D8.8.2 REPORT ON THE IQMULUS PROCESSING CONTEST – YEAR 2

Deliverable D8.8.2

Circulation:	RE - Restricted
Lead partner:	CNR-IMATI-GE
Contributing partners:	CNR-IMATI-GE, TUDelft, UCL, IGN, UBO, FOMI,
Authors:	Silvia Biasotti, Marco Attene (CNR-IMATI-GE); B.G.H. Gorte (TUDelft), Jan Boehm (UCL); Nicolas Paparoditis (IGN), Helen Piete (UBO), Daniel Kristof [FOMI]
Quality Controller:	Ewald Quak (SINTEF)
Version:	1.0
Date:	31.10.2014

Consisting of

SINTEF	STIFTELSEN SINTEF, Department of Applied Mathematics, Oslo, Norway
Fraunhofer	Fraunhofer Institute for Computer Graphics Research, Darmstadt, Germany
CNR-IMATI-GE	Institute for Applied Mathematics and Information Technologies of the National Research Council (CNR-IMATI), Genova, Italy
MOSS	M.O.S.S. Computer Grafik Systeme GmbH (MOSS), Munich, Germany
HRW	HR Wallingford Ltd (HRW), Wallingford, UK
FOMI	Hungarian National Mapping and Cadastral Agency (FÖMI), Institute of Geodesy, Cartography and Remote Sensing, Budapest, Hungary
UCL	University College London (UCL), Research centre for Photogrammetry, 3D Imaging and Metrology, London, UK
TUDelft	Delft University of Technology (TUDelft), Department of Earth and Climate Sciences & Man-Machine Interaction Group, Delft, The Netherlands
IGN	Institut National de l'Information Géographique et Forestière (IGN), Paris, France
UBO	Université de Bretagne Occidentale (UBO), European Institute for Marine Studies, Brest, France
Ifremer	L'Institut Francais de Recherche pour l´Exploitation de la Mer (Ifremer), Brest, France
Liguria	Regione Liguria, Genova, Italy

DOCUMENT HISTORY

Version[1]	Issue Date	Stage	Content and Changes
0.5	Sep 25 2014	Draft	Initial draft of the report and circulation among the project partners
0.8	Oct 12 2014	Draft	Partners' comments are collected and included in the report
0.9	Oct. 16 2014	Draft	Submission to the quality controller
1.0	Oct. 28 2014		Version to be submitted to the Project Officer

EXECUTIVE SUMMARY

The deliverable D8.8.2 is related to the second year activities of the IQmulus Task 8.5, described as follows in the Description of Work (DoW):

Focus of the task will be the dissemination of IQmulus results on data processing by means of setting up and running yearly contests, open to the whole scientific community as well as user groups, on research issues which are relevant to the WP4 area of work.

This report describes the selection process of the most interesting tasks/services proposed as contest tracks and a description of the proposed challenges. In addition, we list the panel of experts, selected among the partners and also involving scientists outside the consortium that have been chosen to support the planning of the contest, establishing the processing tasks to be benchmarked and selecting the appropriate data sets, performance measures and procedures to realize the validation. Finally, we add in the Appendix the reports of the tracks presented during the first IQmulus Workshop on *Processing Large Geospatial Data* held in Cardiff, July the 8th, 2014.

[1] Integers correspond to submitted versions

Table of contents

1 INTRODUCTION

The IQmulus Processing Contest started in 2013 to provide the scientific community and practioners working in the field of 3D point cloud processing with certified benchmarks and ground truths in geospatial data processing. While in 2013 the tracks (and therefore the challenges to be approached) were proposed by partners involved in the project's processing work package (WP4), in the second year (2014) calls for track proposals have been specifically targeted at project stakeholders. Once the tracks proposed had been validated and accepted by the Advisory Board, they were publicized with a specific call for participation and through a set of mailing lists (among others the geometry processing list "geometryprocessing@inria.fr").

The results of the second-year contests have been made publicly available and presented at the IQmulus Workshop on "Processing Large Geospatial Data" held in Cardiff, July the 8th, 2014 and co-located with the Symposium on Geometry Processing (SGP) 2014. In addition, a continuous submission of executable codes or results is possible so as to make the participation in the contests more attractive for scientists.

This report describes the second year's activities related to the IQmulus Processing Contest, especially the challenges faced by the contest tracks. In addition, we list the panel of experts, selected among the partners and also involving scientists outside the consortium that have been chosen to support the planning of the contest, establishing the processing tasks to be benchmarked and selecting the appropriate datasets, performance measures and procedures to realize the validation.

1.1 ADVISORY AND TECHNICAL BOARDS

To allow for a large scientific scope, the IQmulus contest is implemented in collaboration with the research community and with the support of an international advisory board, composed of high level scientific and technical experts in the field.

For the second year, the members of the IQmulus consortium involved in the Committee/ Advisory Board are:

- Silvia Biasotti, CNR-IMATI-GE
- Giuseppe Patanè, CNR-IMATI-GE
- Jan Boehm, UCL
- Nicolas Paparoditis, IGN
- Roderik Lindenbergh, TUDelft

The project-external members of the Advisory Board are:

- Hamish Carr, Visualization and Virtual Reality Group, University of Leeds, UK;
- Debra Laefer, School of Civil, Structural & Environmental Engineering, University College Dublin, Ireland. Debra Laefer is the recipient of the European ERC Starting Grant "RETURN: Rethinking Tunnelling in Urban Neighbourhoods", 2013-2017.

In addition, for the management of the second-year platform, an IQPC'14 Technical Board has been established at IMATI-CNR, Italy. People involved in this activity are:

- Corrado Pizzi, IMATI-CNR, Italy
- Piero Bruno, IMATI-CNR, Italy
- Davide Sobrero, IMATI-CNR, Italy

2 TIMELINE OF THE CONTEST

In the following, we briefly summarize the different steps of the preparation of the second year IQmulus Processing Contest.

Jan 2014: Creation of the mailing list of contacts for the contest, with the first call for the track proposals circulated both inside and outside the IQmulus consortium. The track organizers were asked to send their proposal to the address of the task leader (namely, silvia@ge.imati.cnr.it), describing and motivating the envisioned task, specifying the dataset, ground truth, aspects of the codes to be evaluated and the evaluation method, the expected number of participants and envisaged dissemination channels for advertising the track.

February 2014: Deadline for potential track proposals and selection of possible tracks from the IQmulus consortium.

March 2014: Promising tracks and their organizers were listed on the IQmulus contest web page and the track organizers started working out the details. The official web page was set up at the address: http://www.iqmulus.eu/index.php/iqmulus-contest-2014. An additional web page was set up at the address: http://www.ge.imati.cnr.it/iqmulus_contest2014.

April 2014: The track organizers published the track schedule and details on the participation. Calls for participation were circulated both for the whole IQmulus Processing Contest and the specific single tracks. Among others, the call for participation was circulated to the Eurographics, GIRPR, geometryprocessing and ISPRS lists.

April 2014 – June 2014: Running of the contest tracks, each according to its own schedule. The track organizers collected and evaluated the results.

May 2014: Set up of the contest infrastructure at CNR-IMATI

June 2014: The track organizers delivered their track reports to be included in the IQmulus Workshop Proceedings.

July 08, 2014: Presentation of the IQPC'14 tracks at the IQmulus Workshop on Processing Large Geospatial Data organized by the IQmulus project

September 2014: three papers related to the methods and tracks involved in the IQmulus Processing Contest were submitted for possible publication to the "Special Issue on Processing Large Geospatial Data" of the journal Computers & Graphics.

3 THE CONTEST INFRASTRUCTURE

As detailed in the previous deliverable D8.8.1, the IQmulus Processing Contest is hosted in a CPU-cluster unit related to the Visualisation Virtual Services (VVS) offered by the FP7-INFRASTRUCTURES project VISIONAIR (2011-2015) that has been upgraded by augmenting its data storage and processing capabilities.

The services for running the benchmarking procedures are implemented using the Collage Authoring Workbench, http://collage.ge.imati.cnr.it, developed by the Academic Computer Center Cyfronet AGH in Cracow in Poland, which allows the documentation and the long-term preservation of the outcome of the contests. Figure 1 shows a screenshot of the initial page the participants find when connecting to the contest infrastructure. The infrastructure is accessible

via https protocol at the web address http://collage.ge.imati.cnr.it/workbench and a tutorial with examples on its usage are available at http://collage.ge.imati.cnr.it/tutorials.

In 2014, the operating system used in the Collage Authoring Workbench adopted in the IQmulus Processing Contest is Linux Ubuntu 14.04. There the participants may run their code and executables using the Unix bash shell (http://www.hypexr.org/bash_tutorial.php#emacs), the Windows files .exe can be run using the environment "wine" (http://www.winehq.org/) and the interpreted language Octave 3.2.1 (http://www.gnu.org/software/octave/) is available to run MATLAB-like code. In addition plots are possible using the gnuplot package (http://www.gnuplot.info/). To support geospatial data processing, additional libraries, such as PCL, GDAL, MPI, VDK, FLANN, Boost, etc., are available and already installed in the VVS infrastructure.

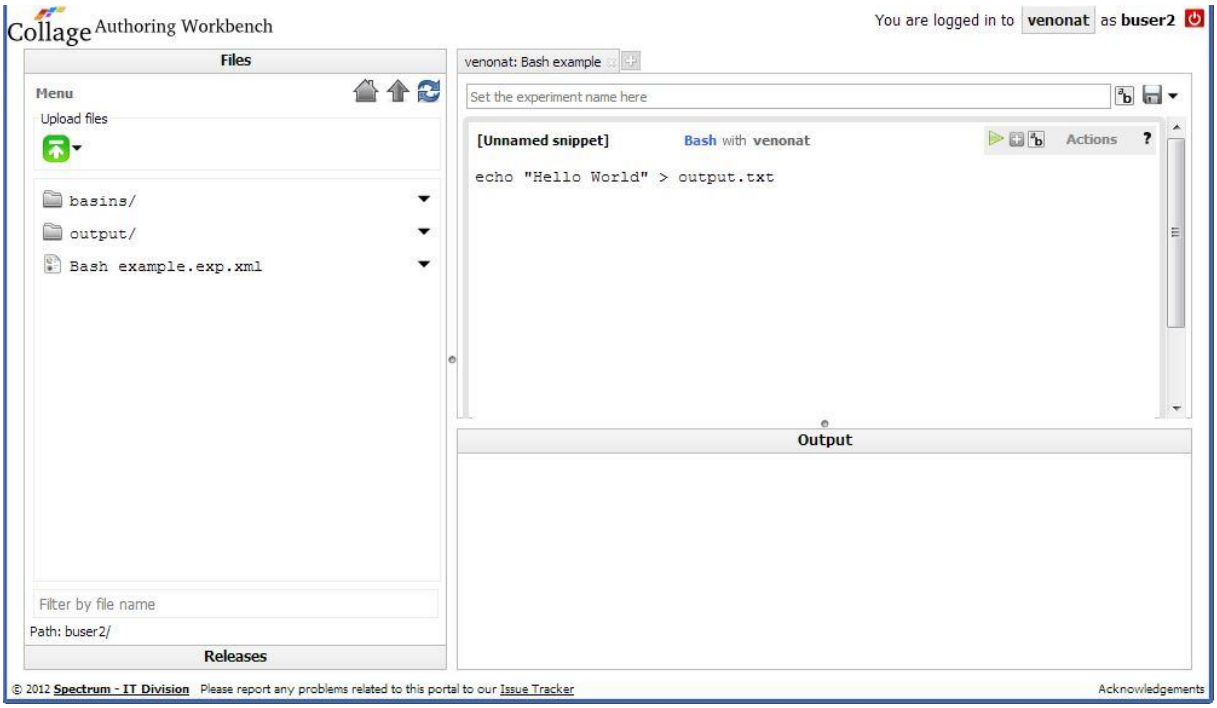


FIGURE 1 A SCREENSHOT OF THE MAIN PAGE OF THE COLLAGE AUTHORING WORKBENCH

In 2014, CNR-IMATI also offered a multi-platform infrastructure available via SSH protocol, including Linux Ubuntu 12.04 64bit, Linux CentOS 64bit and Windows 7 64bit operative systems. In case a participant has any specific request (particular operative systems, 32bit versions of the operative systems and/or someone needs a particular virtual machine), CNR-IMATI foresaw the possibility of providing ad-hoc virtual machines.

The Linux-Ubuntu machine was equipped with Octave (MATLAB-like) packages, wine, gnuplot, PCL, GDAL, VDK, MPI, FLANN, boost and other specific libraries. The Linux-CentOS machine was installed with Torque for parallel computing while the Windows-based machine hosted Visual Studio.

