Airborne Laser Scanning: Efficient data processing and information extraction

PhD Proposal MAP-I

Supervisors

Luís Gonçalves-Seco (Email: lgs@fc.up.pt) and Michel Ferreira (Email: michel@dcc.fc.up.pt)

Research Units

CICGE – Centro de Investigação em Ciências Geo-Espaciais da Universidade do Porto

LIACC – Lab. Inteligência Artificial e Ciência de Computadores da Universidade do Porto

Motivation and goals

At all times spatial information has played a vital role in human life, from knowing where to find food in prehistory up to make a round-the-world trip in a few seconds using Google Earth. Even in times of global networking and Web 2.0, space and time are the major dimensions in human interaction. Hence, spatial information is necessary for driving decisions on all kind of levels (from an individual up to public decision-makers). Nowadays, using remote sensing techniques (terrestrial, airborne or spaceborne) a high amount of spatial datasets in a variety of domains can be acquired in a very short time.

A challenging problem is to transform this huge amount of data into useful information without losing accuracy or information itself. These large amounts of spatial datasets make manual processing impossible and it is therefore necessary to develop efficient and effective techniques for semi or fully automatic information extraction. Taking into account the dimension time, i.e. with the availability of multi-temporal datasets, the need for standardized and repeatable extraction methods even increases. For example, in natural hazard management fast access to spatial information extracted from high resolution spatial data is crucial. Basically, data availability plays a key role. Not only the pure existence of spatial data with high information content is sufficient but data accessibility is even more fundamental.

This proposal concentrates on spatial data acquired by means of Airborne Laser Scanning (ALS) or Light Detection and Ranging (LiDAR). This active remote sensing technique is mainly used for topographic data[1] acquisition resulting in high resolution elevation models of the Earth's surface. Country and State-wide ALS datasets already exists, as Switzerland, Federal State of Vorarlberg/Austria, Autonomous Province of Bolzano/Italy and Federal Emergency Management Agency (FEMA). A wealth of free LiDAR data are also accessible to the public from the websites maintained by governmental agencies such as the U.S. Geological Survey (the Center for LiDAR Information, Coordination and Knowledge: CLICK), National Oceanic and Atmospheric Administration (Coastal Service Center), and U.S. Army Corps of Engineers (the Joint Airborne LiDAR Bathymetry Technical Center of Expertise: JALBTCX).

ALS is characterized by high data density (namely point density) and high accuracy, also in vegetated areas where ALS clearly exceeds the quality of competing acquisition methods. The primary result of ALS, which is

[1] Nowadays was extended to many fields of applications such as floodplain mapping, glaciology, hydrology, geomorphology, and forest inventory, urban planning, and landscape ecology, etc.

also delivered to end-users, is called point cloud. The point cloud consists of single point measurements of x, y, z coordinates and a value for signal strength, called intensity or amplitude. In general, the point cloud is further processed with a high degree of automation to Digital Terrain and Surface Models (DTMs and DSMs) continuously representing space (e.g. raster data). But nevertheless the original point cloud, discontinuously representing space (vector data), contains the highest degree of information, as it is pseudo three-dimensional, not modified by any algorithm or aggregated in any dimension.

The LiDAR point cloud contains the highest degree of information but problems arise if standard Geographic Information Systems (GIS) and remote sensing software are used for processing and analysis of this huge amount of single point measurements. The major problems are i) that only few algorithms working on the LiDAR point cloud are implemented and ii) that the implemented vector data models are not designed for billions of points (e.g. reaching file size limits of the operating system). Not only 2.5D coordinates and intensity are important information sources but also time, number of reflections / echoes and corresponding sensor position are useful additional attributes for each single laser point, which should not be disregarded if available.

So, the scope of the proposed research evolved from the necessity to efficiently extract highly accurate spatial information contained in LiDAR data but not yet operationally used. It was chosen to concentrate on LiDAR data only, to better evaluate the potential given by using just this data source, in comparison to approaches combining LiDAR with other data sources, such as optical data. The goals of the proposal can be pointed out as following:

1. Investigate strategies concerning the LiDAR data management that respond to the many questions remains: which spatial indexing is supported by LiDAR file database – raw laser strips, pretiled files, both? Which data is used for LiDAR processing - point cloud or interpolated raster? It is possible to add or remove features from the data when new information is available? Which neighborhood topology has be chosen for RAM LiDAR data structure?

2. Design an open source library for efficient LiDAR data processing. Several aspects should be taken into account: a) import and export data where the specification of an efficient and versatile LiDAR file format should be discussed; b) convert data from one geodesic coordinate system to another; c) add/remove LiDAR feature and d) perform efficient search query on large data set.

3. Explore and propose a new method to filter LiDAR point cloud using data mining techniques[2]

[2] It does not need any preprocess that could produce errors in the results and it can be easily applied to big datasets as, for example, the LIDAR data clouds.