Enterprise Geospatial Production

by Lewis Graham

Our company was formed several years ago based on some late night sessions brainstorming an industry (geospatial data production) to which we all have dedicated our working lives. As we tossed around ideas for new products in a somewhat crowded but fragmented market, we kept coming back to the fact that the thing that is still missing in the geospatial data creation industry is a *production framework* concept. Why is it that so many promising innovations in our industry never bear fruit? We see many examples of applications of data fusion and knowledge extraction but these algorithms primarily exist in the analysis segment of the geospatial industry and are agonizingly slow in finding their way into the production system segment. We realized that the missing element was a framework that could integrate a collection of disparate tools and make them usable across an enterprise. In other words, what is needed is a way to envelop tools and make them applicable to large production workflows.

We set out to build an extensible framework into which both existing software tools, as well as emerging prototype applications, could be tied together into a cohesive production system that would scale to the level needed for today's typical project sizes. Our true interest is in developing new algorithms and paradigms for geospatial data processing using data fusion at the production phase of the process rather than simply viewing these tools at the analysis end of the business. For example, multi and hyperspectral data should be used as part of the classification process in reducing lidar data rather than simply bringing in a lidar derived elevation model during final visualization.

The Requirements

We had two fundamental requirements in the design of our core geospatial process management framework – *enterprise deployment* and production process *continuity*. Enterprise is surely one of the most overused words in the computer jargon (right up there with synergistic) but it is the term adopted for what we have built. To us, *Enterprise* means:

- True multi-user, simultaneous access to the same production project from any workstation in the production network
- Transaction processing against a central database
- Real-time automatic status updates of client workstations in the network as the project progresses
- Project access and security on par with system domain security
- Rational schemes for managing high volume, highly transient data types such as imagery and lidar.

If you think about it, many of the above capabilities exist in industries outside our own such as banking and global reservation systems. We realized (during our beer-storming sessions) that our entire focus, as an industry, has been very much on what is happening on the *workstation* but not what is collectively happening across all of the workstations involved in a production job. Clearly today's production tools will not scale.

The second major consideration driving our new platform development is *continuity.* This is really a hybrid business/technology decision. Let us explain this point by way of our thought processes in a lidar production management system. When we started designing our production management framework, we realized that it would be a pretty tough sell to go into a potential customer site with our generic framework and suggest that all the customer need do is integrate it into their workflow. No one has the time or money to do this (this is the primary reason why enterprise data management solutions never gained traction in the geospatial market). We needed to develop a "beachhead" workflow solution that could be rapidly deployed and solve a hard problem. When we examined the state of the industry, lidar became the obvious choice. The business factors were:

- Large projects require several cooperating production workers
- Lidar is data heavy
- There are no "800 pound gorilla" lidar software tool vendors
- Nearly every company uses the same lidar data edit tool – TerraScan from Terrasolid, Oy.

Thus our goal became to build a very sophisticated geospatial process management system that could serve us for years to come as the base platform and simultaneously develop a practical, very high performance lidar workflow system.

The Development Strategy

We have developed quite a few specialized tools for geospatial production in past work and several enterprise process management systems. All of these past systems contained elements that we would do differently were we to start over. We discussed these efforts in a great deal of detail with a goal of not having the same development regrets on our new system.

We decided to divide our engineering team into several groups that would be very loosely coupled. These groups are:

- Repository – This group is responsible for the central system architecture. They have built the GeoCue Server, the Software Development Kit and the Web Server.
- Client – This group is responsible for designing the GeoCue Client. This provides the "user experience" with GeoCue when using *Environments.*
- Environment Builders – This team builds the actual application-specific workflow components. We package these collections of tools and workflow entities into what we call *CuePacs*. We are currently developing CuePacs for lidar, ortho and project reporting.

Now there is nothing unusual about dividing up a development team as we have done. However, the novel concept we have employed is that the various teams can only integrate their components using the Software Development Kit. For example, the GeoCue Client software cannot make use of any internal knowledge about the GeoCue Server to solve programming problems. If a problem arises, the Client and Server teams must get together in a design session

continued on page 242

continued from page 241
and modify the SDK. The same goes with the application domain development teams. The net result, we believe, will be a system with a proven ability to act as the central, generic production management platform regardless of workflow discipline.

A bigger challenge in our design was our goal of *wrapping* existing workflows in a managed environment. GeoCue is designed to use existing workflow software tools as well as workflow procedures with the object of *managing* the process rather than *replacing* the process. This means that if you use TerraScan for lidar processing today, you will still use TerraScan for lidar processing within a GeoCue managed environment. Your production operators will still use all of the tricks and techniques that they have developed over the years for data processing. The difference will be that they will execute these processes much more quickly due to tool-to-tool integration and far fewer workflow-related errors.