Credentials to access the infrastructure were provided upon request to the track organizers to run the executables the track participants submitted to them.

4 THE SECOND YEAR'S TRACKS

The contest enables reproducible evaluation through the creation of standardized benchmarks and evaluation methodologies. In the second year, three track proposals have been selected as representatives of acute issues in geospatial data: i) Urban 3D model generation, ii) Urban Point Cloud Classification and iii) Approximation of Rainfall Data.

In the following, we briefly describe the three tracks implemented in 2014. Each track is listed in the IQmulus Processing Contest webpage and described in a dedicated webpage. Details on the challenges faced in each track, the description of the datasets and ground truths, the evaluation criteria and the results of the participants are provided in the Appendix, where we attach the reports of the contest contributions presented at the IQmulus Workshop 2014 on "Processing Large Geospatial Data".

4.1 URBAN 3D MODEL GENERATION

The description of this track is hosted at the following web address: https://sites.google.com/site/iqmuluscontest2014 and has been organized by members of the University College Dublin.

This track focuses on urban three-dimensional (3D) model generation which is a rapidly growing topic for a large range of applications (e.g. environmental planning and monitoring, computational simulation, disaster management, security, telecommunications, location-based services), and Aerial Laser Scanning (ALS) point clouds are a potential source for 3D building reconstruction.

This contest track foresaw two tasks:

1. to find the buildings in the study area (Task A)
2. and/or to reconstruct a subset of them (Task B).

The participants were expected to submit results to the competition organizers for an evaluation based on ground truth. It was expected to have separate winners (selected based on an overall evaluation score based on the specific evaluation criteria described in the Appendix) and two runners up for each task.

Figure 2 shows an overview of the dataset under study. The test area is approximately 1km^2 and contains 205 blocks, each of which may contain in excess of a dozen buildings per block. The buildings in this area generally abut one another, with some buildings sharing an adjoining wall. In general, individual buildings are closely spaced. Approximately 225 million points were acquired within the study area, giving a typical point density of ~225 points/m^2. The data set is organized in 9 files, each representing a 500m x 500m tile.

FIGURE 2 AN OVERVIEW OF THE DUBLIN AREA USED AS DATASET.

Regarding the evaluation of the results, accuracy has been assessed by calculating discrepancies from the ground truth. For Task A, the overall building detection was evaluated in terms of completeness and correctness, as well as the quality of the results. For Task B, the generated building models were mapped onto the reference models by the competition organisers, based upon each building's centroid, as computed using the edge coordinates of the building at the ground level. The evaluation criteria of a building included the building detection, the vertical wall detection, the roof configuration, and opening details.

4.2 URBAN POINT CLOUD CLASSIFICATION

The description of this track is hosted at the web address: http://data.ign.fr/benchmarks/UrbanAnalysis and has been co-organized by members of IGN, France, and CMM - MINES ParisTech, France.

The contest track aimed at stimulating researchers from different fields such as Computer Vision, Computer Graphics, Geomatics and Remote Sensing, to work on the common goal of processing 3D data as well as benchmarking segmentation and classification methods working on 3D MLS data. Indeed, in the literature, most available urban data consist of close-range images, aerial images and satellite images but few MLS datasets (ISPRS data, IGN data). Moreover, manual annotations and algorithm outputs are rarely found in available 3D repositories. Some annotated 3D MLS datasets publicly available are the Oakland 3D point cloud dataset and the Paris-rue-Madame dataset.

The benchmarking was carried out by the MATIS Lab of the French National Mapping Agency (IGN) and the Center for Mathematical Morphology (CMM-MINES ParisTech) in conjunction with the TerraMobilita project.

The database contains 3D MLS data from a dense urban environment in Paris (France), composed of 300 million points. The acquisition was made in January 2013 within the framework of the

IQmulus and TerraMobilita projects using Stereopolis II, a MLS system developed at the French National Mapping Agency (IGN). Annotation will be carried out in a manually assisted way by the MATIS laboratory at IGN. In this database, the entire 3D point cloud is segmented and classified, i.e. each point contains a label and a class. Thus, point-wise evaluation of detection-segmentation-classification methods becomes possible.

The datasets and their processing results must be presented in PLY format with little endian encoding. All coordinates are geo-referenced (E, N, U) in the Lambert 93 and altitude IGN1969 (grid RAF09) reference system, reflectance is the laser intensity. The output expected from each participant is a PLY file containing the original points, their original attributes, an identifier/label for each segmented object and its classification.

Each individual object has been manually annotated in the 3D dataset by IGN personnel using a web based editor developed by IGN. All the 3D points belonging to the same object will have the same object identifier. Thus, the number of different ids in the 3D point cloud corresponds to the number of objects. In the classification step, a category is assigned to each segmented object. Each class represents an urban semantic entity, as defined in the following subsection.

Figure 3 shows examples of labels and classes on a 3D point cloud. In the left image, note that ground, facades, pedestrians and cars are represented with different colors because they are different objects and have different labels. In the right image, the colors represent the object classes: ground in gray, facades in blue, cars in yellow, pedestrians in skin color, and furniture in red and cyan. If two points have the same label, they must have the same class. In this contest, a hierarchy of semantic classes has been defined and stored in a class tree downloadable as an xml file.

Participants were asked to choose the classes that they want in the tree and to recognize the elements they selected; then the evaluation has been performed according to their choices.

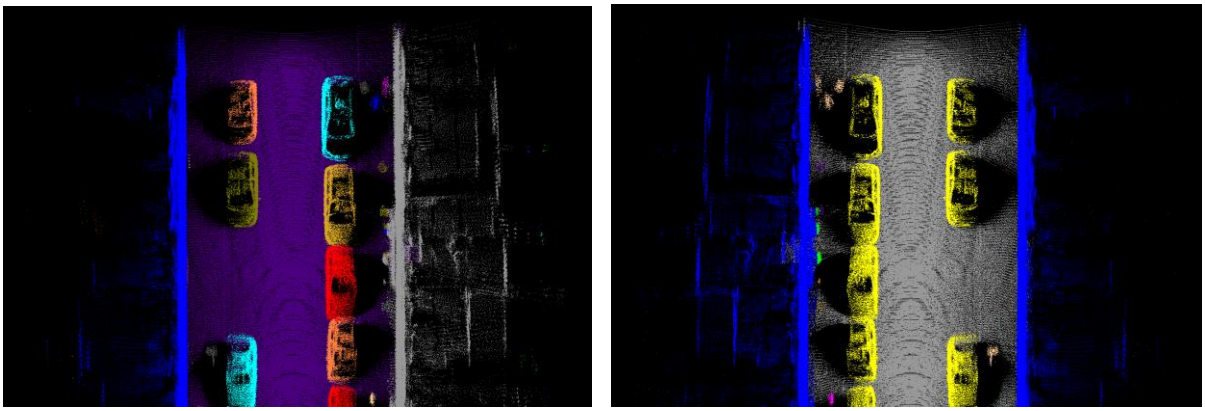


FIGURE 3 EXAMPLES OF CLASSES, I.E., SEGMENTED OBJECTS, AND LABELS, I.E., OBJECTS WITH THE SAME FUNCTION.

The benchmark does not aim at ranking the participants but at providing insights on the strengths and weaknesses of each method. We consider that the quality of a method is subjective and application dependent, and the results of this benchmark should only help a user choosing one approach depending on its own specific requirements. Results were evaluated at three levels: as a classification, as a detection and as a segmentation. Details of the evaluation metrics used are given in the Appendix.

4.3 APPROXIMATION OF RAINFALL DATA

The description of the track is hosted at the address: http://www.ge.imati.cnr.it/node/246 and has been organized by members of CNR-IMATI, Italy.

The goal of this contest track was to compare the accuracy and performances of generic and specialized approximation techniques (e.g., RBFs, Splines, Kriging) for rainfall data and to select the most suitable algorithms.

The datasets for the evaluation include a DEM model and the rainfall data of a day over that DEM, in details:

- *Rainfall data [29 Sept. 2013, whole day, filename: observed_20130929_GB.csv]*
  The data set is gathered from two different rain gauges networks with a different spatial distribution. The first one is the network developed by Regione Liguria over the whole region, with 143 measure stations. The second one is the measure system deployed by the Genova municipality within the city boundary, with about 35 measure stations. Since the temporal interval is different for each network, station rainfalls are cumulated to a step of 30 minutes. The selected rainy day is September 29, 2013 because it was characterized by light rain over Regione Liguria with 2 different thunderstorms that do not cause damages (only local floodings and landslides).

- *DEM of Regione Liguria [filename: Liguria100.concave.ply]*
  The interpolation process will be performed over DEMs of Regione Liguria at different resolutions: (i) the SRTM model (downscaled at 1 point every 100m). The interpolations' accuracy will be tested over the (143+25) stations and the extrapolation test (approximation accuracy with respect to the ground truth and speed) will be performed over the DEM of Regione Liguria, see Figure 4. The coordinate reference system is the Gauss-Boaga west (epsg 3003).

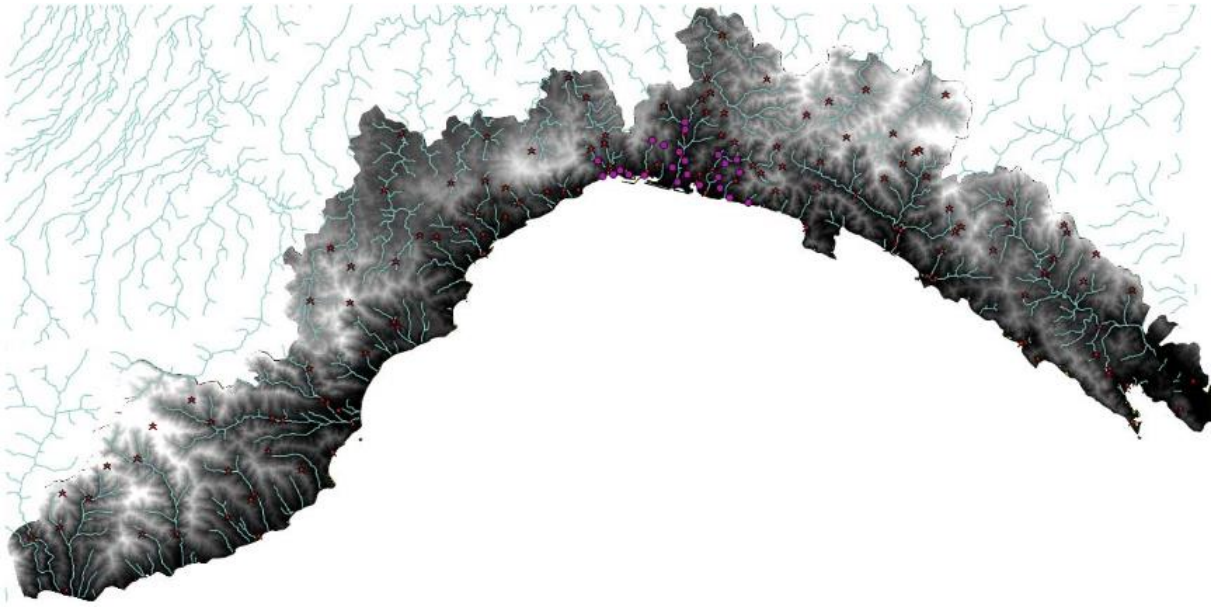


FIGURE 4 INPUT: RAINFALL MEASURES AT 143 STATIONS (REGIONAL LEVEL, RED POINTS) AND 25 STATIONS (MUNICIPALITY LEVEL, PURPLE CIRCLES).

The evaluation criteria for the methods which participated in the contest are:

- *Evaluation of the accuracy of each approximation scheme* through the comparison between the approximated and true values at every sample point. The expected result is a matrix of (143+35) x 48 elements representing the value of the approximating function at each station every 30 minutes.
- *Cross-validation of the method.* Every 30 minutes, each rain gauge will be turned off and the approximation function will be sampled at this position. Then, the value obtained will be compared with the true value. The expected results is a matrix of dimension (143+35) x 48, whose elements represent the sampled value.
- *Evaluation of the accuracy of the algorithms through the comparison with the ground truth*. For the validation of the approximation techniques, the rainfall data measured by the Genova municipality will be used as ground truth to validate the values interpolated from the regional data set. The two observation networks cover an overlapping region of the study area. The network of the Genova municipality is located within the boundary of the city and is denser than the one of Regione Liguria that covers the whole study area. Some stations of the regional network are located in the Genova area and allow us to interpolate the rainfall field over the city within a certain accuracy, which will be further estimated comparing the approximation results at these two scales. The results of this test will be the interpolated values at the position of the municipal weather stations. The expected result is a 35x48 matrix.

