REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
22-09-2009	Final Report	1-Aug-2004 - 31-Jul-2008

4. TITLE AND SUBTITLE
Modeling and Analyzing Terrain Data Acquired by Modern Mapping Techniques

5a. CONTRACT NUMBER: W911NF-04-1-0278
5b. GRANT NUMBER:
5c. PROGRAM ELEMENT NUMBER: 611102
5d. PROJECT NUMBER:
5e. TASK NUMBER:
5f. WORK UNIT NUMBER:

6. AUTHORS
Pankaj K. Agarwal, Lars Arge, Helena Mitasova

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Duke University
Office of Research Support
Duke University
Durham, NC 27705 -

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
46912-CS.1

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Modern remote sensing methods such as LIDAR readily generate high-resolution elevation data, which can be tens or hundreds of gigabytes in size. Several applications including erosion modeling, landslide risk assessment, stream mapping, and hydrologic modeling can benefit from this high-resolution data but elevation data point sets must

15. SUBJECT TERMS
GIS, terrain modeling and analysis, large scale algorithms

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Pankaj Agarwal
U	U	U	SAR		19b. TELEPHONE NUMBER 919-660-6540

Annual Report

ARO Grant W911NF-04-1-0278 (project# 46913-MA)
Modeling and Analyzing Terrain Data Acquired by Modern Mapping Techniques

1 Publications

The following publications acknowledge ARO Grant W911NF-04-1-0278. For brevity, we do not list papers published in conference proceedings for which a journal version is also available.

Journal publications.

[1] P. K. Agarwal, B. Aronov and V. Koltun, Efficient algorithms for bichromatic separability, *ACM Trans. Algorithms*, 2 (2006), 209–227.

[2] P. K. Agarwal, D. Brady and J. Matoušek, Object space segmentation by geometric reference structures, in *ACM Trans. Sensor Networks*, 2 (2006), 455–465.

[3] P. K. Agarwal, H. Edelsbrunner, J. Harer, and Y. Wang, Extreme elevation on a 2-manifold, *Discrete Comput. Geom.*, 36 (2006), 553–572.

[4] P. K. Agarwal, R. Klein, C. Knauer, S. Langerman, P. Morin, M. Sharir, and M. Soss, Computing the detour and spanning ratio of paths, trees, and cycles in 2D and 3D, *Discrete Comput. Geom.*, 39 (2008), 17–37.

[5] P. K. Agarwal and N. H. Mustafa, "Independent set of intersection graphs of convex objects in 2D," *Comp. Geom.: Theory Appls.*, 34 (2006), 83–95.

[6] P. K. Agarwal, M. Overmars and M. Sharir, Computing maximally separated sets in the plane and independent sets in the intersection graph of unit disks, in *SIAM J. Comput.*, 36 (2006), 815–834.

[7] P. K. Agarwal, M. Sharir and E. Welzl, Algorithms for center and Tverberg points, *ACM Trans. Algorithms*, 5, 1 (2008), Article 5, (20 pp.).

[8] P. K. Agarwal, Y. Wang and H. Yu, A 2D Triangulation with near-quadratic topological changes, *Discrete Comput. Geom.*, 36 (2006), 573–592.

[9] P. K. Agarwal, J. Pach and M. Sharir, "State of the union (of geometric objects): A Review," in *Surveys on Computational Geometry: Twenty Years Later* (J. Goodman, J. Pach, and R. Pollack, eds.), American Mathematical Society, Providence, 2008, pp. 9–48.

[10] S. Govindrajan, M. Dietze, P. K. Agarwal, and J. Clark, "A scalable algorithm for dispersing population," (invited) *Journal of Intelligent Information Systems*, 29 (2007), 39–61.

[11] P. K. Agarwal, N. Mustafa and Y. Wang, "Fast molecular shape matching using contact maps," *J. Comp. Bio*, 14 (2007), 131–143.

[12] Y. Zheng, D. J. Brady, and P. K. Agarwal, "Localization using boundary sensors: an analysis based on graph theory," *ACM Trans. Sensor Systems*, 3 (2007), Article 21 (19 pp.).