We examined particular workflows, not from the individual tools point of view, but from the production manger's point of view. For example, we did not look at a specific processing tool such as TerraScan and try to say "are there elements missing in this tool?" This sort of question placed the focus at too low a level for what we were trying to achieve. Rather we examined workflows from the macro point of view and asked "how can we make this entire process go faster and have fewer errors?"

Reversing the Production Paradigm

The way our industry approaches production has not changed since we moved to digital production. We open a tool (a software application) and then we browse for the data we intend to process (usually by browsing for a file). The biggest change on the horizon is to store large data elements, such as images, directly in databases but we will still open a tool and browse, now in the database, for the data!

In GeoCue, we turn this concept completely around. We *always* maintain a central, synoptic project view and directly launch the tools *from* the data. We call this a *data centric* approach to production. Figure 1 shows a county-wide lidar project within a GeoCue Client. Notice that you can see the production state of the entire data set within this single view. Note the color coding of the lidar working segments with the associated production *key* displayed in the upper right *checklist* pane.

The production concept in GeoCue is to always focus on the state of the data and use the data itself to drive production.

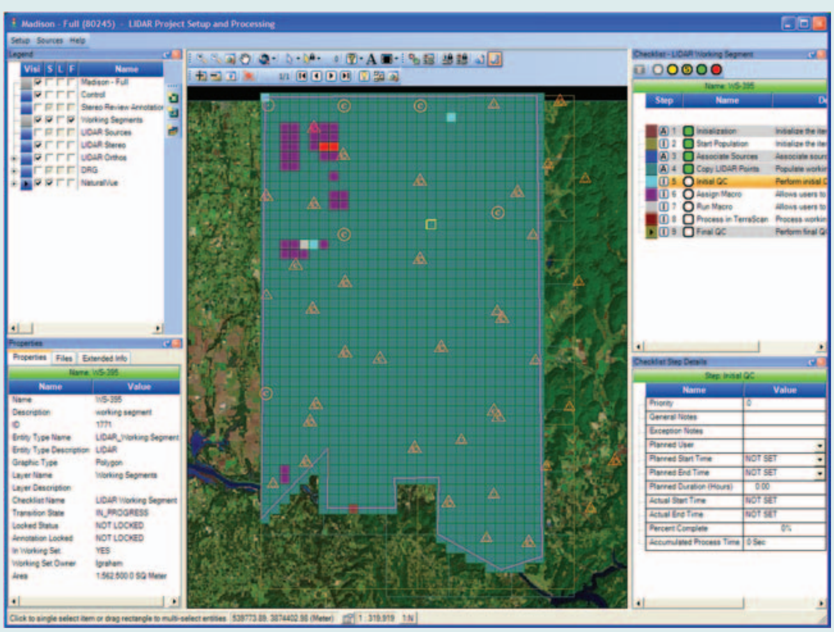

Figure 1. Data-centric production

A Managed Workflow

GeoCue is a platform for geospatial process management that is very generic in construct. This means that you can tailor the system to do anything from managing the review of intelligence imagery to generating county-wide orthos simply by loading an appropriate environment. We will illustrate the capabilities of the architecture by touring a specific environment – the Lidar 1 CuePac. Throughout the following discussions we will be using a county-wide lidar data set provided courtesy of MD Atlantic Technologies, a Macdonald Dettwiler company. This data set represents county-wide coverage of Madison Country, Alabama comprising about 1,800 square kilometers of lidar data flown with an average lidar ground sample distance (GSD) of 2 meters. Due to a high degree of lidar flight-line overlap, the entire data set comprises about 25 GB of source data representing about 900 million lidar points. The project was flown with an Optech ALTM 1210.

Where to start?

Where does production really start? We believe that it is probably when the ultimate customer (for example, a county GIS department) does a needs assessment that drives a Request For Proposal (RFP). This RFP is evaluated by our customer, the geospatial data production company. Thus our ultimate goal is to extend our framework such that data consumers can make digital RFPs that can directly feed into production systems. Our approach is to start at the point where lidar has been acquired and post-processed (geocoded) and slowly extend the framework in both directions along the production timeline.

Importing Lidar Strips

We have just begun work with Leica Geosystems to integrate the mission planning and post-processing aspects of the Airborne Laser Scanner (ALS) products into GeoCue (and hope to soon be doing the same with Optech's ALTM series) but at this point, we bring lidar data into GeoCue immediately after post-processing (geocoding). Referencing lidar strips into GeoCue involves creating a layer and browsing for the collected strip files. GeoCue then reads all of the strips, computing and storing into the GeoCue database information for later processing and finally displaying the convex hulls (the minimum convex bounding polygon encompassing all of the points in a strip) of all of the project strips (see Figure 2).