From the computational point of view, the computation time and scalability of the algorithms were compared over the two DTMs and the executables run on the CNR-IMATI high-performance cluster. Participants were further asked to briefly discuss (i) the robustness to noise of the algorithm and (ii) user-selected parameters. Details on the participants and the evaluation metrics are shown in the Appendix.

5 FUTURE PLANS

The set up and the yearly run of open contests are important for the dissemination of the IQmulus results on data processing to the whole scientific community and user groups. Besides the presentation of the IQmulus Processing Contest tracks during the IQmulus Workshop on Processing Large Geospatial Data and the preparation of the track reports (see the Appendix), three papers related to the contest tracks have been submitted in September to the upcoming special issue of the journal Computers & Graphics on Processing Large Geospatial Data, namely one from the Urban Reconstruction scenario and two from the Urban Classification scenario.

Thanks to the feedback received from the members of the Advisory Board and the participants of the tracks of the first IQmulus contest, new challenges are planned for the next editions of the contest; among others we list:

- Feature detection: extraction of significant shape characteristics (either geometric or semantic);
- Change detection: an acute issue in geospatial data and related to the task 4.5 of the IQmulus project.

For the next years we plan to strengthen the links with the panel of experts, even leaving the coordination of some tracks to members of the Advisory Board. In addition, we plan to increase the number of members of the Advisory Board.

In addition, we are exploring the organization of special issues of journals with executable code, following the example of the "Elsevier executable paper" based on the Collage Authoring Workbench published in 2013 by the Computers & Graphics journal.

6 CHANGE OF THE TASK LEADER

Since October, the 1st, 2014, the original task leader Dr. Silvia Biasotti (CNR-IMATI) is on maternity leave and the leadership of the task has been taken over by Dr. Marco Attene, also from CNR-IMATI.

7 APPENDIX

The Appendix contains the three reports of the tracks presented at the IQmulus Workshop on Processing Large Geospatial Data held in Cardiff, July the 8th, 2014 co-located with the Symposium on Geometry Processing (SGP) 2014.

IQmulus Processing Contest (IQPC) 2014: Urban 3D Model Generation

Linh Truong-Hong and Debra F. Laefer

Abstract Over last decade, several automatic approaches have been proposed to automatically reconstruct 3D building models from aerial laser scanning (ALS) data and those have been benchmarked with data sets typically having less than 25 point/m^2. However, these test data lack of geometric points on vertical surfaces. With a recent improvement of sensors integrated into airborne laser scanners and change to flight path planning, the quality and density of ALS data have been improved significantly. The paper presents the objective of the track Urban 3D Model Generation in the IQmulus contest 2014 2014 with such a high quality data set. This tract encouraged participants to test automatic approaches on building detection and 3D building model reconstruction based on ALS data captured over a 1 km^2 area of the Dublins city centre with density up to 225 points/m^2. Furthermore, evaluation strategies have been proposed to benchmark the results in terms of quality of the models involving geometrical accuracy and destined levels of detail of the model.

1 Introduction

Urban three-dimensional (3D) model generation is a rapidly growing topic for a large range of applications (e.g. environmental planning and monitoring, computational simulation, disaster management, security, telecommunications, location-based services). Currently, the raw data used for 3D building reconstruction at a city scale is obtained from various resources through a wide range of techniques. One of these involved, a recent development of laser technology offering airborne sensors

Linh Truong-Hong
Urban Modelling Group (UMG), School of Civil, Structural and Enviromental Engineering (CSEE), University College Dublin (UCD), Belfield, Dublin 4, Ireland, e-mail: linh.truonghong@ucd.ie

Debra F. Laefer
UMG, CSEE, UCD, Belfield, DUblin4, Ireland e-mail: debra.laefer@ucd.ie

can acquire high density data for building detection and reconstruction. Although working on 3D building reconstructions has been active for over a decade, building detection and reconstruction still remain an active research topic because the raw data are constantly improving, as well as user expectations particularly respect to facade details [1].

Currently, existing methods in 3D building model reconstruction from airborne laser scanning (ALS) data first extract data points of roofs and reconstruct roof models [2, 3]. A complete building model is then generated by extruding roofs outlines to a ground plane or extruding building footprint with the building heights [4, 5]. These methods have the disadvantages of assuming the availability of accurate a priori data and no overhanging roofs. Unfortunately neither assumption can be relied upon. The current methods have been benchmarked through testing on the same data sets [6, 7]. However, during collecting these data sets, geometric data points on vertical surfaces of buildings have not been considered. This causes difficulty in automatically extracting and generating buildings with more facade details. Thus, this contest challenges researchers to propose and test 3D building reconstruction algorithms to detect building outlines and to achieve accurate building models for a $1km^2$ area of Dublin city centre, Irelands capital. The competition will be based on a highly dense ALS data set (typically 225 pt/m^2) acquired by the Urban Modelling Group at the University College Dublin. The data acquisition has been designed to get unprecedented density of data points on vertical surfaces of the buildings [8].

The participants are encourage to run Task A and/or Task B of this track. The goal of Task A is to extract a point cloud of individual buildings in the study area. The evaluation will focus on geometrical accuracy of the building outlines. On the other hand, in Task B, the participants are encouraged to reconstruct 3D building models from a sub-data set. The evaluation goal is benchmarked not only level of detail (LoD) of the submitted 3D building models but also their quality involving the roof shape and geometrical accuracy of the models. Details of the evaluation strategy are described in Section 4. The participants should submit results to the competition organizers for evaluation based on ground truth. There will be separate winners and two runners up for each task. The winners of each task will be selected based on an overall evaluation score, using the criteria described in Section 4. As the submitted results is in evaluation, this paper reports information about the data, task objectives, and evaluation strategies.

2 Data

The test area is approximately $1km^2$ and contains 205 blocks each of which may contain in excess of a dozen buildings per block (Fig. 1). The buildings in this area generally abut one another, with some buildings sharing an adjoining wall. In general, individual buildings are closely spaced. The area contains mostly four-storey masonry (brick and stone) buildings. The typical building is 11-15m in height, less than 5m in wide and 6-10m in long.

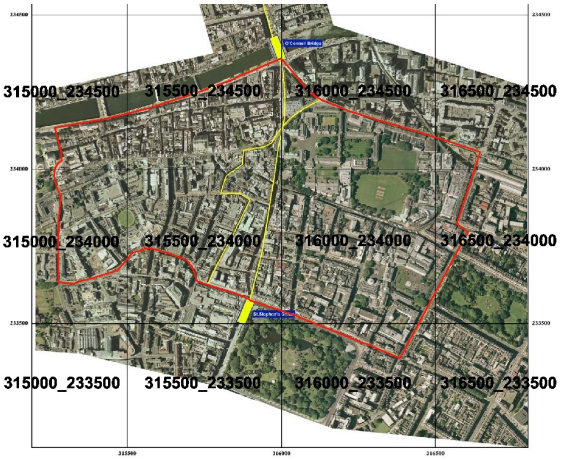


Fig. 1 Acquired ALS area in Dublin central and ALS tiles (contest area outlined in red)

The data was gathered by ALS using the FLI-MAP 2 system, which generated 1000 pulses for each scan line. A total of 370,154 scan lines were acquired. The system operated at a scan angle of 60 degrees, with an angular spacing of 60/1000 degrees between pulses, which is roughly equal to one milli-radian. The quoted accuracy of the FLI-MAP 2 system is 8cm in the horizontal plane and 5cm in the vertical direction, including both laser range and navigational errors. Acquired points were provided in a global coordinate system with reference to the National Irish Grid (Irish Grid), relating to the use of a Global Navigation Satellite System (GNSS) to determine the aircraft position during scanning. The FLI-MAP 2 system is capable of recording up to four echoes for each emitted pulse with the echo distribution in Table 1. In this data set, the vast majority of points are first echoes. Secondary echoes constitute only a small portion of the points, as the overwhelming majority of surfaces in the study area are solid, in the form of streets and buildings. The FLI-MAP 2 system also provides spectral data with intensity values. The RGB dataset is unfortunately not available due to an equipment malfunction during the data capture.

Table 1 Echo distribution of acquired ALS points

Echo	Count	Percentage (%)
1st	217,497,975	96.33
2nd	7,902,595	3.5
3rd	383,840	0.16
4th	4,028	0.001784
Total	225,788,438	100

The flight path was designed to maximize vertical building wall data capture (Fig. 2). The dominant directions of the flight tracks were chosen to be north-east, north-west, south-east and south-west. In total, 44 flight strips were acquired in

which 2823 flight path points were recorded. The flying altitude was selected to be as low as possible (with respect to approval by the Irish Aviation Authority) and varied between 380 to 480m, with an average value of 400m. As a result, approximately 225 million points were acquired within the study area, giving a typical point density of 225 points/m^2. For further information about this ALS data, participants are referred to [9]. The data set is organized in 9 files, each representing a 500m x 500m tile. The 5.8 Gb of data have an LAS extension. The data for both Task A and B can be found at: https://sites.google.com/site/iqmuluscontest2014/

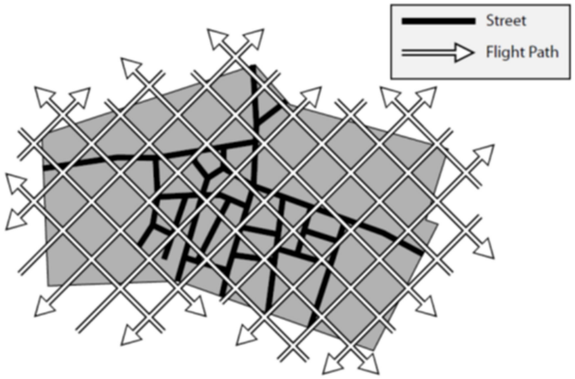


Fig. 2 Designated flight path

3 Task

The participants are asked to run Tasks A and/or B. Task A is to determine the total number of individual buildings in the study area. In an urban region, city blocks are dominated by closely space or abutted buildings (Fig. 3). This makes automated detection challenging as one cannot rely upon a distinctive change in elevation around each building.

The results of Task A will be an ALS dataset of each building and should be submitted in ASCII file, where each row represents the x, y, and z coordinates of the data point. The file name should be Building_X1_Y1_X2_Y2, where the pairs X1 and Y1 and X2 and Y2 are two opposite corners of the bounding box of the dataset on a horizontal plane.

Task B is to undertake three-dimensional building reconstruction of the buildings on one side of one block from the provided data file (Fig. 4). at a minimum of a Level of Detail 2 (LoD2) as defined according to the CityGML standard defined [11]. With LoD2, the generated building models should consist of all visible exterior walls, as well as major roof components (e.g. roof surfaces and small dormers) but without roof overhangs or facade details. For those attempting LoD3, for this competition

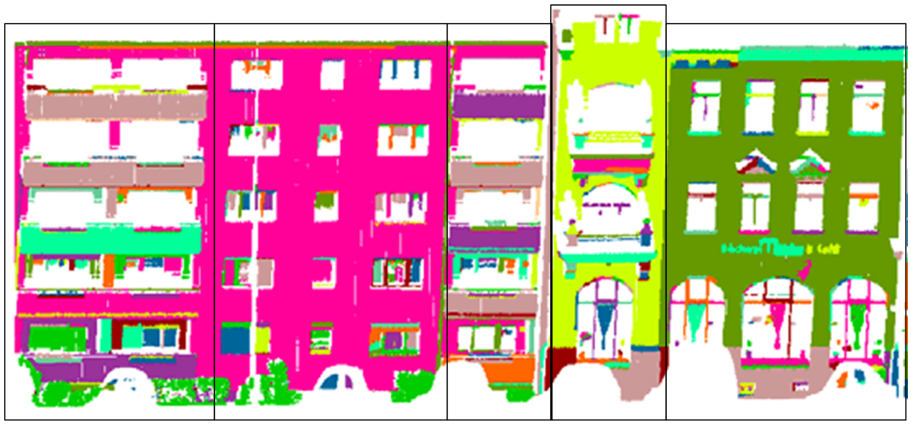


Fig. 3 Extract data of an individual building (after [10])

the reconstructed building models must generate all facade openings with respect to doors and windows. The evaluation score will be determined based on quantity and quality of the reconstructed buildings, as describe in Section 4.