[13] P. K. Agarwal, F. Hurtado, G. T. Toussaint, and J. Trias, "On polyhedra induced by point sets in space," *Discrete Appl. Math.*, 156 (2008), 42–54.

[14] P. K. Agarwal, R. Klein, C. Knauer, S. Langerman, P. Morin, M. Sharir, and M. Soss, "Computing the detour and spanning ratio of paths, trees, and cycles in 2D and 3D," *Discrete Comput. Geom.*, 39 (2008), 17–37.

[15] P. K. Agarwal, S. Har-Peled and H. Yu, "Robust shape fitting via peeling and grating coresets," *Discrete Comput. Geom.*, 39 (2008), 38–58.

[16] P. K. Agarwal, R. Poreddy, K. Varadarajan, and H. Yu, "Practical methods for shape fitting and kinetic data structures using coresets," *Algorithmica*, 52 (2008), 378–402.

[17] L. Arge, M. de Berg, H. J. Haverkort, and K. Yi. The priority R-tree: A practically efficient and worst-case optimal R-tree, in *ACM Transactions on Algorithms*, 2006.

[18] D. Kinner, H. Mitasova, R. Harmon, L. Toma, R. Stallard, GIS-based stream network analysis for the chagres river Basin, Republic of Panama. In: *The Rio Chagres: A Multidisciplinary Profile of a Tropical Watershed*, R. Harmon (Ed.), Springer/Kluwer, 2005, p. 83-95.

[19] H. Mitasova, M. Overton, and R. S. Harmon, Geospatial analysis of a coastal sand dune field evolution: Jockey's Ridge, North Carolina, *Geomorphology*, 2005.

[20] Mitasova, H., Overton, M., Recalde, J.J., Bernstein, D., and Freeman C., 2009, Raster-based Analysis of Coastal Terrain Dynamics from Multitemporal Lidar Data, *Journal of Coastal Research* 25(2), p. 507-514.

[21] H. Mitasova, L. Mitas, C., Ratti, H., Ishii, J. Alonso, and R.S. Harmon, Real-time human interaction with landscape models using a tangible geospatial modeling environment, *IEEE Computer Graphics and Applications* 26(4), 2006, pp. 55-63.

[22] H., Mitasova, L. Mitas, and R.S. Harmon, Simultaneous spline interpolation and topographic analysis for lidar elevation data: methods for Open source GIS, IEEE GRSL, 2(4), 2005, pp. 375- 379.

[23] H., Mitasova, M., Overton, R.S. Harmon, Geospatial analysis of a coastal sand dune field evolution: Jockey's Ridge, North Carolina, *Geomorphology*, 72, 2005, pp. 204-221.

Conference publications.

[1] M. A. Abam, P. K. Agarwal, M. de Berg, and H. Yu, Out-of-order event processing in kinetic data structures, in *Proc. 14th European Sympos. Algorithms*, 2006.

[2] P. K. Agarwal, L. Arge, and K. Yi. I/O-efficient Construction of constrained Delaunay triangulations, in *Proc. of 13th Annu. European Sympos. on Algorithm*, 2005.

[3] P. K. Agarwal, S. Bereg, O. Daescu, H. Kaplan, S. Ntafos, and B. Zhu, Guarding a terrain by two watchtowers, in *Proc. 21st Annu. Sympos. Comput. Geom.*, 2005.

[4] P. K. Agarwal, M. de Berg, J. Gao, L. Guibas, and S. Har-Peled, Staying in the middle: Exact and approximate medians in $\mathbb{R}^1$ and $\mathbb{R}^2$ for moving points, in *Proc. 18th Canadian Conf. Comput. Geom.*, 2005.

[5] P. K. Agarwal, S. Har-Peled and H. Yu, in Robust shape fitting via peeling and grating coresets, *Proc. 17th ACM-SIAM Sympos. Discrete Algorithms*, 2006.

[6] P. K. Agarwal and J. Phillips, On bipartite matching under the rms distance, to appear in *Proc. 17th Canadian Conf. COmput. Geom.*, 2006.

[7] P. K. Agarwal, J. Xie, J. Yang, and H. Yu, Monitoring continuous band-join queries over dynamic data, in *Proc. 26th Intl. Sympos. on Algo. Comput.*, 2005.