Tools are provided in lidar 1 to compute the exact ground footprint of the lidar strips. A comparison of a convex hull to a detailed footprint is shown in Figure 3.

Lidar Working Segments

The most effective way to approach a large task is to divide the work into some sort of digestible units and parcel those units out to the production staff. This

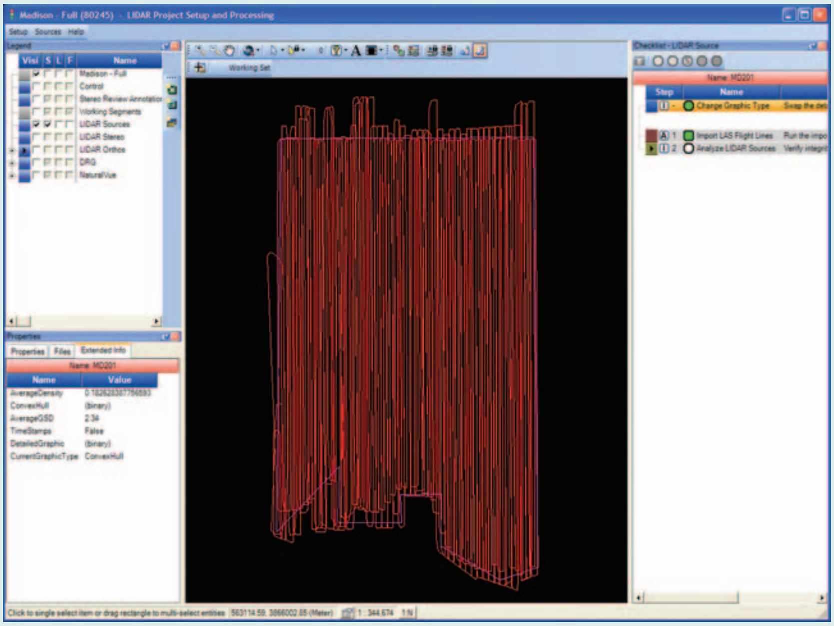

Figure 2. Lidar strips following import

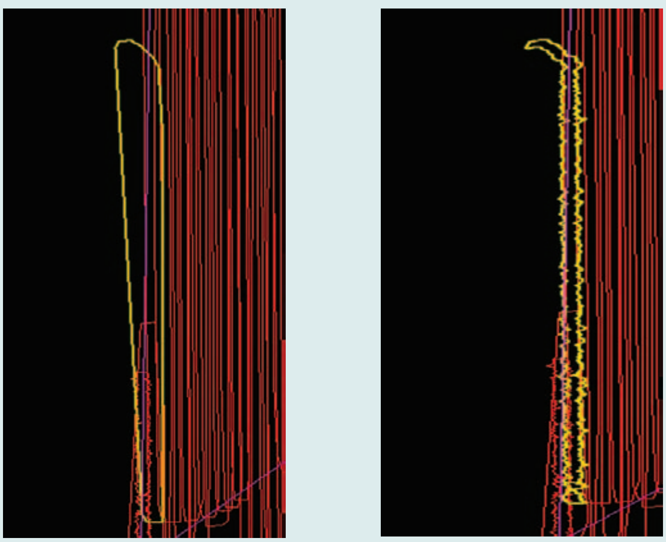

Figure 3. Convex Hull vs. Detailed Footprint

division of a project occurs naturally when the acquisition sensor is a framing technology (such as a film or digital framing camera) but must be accomplished in the production work setup when data do not fall into these natural partitions (e.g. data collected by scanning sensors such as line cameras and lidar systems). There are several reasons to partition data. The first is simply to organize the data into units that are small enough to be acceptable to your processing tools. For example, most lidar editing tools are most effective if the point volume is kept below about 4 million data points. A second reason is to distribute the work among multiple production technicians (you may want to divide up source data for this reason alone). A third reason is that source data is often not in an arrangement that is suitable to production processing. This is certainly the case in lidar where strips overlap and do not fit delivery geometries.

Lidar 1 contains a rich set of tools for segmenting a lidar project into *working segments* suitable to production. These tools allow a range of segmenting options from quick gridding to individual block creation. When working segments are created, entries are entered into the GeoCue database that track a variety of metadata about each segment such as geometry, associated lidar and mapping files, production state, coordinate system and so forth. A section of the working segments created for the Madison project are shown in Figure 4.