For Task B, the results should be submitted in a DXF file format, compatibile with a standard CAD platform. The geometry of each building model should be stored as either a polygon or a 3D solid. Within this file, components of the building model (e.g. exterior vertical walls or roofs) should be arranged in a separate layer as part of the DFX file. For example, a building roof may be stored in a layer entitled *roof*. In addition, the participants are asked to submit a brief description of the approach implemented to reconstruct 3D building models.

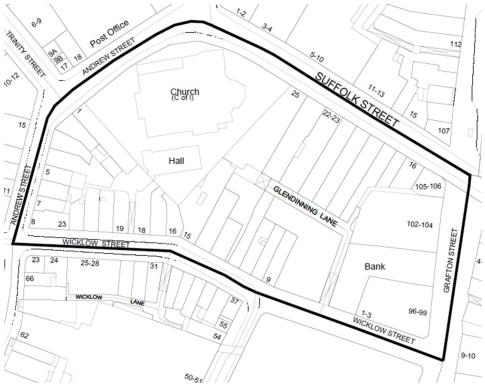


Fig. 4 Plan view of the Task B study area

The participants have asked to submit the results at:
http://collage.ge.imati.cnr.it/workbench

4 Evaluation

4.1 Ground truth

For Task A, the *ground truth* will be 2D footprints provided by the Ordnance Survey Ireland (OSI).

For Task B, the *ground truth* reference models will be a set of independently measured architectural drawings. The generated building models will be evaluated by checking roof surfaces, as well as vertical faade surfaces. Please note that the vertical facade surfaces should all openings (e.g. doors and windows), however details of these openings (e.g. window frames or lintels) are not required. Other architectural detailing may also be neglected.

4.2 Methodology

Accuracy will be assessed as by calculating discrepancies from the ground truth. For Task A, the overall building detection will be evaluated in terms of completeness and correctness, as well as the quality of the results. The concept of true positives, false negatives, and false positives (as proposed by [12]) will be applied by using group classification into either *object* or *background*. A True Positive (TP) is an entity which is classified as the *object group* that also corresponds to an object in the reference model. A False Negative (FN) is an entity classified as *object group* in the reference model corresponding to *background group*, while a False Positive (FP) is an entity classified as *object group* that does not correspond to an object in the reference model [12]. Therefore, evaluation quantities can be expressed using Eq.s 1-3.

$$Comp = \frac{||TP||}{||TP|| + ||FN||} \tag{1}$$

$$Corr = \frac{||TP||}{||TP|| + ||FP||} \tag{2}$$

$$Quality = \frac{||TP||}{||TP|| + ||FP|| + ||FN||} \tag{3}$$

For Task B, the generated building models will be mapped onto the reference models by the competition organisers based upon each buildings centroid, as computed using the edge coordinates of the building at the ground level (Fig. 5). Evaluation of the building model will be assessed according to the following: (1) Building detection, (2) Vertical wall detection, (3) Roof configuration, and (4) Opening details.

1) Building Detection will be determined as per the evaluation in Task A

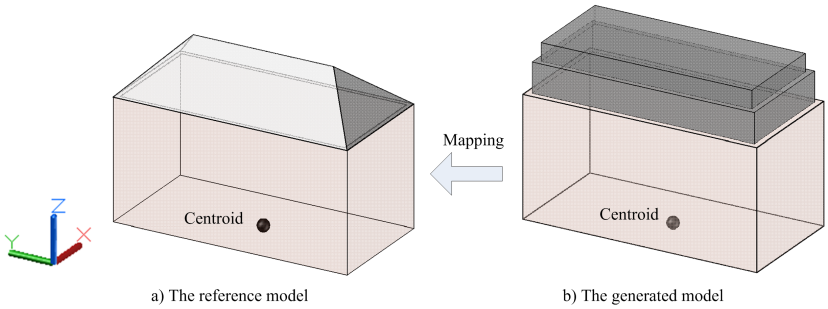


Fig. 5 Mapping the sample model to the reference one for building evaluation

2) Vertical Wall Detection - To determine the accuracy of the vertical walls, a distance metric will be used. Specifically, the Euclidean distance between corners of the building surfaces from the reference and sample models will be computed using their coordinates. To provide a comparison, the distance mean, standard deviation (STD), and root mean square error (RMSE) will be evaluted.

(3) Roof Models - The Euclidian distance metric calculated in the previous section will also be used to evaluate the roof model. In addition however, specialty shapes or major elements (e.g. water tanks) must also be visually recognizable. Notably, if the generated roof model is incomplete, the maximum difference between the completely generated roof model and the reference ones will be adopted for the incomplete roof model. This value will be used to calculate RMSE.

(4) Opening details - To evaluate the building openings (e.g. doors and windows), the generated building surfaces will be mapped onto those of the reference model, using the centroid surface at the ground level, which is computed from edges coordinates at the ground, and the normal vector of the surface (Fig. 6). Opening positions and dimensions will be evaluated. The opening position will be assesed according to a single corner of the opening. This comparison will also be assessed using the mean, STD ,and RMSE (according to [13]).

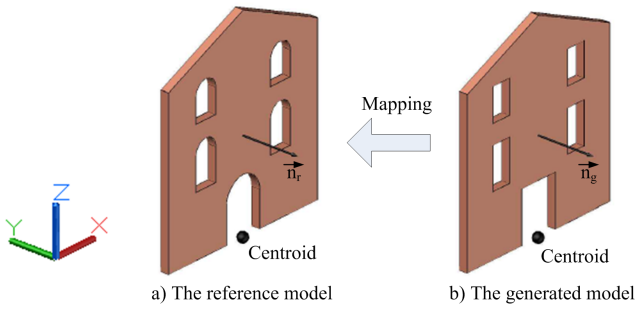


Fig. 6 Mapping openings to evaluate their accuracy

4.3 Scoring

For Task A, submission will be scored for a maximum of 10 points each for the three quantities of completeness, correctness and the quality (Table 2). In the case of a tie, the score derived from the completeness will be used to decide the award.

Table 2 Score range for the comparative quantities

Percentage of comparative results	Score	Percentage of comparative results	Score
40%-49%	1	75%-79%	6
50%-59%	2	80%-84%	7
60%-64%	3	85%-89%	8
65%-69%	4	90%-94%	9
70%-74%	5	$\geq$95%	10

For Task B, 4 of the possible 13 buildings needing reconstruction will be used to check the submissions. Exactly which 4 buildings will not be pre-disclosed to the participants. The total score for this task is 25 points in accordance with the information listed below:

a) For overall building detection, a total score is 10: 4 for the completeness, 3 for correctness, and 3 for the quality of the results. The score range for each quantity is listed in Table 3-5.

Table 3 Score range for the completeness

Percentage of completeness	Score	Percentage of completeness	Score
61%-70%	1	81%-85%	3
71%-80%	2	$\geq$86%	4

Table 4 Score range for the correctness

Percentage of correctness	Score
71%-80%	1
81%-87%	2
$\geq$ 87%	3

b) Vertical wall detection will be scored from 1 to 5 based on the average RMSE of all generated buildings (Table 6).

c) The roof model will also be scored from 1 to 5 based on the average RMSE between the reference roof model and the generated ones (Table 7).

Table 5 Score range for the quality of the results

Percentage of the quality of the results	Score
51%-65%	1
66%-70%	2
$\geq$71%	3

Table 6 Score range for the vertical wall detection

Average RMSE (m)	Score	Average RMSE (m)	Score
1.20-1.39	1	0.67-0.74	4
0.90-1.19	2	$\leq$ 0.67	5
0.75-0.89	3		

Table 7 Score range for the roof model

Description	Score	Description	Score
No roof model or RMSE $\geq$ 0.70m	0	0.40m $\leq$ RMSE $\leq$ 0.49m	3
0.60m $\leq$ RMSE $\leq$ 0.69m	1	0.35m $\leq$ RMSE $\leq$ 0.39m	4
0.50m $\leq$ RMSE $\leq$ 0.59m	2	RMSE $\leq$ 0.34m	5

d) Opening details will be scored from 0 to 5. If an opening is not detected, the total score will be reduced proportionally (e.g. if 4 of the 5 openings were detected with an average RMSE resulting in a score of 3, only 80% of the score 3 will be awarded). For Task B, if the participants have the same score, the score from the evaluation of the opening details will be used to decide the award.

5 Conclusion

Over last decade, automatic approaches have been proposed to reconstruct 3D building models and those have been benchmarked. However, these evaluation processes are tested on data sets with low density and minimal points on the building facades. The paper presents the objective of the track Urban 3D Model Generation in the IQmulus contest 2014. This tests the automatic approaches focussing on building detection and reconstruction. The test data set is consisted of ALS data captured over 1km^2 of the Dublins city centre with density up to 225 points/m^2, acquired by Urban Modelling Group at the University College Dublin in 2007. The data acquisition was designed to maximize data points on vertical surfaces of the buildings. Furthermore, an evaluation strategy has been proposed to benchmark the results in terms of quality of the models involving geometrical accuracy and a level of detail of the model. The test data sets still remain on the webpage of the track after the

IQmulus contest opening in Cardiff, UK. Results are welcome to be submitted for future evaluation which will be included in the final report, which will appear in a special issue of the Computer and Graphics Journal.

Acknowledgements Data acquisition was generously supported by Science Foundation Ireland Grant 05/PICA/I830 and Ireland's Environmental Protection Agency Grant 2005-CD-U1-M1. This work was funded by the European Unions ERC-2012-StG_20111012 Project 307836. The authors gratefully acknowledge this support.

References

1. Brenner C (2013) Building Extraction. Chapter In Airborne and Terrestrial laser scanning, edited by Vosselman G, and Maas H-G, 169-207. Place Published Whittles Publishing, Scotland, UK.
2. Perera, GSN, Maas H-G. (2014) Cycle graph analysis for 3D roof structure modelling: Concepts and performance. ISPRS Journal of Photogrammetry and Remote Sensing 93:213–226
3. Xiong B, Elberink SO, Vosselman G (2014) A graph edit dictionary for correcting errors in roof topology graphs reconstructed from point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 93:227–242
4. Zhou QY, Neumann U (2010) 2.5 D dual contouring: a robust approach to creating building models from Aerial LiDAR point clouds. Computer Vision–ECCV 2010, 115–128
5. Laycock RG, Day AM (2003) Automatically generating large urban environments based on the footprint data of buildings. Proceedings of the eighth ACM symposium on Solid modeling and applications 346–351
6. Kaartinen H, Hyyppä J, Gülch E, Vosselman G, Hyyppä H, Matikainen L, Hofmann AD, Mäder U, Persson Å, Söderman U, others (2005) Accuracy of 3D city models: EuroSDR comparison. International archives of photogrammetry, remote sensing and spatial information sciences 36(3/W19):227–232
7. Rottensteiner F, Sohn G, Jung J, Gerke M, Baillard C, Benitez S, Breitkopf U (2012) The ISPRS benchmark on urban object classification and 3D building reconstruction. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences I-3 293–298
8. Hinks T, Carr H, Laefer DF (2009) Flight optimization algorithms for aerial LiDAR capture for urban infrastructure model generation. Journal of Computing in Civil Engineering 23(6):330–339
9. Hinks T (2011) Geometric Processing Techniques for Urban Aerial Laser Scan Data. PhD thesis, University College Dublin, Dublin, Ireland.
10. Arachchige NH, Maas H-G (2012) Automatic Building Facade Detection in Mobile Laser Scanner point Clouds
11. Gröger G, Kolbe TH, Nagel C, Hafele KH (2009)OpenGIS City Geography Markup Language (CityGML) Encoding Standard (OGC 12-019). Version 2.0.0. OGC 12-019. Open Geospatial Consortium. Available via OGC.
http://www.opengeospatial.org/standards/citygml. Cited 11 March 2014
12. Rutzinger M, Rottensteiner F, Pfeifer N (2009) A comparison of evaluation techniques for building extraction from airborne laser scanning. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE 2(1):11–20
13. Truong-Hong L, Laefer DF (2013) Validating Computational Models from Laser Scanning Data for Historic Facades. ASTM Journal of Testing and Evaluation 41(3):16

TERRAMOBILITA/IQMULUS URBAN POINT CLOUD CLASSIFICATION BENCHMARK

M. Brédif[1], B. Vallet[1], A. Serna[2], B. Marcotegui[2], N. Paparoditis[1]

[1] Université Paris-Est, IGN/SR, MATIS, 73 avenue de Paris, 94160 Saint Mandé, France
[2] Centre de Morphologie Mathmatique (CMM), 35 rue Saint Honor, 77305 Fontainebleau, France

KEY WORDS: Benchmark, urban scene, laser scanning, mobile mapping, classification, segmentation, detection, analysis

ABSTRACT:

The object of the TerraMobilita/iQmulus 3D urban analysis benchmark is to evaluate the current state of the art in urban scene analysis from mobile laser scanning (MLS). A very detailed semantic tree for urban scenes is proposed (cf Figure 1). We call analysis the capacity of a method to separate the points of the scene into these categories (classification), and to separate the different objects of the same type for object classes (detection). The ground truth is produced manually in two steps using advanced editing tools developed especially for this benchmark. Base on this ground truth, the benchmark will aim at evaluating both the classification, detection and segmentation quality of the submitted results.