[8] P. K. Agarwal, J. Xie, J. Yang, and H. Yu, Scalable continuous query processing by tracking hotspots, in *Intl. Conf. Very Large Databases*, 2006.

[9] P. K. Agarwal, Y. Wang, and P. Yin. A lower bound on weighted spanners, in *Proc. 16th ACM-SIAM Sympos. Discrete Algorithms*, 2005.

[10] P. K. Agarwal, L. Arge, and A. Danner. From point cloud to grid DEM: A scalable approach, in *Proc. International Symposium on Spatial Data Handling*, 2006.

[11] P. K. Agarwal, L. Arge, and K. Yi. I/O-Efficient batched union-find and its applications to terrain analysis, in *Proc. 22nd Annu. Symposium on Computational Geometry*, 2006.

[12] P. K. Agarwal, H. Kaplan and M. Sharir, "Computing the volume of the union of cubes," in *Proc. 23rd Annu. Symposium on Computational Geometry*, 2007.

[13] P. K. Agarwal, R. Apfelbaum, G. Purdy, and M. Sharir, "Similar simplices in a d-dimensional point set," in *Proc. 23rd Annu. Symposium on Computational Geometry*, 2007.

[14] P. K. Agarwal, S. Har-Peled and H. Yu, "Embeddings of surfaces, curves, and moving points in Euclidean space ," in *Proc. 23rd Annu. Symposium on Computational Geometry*, 2007.

[15] P. K. Agarwal and H. Yu, "A space-optimal data-stream algorithm for coresets in the plane," in *Proc. 23rd Annu. Symposium on Computational Geometry*, 2007.

[16] P. K. Agarwal, L. Arge, T. Mølhave, and B. Sadri, "I/O efficient algorithms for computing contour lines in a terrain," in *Proc. 24th Annu. Symposium on Computational Geometry*, 2008.

[17] P. K. Agarwal, B. Sadri, and H. Yu, “A geometric approach for untangling a mesh using local surgery,” in *Proc. 24th Annu. Symposium on Computational Geometry*, 2008.

[18] L. Arge, M. de Berg, and H. Haverkort. Cache-oblivious R-Trees, in *Proc. 21st Annu. Sympos. Comput. Geom.*, 2005.

[19] L. Arge, G. S. Brodal, R. Fagerberg, and M Laustsen Cache-oblivious planar orthogonal range searching and counting, in *Proc. 21st Annu. Sympos. Comput. Geom.*, 2005.

[20] L. Arge, A. Danner and N. Zeh. Computing Pfafstetter labellings I/O-efficiently, in *Proc. 1st Workshop on Massive Geometric Data Sets*, 2005.

[21] L. Arge, D. Eppstein and M. Goodrich. Skip-Webs: Efficient distributed data structures for multi-dimensional data sets, in *Proc. 24th ACM SIGACT-SIGOPS Sympos. Principles Of Distr. Comput.*, 2005.

[22] L. Arge, G. S. Brodal, and L. Georgiadis. Improved dynamic planar point location, to appear in *47th Annual IEEE Symposium on Foundations of Computer Science*, 2006.

[23] L. Arge, A. Danner, H. Haverkort, and N. Zeh. I/O-efficient hierarchical watershed decomposition of grid terrain models, in *Proc. International Symposium on Spatial Data Handling*, 2006.

[24] L. Arge and N. Zeh. Simple and semi-dynamic structures for cache-oblivious planar orthogonal range searching, in *Proc. 22nd Annu. Symposium on Computational Geometry*, 2006.

[25] T.P., Colson, J.D., Gregory, H. Mitasova, S.A.C. Nelson, Comparison of stream extraction models using lidar DEMs, *Proc. GIS and Water Resources IV*, AWRA, Houston, TX, May, 2006.

[26] P. Flikkema, P. K. Agarwal, J. Clark, C. Ellis, A. Gelfand, K. Munagala, and J. Yang, Model-driven dynamic control of embedded wireless sensor networks, *Proc. of the Third International Conference on Computational Science*, 2006.