Following creation of the working segments, they are *populated* by copying lidar points from the source strips to the segments. Each segment has a separate associated lidar file as well as optional mapping files (such as a MicroStation design file). Our design focus is always on high performance operations so GeoCue contains a number of speed enhancing features such as project-wide Quad Tress for working segment access. All of these actions are automatically performed by GeoCue. In fact, a user need never be concerned with actual file paths, directories and so forth since data access in GeoCue is accomplished by launching tools from the graphic view of the project data.

Context Data

Providing context for visualizing and managing projects can add a great deal of value in the form of information that makes navigation through the project quick and efficient. Thus we included the ability to load *reference* layers into GeoCue in the form of both raster and vector data. Figure 5 shows the lidar source strips (the wavy, red vertical lines) and several working segments (*Selected* in yellow) superimposed over scanned USGS digitized raster graphics and a section of a USGS ortho (the Huntsville airport image inset) all superimposed on a backdrop of EarthSat NaturalVue 15m LandSat data.

Lidar Orthos

One of the big problems that lidar data processing companies faced prior to the advent of GeoCue was the overall visualization of a large project. Simply figuring out if the project were fully covered by acquired data was a very labor intensive process that could require days of analysis. Discovering a missing area of lidar coverage weeks after flying a project is a very expensive revelation!

Since we already had tools to synoptically view an entire project as well as the ability to manage and display raster context imagery, we decided to experiment with lidar orthographic rasters derived from the lidar point cloud. We call these images lidar Orthos since we compute them with each point in its correct planimetric position regardless of its elevation. Lidar 1 includes the ability to generate synoptic ortho coverage with the orthos generated from any combination of the lidar data's attributes such as laser return energy (intensity), thematic classification and return number. Using this feature provides very powerful analysis tools throughout the production process. For example, generating a project-wide ortho cover based on intensity provides an immediate analysis of lidar ground coverage. Tools in Lidar 1 allow the user to specify the ortho GSD in the ground units of the processing coordinate system. As we generate the ortho images, we detect areas where the lidar data are too sparse to fill the ortho pixels. We mark these pixels as *Voids*. The voids can then be rendered as transparent in GeoCue, allowing the user to see through the voids to a backdrop raster or a colored background. Figure 6 is the lidar ortho coverage (superimposed over

