://www.cs.unc.edu/~isenburg/googleearth/) can provide education and outreach beyond the scientific community. Visualizing LiDAR data in Google Earth can provide opportunities for managers to understand the value in high resolution topographic data, further promoting its use.

Table 1: Selected software tools commonly used for LiDAR point and DEM data. The LiDAR community needs some common functionality in both web-based and desktop applications and then specific tools for both platforms. These tools need to be free and open source to the extent possible.

Name	Provider	Comercial vs. free	link
Points2Grid	GEON	Free	http://lidar.asu.edu/points2grid.html
LViz	GEON	Free	http://lidar.asu.edu/LViz.html
LidarViewer	KeckCAVES	Free	http://keckcaves.org/software/lidar/index.html
Real-time Interactive Mapping System	KeckCAVES	Free	http://keckcaves.org/software/RIMSG3/
GRASS		Free	grass.itc.it
LiDAR Tools (ENVI extensions)	Nancy Glenn & BCAL	Free on commercial	http://geology.isu.edu/BCAL/tools/EnviTools/index.html
MATLAB extensions	George Hilley	Free on commercial	
GeomorphTools (MATLAB+ ArcGIS extensions)	Geomorph Tools & Kelin Whipple	Free on commercial	http://www.geomorphtools.org/Tools/StPro/StPro.htm
ArcGIS	ESRI	Commercial (typical academic site license)	www.esri.com
ENVI/IDL	ITTvis	Commercial (typical academic site license)	www.ittvis.com/envi
MATLAB	Mathworks	Commercial (typical academic site license)	www.mathworks.com
Polyworks	InnovMetric	Commercial	www.innovmetric.com
TerraScan	TerraSolid	Commercial	www.terrasolid.fi

Acknowledgements

We formed this writing group at the suggestion of co-author Nancy Glenn at the June 2008 Studying Earth Surface Processes with High-Resolution Topographic Data Workshop. Thank you very much to Dorothy Merritts and colleagues for leading that workshop and to all the participants for a very stimulating set of presentations and discussions. Thank you to Bill Dietrich for encouraging us to write something.

References

Bernardin, T., Cowgill, E., Gold, R., Hamann, B., Kreylos, O., and Schmitt, A., 2006, Interactive mapping on 3-D terrain models: Geochemistry, Geophysics, Geosystems, v. 7, p. Q10013, doi:10.1029/2006GC001335.

Bernardin, T., Cowgill, E. S., Gold, R. D., Hamann, B., Kreylos, O., and Schmitt, A., 2008, Real-time terrain mapping, *in* Hagen, H., ed., Scientific Visualization: Challenges for the Future: Los Alamitos, California, IEEE Computer Society Press.

Crosby, C.J., Arrowsmith, J R., Frank, E., Nandigam, V., Kim, H.S., Conner, J., Memon, A., Baru, C., 2006, Enhanced Access to High-Resolution LiDAR Topography through Cyberinfrastructure- Based Data Distribution and Processing, Eos Trans. AGU, 87(52), Fall Meet. Suppl., Abstract IN41C-04.

Efrat Jaeger-Frank, Christopher J. Crosby, Ashraf Memon, Viswanath Nandigam, J. Ramon Arrowsmith, Jeffery Conner, Ilkay Altintas, Chaitan Baru, 2006, A Three Tier Architecture for LiDAR Interpolation and Analysis, Lecture Notes in Computer Science, Volume 3993, p. 920-927, DOI: 10.1007/11758532_123.

Streutker, D. and Glenn, N., 2006. LiDAR measurement of sagebrush steppe vegetation heights. *Remote Sensing of Environment*, 102, 135-145.

Glenn, N.F., Streutker, D., Chadwick, J., Thackray, G., Dorsch, S., 2006. Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology*, 73, 131-148.

Gold, P., Gold, R., Cowgill, E., Kreylos, O., and Hamann, B., 2007, Efficient, off-grid LiDAR scanning of remote field sites: Eos Transactions, American Gephysical Union, v. 88, no. 52, Abstract G51B-0435.

Kellogg, L. H., Bawden, G. W., Bernardin, T., Billen, M., Cowgill, E., Hamann, B., Jadamec, M., Kreylos, O., Staadt, O., and Sumner, D., 2008, Interactive Visualization to Advance Earthquake Simulation: Pure and Applied Geophysics, v. 165, no. 3/4, p. 621-633.

Kreylos, O., Bawden, G. W., and Kellogg, L. H., 2007, New Visualization Techniques to Analyze Ultra-High Resolution Three- and Four-Dimensional Airborne and Tripod LiDAR Point-Cloud Data: Eos Transactions, American Gephysical Union, v. 88, no. 52, Abstract G51B-0434.