[27] P. G. Flikkema, P. K. Agarwal, J. S. Clark, C. S. Ellis, A. Gelfand, K. Munagala, and J. Yang. “From data reverence to data relevance: Model-mediated wireless sensing of the physical environment,” *Proceedings of the Fourth International Conference on Computational Science*, 2007.

[28] M. Overton, H., Mitasova, J.J. Recalde, and N. Vanderbeke, Morphological evolution of a shoreline on a decadal time scale, *Proc. of ICCE 2006 meeting*, San Diego, CA, Sept. 2006

[29] J. Phillips, J. Rudolph, and P. K. Agarwal, Segmenting motifs in protein-protein interface surfaces, to appear in *Proc. 6th Workshop on Algorithms in Bioinformatics*, 2006.

[30] Y. Wang, P. K. Agarwal, P. Brown, H. Edelsbrunner, and J. Rudolph, Coarse and reliable geometric alignment for protein docking, in *Proc. Pacific Sympos. Biocomputing*, 2005.

2 Objectives

Modern remote sensing methods such as LIDAR readily generate high-resolution *elevation data*, which can be tens or hundreds of gigabytes in size. Several applications including erosion modeling, landslide risk assessment, stream mapping, and hydrologic modeling can benefit from this high-resolution data but elevation data point sets must first be transformed into a digital elevation models (DEMs) and derived products such a river networks or watersheds before users can conduct relevant studies. Processing these massive data sets poses a number of algorithmic challenges.

The goal of this project is to provide enhanced terrain modeling and analysis capabilities by developing sophisticated algorithms that function with massive non-standard datasets, such as point clouds, and that produce a confidence level for the results. We are developing algorithmic techniques to overcome the computational challenges encountered when processing the massive, dynamic, and heterogeneous geospatial data acquired today: We utilize *approximation techniques* to trade efficiency with accuracy—use hierarchies to represent the data at varying levels of detail, and rely on approximation algorithms to solve various terrain-analysis problems efficiently. To handle the massive amounts of data efficiently, we utilize recent advances in *memory-aware algorithms*, that is, algorithms that are specifically designed to handle data sets that do not fit in main memory of underlying devices.

3 Approach

We followed a comprehensive approach to design new techniques for terrain modeling and analysis that can handle massive amounts of heterogeneous data sets that are being updated dynamically. We rely on the following approaches:

Approximation methods. Relying on approximation techniques that effectively trade efficiency with accuracy is not only desirable but a necessity when dealing with massive datasets, especially in time critical applications. We use hierarchies to represent the 3D digital models of terrains at varying level of detail and rely on approximation algorithms to solve modeling and analysis problems efficiently. Approximation techniques often provide efficient, robust, and simple algorithms, at the cost of sacrificing the accuracy within a certain acceptable threshold. In many geometric-optimization problems, approximation is desirable and appropriate since the underlying data, acquired through physical sensors, carry certain level of uncertainty and noise. In time-critical missions, approximate solutions are actually all one can hope for because the time required to find an optimal solution is unaffordable.

Memory-aware methods. When processing massive datasets larger than the main (random-access) memory of the computation platform, the input/output communication (I/O) between fast main memory and slow disk is often the bottleneck in the computation; disks accesses are often around 106 times slower than main memory accesses. In such cases, using so-called I/O-efficient algorithms can lead to tremendous runtime improvements. I/O-efficient algorithms obtain these improvements by explicitly managing I/O between main memory and disk, and more importantly by taking advantage of the fact that data is moved between memory and disk in large contiguous blocks; by making sure all data in a block being brought into main memory is put to good use, the large disk access time is amortized over a lot of data access.

Dynamic techniques. Most of the current systems are tailored to batched processing of spatial data. Instead, data could be regarded as providing partial information about dynamic geometry, topography of a geographic area. Despite much work on dynamic data structures in computer science, little progress has been made in developing dynamic techniques for rapidly updating geospatial models as new data arrives. Of course, one can construct the entire structure after each update, but it will be extremely inefficient. One of the difficulties in dynamically updating digital 3D models of terrains stems from the non-local dependencies. Multidimensional analysis. Increasing efficiency of mapping technologies supports repeated surveys of topography in rapidly changing regions such as coastal areas, dune fields or battlefields. We employ multivariate approximation to generate multidimensional models of terrain dynamics from time series of point cloud data and develop new type of maps that characterize the spatial patterns of terrain dynamics/stability, including terrain evolution gradient maps and core surface maps.