continued on page 244

continued from page 243

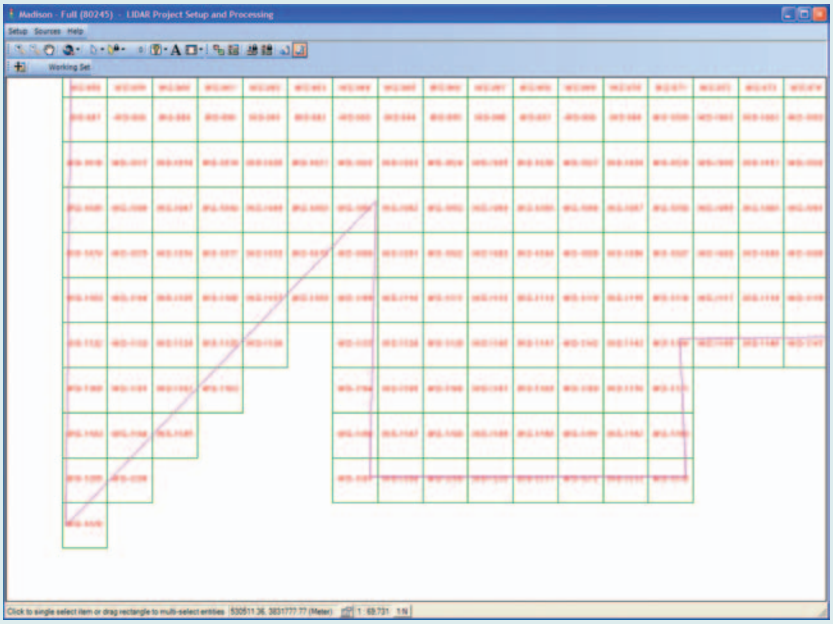

Figure 4. An area of lidar working segments

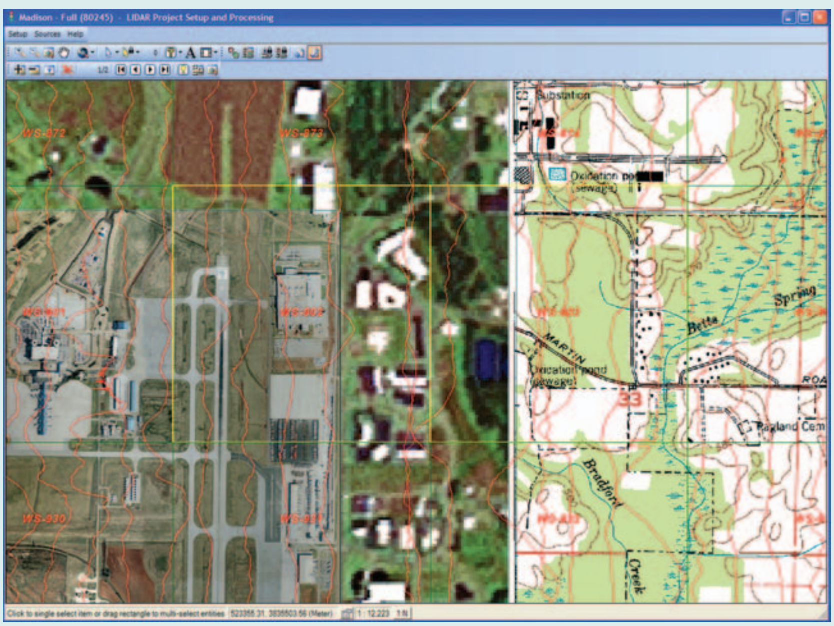

Figure 5. DRG, Ortho, Satellite Imagery Reference Data

a red backdrop) of an area of the project near Huntsville airport. The working segment on the right exhibits voids which are caused by the mirror effect of water bodies (a normal condition). However, the left working segment clearly shows a series of "roll gaps" caused by roll of the aircraft during acquisition (an abnormal condition).

Our customers have found the ortho coverage analysis system to be an extremely valuable feature of the Lidar 1 CuePac tool collection. For example, the time from lidar import to complete coverage analysis of the Madison project (some 900 million lidar points over an area of 1,800 square kilometers) was about eight hours! Coverage analysis of a project of this size typically required several weeks using other tools and techniques.

Processing Tools and Data Management

The purpose of GeoCue and its domain environments such as Lidar 1 CuePac are not to replace existing production tools but rather to host them in an enterprise management system. For example, TerraScan (from Terrasolid, Oy) is the most versatile and widely used commercial lidar editing tool today. Our goal in GeoCue Lidar 1 CuePac is not to replace functions in TerraScan but rather to integrate this (and other tools) into a framework that manages the multiuser, project level details of a production workflow. This will allow domain experts such as Arttu Soininen (the principal developer of TerraScan) to focus his efforts on breakthrough areas such as automatic feature extraction from lidar rather than the minutia of data and process management. This approach also has a very positive financial impact on companies who move from workstation-centric to enterprise production via GeoCue deployments; their investments in software tools and staff training are preserved.

The process management aspects of GeoCue are concerned primarily with guiding the user through the steps necessary to achieve production. This role of Cuing the user through the production process is the "Cue" in GeoCue. The key elements of the process management aspects of GeoCue are:

- Production Step Cuing
- Automatic software application launching
- Automatic data loading into launched applications
- Multiuser access control
- Synoptic project status viewing

This is illustrated by examples of editing of the lidar data in our Madison project. In Figure 7, the user *lgraham* has attempted to add a working segment to a production set queue (we call this the Working Set). However, this particular working segment is currently being processed by another user, *pmitester.* GeoCue automatically controls multiuser conflicts such as this, not with an error message but rather with a complete set of information such as who is currently using the desired data element, what time they locked the element for production and exactly which production step they are currently executing. In actual use, the user would never really progress to this level of control by GeoCue because the Client display graphically displays information such as short and long term locks as well as production status.

Perhaps one of the most powerful features of GeoCue is its Cuing feature. *Environments* establish the workflow process such as which steps are to be performed against which data elements and by what processing software. All of these production specifications are user encodable and modifiable so that GeoCue can be individually tailored.

Figure 8 shows Working Segment 845 selected into the Working Set (this places a multiuser lock on the working segment). The Checklist pane automatically loads the processing step sequence for this

