Towards Generalization Processes of LiDAR Data based on GRID and OGC Web Processing Services

Sandra Lanig¹, Alexander Zipf¹

¹Department of Geography, Chair of Cartography, University of Bonn
{lanig.zipf}@geographie.uni-bonn.de

Abstract. The use of Airborne Laserscanner Data (ALS or LiDAR, Light Detection And Ranging) is increasing for generating very detailed high resolution Digital Elevation Models (DEM). However, massive amount of data result by the enormous data acquisition rate and the high spatial point density. To reduce the number of the massive data points of the high quality surface and make it manageable, we need methods of generalization and simplification. Nevertheless, data storage and processing of the high quality LiDAR data sets could be hardly managed within current classical desktop Geographic Information Systems (GIS). Both computing power for data processing and storage organization supporting data access as well as simplification algorithm for the LiDAR point cloud data are required. To achieve high processing performance and storage load the use of GRID Computing is a fine option. This paper propose a Web Processing Service (WPS) for Generalization based on the specification of the Open Geospatial Consortium (OGC) that access pre-processing generalization functionalities of massive LiDAR data to a standardized Web interface. In order to use the advantages of the GRID, the WPS Generalization Service is made accessible as GRID service.

1 INTRODUCTION

High quality surface models play an important role in different domains like 3D city models or disaster management. Through newly available data acquisition techniques such as Airborne Laserscanner Data (ALS or LiDAR) it is possible to gain high-precision and spatial densely surface data. The questions related to this new acquisition technique are: 1) Where can these massive data sets be stored; and 2) How can these amounts of surface data sets be processed. The outcome of this is the main question: How can we manage the larger and lager 3D surface data sets. Such data can be several giga-bytes in size, or may contain millions of measurement points.

To minimize the number of the massive laser scanning points in order to reduce the huge amount of data and optimize processing time, we need methods for approximations of the original high resolution LiDAR data sets. For this purpose, diverse generalization or simplification algorithms and,
in other words for terrain visualization in computer graphics, different Level of Detail (LOD) algorithms are available.

In order to share and access distributed spatial data and resources, Spatial Data Infrastructures (SDIs) are generally used. Based on Service Oriented Architectures (SOA), SDI supports discovery, access to distributed geospatial data and use of geographic information (Nebert, D. (Ed.) 2004). Nevertheless, data storage and processing of raw LiDAR data sets could hardly manage within classical SOA based SDIs. LiDAR data processing requires access to computing power and storage capacity as well as several simplification processing algorithms for the high point cloud density. To achieve high processing performance and storage load of current existing IT resources the use of GRID Computing is a fine option.

This paper presents an approach to offer a Web Processing Service (WPS) for different surface simplification algorithms based on the specification of the Open Geospatial Consortium (OGC) that access preprocessing generalization functionalities of massive LiDAR data to a standardized Web interface. In order to use the advantages of the GRID, the WPS Generalization Service is made accessible as GRID service. In the following we give a short introduction to LiDAR and describe basic generalization techniques as well as the fundamentals of SOA, starting with a description of the WPS interface and an introduction to GRID services in general. The following sections explain the architecture and the prototypical implementation. We finish with a conclusion and an outlook.

2 BACKGROUND

2.1 Airborne Laserscanner Data

Airborne LiDAR (Light Detection And Ranging) data is one of the most important techniques for studying the earth’s surface of a wide range with up-to-date DEMs with a high degree of accuracy (Wever, 1999). The LiDAR system consists of three components: a laserscanner, an Intertial Measurement Unit (IMU) and a Global Positioning System (GPS). The laserscanner gauging the earth surface and classify the measurement points into first and last pulse data sets. Typically the system returns up to four points per square meter with an absolute vertical accuracy of better than 0,15 m. Due to this high point density, LiDAR data offers geo-scientist now the possibility to study processes based on DEMs at very detailed resolutions. However, managing and analyzing these massive LiDAR DEMs in classical desktop GIS is currently difficult.
2.2 The OGC Web Processing Service (WPS) Interface

Based on *Service Oriented Architectures* (SOA), SDIs are primarily used to discovery, access and share distributed geospatial data. However, the core of SDIs are processing and analysing functions. Therefore a recognized standard for distributed spatial data processing was missing for a long time. This issue was addressed by the development of a *Web Processing Service* (WPS), which has been approved as official standard by the OGC in December 2007 (Schut, 2007). The WPS makes real processing functionality available. The rather generic WPS standard defines a web-based interface to distribute and execute geospatial processes in different complexity. This allows developing more complex services using standardized interfaces via the Internet by providing the possibility of performing arbitrary geospatial analysis and manipulation on geospatial data. According to the specification a “WPS defines a standardized interface that facilitates the publishing of geospatial processes, and the discovery of and binding to those processes by clients. “Processes” include any algorithm, calculation or model that operates on spatially referenced data“. The WPS standard is quite open and so a “WPS may offer calculations as simple as subtracting one set of spatially referenced numbers from another (e.g., determining the difference in influenza cases between two different seasons), or as complicated as a global climate change model. The data required by the WPS can be delivered across a network or available at the server“ (Schut, 2007).