4 Significance and Army Value

Terrain analysis is an integral part of the military intelligence preparation of the battlefield, commonly used to support both defensive and offensive operations. It consists of interpreting natural and man-made terrain features, together with the influences of weather, to determine their effects on military operations. In recent years, the potential of combat terrain information systems have been greatly enhanced by new terrain mapping technologies such as Laser altimetry (LIDAR), ground based laser scanning and Real Time Kinematic GPS (RTK-GPS) that are capable of acquiring millions of georeferenced points within short periods of time (minutes to hours). However, while acquiring and georeferencing the data has become extremely efficient, transforming the resulting massive amounts of heterogeneous data to useful information for different types of users and applications is lagging behind. Thus the full potential of new mapping technologies has not been utilized.

One of the main reasons of the large chasm between the performance of the mapping technologies and the topographic support systems is the scarcity of robust, efficient methods for terrain modeling and analysis that can handle massive datasets acquired by different technologies (with different properties such as accuracy, density, spatial distribution) and that can rapidly detect and predict changes in the model as the new data is acquired. In addition, the existing algorithms are unable to: (i) calibrate their behavior according to the constraints of the underlying machine (from high-end desktops in command centers to laptops or even handheld devices in the field), and (ii) provide information at an appropriate level of detail best suited to the user (from commanders to the soldiers in the field) *while* maintaining data consistency.

This project followed a vertically integrated approach that addressed the issues at all levels (from modeling to optimizing algorithms for given devices) in a unified manner in order to meet the challenges in terrain modeling and analysis.

The new algorithms provide Army with new capabilities to extract critical information such as stream networks and watershed hierarchies from large DEMs at significantly higher level of efficiency, accuracy and flexibility, and reduce the time between the data acquisition and delivery of core derived information for planning and decision making.

The multidimensional analysis of elevation data time series provides new approach to identification locations with highly unstable topography that may influence military activities. Classification of topography based on its sensitivity to change also provides important information on locations that need to be mapped with increased accuracy and shorter time interval.

5 Scientific Accomplishments

This report summarizes the scientific accomplishments of the project. The main focus of the project was designing and developing a scalabe modular system to construct watershed hierarchies from massive high-resolution elevation data such as LIDAR data. In addition we also studied a number of fundamental geometric data structure problems. Further details can be found at the project web page: `http://terrain.cs.duke.edu`

5.1 From elevation data to watershed hierarchies

We considered the problem of constructing a watershed hierarchy from given elevation data—a point cloud in 3-space. We have developed and implemented an approach that has the following features: it is modular so that a user can use different models for each of the modules; it works for both TIN and grid DEMs; it is scalable to massive data sets; it allows a user to run the whole "pipeline" without any manual intervention between different stages. The *pipeline* or *work-flow* approach for developing terrain-analysis software is common in GIS. Most software packages support some way of connecting separate modules together to form pipelines, however this requires manual intervention. While a typical GIS can manage many Gigabytes of data in a collection of layers, most of the GIS modules are not designed to handle single multi-gigabyte input layers, especially the modules that must compute a global property of the input terrain, such as the river network. We are not aware of system for constructing watershed hierarchy that supports the above four features.

Our pipeline consists of four main stages: DEM construction, removal of topological noise, extraction of river networks, and construction of the watershed hierarchy. Figure 1 shows various stages of our pipeline.