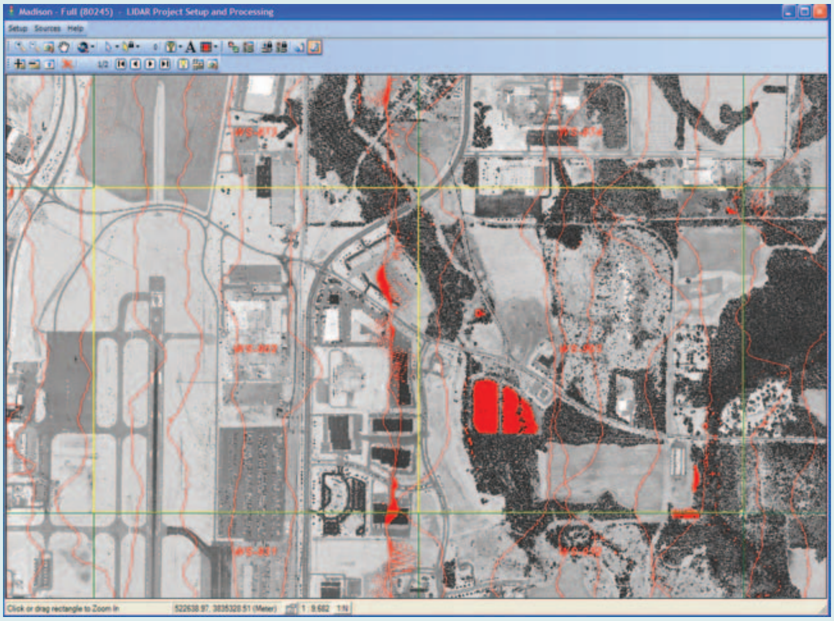

Figure 6. Lidar Ortho coverage showing water bodies and roll gaps

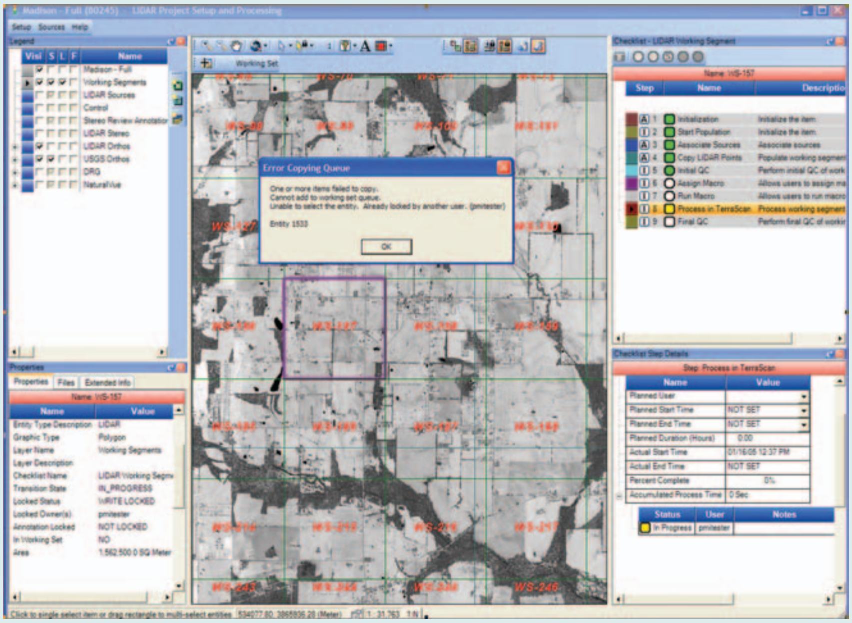

Figure 7. Multiuser Access Control in GeoCue

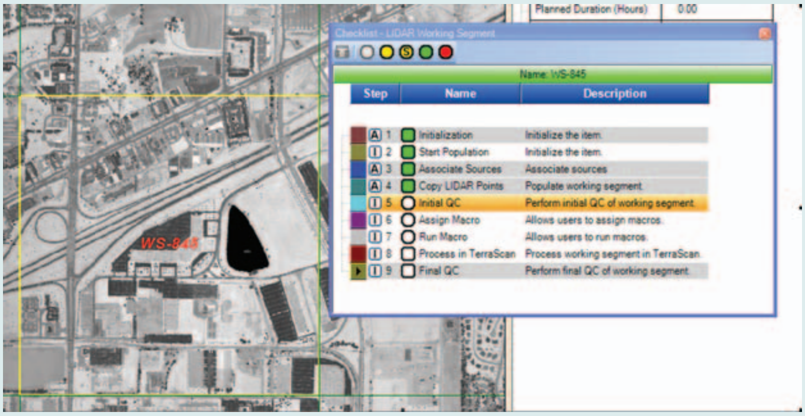

Figure 8. The GeoCue Checklist Processor

working set. The user need only select the next step in the list (highlighted in orange) and press the yellow "process" button at the top of the dialog to launch the associated software tool (in our example, the step is Initial QC to which we have mapped the lidar QC tool PointVue).

Pressing the Process button starts several activities in GeoCue:

- The software tool associated with the step is invoked (this can be any software application from any vendor – the instructions on what to launch, where the software is located and how it is to be invoked are contained in user-settable database tables).
- The production log for the data object is updated with the login name of the user invoking the step and the time that the step was started.
- The data associated with the production element is automatically located and mapped to the processing tool. In our example, an external file in LAS format containing the lidar data associated with Working Segment 845 will be mapped to the PointVue application.
- An event is sent to all other GeoCue Clients who have this same project loaded. This event updates all displays to indicate the current production state of WS-845.