1 INTRODUCTION

Nowadays, LiDAR technology (Light Detection And Ranging) has been prospering in the remote sensing community. We can find several developments such as: Aerial Laser Scanning (ALS), useful for large scale buildings survey, roads and forests; Terrestrial Laser Scanning (TLS), for more detailed but slower urban surveys in outdoor and indoor environments; Mobile Laser Scanning (MLS), less precise than TLS but much more productive since the sensors are mounted on a vehicle; and more recently, "stop and go" systems, easily transportable TLS systems making a trade off between precision and productivity.

Thanks to all these technologies, the amount of available 3D geographical data and processing techniques has bloomed in recent years. Many semi-automatic and automatic methods aiming at analyzing 3D urban point clouds can be found in the literature. It is an active research area. However, there is not a general consensus about the best detection, segmentation and classification methods. This choice is application dependent. One of the main drawbacks is the lack of publicly available databases allowing benchmarking.

In the literature, most available urban data consist in close-range images, aerial images, satellite images but few MLS datasets. Moreover, manual annotations and algorithm outputs are rarely found in available 3D repositories. Some annotated 3D MLS datasets publicly available are the Oakland 3D point cloud dataset **?**, and the Paris-rue-Madame dataset Serna et al. (2014)

In this context, this paper presents a benchmark that aims at stimulating researchers from different fields such as Computer Vision, Computer Graphics, Geomatics and Remote Sensing, working on the common goal of processing 3D data, benchmarking segmentation and classification methods working on 3D MLS data. This will provide a ground for cross-fertilization and discussions on the future challenges in this important research area. More information about this benchmark are available on the webpage: `http://data.ign.fr/benchmarks/UrbanAnalysis/index.html`

2 EXPERIMENTAL DATASET

The database contains 3D MLS data from a dense urban environment in Paris (France), approximately 10 km long. The acquisition was made in January 2013.

Figure 1: 3D View of the dataset.

This database is produced in the framework of the iQmulus and TerraMobilita projects. It has been acquired by Stereopolis II, a MLS system developed at the French National Mapping Agency (IGN). Annotation will be carried out in a manually assisted way by the MATIS laboratory at IGN.

In this database, the entire 3D point cloud is segmented and classified, i.e. each point contains a label and a class. Thus, point-wise evaluation of detection-segmentation-classification methods becomes possible. The datasets and their processing results must be presented in PLY format with little endian encoding. All coordinates are geo-referenced (E,N,U) in Lambert 93 and altitude IGN1969 (grid RAF09) reference system, reflectance is the laser intensity. An offset has been subtracted from the XY coordinates with the aim of increasing data precision: X0 = 649000 m and Y0 = 6840000 m. Each vertex contains the attributes presented in table 1.

3 ANALYSIS PROBLEM STATEMENT

The problem adressed by this benchmark is to perform a point-wise segmentation and classification. Each processed file, provided by each participant, must be a PLY file containing the original points (in the original order), their original attributes and two additionnal attributes: id and class. All the 3D points belonging to the same object will have the same object identifier (id). Thus, the number of different ids in the 3D point cloud corresponds to

Type	Properties	Description
`float32`	x,y,z	Mesured position
`float32`	x0,y0,z0	Sensor position
`float32`	reflectance	backscattered intensity corrected for distance
`uint8`	num_echo	number of the echo (to handle multiple echoes)
`uint32`	id	object identifier in the segmentation
`uint32`	class	class label assigned to its segmented object.

Table 1: Vertex Properties. Positions are expressed in the Lambert 93 system. Two points having the same id must have the same class. Since each point of the dataset contains an id and a class, the evaluation will be carried out in a point-wise way.

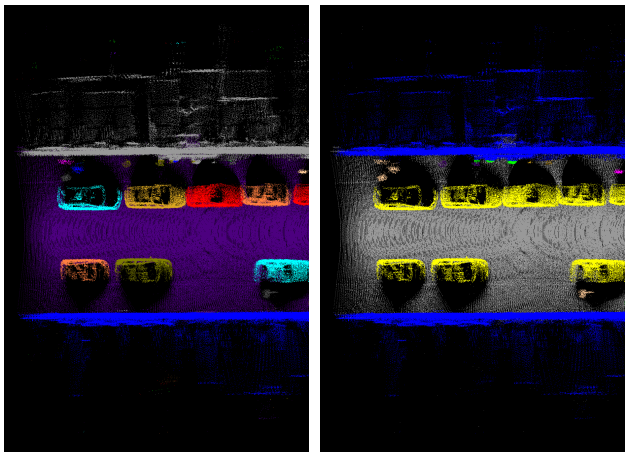


Figure 2: (left) Ids. (right) Classes.

the number of objects. In the classification step, a category is assigned to each segmented object. Each class represents a semantic urban entity, as defined in the following subsection. If two points have the same label, they must have the same class.

Figure 2 shows an example of labels and classes on a 3D point cloud. In the left image, note that ground, facades, pedestrians and cars are represented with different colors because they are different objects and have different labels. In the right image, the colors represent the object classes: ground in gray, facades in blue, cars in yellow, pedestrians in skin color, and furniture in red and cyan.

3.1 Classification Ontology

In this contest, a hierarchy of semantic classes has been defined. The class tree (figure 3) is downloadable as an xml file from `http://data.ign.fr/benchmarks/UrbanAnalysis/download/classes.xml` and is composed as follows:

Surface Surface of unbounded or very large objects

- **Ground** Ground surface
- **Building** All points on the outside of the building
- **Other Surface** Surface of unbounded objects that does not fit in one of the following categories

Object All semantic objects

- **Static** Objects that are not made to be easily moved
- **Dynamic** Individuals or objects that can move easily

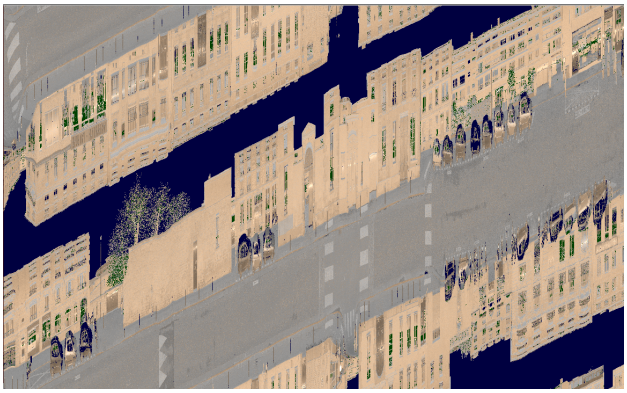


Figure 4: Lidar point cloud viewed in sensor-space : horizontal axis corresponds to time and vertical axis corresponds to rotation angle.

- **Natural** Natural objects or vegetation
- **Other Object** Objects that do not fit in the other categories

Other Undefined points, outliers, points inside buildings...

Unclassified Not classified yet

The tree is voluntarily very detailed as we aim at producing a ground truth that can be usefull to a wide range of methods. Participants can choose the classes that they want in this tree and the evaluation will be performed accordingly. The **Unclassified** label is used to only focus the evaluation on the portion of the point cloud that has been classified only. **Other X** labels are scattered throughout the hierarchy so that the classifications may differentiate between a classification that did not try to distinguish among the more specialized child labels of an inner node of the classification hierarchy, and a classification that expresses that the relevant label is not part of the classification hierarchy.

4 SEMI-AUTOMATIC GROUND TRUTH PRODUCTION

As soon as any form of algorithm is used to produce the Ground Truth that will be used to evaluate a classification result, this particular algorithm will bias the results to favor similar approaches. That is why we have strived to propose an approach that is as manual as possible, thereby reducing algorithmic bias, while providing an editing tool that enables an efficient segmentation and classification of the benchmark dataset. Therefore, we have set up a semi-automatic approach where the user has a full control over the resulting segmentation and classification.

4.1 Segmentation

Navigating and selecting through a point cloud is a counter-intuitive task, due to the absence of proper occlusion and its sparse nature. As the segmentation was to be performed at the point level rather than alternative representations such as bounding boxes, an efficient browsing and segmentation of the point cloud was a key issue. We tackled this problem by proposing an interface that shows the pointcloud in sensor space (figure 4). This was made possible and convenient by the geometry of acquisition and the availability of its two parameters in the raw dataset : the constant time step dt_{pulse} between each emitted pulse and the constant time step dt_{turn} between each rotation of the lidar. Please not that we do not require that each turn is composed of an integral number of pulses, which yield overall shear in figure 4. This

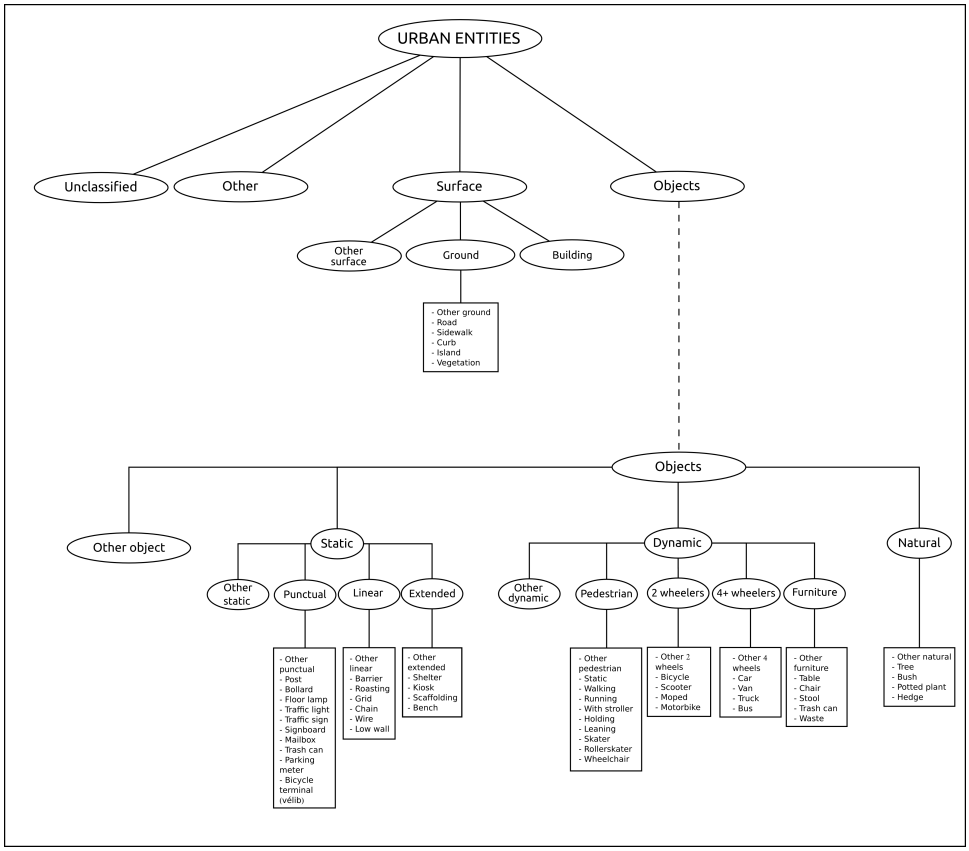


Figure 3: Hierarchy of class annotations.

parameter allow to recover a regular topology out of the point cloud stream : the pulse neighbors are the immediately preceding and succeeding pulses and the closest ones on the preceding and succeeding rotations. The recovery of such a 2D topology with 4-connectivity enables the use of traditionnal image processing techniques. A particularity is on its topological boundary : it is composed of two topological circles corresponding to the pulses of the first and last rotations. This is due to the continuous sampling of the particular scanner used to acquire this dataset : the last pulse of a rotation is connected to the first pulse of the following one.

In this 2D sensor-space, it is then trivial to create, maintain a 2D segmentation by providing to the user a graph edition tool : node creation and deletion, edge creation between existing nodes, polyline input, insertion of a node by splitting an edge... In order for this graph to define a partition of the sensor measurements, nodes are placed at pixel corners and edges are manhattan curves along pixel boundaries. The user experience is improved by allowing zooming, panning, and snapping.