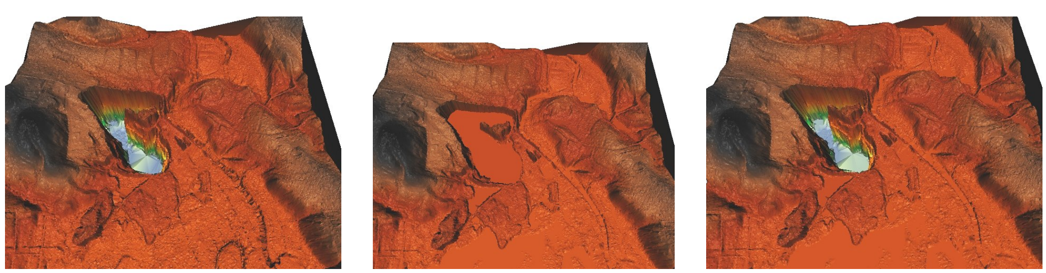
The first stage of the pipeline can construct a grid or TIN DEM —the two most widely used terrain DEMs. The subsequent stages of the pipeline work for height graphs and thus for both grid and TIN DEMs. We developed a GIS-based workflow for computation of seamless elevation grids from diverse, massive point and profile data sets acquired by modern technologies such as LIDAR, IFSARE, single and multibeam sonar and Real Time Kinematic GPS. that includes: (a) analysis of spatial properties of data in individual subsets with different sampling patterns, point densities, accuracies; (b) geometry-based point thinning; (c) identification and reduction of systematic errors, such as vertical shifts in LIDAR point clouds; (d) simultaneous computation of elevation grid, topographic parameters and random noise smoothing using spline-based approximation technique; (e) iterative approximation for areas with large data gaps.

The workflow has been developed and tested for diverse applications including computation of time series of lidar-based 0.5m resolution DEMs for sections of North Carolina (NC) coast, seamless integration of IFSARE and SRTM elevation models and gap filling for Panama, computation of seamless topobathy from lidar points clouds, single beam and RTKGPS profiles and multibeam sonar for North Carolina and Virginia coast and other applications. We have demonstrated that the predictive error of our interpolation is dependent on land cover and on bare earth surface it is lower than measurement error for wide range of interpolation parameters. The applications of our methodology has also shown that investigation of possible systematic errors, especially vertical shifts, in multi-temporal elevation data sets is of critical importance for correct quantification of terrain change and its spatial pattern.

Figure 1. Various stages in our pipeline constructed on the Neuse river basin: (a) LIDAR data (500M points), (b) grid DEM at 10ft resolution, (c) flow network, (d) watershed hierarchy.

5.2 Noise removal on a DEM

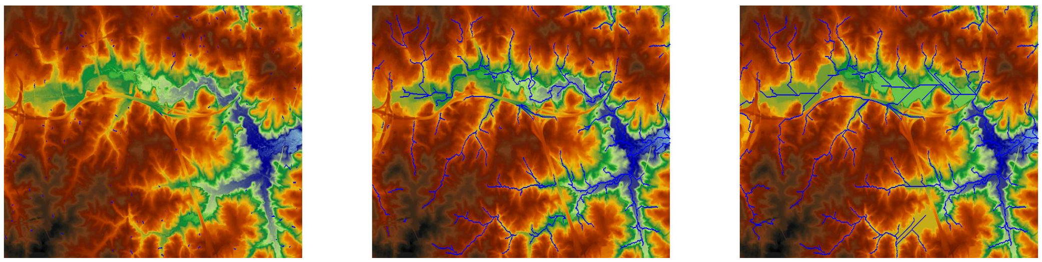
The DEMs constructed (grids or TINs) may have many local minima or *sinks*. Some sinks in the DEM are due to noise in either the elevation data point sample or the construction method used, while other sinks correspond to real geographic features such as quarries, sinkholes or close water basins with no drainage outlet. Typical flow-modeling algorithms assume that water flows downhill until it reaches a sink. However, sinks due to noise impede the correct flow of water and result in artificially disconnected hydrological networks. Therefore, it is important to construct "hydrologically correct" (or hydrologically conditioned) DEMs that remove sinks due to noise. Ideally, only those sinks due to noise should be removed while genuine sinks should be preserved.



(a) (b) (c)

Figure 2. (a) The terrain. (b) Flood the terrain until a steady-state is reached. (c) Partially flood the terrain.