The net effect of this is that users never search cryptically named file directories on a maze of networks for oddly named data files. In fact, the user generally does not know (although they can certainly find out) where the data is located or even its storage paradigm (flat files on disk or entries in a database). Additionally, they do not need to try to remember that initial QC is performed in PointVue and editing is performed in TerraScan. The appropriate executable program is automatically loaded.

This is the area of GeoCue where the data centric approach is most apparent. The conventional sequence to carry out the step described above would be to know from a written procedural workflow document (in reality, this is usually not even written down but located in the "knowledge expert's" head) that Initial QC is performed after processing segments are created. The user would start up the associated program (in our example, PointVue) and then browse for the correct file. The files are usually arranged by some naming convention of files and folders. Of course, in many instances folks make local copies of files for processing (to "speed things up") and forget to restore them to the well known locations. The amount of time wasted in trying to find software and data files and then trying to fix the problems that inevitably occur by selecting the wrong file are well

continued on page 246

continued from page 245
know in our industry. In GeoCue, the data *is* the driver and thus all of the above problems disappear.

NIIRS10 Unique Tools

As we develop production environments such as lidar, we are always on the lookout for ways that we can speed up production or add useful tools without duplicating the applications that are already in common use. For example, nearly everyone who processes lidar data uses TerraScan as either their primary production tool or as an augmentation to in-house developed algorithms. Thus it is of little value to our customers for us to try to duplicate tools already contained in TerraScan. Our approach is to look at the process from the project-wide point of view and ask what tools we can develop to significantly improve production or to derive more value from the data. Our ability to very rapidly generate project-wide lidar orthos is certainly in this category.

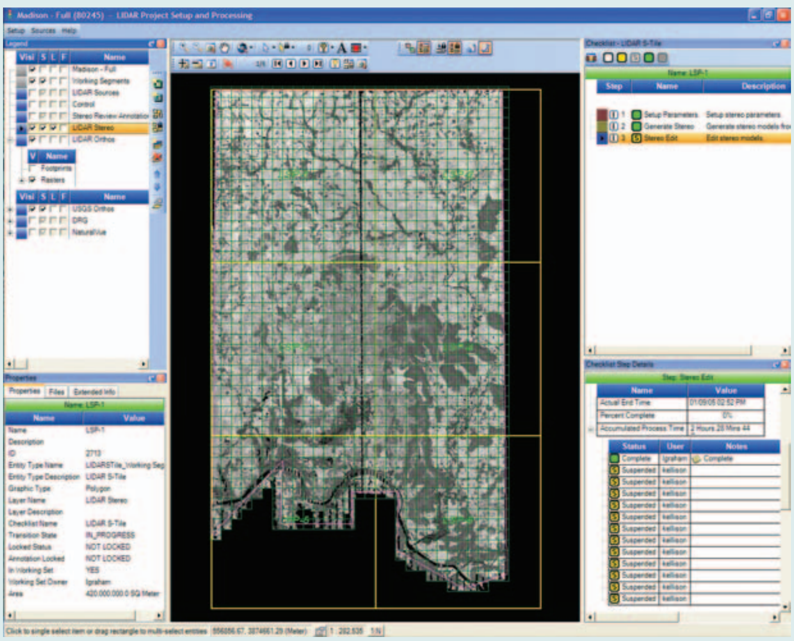

Figure 9. Lidar Stereo Pairs for the Madison Project.

Checklist Step Details

Step: Process in TerraScan

Name	Value
Priority	0
General Notes	
Exception Notes	
Planned User	ttaylor
Planned Start Time	01/11/05 12:00 PM
Planned End Time	01/11/05 12:10 PM
Planned Duration (Hours)	0.10
Actual Start Time	01/11/05 01:20 PM
Actual End Time	01/15/05 04:36 PM
Percent Complete	100%
Accumulated Process Time	18 Mins 13 Secs

Status	User	Notes	Start Time	End Time	Duration
Complete	ttaylor	Standard radial filter	01/11/05 01:20 PM	01/11/05 01:22 PM	2 Mins 40 Secs
Suspended	ttaylor	I inadvertently marked this step complete. It needs additional classification.	01/11/05 03:06 PM	01/11/05 03:06 PM	0 Sec
Suspended	lgraham	Requires building classificaiton - handing off to S Riddell	01/15/05 03:40 PM	01/15/05 03:54 PM	13 Mins 48 Secs
Complete	sriddell	Classified 2 buildings	01/15/05 04:34 PM	01/15/05 04:36 PM	1 Min 44 Secs

Figure 10. Production history for one Segment for one Step.