The only automatic task used in this editing tool applies when a user adds an edge between two nodes or when one is modified by a moving end node. Two modes are available to rasterize a path between the two end nodes of the path on the pixel boundaries :

Sensor-space Line This is a simple rasterization of the sensor-space line segment as a Manhattan curve passing through pixel corners.

Optimal Path This mode performs an $A*$ search to provide the rasterized path which is optimal with respect to the sum of weights of the traversed pixel boundaries. This $cost$ is expressed in terms of the absolute differences of the measured quantities $\Delta_\alpha, \Delta_d, \Delta_r$, where α is the angle between the normal and the vertical direction, d is the measured distance and r is the measured reflectance, and $\lambda, \alpha_0, d_0, r_0$ are empirical weights :

$$cost = \max\left(1, \lambda\left(1 - \max\left(\frac{\Delta_\alpha}{\alpha_0}, \frac{\Delta_d}{d_0}, \frac{\Delta_r}{r_0}\right)\right)\right) \quad (1)$$

The user experience is improved by preprocessing the shortest path from the selected node as a background process, so that the optimal path may be updated efficiently as the mouse moves. The user click then only validates the already optimized and displayed path, increasing the productivity.

4.2 Classification

As the number of objects is limited compared to the massive size of the point cloud, assignation of class labels to an already segmented lidar point cloud is performed by a paint-bucket-like interface : the user selects a class label from a drop-down list of available labels and then clicks on the pointcloud to assign this label to the segmented object id. The display mode may be either set to segmentation mode that applies a random palette according to the segment ids or a classification mode where the visualization depends on the class label. To further help this process, points to be rendered may be filtered base on there class labels.

5 SUBMISSION EVALUATION

The proposed semantic tree is very detailed and probably no existing method treats the whole problem. This is why, the participants to the benchmark can choose to analyse the scene using any

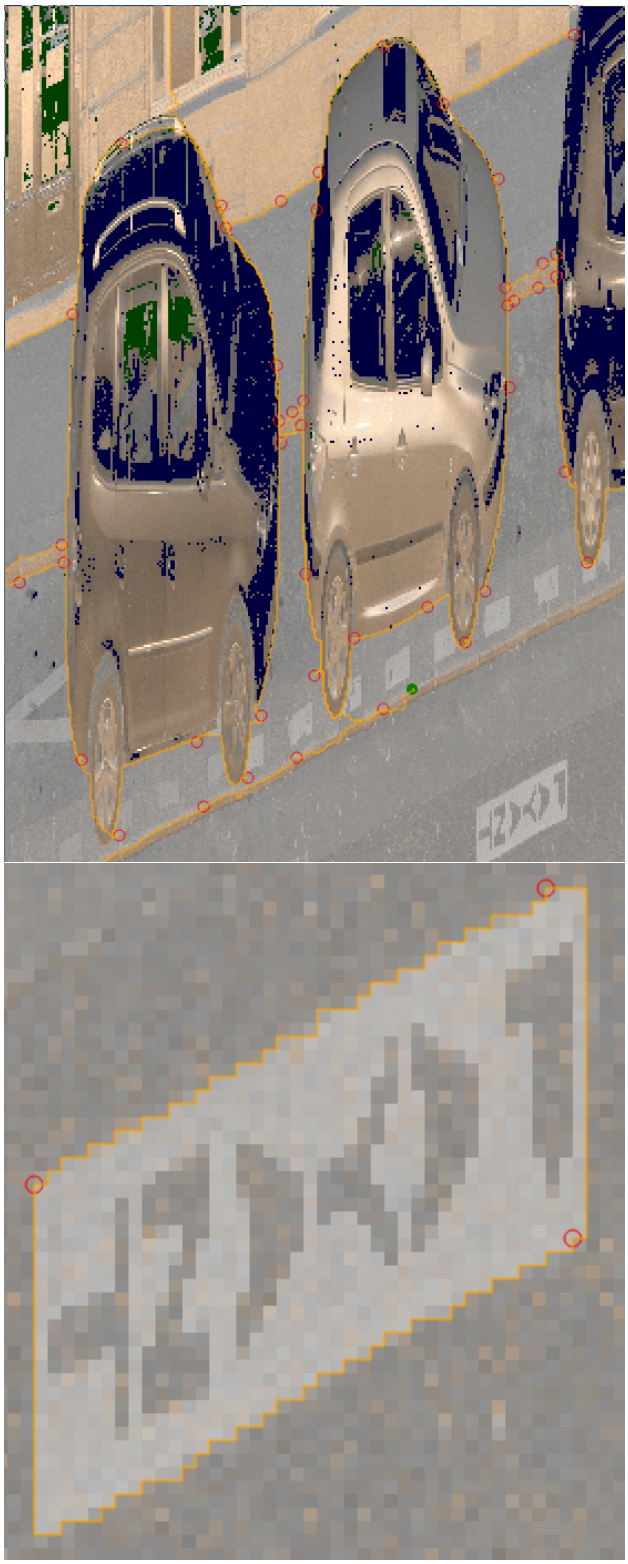


Figure 5: Segmentation-inducing graph.

subtree of the tree. In this case, they will simply apply the "other" label to the nodes that they do not wish to detail. The evaluation will be performed accordingly and only the relevant metrics will be given.

The benchmark does not aim at ranking the participants but at providing insights on the strengths and weaknesses of each methods. We consider that the quality of a method is subjective and application dependent, and the results of this benchmark should only help a user choosing one approach depending on its own specific requirements. Results will be evaluated at three levels: as a classification, as a detection and as a segmentation. Details of the evaluation metrics used are given in the Evaluation Protocol document.

5.1 Classification quality

The classification quality will be evaluated point-wise. The results of the evaluation will be a confusion matrix for each node of the tree that the evaluated method handles. Rows and lines will be the subclass labels from the ground truth and the evaluated method respectively, and matrix values are the percentage of points with the corresponding labels. All nodes from the semantic tree have an "other" class, so participants can classify into less classes than what is given in the tree. For non root nodes, an additional label "not in class" will be given for point that were not classified correctly at a lower level.

5.2 Detection quality

The detection quality work measures the capacity of the method to detect the objects present in the scene. Thus it requires to choose a criterion to decide if an object from the ground truth is detected or not. This biases the evaluation as this choice will impact the result. The solution that we propose is to give the evaluation result for a varying threshold m on the minimum object overlap. In this benchmark, an object is defined by the subset of points with the same object identifier. For a such subsets S^{GT} of the ground truth and S^{AR} of the evaluated algorithm result, we will validate S^{AR} as a correct detection of S^{GT} (a match) iff:

$$\frac{|S^{GT}|}{|S^{GT} \cup S^{AR}|} > m \text{ and } \frac{|S^{AR}|}{|S^{GT} \cup S^{AR}|} > m \qquad (2)$$

where $|\cdot|$ denotes the cardinal (number of objects) of a set. The standard precision/recall are then functions of m:

$$precision(m) = \frac{\text{number of detected objects matched}}{\text{number of detected objects}}$$

$$recall(m) = \frac{\text{number of detected objects matched}}{\text{number of ground truth objects}}$$

Precision/Recall will be evaluated for each object types at each level of the semantic tree that the participants have handled and results will be presented as two curves. Precision/Recall are decreasing in m and this decay indicates the geometric quality of the detection (good geometry implies slower decay).

5.3 Segmentation quality

When the threshold m is below 0.5, the criterion (2) does not guarantee that objects are uniquely matched. When $m < 1/n$, n objects from the ground truth (GT) can be matched to a single object of the algorithm result (AR), or the opposite. Thus for $m < 0.5$ we will also give the curves of over-segmentation (1-to-n) and under-segmentation (n-to-1) by averaging n over the matches defined by (2).

6 PARTICIPANTS

For the moment, two participants have shown interest, but the benchmark will stay open at least until the end of 2014.

6.1 IPF - KIT

The institute of Photogrammetry and Remote Sensing (IPF) from the Karlsruhe Institute of Technology (KIT) will participate with a method that they present as follows:

We propose a new methodology for large-scale urban 3D scene analysis in terms of automatically assigning 3D points respective semantic labels. The methodology focuses on simplicity and reproducibility of the involved components as well as performance in terms of accuracy and computational effort. Exploiting a variety of low-level geometric features and considering recent advancements in order to improve their distinctiveness, the methodology is in principal designed to process point clouds with a few millions of 3D points. For analysing huge 3D point clouds with possibly billions of points for a whole city like Paris, however, an adaptation has to be introduced. For this purpose, we propose an adaptation which is based on a tiling of the scene and thus allows a successive processing in reasonable time without affecting the quality of the classification results. We demonstrate the performance of our methodology on two adequate, labelled 3D point cloud datasets with respect to robustness, efficiency and scalability.

For further details, the reader is encouraged to review Weinmann et al. (2013).

6.2 CMM - MINES ParisTech

The Centre de Morphologie Mathématiques (CMM) from MINES ParisTech will participate with a method that they present as follows:

Our method is based on elevation images and it uses image processing techniques, specially Mathematical Morphology, and machine learning techniques. Our general workflow is presented in Fig. 6 and it consists in the following steps: i) the 3D point cloud is projected onto an elevation image; ii) ground is segmented using the λ–flat zones labeling algorithm, curbs are segmented and their accessibility is analyzed; iii) facades are segmented as the highest vertical structures in the urban scenario; iv) objects are detected as discontinuities on the ground; v) small and isolated structures are eliminated and connected objects are separated using a constrained watershed; vi) objects are classified in several categories using a SVM approach with geometrical and contextual features; and, vii) the segmented and classified images can be reprojected to the 3D point cloud for visualization purposes.

For further details and complete analyses in each step, the reader is encouraged to review Hernández, J. and Marcotegui, B. (2009); Serna, A., Marcotegui, B. (2013a,b, 2014).

Fig. 7 presents a results of our proposed method on a datasets in the *rue Cassette* in Paris.

7 CONCLUSIONS

Results of the contest will be presented at the IQmulus workshop taking place on July 8th, 2014 in Cardiff (UK), in conjunction with SGP'14.

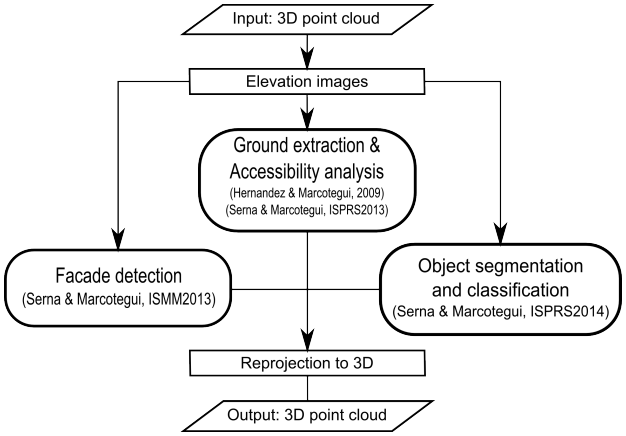
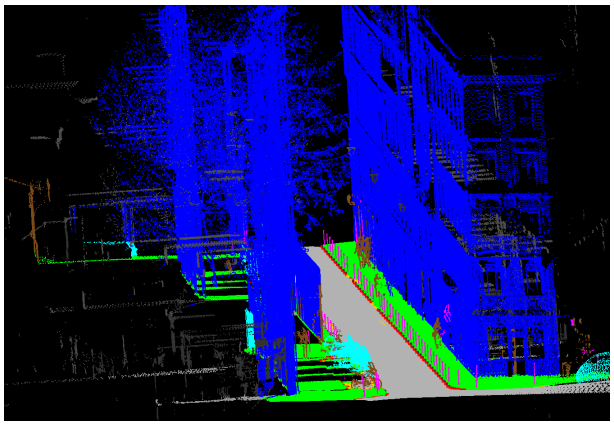


Figure 6: Work-flow of our proposed segmentation methodology.



(a) Rue Cassette, Paris. Acquired by Stereopolis II, IGN France.

Figure 7: 3D object segmentation and classification using our proposed methodology. Facades (blue), road (gray), sidewalk (green), accessible curbs (yellow), non-accessible curbs (red), cars (cyan), pole-like objects (magenta), other (brown) and undefined (dark-gray).

References

Hernández, J., Marcotegui, B.. Filtering of artifacts and pavement segmentation from mobile LiDAR data. In: Bretar, F., Pierrot-Deseilligny, M., Vosselman, M.G., editors. ISPRS workshop Laser scanning '09; vol. XXXVIII-3/W8 of *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Paris, France; 2009, p. 329–333.