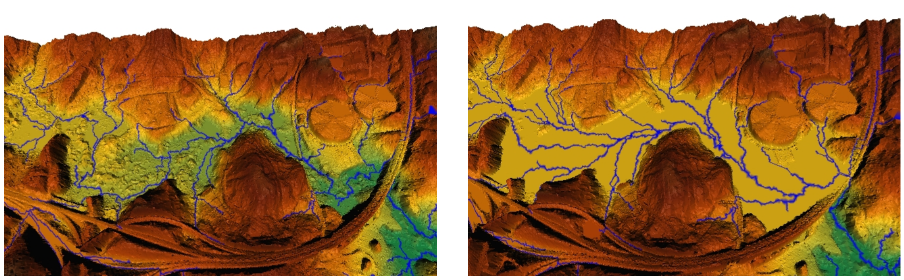
Flooding or *pit-filling* is a popular method for removing sinks, which simulates uniformly pouring water on the terrain until all sinks are filled and a steady-state is reached. Refer to Figure 2(a) and (b). Typically, all sinks are filled (irrespective of their importance) so that the only remaining sink is the "outside", which corresponds to the ocean or a global minimum. To preserve important sinks, users typically have to manually mark the important sinks as "real sinks" to distinguish them from the spurious ones. Consider the case shown in Figure 2(a). On the high level suppose there are three significant real sinks, but a number of small sinks due to noise. Traditional flooding that removes all sinks would result in Figure 2(b), but we would prefer something more like Figure 2(c) where only the small sinks are flooded.



(a) (b) (c)

Figure 3. Flow network at various persistence values: (a) persistence=0 (no flooding), (b) persistence=10 (small pits filled), (c) persistence=100 (many real pits also filled). Flow network with high persistence does not look very realistic.

We use a method based on *topological persistence*, oroginally proposed by Edelsbrunner *et al.*[1] that ranks the significance of each sink in a DEM and allows the user to remove sinks below a certain threshold. The original algorithm is not scalable as it accesses memory in a random order. We have designed and implemented an I/O-efficient algorithm for computing the topological persistence on a DEM. Through experiments on some of the LIDAR data sets, our new algorithm is shown to be more than a hundred times faster than the previous approach. We then use this algorithm for developing a scalable flooding algorithm. See Figures 3 and 4.



(a) (b)

Figure 4. Flow network on a TIN DEM: (a) persistence=10, (b) persistence=100.

[1]H. Edelsbrunner, J. Harer, and A. Zomorodian. Hierarchical morse complexes for piecewise linear 2-manifolds. *Proc. 16th ANnu. Sympos. Comput. Geom.* 70–79, 2001.

5.3 Flow modeling

Two of the most important concepts in terrain flow modeling are *flow routing* and *flow accumulation*. Intuitively, flow routing is the assignment of flow directions to every point in a terrain in order to globally model how water flows through it. Flow accumulation then quantifies how much water flows through each point of the terrain if poured uniformly onto it. Flow routing and flow accumulation are the basic for computing other attributes such as drainage networks and watersheds.

We have developed scalable algorithms for a few most widely used flow routing methods (e.g., steepest descent method, D_∞ method). One major issue in flow modeling is flow routing on flat areas, that is, assignment of flow direction to cells, vertices or triangles without any strict downslope neighbors. Flat areas can either be naturally present in a DEM, or (as commonly is the case) have been introduced by denoising or flooding. We are exploring a number of approaches (also in conjunction with the noise-removal step) and have implemented a few scalable algorithms. The existing algorithm that produce reasonable results are not scalable.

Building upon our TerraFlow project (`http://www.cs.duke.edu/geo*/terraflow/`), we have designed a developed a new version of TerraFlow to increase efficiency (both I/O and CPU computation) and stability. We have also increased the functionality in a number of ways. For example, we have added support for TIN DEMs (where flow directions are assigned to vertices), just as we have added support for flow modeling on terrains with sinks (i.e., the DEM is not flooded before routing and accumulation). We have also improved the SFD and MFD routing methods and added support for further user-specified routing methods; we are currently implementing D_∞ using this feature. Support for user-specified flat area routing has also been added and we are in the processes of implementing other than the simple shortest spill-point path method.

