

# OVERCOMING LIDAR'S BIG DATA PROBLEM

LIDAR CAN MEASURE SURFACES INCREDIBLY ACCURATELY, BUT IT CAN PRODUCE HUGE AMOUNTS OF DATA AS A RESULT, SLOWING DOWN SOFTWARE, SOMETIMES TO THE POINT OF UNUSABILITY. HERE, RICK HARRISON REVEALS EIGHT SIMPLE STEPS TO CUT THE PROBLEM DOWN TO SIZE

It was just 40 years ago when a contour map of a site was generated by one of two ways. The first was to have a four-man survey crew locate 50 to 60 points a day and for an office technician to then interpret the data. The second was to painstakingly trace orthophotos manually.

Today, a two-man survey crew can gather over 1,000 precision points. Aerial mapping data was also efficient... until they replaced orthophotos with LiDAR. Instead of determining critical topographic data such as the top of a curb and 15cm away the bottom of that curb to accurately show the contour lines along streets, LiDAR randomly measures points along the ground, accurately but producing a huge amount of data.

Aerial mapping companies promote LiDAR as it significantly reduces manual labour. However, they also transfer the data overload to the end user, who either cannot use the excessive data as their software cannot handle it or somehow has to make the data useable.

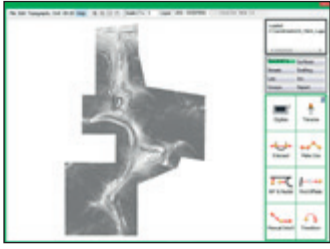
Ultimately, a representation of a ground surface is used to com-

municate the terrain for a purpose – typically, civil engineering such as street, lot, structure, drainage design or earthwork. In the near future, interactive visualisation and virtual reality will become as common as the paper plan. But before that can occur, a major roadblock to both design productivity and virtual performance is excessive LiDAR data.

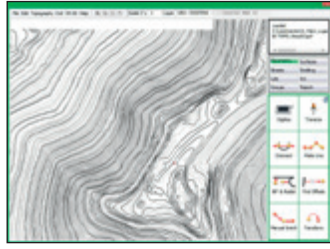
What can be done to reduce LiDAR information to a useable form? The simplest solution is to go to the original source (the aerial mapping firm) and demand filtered 'breakline' data that somewhat emulates the actions of a ground survey that delivers points where the land surface changes slope. If the company refuses to provide data in that format, the next steps can be taken in the example below.

## A solution

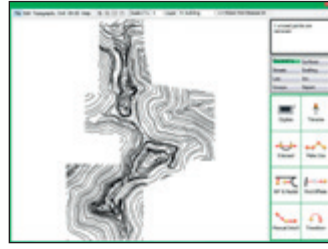
To define the rural terrain in the following example, the original LiDAR-generated, 5cm-interval contour lines contained more than one million



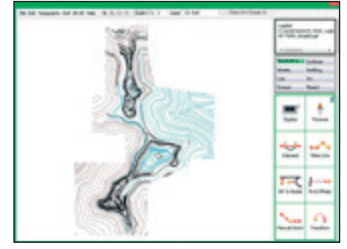
Step 1



Step 2



Step 3



Step 4

points. There is no viable way to use a million points within DTM-based engineering or surveying software running on even the most powerful computers without significantly slowing down the design process. This amount of data would also render 3D interactive visualisation useless.

The following editing process was used to create useable data from the excessive contour line information.

#### Step 1: Initial filtering

The unedited contour lines as delivered from the mapping company to the client in CAD format, forming polylines after a stripping factor of eight was applied (using one out of eight points). This reduced the polylines to a more manageable 139,000 points. The filtering was determined by experimenting to make sure no visible changes to the original data were obvious.

#### Step 2: Isolating contours of surface change

The next step was to change the layer of the lines where there was a change in progression. In other words you do not need the contour polylines that are evenly spaced, only those where there is a change in shape or trend in interval. Most DTM software will reproduce the contours in between.

