Heterogenous Spatial Data

Fusion, Modeling, and Analysis for GIS Applications

Giuseppe Patanè and Michela Spagnuolo, *Editors*
CNR-IMATI

SYNTHESIS LECTURES ON VISUAL COMPUTING: COMPUTER GRAPHICS, ANIMATION, COMPUTATIONAL PHOTOGRAPHY, AND IMAGING #24

MORGAN & CLAYPOOL PUBLISHERS

Heterogenous Spatial Data: Fusion, Modeling, and Analysis for GIS Applications

Giuseppe Patanè and Michela Spagnuolo, *Editors*

www.morganclaypool.com

ISBN: 9781627054621 paperback
ISBN: 9781627056700 ebook

DOI 10.2200/S00711ED1V01Y201603VCP024

A Publication in the Morgan & Claypool Publishers series
SYNTHESIS LECTURES ON VISUAL COMPUTING: COMPUTER GRAPHICS, ANIMATION, COMPUTATIONAL PHOTOGRAPHY, AND IMAGING

Lecture #24
Series Editor: Brian A. Barsky, *University of California, Berkeley*
Series ISSN
Print 2469-4215 Electronic 2469-4223

ABSTRACT

New data acquisition techniques are emerging and are providing fast and efficient means for multidimensional spatial data collection. Airborne LIDAR surveys, SAR satellites, stereo-photogrammetry and mobile mapping systems are increasingly used for the digital reconstruction of the environment. All these systems provide extremely high volumes of raw data, often enriched with other sensor data (e.g., beam intensity). Improving methods to process and visually analyze this massive amount of geospatial and user-generated data is crucial to increase the efficiency of organizations and to better manage societal challenges.

Within this context, this book proposes an up-to-date view of computational methods and tools for spatio-temporal data fusion, multivariate surface generation, and feature extraction, along with their main applications for surface approximation and rainfall analysis. The book is intended to attract interest from different fields, such as computer vision, computer graphics, geomatics, and remote sensing, working on the common goal of processing 3D data. To this end, it presents and compares methods that process and analyze the massive amount of geospatial data in order to support better management of societal challenges through more timely and better decision making, independent of a specific data modeling paradigm (e.g., 2D vector data, regular grids or 3D point clouds).

We also show how current research is developing from the traditional layered approach, adopted by most GIS softwares, to intelligent methods for integrating existing data sets that might contain important information on a geographical area and environmental phenomenon. These services combine traditional map-oriented visualization with fully 3D visual decision support methods and exploit semantics-oriented information (e.g., a-priori knowledge, annotations, segmentations) when processing, merging, and integrating big pre-existing data sets.

KEYWORDS

heterogeneous spatial data, spatio-temporal data fusion, multi-variate surface generation, feature extraction, GIS applications

Contents

Preface

Nowadays, airborne LIDAR surveys, SAR satellites, stereo-photogrammetry, and mobile mapping systems are providing fast and efficient means for multidimensional spatial data collection and are increasingly applied to the digital reconstruction of environment. The aforementioned and new data acquisition techniques are also providing fast and efficient means for the collection of extremely high volumes of raw data, which are often enriched with other sensor data (e.g., beam intensity). These volumes increase significantly faster than computing speeds; for instance, LIDAR acquisition speed has increased by a factor of 100 to more than 1 million points/second in just four years. Other data sets for topography or bathymetry apply different modeling paradigms, such as raster formats or boundary representations. While the growth of these data sets is not as fast as with point cloud data, there exists a giant stockpile of such data which cannot be fully used due to its size. With each acquisition approach, lots of data are collected with different modalities, thereby producing highly *heterogeneous* data sets which require harmonization and integration before being really useful.

Realistic simulations, *real-time* information, and timeliness are key factors for the management of environmental events. Flooding, flash floods, and industrial accidents, among others, are happening at time-scales where instant actions and decisions are necessary. However, taking decisions requires the availability of adequate information derived from up-to-date, harmonized data. On the one hand, preliminary planning based on scenarios and simulations gives crucial information for decisions, and the capacity to integrate them with data captured in real-time is needed in quickly changing situations. On the other hand, planning applications and also simulations require the identification, measurement and assessment of specific objects. However, the amount and complexity of such data makes the extraction of objects of interest extremely delicate. Information extraction is therefore often carried out manually, making this process time-consuming and expensive. Consequently, methods for automated feature extraction are being developed and are now available in many processing software but are often limited to a set of pre-programmed objects.

A large amount of pre-existing ancillary data is most often available and could be used for the identification, extraction, and mapping of objects of interest. In the marine case, this situation refers to vector maps and engineering drawings that can both provide the shape of the reference target feature and/or the position of a particular target feature. There are data collected by local authorities, and increasingly even data collected from the ground by smartphones and other GPS-equipped devices. Typical objects can be identified and the system can be trained by using the information already available. Furthermore, semantic information, which is explicit in individual data sets, may be exploited to make integration more efficient and effective. Individual data sets may show a varying depth of information resulting from their nature, and even give a slightly different view of the general structure of reality. For instance, LIDAR technology can be very useful to improve our knowledge of the seabed; beyond the detection of bottom elevation in clear enough water, a wealth of information is latent in the LIDAR waveform and this is a key alternative to complex acoustics in shallow waters. The LIDAR waveform can be used to infer other information than simply the air-water interface and the distance from the seabed to the water surface. The signal can also be analyzed to provide information on sea bed cover type (rocks, silt, sand, gavel, seagrass, etc.), water surface state, and water column information (suspended materials, physical properties). Engineering features such as trenches and cables can also be detected. Often, such features are removed in order to give a DEM.