Serna, A., Marcotegui, B.. Attribute controlled reconstruction and adaptive mathematical morphology. In: 11th International Symposium on Mathematical Morphology. Uppsala, Sweden; 2013a, p. 205–216.

Serna, A., Marcotegui, B.. Urban accessibility diagnosis from mobile laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing 2013b;84:23–32.

Serna, A., Marcotegui, B.. Detection, segmentation and classification of 3D urban objects using mathematical morphology and supervised learning. ISPRS Journal of Photogrammetry and Remote Sensing 2014;93:243–255.

Serna, A., Marcotegui, B., Goulette, F., Deschaud, J.-E. Paris-rue-Madame database: a 3D mobile laser scanner dataset for

benchmarking urban detection, segmentation and classification methods. ICPRAM 2014. http://cmm.ensmp.fr/~serna/rueMadameDataset.html

Weinmann, M., Jutzi, B., Mallet, C. Feature relevance assessment for the semantic interpretation of 3D point cloud data. ISPRS Workshop on Laser Scanning, Antalya, 2013.

A Comparison of Approximation Methods for Rainfall Data

Giuseppe Patané and Simone Pittaluga and Vibeke Skytt

Abstract The extrapolation of the behavior of rainfall from a sparse set of measured data and its integration with other data sources (e.g., radar data) are crucial to support real-time monitoring, historical data analysis, and effective visualization of the precipitation field over 3D terrain models. The goal of this paper is to compare the accuracy and performances of generic and specialized approximation techniques on rainfall data. To this end, the proposed comparison takes into account the pointwise approximation of rainfall data with LR B-splines, ordinary kriging, radial basis functions and compares their approximation accuracy at a single scale, their cross validation, and their extrapolation capabilities.

1 Introduction

The extrapolation of the behavior of rainfall from a sparse set of measured data and its integration with other data sources such as radar data are crucial to support real-time monitoring and historical data analysis. To this end, approximation techniques provide a standard way to handle sparse and heterogeneous data, to code the uncertainty/reliability of observed measures, and to adapt corrections in case of failures of stations. Furthermore, approximation schemes can be used to define an effective visualization of the precipitation field over 3D terrain models and the rendering of the corresponding accuracy.

In this context, the goal of this paper is to compare the accuracy and performances of generic and specialized approximation techniques on rainfall data, such as LR B-splines, ordinary kriging, and radial basis functions (Sect. 2). To this end,

Giuseppe Patané and Simone Pittaluga
CNR-IMATI, Genova, Italy e-mail: {patane,pittaluga}@ge.imati.cnr.it

Vibeke Skytt
SINTEF, Oslo, Norway e-mail: Vibeke.Skytt@sintef.no

we measure their approximation accuracy at a single scale, their cross validation, and their extrapolation capabilities.

In our experiments (Sect. 3), the *input data set* is gathered from two different rain gauges networks with a different spatial distribution, and the main outcomes of our comparison. The first one is the network developed by Regione Liguria over the whole region, with 143 measure stations. The second one is the measure system deployed by the Genova municipality within the city boundary, with 25 measure stations. Since the temporal interval is different for each network, station rainfalls have been cumulated to a step of 30 minutes. The selected rainy day is September 29, 2013, which was characterized by light rain over Regione Liguria with 2 different thunderstorms that caused local flooding and landslides, without damages.

For the *evaluation of the accuracy of each approximation scheme*, the corresponding approximating function has been sampled at each station and the resulting value has been compared with the ground-truth. All the algorithms show an accuracy under the instrumental error and the best results are achieved by the ordinary kriging.

For the *cross-validation of the each approximation method*, each rain gauge has been turned off and the approximation function has been sampled at this position. In this case, the data set is time-varying and over the different time steps the rainfall field is not homogeneous. All the different methods have the same trend and follow the shape of rain intensity over the different time steps. The ordinary kriging and implicit approximation with RBFs have achieved similar results, which are superior to LR-Splines.

For the *evaluation of the approximation accuracy at different scales*, the rainfall data measured by the Genova municipality has been used as ground truth to validate the values interpolated from the regional data set. The test result is analogous to the cross validation but the aim of this test is to get the capability of different methods to estimate the local features of rain fields interpolated over a sparser data set. Ordinary kriging and LR-Spline have the smaller maximum approximation error but the average error of implicit approximation with RBFs is reduced by the half with respect to the previous methods.

2 Approximation schemes

We briefly review the main properties the following approximation techniques: LR B-splines, ordinary kriging, and radial basis functions.

LR-Splines

An *LR B-spline surface* is a spline surface defined over a box partition [4]. The surface possesses, in contrast to tensor product spline surfaces, the property of local

refinement. The LR B-spline surface also satisfies most of the properties of tensor product spline surfaces, such as non-negative basis functions with local support and partition of unity. More precisely, an LR B-spline surface is a parametric surface and can be represented as

$$F(x,y) = \sum_{i=1}^{K} s_i P_i N_i^{d_1,d_2}(x,y)$$

where P_i, $i = 1,\ldots,K$ are the surface coefficients, s_i, $i = 1,\ldots,K$, are scaling factors to ensure partition of unity and N_i, $i = 1,\ldots,K$ are the corresponding B-splines. The B-splines are piecewise polynomials in two parameter directions with polynomial degree d_1 and d_2. Each B-spline is defined over a regular grid of knots, but the union of all knots in all B-splines does not need to be regular. This property allows the surface to have many degrees of freedom in areas where the data represented by the surface has large variation and few degrees of freedom in other areas.

An LR B-spline surface is a smooth surface and in this paper the surface is approximating rainfall data. The approximating surface is parameterized on the xy-coordinates of the rainfall data, letting $F(x,y)$ be a function approximating the precipitation. The surface is computed in an iterative process. A lean initial, tensor product B-spline surface over the domain covered by the rainfall data is set to zero. At each step, an LR B-spline adaption to the MBA algorithm [7] is used to update the surface and the surface is refined where the accuracy is not within a prescribed tolerance.

Multilevel B-spline approximation (MBA) is a local method where each surface coefficient is updated with respect to the distance between the data points in the support of the corresponding B-spline and the surface. Let (x_c, y_c, z_c), $c = 1,\ldots,C$, be the data points in the support of a B-spline. For each data point the corresponding new coefficient P_i is determined by

$$P_i = \frac{\sum_c (s_i N_i(x_c,y_c))^2 \phi_c}{\sum_c (s_i N_i(x_c,y_c))^2}$$

$$\phi_c = \frac{s_i N_i(x_c,y_c) z_c}{\sum_{j=1}^{J} (s_i N_j(x_c,y_c))^2}$$

and J is the number of B-splines having support at the data point. At each iteration level, the difference between the surface and the data point is approximated by a difference surface, which is then added to the initial surface. It is also ensured that the surface will be non-negative. The refinement step is performed by adding new knot intervals to the surface description. The knot intervals do not need to cover the entire parameter domain, but must divide at least one B-spline. The surface coefficients after refinement is computed by the Oslo algorithm [3].

The approximation is performed in an iterative process and typically about 10 iteration steps are used. At each step, the evaluation and update of the surface is linear in the number of data points. In both cases, $(d_1+1)\times(d_2+1)$ basis functions

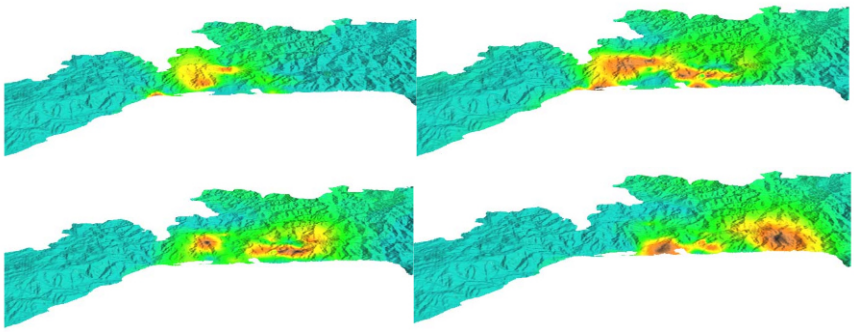


Fig. 1: Evolution (every 30 minutes) of the rain field associated to a thunderstorm over the city of Genoa. The color map begins with blue (smallest rain rate) and then passes through green, yellow, orange, and ends up with brown (highest rain rate).

are computed recursively. In the rainfall approximation, we select $d_1 = d_2 = 2$. The number of refinements performed during the approximation varies with the actual behavior of the rainfall data and less with the number of points. For each new B-spline that is computed, the number of operations is 6 times the degree+1. For the data used in this comparison, a typical accumulated number of B-spline subdivision for all iterations is 1000. If the number of B-spline grow large, then the cost of traversing the data structure outnumbers the computational cost. This is not the case in this context.

The rainfall data is scattered, sparse and the behavior is not overall smooth. In this context, the local behavior of the MBA algorithm ensures that only nearby data points influence the surface and that the surface is well behaved. Moreover, the method of adaptive, local refinements enables degrees of freedom only where it is needed.

Ordinary kriging

Ordinary kriging is a point estimator algorithm in the best linear unbiased estimator family. The estimate is a linear combination of the available data; it tries to be unbiased by having the residual mean equal to zero to minimize the residual error. The point estimate is expressed as $\mathbf{p}_0 = \sum_{i=1}^{n} w_i \mathbf{p}_i$, where $\mathbf{p}_0$ is the estimated value at position "$\mathbf{0}$", $\mathscr{P} := \{\mathbf{p}_i\}_{i=1}^{n}$ are the known samples, and $\{w_i\}_{i=1}^{n}$ are the corresponding weights. These weights are computed as $\mathbf{C}^{-1}\mathbf{D}$, where $\mathbf{C}$, $\mathbf{D}$ are the covariance matrices calculated (i) among all the input points or (ii) among the points to be calculated and all the known data, respectively. In our approach, the model variogram for the computation of the covariance matrix is constant; our implementation can use all the measured rainfall values to compute the covariance

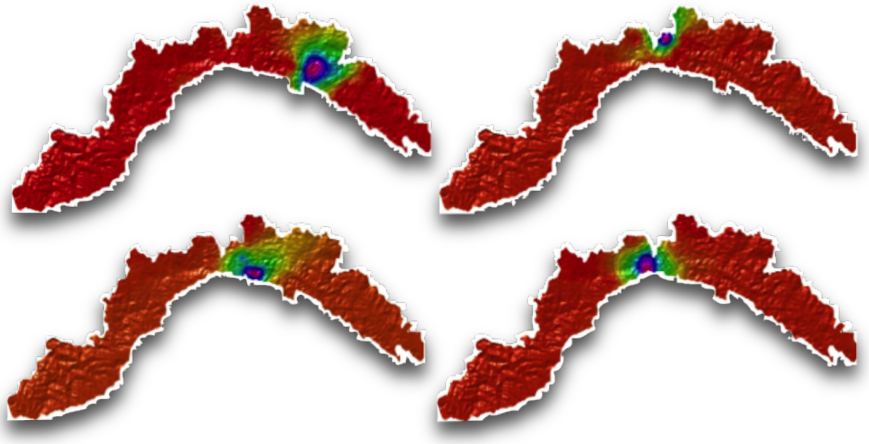


Fig. 2: Color map of the approximation of the rainfall data with RBFs and at 4 different times. The color map varies the hue component of the hue-saturation-value color model; the colors begin with red, pass through yellow, green, cyan, blue, and magenta, and return to red.

matrices or it can be limited to the closest known values, thus reducing the computational cost for inversion of **C**. The matrices **C** and **D** are calculated as functions of spatial continuity, which is measured with the spatial variance expressed as $\gamma(h) = (2N(h))^{-1} \sum_{(i,j)\approx h} |f(\mathbf{p}_i) - f(\mathbf{p}_j)|^2$, where $f(\mathbf{p}_i)$ is the input value at $\mathbf{p}_i$ and the points $\mathbf{p}_i$, $\mathbf{p}_j$ are separated by $N(h)$. The γ values can be plotted against h to get the variogram plot that shows the spatial dissimilarity. The experimental variogram is fitted using a linear combination of basic variogram models. In addition to the spherical, Gaussian, and exponential variorums, we can consider vertical shifts, known as nugget effect, representing the local variance of the data. For each sample, the ordinary kriging estimator needs to solve a linear system whose coefficient matrix. In the worst case, the solution of linear system is computed using the LU decomposition that takes $\mathcal{O}(2/3n^3 + n^2)$. Fig. 1 shows the resampling of the ordinary kriging approximation of a set of sparse rainfall data on a 3D model of Regione Liguria.