Unique to the Lidar 1 CuePac tools is the integration of a Lidar Synthetic Stereo workflow. Since lidar data are three-dimensional, we can place a virtual pair of stereo image planes anywhere in space above the project that we desire. We can then "back project" the lidar data to these virtual image planes and interpolate an image. Finally, if we construct the necessary metadata, we can feed this synthetically generated "Lidar Stereo Pair" to a standard stereo photogrammetric workstation as if the data were conventional stereo imagery. This exact process is built directly into the Lidar 1 CuePac.

Shown in Figure 9 are the Lidar Stereo Pairs for the Madison County project. Notice that we have covered the entire project with only six models (of course, Lidar 1 will allow you to make these any size you desire). There are several major benefits of using Lidar Stereo Pairs. Perhaps the most exciting of these is that you can use the lidar data effectively as imagery in your existing stereo exploitation workstation (Lidar 1 currently supports SOCET SET, ImageStation and Summit Evolution). We have customers today who use this capability to directly collect features from lidar, obviating the need for ancillary imagery.

An even more novel use of the Lidar Stereo Pairs is as a very rapid lidar point classification assessment tool. Quality checking individual working segments for classification accuracy is a daunting task. Our Madison County project contains 1,281 lidar working segments. Each of these segments must be classified and inspected. With Lidar Stereo coverage, these 1,281 segments have been encoded into 6 stereo models. Using a conventional photogrammetric workstation, the entire project can be reviewed in a few hours time! We have just released a new set of tools included with the Lidar 1 CuePac called the *Stereo Annotation System* (SAS). SAS allows an operator to make notations directly within the stereo environment that are automatically mapped by GeoCue to the appropriate lidar working segments, automatically loading segment metadata tables with the name of a TerraScan processing macro. Using an iterative approach, a user can review the project using Lidar Stereo Pairs, make annotations regarding problems and batch reclassify the data. This new approach to project review/correction promises to rapidly reduce the processing time needed to derive lidar models while using a customer's existing photogrammetric production tools.

"When Will it be Done?!"

When we discuss production with our customers, an always repeated frustration is the difficulty in measuring the true status of production. We are trying to make this easy in GeoCue by automatically generating graphical depictions of the production status. The opening figure of this article (Figure 1) depicts the 1,281 Working Segments in the Madison Project. In a single project-wide view, the production manager has an instant picture of the status of the project. Since this status is an attribute of the production data objects themselves, it is always exactly up to date.

Production History

A very important element of the production process is planning the production process and keeping a very accurate record of what actually occurred during production. This is needed not only for costing a project but also for estimating the next RFP that comes through the door. With profit margins constantly being squeezed, accurate histories of what production actually costs are invaluable.

GeoCue automatically maintains a complete history for the project at a very fine-grained level. Depicted in Figure 10 are both the planning data and processing history for a single step in the life of Working Segment WS-440. Every aspect of this history except for the processing notes is automatically recorded by GeoCue. Our goal with this automatic approach is to eliminate the common procedure in which production technicians are expected to record their actions in a "side-car" tracking application such as an Excel spreadsheet.

We will be releasing a module in the first half of 2005 that will generate comprehensive reports based on the GeoCue project tracking information as well as bulk ways to load project planning data.

The Future

Our goal with GeoCue is to develop a universal geospatial production framework that will be used for all types of production. Lidar is a beachhead because the enterprise tools in this area are nonexistent. As we move forward in time, we will release other *CuePacs* aimed at a variety of different geospatial production workflows. We are also working with third parties to develop CuePacs that they will sell in discipline areas with which we are not well versed.

To us, one of the most exciting prospects of GeoCue is as the foundation for *fusion*-based production. By this we mean that different layers in GeoCue can hold different types of data and GeoCue-driven applications can access these layers as they desire. We are beginning to experiment with thematically classified imagery on one layer in GeoCue being used as a pre-classifier for lidar data held on another layer. The advantage of GeoCue, of course, is that the resultant algorithms are hosted in a production system rather than an analysis system.

Acknowledgements

The following companies and organizations have been instrumental in creating GeoCue through their close cooperation with us and their true desire to see mutual success on this project:

ASPRS Lidar Committee
BAE Systems
Bentley Systems Incorporated
DAT/EM
Earth Satellite Corporation
ESRI
Intergraph
Leica Geosystems
MD Atlantic Technologies (a Macdonald Dettwiler company)
Optech
Terrasolid, oy
Woolpert, Inc.

Author

Lewis Graham, President, NIIRS10 Incorporated, 132 Westchester Drive, Madison, Alabama 35758, USA, 01-256-461-8289, lgraham@niirs10.com, www.niirs10.com.