Improving methods to process and visually analyze this massive amount of geospatial data, including user-generated one, is crucial to increase the efficiency of organizations and better management of societal challenges through more timely and better decision making. While processing and visualization of spatial data have been thoroughly investigated, the issues listed below are still open:

- to define methods and protocols to cope with heterogeneous quality in raw data and to provide information about the quality of the integrated data sets, to deliver appropriate knowledge on how applications will behave when making use of these data set;
- to develop a processing framework that is independent of a specific data modeling paradigm (e.g., 2D vector data, regular grids or 3D point clouds) and of a specific distributed execution architecture; and
- to devise methodologies and tools which exploit semantics-oriented information (e.g., *a priori* knowledge, annotations, segmentation results, features) when processing, merging, and integrating big pre-existing data sets.

All these elements also show the need to move from the traditional layered approach, adopted by most GIS softwares, to services for integrating on demand existing data sets that might contain important information on a geographical area, environmental problem or phenomenon, also coupling traditional map-oriented visualization to fully 3D visual decision support methods.

In this context, this book aims at proposing an up-to-date view of computational methods and tools for the following tasks:

- **spatio-temporal data fusion** (Chapter 1), by presenting a survey of the data types that are processed from the perspective of the sensors that acquired the data (e.g., satellite images, radar, sonar, echosounding, and LIDAR data) and the pre-processing that the data undergo. In particular, we will discuss the principles and processes that underlie data fusion and integration and the management of spatial data;
- **multivariate surface generation** (Chapter 2), by discussing the main reconstruction methods of surfaces with discrete, spline-based, and implicit approximations;

- **feature extraction and classification** (Chapter 3), by presenting the main methods for feature extraction, segmentation, and classification through random forests and topological methods for characterization and description of digital terrains and urban data; and
- **applications** to surface approximation and rainfall analysis (Chapter 4), where we select a set of case studies related to GIS.

This book is intended to attract the interest from different fields such as Computer Vision, Computer Graphics, Geomatics, Remote Sensing, High Performance Computing, and Grid Computing working on the common goal of processing 3D data. To this end, it presents and compares methods that process and analyze the massive amount of geospatial data in order to support a better management of societal challenges through more timely and better decision making and independently of a specific data modeling paradigm (e.g., 2D vector data, regular grids or 3D point clouds). We show how the current research is developing from the traditional layered approach, adopted by most GISs, with "intelligent" methods for integrating existing data sets that might contain important information on a geographical area and environmental phenomenon by coupling traditional map-oriented visualization to fully 3D visual decision support methods and exploiting semantics-oriented information (e.g., a-priori knowledge, annotations, segmentations) when processing, merging and integrating big pre-existing data sets.

Finally, this book covers the main research topics and scientific results of the European FP7 Integrated Project IQmulus; for more details, we refer the reader to the `http://www.iqmulus.eu`.

Giuseppe Patanè and Michela Spagnuolo, *Editors*
March 2016

Acknowledgments

This book was supported by the **European FP7 Integrated Project IQmulus** "*A High-volume Fusion and Analysis Platform for Geospatial Point Clouds, Coverages and Volumetric Data Sets*."

We thank the IQmulus Partners for interesting discussion on the book topics during the project meetings. In particular, we mention: Bianca Falcidieno, Silvia Biasotti, Andrea Cerri (**CNR-IMATI**); Jan Boehm, Simon Julier, Dietmar Backes (**UCL**); Dániel Kristóf, Márta Belényesi, István László, Roberto Giachetta, Gizella Nádor, István Fekete, Dávid Gera (**FÖMI**); Tor Dokken, Vibeke Skytt (**SINTEF**); Roderik Lindenbergh, Beril Sirmacek (**TUDelft**); Romain Cancouet, Andre Stumpf (**UBO**); Jacques Populus, Touria Bajjouk (**Ifremer**); Michel Kraemer, Sebastian Pena Serna (**Fraunhofer**); Laura Muraglia, Francesca Giannoni, Federica Martina (**Regione Liguria**); Nicolas Paparoditis, Mathieu Bredif (**IGN**); and Daniel Holwegg (**MOSS**), Keiran Millard (**HR Wallingford**).

Finally, we acknowledge the contribution of Hans van der Marel (TUDleft), Beril Sirmacek (TUDelft), István László (FÖMI), and Sebastian Pena Serna (Fraunhofer) to Chapter 1, and Regione Liguria and Genova Municipality for the rainfall data in Chapter 4.

Giuseppe Patanè and Michela Spagnuolo, *Editors*
March 2016