The specification defines three mandatory operations performed by a WPS. The *getCapabilities* operation returns a brief service metadata XML document describing the resources of the specific server implementation, and gives a short description of each process offered by the WPS instance. The *describeProcess* operation returns a detailed description of a process including its required input- (including the allowed formats) and output-parameter. Finally the *execute* operation runs the offered process. These operations can be requested by a client via HTTP and performed by a WPS server implementation.

The data input and output can be simple literal values such as integer or double numbers as well as character strings (*LiteralData*), BoundingBox data structure like two pairs of coordinates or *ComplexValue* and *ComplexValueReference* that indicates complex data sets such as GML (*Geography Markup Language*) encoded fragments and raster, vector or other (large) data files. The data in the *ComplexValue* is part of the request/response document, whereas the *ComplexValueReference* handover the URL reference to the location, where the data can be downloaded. The
input data can be managed by storing the data on the server and as part of or only referenced at the request document.

2.3 GRID Services

For determining whether system is a GRID, Forster proposes a three-point checklist. According to this a GRID system “coordinates resources that are not subject to centralized control using standard, open, general-purpose protocols and interfaces to deliver nontrivial qualities of services” (Foster, 2002).

The objective of the Open Grid Service Architecture (OGSA) is to define an open and standardized architecture for grid-based applications to standardize all the services for job and resource management, security etc. by specifying a set of standard interfaces. Normally a Web service is stateless. The Web Service Resource Framework (WSRF) identifies and provides a set of operation that Web services may implement to become stateful services using Web service technologies (WS-Resource). The WSRF is an open GRID standard to define Web services for GRID Computing, or in other words is the infrastructure on which the OGSA is built on. Encapsulated within the WS-Adressing endpoint reference and the WS-ResourceProperties, the identifier of the specific resources are included inside the client request as well as a simple URI address or complex XML content.

Following the IT mainstream, the underlying standards of the WSRF are Web Service Description Language (WSDL) and Simple Object Access Protocol (SOAP) which are publicized by the W3C. Grid services are WSRF-based Web services integrated in GRID Computing infrastructure. To create WSRF compliant GRID applications the Globus Toolkit 4 (GT4) includes some high-level services for resource monitoring, discovery, job submission, security, and data management that we can use.

2.4 Linking OGC SOA to WSRF-based GRID Services

In order to connect the GI community with the GRID community the OGC and the OGF (Open Grid Forum) are working together to specify interfaces, best practices and enhance existing GRID Computing and OGC standards. Therefore the OGF signed a memorandum of understanding to collaborate with the OGC. One objective of the collaboration is to integrate the OGC WPS standard with a range of back-end processing environments to enable large-scale processing. In order of particular interest of the Grid community we focus on the OGC WPS standard.
3 OVERVIEW GENERALIZATION TECHNIQUES

There are numerous reasons for the simplification of surface models. On the one hand it is essential to reduce the storage requirement as well as the data set in particular for Web applications. The aim is to remove irrelevant details, in order to give better overview and the focus to important details and produce different LOD for visualization issues. Further reasons for DEM generalization are speeding display, reducing redundancy and transmitting time or aesthetic causes.

There are different methods for simplification. Firstly, lines can be simplified (line simplification). The Douglas Peucker (DP) algorithm (Douglas, 1973) is the most practiced algorithm used for curve smoothing and point thinning by distance measurement. Secondly, there are simplification algorithms for 3D buildings (Glander and Döllner, 2007, Rau et al., 2006) and finally for surface simplifications. Heckbert and Garland (1997) give a good survey of polygonal surface generalization algorithm.

<table>
<thead>
<tr>
<th>Simplification Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line simplification</td>
</tr>
<tr>
<td>3D City Generalisation</td>
</tr>
<tr>
<td>3D Surface Simplification</td>
</tr>
</tbody>
</table>

Figure 1 Simplification Methods

There exist different operations for simplification or aggregation which determine the features for elimination. These are vertex removal, vertex pair collapse, vertex clustering, edge collapse, triangle collapse, and triangle removal.

4 SYSTEM ARCHITECTURE

The system architecture of the grid-enabled WPS Terrain Generalization service is illustrated in Figure 2. The design follows the three tier architecture. Clients access the generalization service by the standardized OGC WPS interface operations `getCapabilities`, `describeProcess` and `execute`.