Radial basis functions

For the approximation of rainfall data, which are represented as a scalar function $f : \mathscr{P} \to \mathbb{R}$ sampled at a discrete set of points $\mathscr{P} := \{\mathbf{p}_i\}_{i=1}^n$, we apply a global approximation scheme with radial basis functions (RBFs, for short) and globally-supported kernels [2]. According to [1, 8] and choosing a kernel $\varphi : \mathbb{R} \to \mathbb{R}$, the

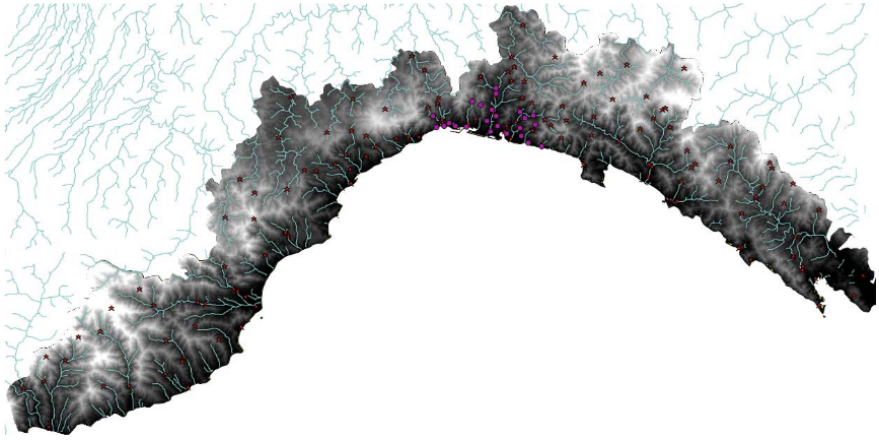


Fig. 3: Input: rainfall measures at 143 stations (regional level, red points) and 25 stations (municipality level, purple circles).

approximation $F:\mathbb{R}^3\to\mathbb{R}$, $F(\mathbf{p}):=\sum_{i=1}^{n}\alpha_i\varphi_i(\mathbf{p})$, $\mathbf{p}=(x,y,z)$, is a linear combination of the radial basis functions $\varphi_i(\mathbf{p}):=\varphi(\|\mathbf{p}-\mathbf{p}_i\|_2)$, centered at $\{\mathbf{p}_i\}_{i=1}^{n}$. Then, the coefficients $\alpha=(\alpha_i)_{i=1}^{n}$ that uniquely satisfy the interpolating conditions $F(\mathbf{p}_i)=f(\mathbf{p}_i)$, $i=1,\ldots,n$, are the solution of the linear system $\mathbf{A}\alpha=\mathbf{f}$, with $a_{ij}:=\varphi(\|\mathbf{p}_i-\mathbf{p}_j\|_2)$ and $\mathbf{p}_i:=(p_{xi},p_{yi},p_{zi})$. We briefly remind that the support of an arbitrary map $g:\mathbb{R}^3\to\mathbb{R}$ is defined as the set $\mathrm{supp}(g):=\{\mathbf{p}:g(\mathbf{p})\neq 0\}$. If $\mathrm{supp}(g):=\mathbb{R}^3$, then g has global support. Common choices of kernels with global support are the Gaussian $\varphi(t):=\exp(-t)$, the harmonic $\varphi(t):=|t|^{-1}$, and the biharmonic $\varphi(t):=|t|^3$ kernel. Fig. 2 shows the resampling of the approximation of a set of sparse rainfall data on a 3D model of Regione Liguria.

For the approximation with RBFs, we solve a linear system whose coefficient matrix is symmetric, positive-definite. Its unique solution is computed through direct solver or pre-factorization of the coefficient matrix. In the current implementation, we apply a direct solver. The computational cost for the evaluation of F is $\mathcal{O}(n^3)$, where n is the number of rainfall stations. The resampling of F on s samples is linear in the number s of samples. Since $s>>n$, the overall computational cost is linear in the number of samples. For the computation of the global approximation, the most time-consuming part is the evaluation of the implicit function at the sample points.

3 Discussion

The data set is gathered from two different rain gauges networks with a different spatial distribution (Fig. 3). The first one is the network developed by Regione Liguria over the whole region, with 143 measure stations. The second one is the mea-

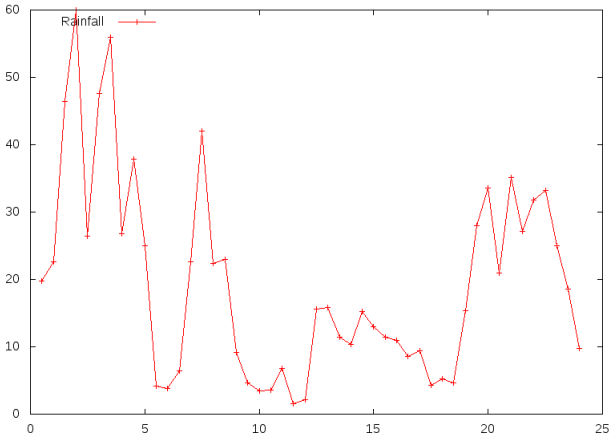


Fig. 4: The maximum rainfall rate [*mm*/30'] (y-axis) recorded over each time step (x-axis).

sure system deployed by the Genova municipality within the city boundary, with 25 measure stations. Since the temporal interval is different for each network, station rainfalls have been cumulated to a step of 30 minutes. The selected rainy day is September 29, 2013, which was characterized by light rain over Regione Liguria with 2 different thunderstorms that caused local flooding and landslides, without damages. The maximum rain-rate over all time step is $60mm/30'$ and the average rain-rate is $19.56mm/30'$.

For each evaluation schema we evaluate (i) the statistics of the error distribution and (ii) the global Mean Squared Error (MSE, for short) expressed as $\mathbf{MSE} = 1/n \sum_{i=1}^{n} (\mathbf{p}_i^* - \mathbf{p}_i)^2$ where $\mathbf{p}^*$ is the estimated values, $\mathbf{p}$ the true values, and n the number of true values. The MSE measures the average square difference between the measured rainfall and its estimate. The MSE value are close to 0 if the approximation is accurate [5].

Accuracy evaluation at a single scale. For the evaluation of the accuracy of each approximation scheme, the corresponding approximating function has been sampled at each station (every 30 minutes) and the resulting value has been compared with the ground-truth. Indeed, the accuracy of the algorithm is understood as the ability to interpolate exactly the measured data set. Our tests (Table 1) outline that all the algorithms submitted have an accuracy under the instrumental error. More precisely, the best results are achieved by the ordinary kriging with a difference between estimated and measured value that is less than $10^{-7}mm$. This result can be explained because the ordinary kriging is an exact estimator. For the LR B-spline surfaces, a tolerance of 0.05 was given as the required approximation accuracy and it has been met for all time steps.

Method	**Max** [mm]	**Mean** [mm]	**Median** [mm]	**Std. dev.** [mm]	**MSE** [mm^2]
Ordinary kriging	3.3E-07	4.3E-11	0	3.8E-09	1.44E-17
RBF	60	1.11	7.1E-14	3.48	12.67
Spline	0.21	0.003	0	0.01	6.99E-5

Table 1: Statistics for the error distribution of the approximation accuracy at a single scale.

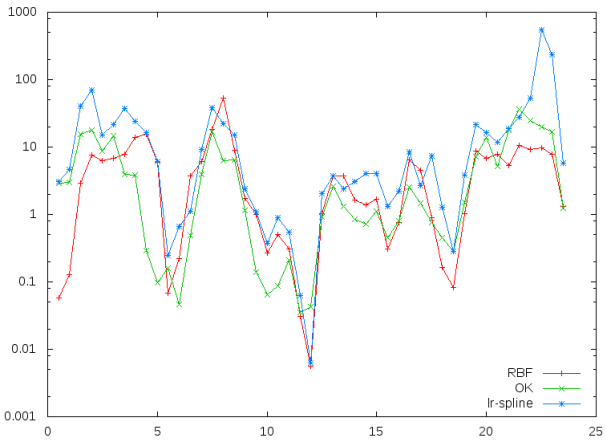


Fig. 5: Cross validation MSE: y-axis reports the MSE [mm^2] for each time step x-axis.

Method	**Max** [mm]	**Mean** [mm]	**Median** [mm]	**Std. dev.** [mm]	**MSE** [mm^2]
Ordinary kriging	32.44	0.02	5.85E-05	2.38	5.64
RBF	37.8	0.97	0.34	2.12	5.44
Spline	57.82	-0.18	3.57E-5	5.29	28.03

Table 2: Summary statistics for the error distribution of the cross validation.

Cross validation. For the cross-validation of the each approximation method and according to Isaaks [6], each rain gauge has been turned off and the approximation function has been sampled at this position. Comparing the input value at the turned-off station with the approximated value at the same location, we have evaluated the extrapolation capability of the selected approximation scheme. The result values are compared using the Mean Squared Error criterion. The MSE is expressed as $\mathbf{MSE} = 1/n \sum_{i=1}^{n} (\mathbf{p}_i^* - \mathbf{p}_i)^2$, where $\mathbf{p}^*$ are the estimated values. The MSE measures the average square difference between the measured rainfall and its estimate.

The data set is time-varying and over the different time steps the rainfall field is not homogeneous. During the whole temporal interval, two local thunderstorms occurred and only few weather stations recorded this heavy event. If the rain gauges that recorded the rain peak are turned off for the cross validation task, then the approximated rain field will underestimate the true value because the data set loses

Method	**Max** [*mm*]	**Mean** [*mm*]	**Median** [*mm*]	**Std. dev.** [*mm*]	**MSE** [mm^2]
Ordinary kringing	28.62	0.59	3.26E-3	4.45	20.21
RBF	36.77	1.41	0.44	3.25	12.58
Spline	31.02	0.6	8.97E-4	4.54	21.02

Table 3: Summary statistics for the error didtribution of the accuracy evaluation at different scales.

details. If the rainfall rate are continuous over the whole region, then the estimate have small error. Comparing Fig. 4 with Fig. 5, we get that all the different methods have the same trend and follow the shape of rain intensity over the different time steps. The ordinary kriging and implicit approximation with RBFs achieve similar results, both have a maximum error lower than LR-Spline and the corresponding MSE is about five times smaller than the one calculated with LR-Spline (Table 2).

Accuracy evaluation at different scales. For the evaluation of the approximation accuracy at different scales, the rainfall data measured by the Genova municipality have been used as ground truth to validate the values interpolated from the regional data set. The two observation networks cover an overlapped region of the study area. The network from Genova municipality is located within the boundary of the city and is denser than the one from Regione Liguria that covers the whole study area. Some stations of the regional networks are located in the Genova area and allow us to interpolate the rainfall field over the city with a given accuracy. Comparing the approximation results at these two scales, we have evaluated the sensitivity of the approximation with respect to the ground truth and their capability of different methods to estimate the local features of rain fields interpolated over a sparser data set. The test and its results (Table 3) are analogous to the cross validation but the actual aim of this test is to get the capability of different methods to estimate the local features of rain fields interpolated over a sparser data set. In this case, ordinary kriging and LR-Spline have the smaller maximum error but the RBFs have a MSE smaller than the previous methods.

Acknowledgements This work has been supported by the FP7 Integrated Project IQMULUS - "*A High Volume Fusion and Analysis Platform for Geospatial Point Clouds, Coverages, and Volumetric Data Sets*". Rain fall data are courtesy of Regione Liguria and ARPAL.

References

1. N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68, 1950.
2. J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3D objects with radial basis functions. In *ACM*

Siggraph, pages 67–76, 2001.
3. E. Cohen, T. Lyche, and R. Riesenfeld. Discrete b-splines and subdivision techniques in computer-aided geometric design and computer graphics. *Computer graphics and image processing*, 14(2):87–111, 1980.
4. T. Dokken, T. Lyche, and K. F. Pettersen. Polynomial splines over locally refined box-partitions. *Computer Aided Geometric Design*, 30(3):331–356, 2013.
5. P. Goovaerts. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of hydrology*, 228(1):113–129, 2000.
6. E. Isaaks and R. Srivastava. *An Introduction to Applied Geostatistics*. Oxford University Press, 1989.
7. S. Lee, G. Wolberg, and S. Y. Shin. Scattered data interpolation with multilevel b-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–244, 1997.
8. T. Poggio and F. Girosi. Networks for approximation and learning. *Proceedings of the IEEE*, 78(9):1481–1497, 1990.