#### Step 3: Erasing excess data

Once the major contours are determined to define the surface, we erased the layer of the polylines we did not need.

#### Step 4: Filtering remaining contours of excessive data

Typically, there will be contour lines that still contain excessive data (too many points along the polyline) and some will have just the right amount of data. By changing the layers of groups (shown by different colours), the layer can be isolated to strip down the excessive polyline points without stripping those that have the correct information.

#### Step 5: Merging a proposed surface

In some cases, there will be proposed and existing contours that must be merged. To accomplish this manually, we exploded all the polylines, then imported the proposed polylines and tagged them for overlay only, so they would not be deleted during editing. We then erased all natural ground information under the proposed polylines so no data overlapped. Of course, if there are no extra surfaces, this step is not needed.

#### Step 6: Creating breaklines

DTM software will need additional guidance where there is an elongated flat surface, or high or low points along roadways and ditches or

areas of drainage. Additional lines may be added to make sure the new DTM replicates the original surface.

#### Step 7: Erasing data outside the scope of the surface

In most cases, there will be edge triangles of a digital terrain model that will produce excess topography in areas of convex shapes, and most DTM creating software can limit or erase this data.

#### Step 8: Producing the filtered DTM

The final contours should be verified before saving the edited DTM.

#### The end result

The above steps took no more than a few hours to reduce the original 1,100,000 coordinates along 5cm contour lines to just 17,948 points to define the same surface to create a DTM. In general, the goal should be to maintain fewer than 30,000 points on any DTM.

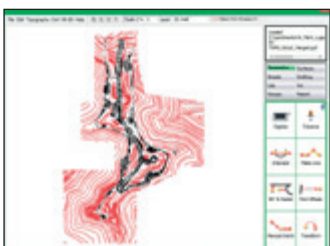
In this example, using LandMentor software, red indicates low areas and the transition to green indicated high ground. There are two forms of software: those that internally carry the DTM and can generate the contours at any interval on demand, thus needing no contour polyline data; and those that require the polylines in a 3D coordinate format. When in doubt, it's probably therefore best to deliver both formats.

#### Pay now or pay later

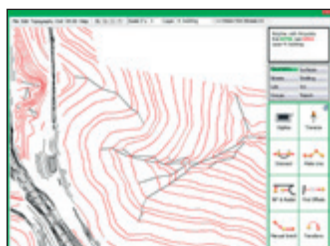
In conclusion, the best way to receive LiDAR data is in its already filtered breakline format, which can be easily accomplished with the software aerial mapping firms already have. Those paying for LiDAR information can refuse to pay for the data unless it's in a usable format, otherwise they'll suffer greater costs, because either every process the data is used for (if it's usable at all) will be slower or someone will need to make it usable by following the steps above to filter only the data that is needed.

**THOSE PAYING FOR LIDAR INFORMATION  
CAN REFUSE TO PAY FOR THE DATA UNLESS  
IT'S IN A USABLE FORMAT, OTHERWISE  
THEY'LL SUFFER GREATER COSTS**

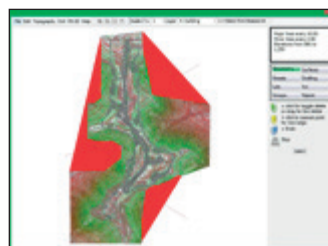
*Rick Harrison is president of Neighborhood Innovations  
([www.neighborhoodinnovations.com](http://www.neighborhoodinnovations.com))*



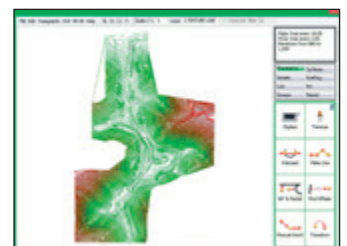
Step 5



Step 6



Step 7



Step 8