As soon as a client have acquire sufficient information by the requested `getCapabilities` and `describeProcess` XML documents, the selected generalization process can be submitted to the server by the `execute` operation. In order to execute the simplification processes at the GRID, each implemented OGC WPS process is mapped to a corresponding GRID service. In order to ensure security, all authentication credentials from the client must be delegated by creating temporary proxy certificate to the GRID service.
To build WSRF compliant Grid applications the *Globus Toolkit 4* (GT4) includes some high-level services for resource monitoring, discovery, job submission, security, and data management that we can use. The WSRF-based GRID service integrates a WSDL description and can be addressed by SOAP. The WPS server that acts as client component calls the corresponding WSRF-based 3D Generalization Web service using a SOAP request. Therefore, the WPS *execute* operation is wrapped as SOAP request.

The simplification of the surface data takes place as job execution on GRID clusters. Each cluster possesses a GRAM (*Grid Resource Allocation and Management*) service, which receives the jobs and distributes it on the free *worker nodes* (WN) in the cluster. The GT4 *Web Service GRAM* (WS GRAM) is addressed like a normal GRID service. On the individual worker nodes the GRID jobs run as executable files. First, before the actual job starts, the necessary data have to be transported to the node (stage-in). The input data is managed by storing on the server and referenced at the request document as URL. If the terrain processing is successfully, the result is sent back to the WPS as GML or in facts of huge terrain data, only the
5 WPS Surface Generalization Service

The goal of the WPS Generalization Service is to offer multi-scale 3D-models with different Levels of Details (LOD). Especially in a real-time environment, the dense surface meshes from raw LiDAR data must be generalized in a pre-processing step to provide adaptive mesh resolution and to use the main advantages of LODs, data reduction, fast data access and hierarchical processing. For example, view-dependent generalization using high resolution objects in the foreground and lower LODs for distant improve performance for visualization objectives. The surface geometry is stored as Triangulated Irregular Network (TIN). Therefore different surface simplification algorithms are developed e.g. by DeFloriani (1996) or Lindstrom (1996).

Depending on specific application requirements, we implement as proof of concept diverse generalization algorithms to handle geometric data efficiently. Firstly, we implement a surface simplification algorithm, which used quadric error metrics, provided by Garland and Heckbert (1997). The algorithm is based on an iterative generalization of edge aggregation by vertex pair contraction. The error approximation for simplification of each vertex is the sum of squared distances to the planes. Secondly, we implement the Douglas Peucker (DP) algorithm (Douglas, 1973) that is the most practiced algorithm used for curve smoothing and point thinning by distance measurement.

Currently, several processing frameworks and implementations based on the OGC WPS standard exist. A first version of a generic WPS Generalization Service has been developed within the degree framework (Heier, 2005). Recently, this Java framework is upgraded to deegree3 and offering the main building blocks for SDI based on standards of the OGC and ISO/TC 211 (ISO Technical Committee 211 – Geographic Information/Geomatics) and implementing the WPS specification version 1.0.0 (Schut 2007).

6 Conclusion and Future Work

The need for computing power and storage capacity is steadily rising within geo-community. In particularly LiDAR data is being used to create a high-resolution 3D model of the Earth’s surface. But to process these vast LiDAR data sets means to compute million of raw data points and running
computationally intensive algorithms. We have presented in this paper a possibility to make the processing of massive LiDAR data easier using standardized WPS in relation with GRID Computing. Scientists can access the WPS Surface Generalization Service choosing their DEM data source and the generalization algorithm via Web interface. Users can swap computationally intensive LiDAR data processing tasks on GRID computing resources by standardized WPS interface.

An additional optimization is the integration of the WSRF- and WPS-based services into workflows. Fleuren (2008) discusses the problems of designing a workflow based on standard OGC Web services and WSRF-based services. For orchestrating services the use of the de facto standard BPEL (*Business Process Execution Language*) allows the construction of complex workflows. After the execution the user could save and reuse the workflow description. Our future work will focus on the integration of our WPS and WSRF-based services into the geospatial workflow engine that based on the BPEL. The client can combine different LiDAR processing algorithms by creating simple XML documents.

7 **ACKNOWLEDGMENTS**

This work is part of the Spatial Data Infrastructure Grid (GDI-Grid) funded by the German Federal Ministry of Education and Research (BMBF) and is still work in process. The main goal of the project is the efficient mining and processing of spatial data for the simulation of noise, dispersion and disaster management.

8 **REFERENCES**


15th annual ACM international symposium on Advances in geographic information systems, ACM, New York, NY, USA, pp. 1–4.