5.4 Geospatial analysis and terrain change

Rapid evolution of laser scanning technology is dramatically changing the level of detail and type of information about the Earth surface that can be acquired repeatedly over large areas. Multiyear lidar surveys are becoming available for many regions, offering unprecedented insight into topographic change and land surface evolution. We have investigated and developed three approaches to analysis of terrain change from time series of raster DEMs: feature based method extracts geomorphological features from raster DEMs for each time snapshot using combination of parameters derived from gradients, curvatures and slope lines, and measures their vertical and horizontal migration; raster based method applies per-cell statistical analysis to raster DEM time series, generating new type of maps that characterize evolution of land surface while preserving the original spatial resolution of the DEM. multivariate functions are used to create a continuous spatio-temporal model of land surface evolution and analyze its rate of change in space and time.

Extraction and tracking of features. We use generalized bivariate smoothing spline with tension to create high resolution set of DEMs from different types of point cloud elevation data. Terrain parameters are derived simultaneously with approximation, at a level of detail needed for the identification and tracking of dynamic terrain features by selecting appropriate tension parameters. Topographic change measures are defined based on geomorphology of the studied site (e.g. coastal erosion, dune migration, floodplain) and extracted from all DEMs using mathematical description of each selected landform, resulting in a multitemporal series of maps representing evolution of individual topographic features. The methodology was demonstrated for a case study that involved a complex sand dune migration; see Figure 5.

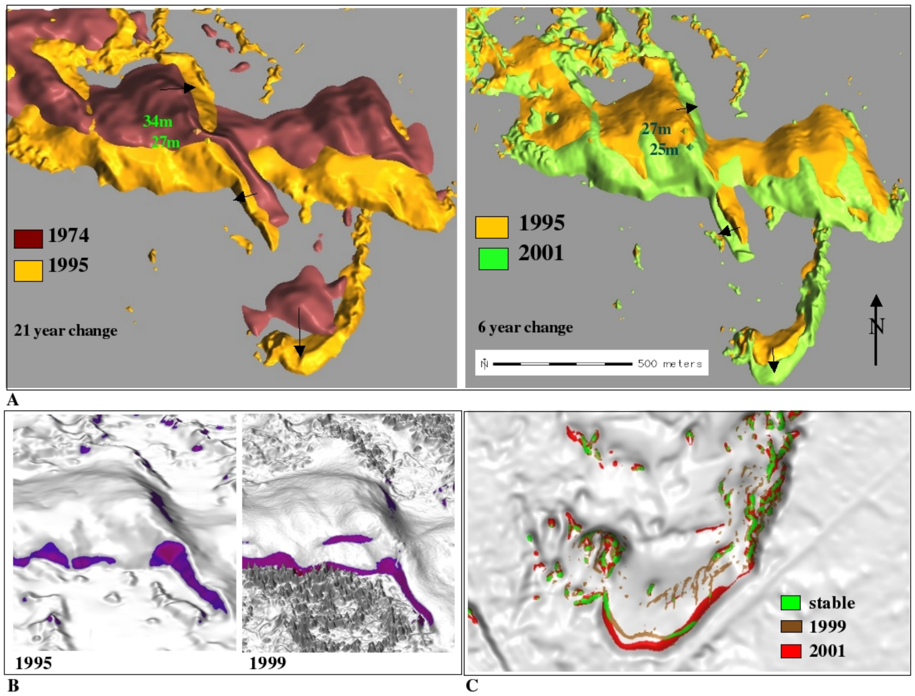


Figure 5. Identification and quantification of a coastal dune terrain change: a) overlayed elevation surfaces showing SE migration (3-6m/y), deflation (0.3m/y) and rotation of the main dune ridge; b) change in Main dune slip-face - emergence of a new slip-face inside dune may affect mobility patterns; c) identification of stable and migrating dune crests - important for planning and management of areas in dune vicinity.

Raster based analysis using per-cell statistics and map algebra. Using univariate per-cell statistics we derive new raster maps characterizing spatial or temporal pattern of land surface evolution from series of DEMs z(i,j,tk), measured at time snapshots tk, k=1,n. We define the core surface as a boundary between stable volume and a dynamic layer, while envelope represents outer boundary of the surface evolution within the given time period. Temporal aspect is captured by time of minimum elevation and time of maximum elevation raster maps. Spatial distribution of linear rate of change is represented by a map of regression line slope computed for each cell along with associated map of correlation coefficient value, mean elevation map and standard deviation map are more standard measures. We explored the possibilities to use these summary maps for highly efficient, automated extraction of information about structures, such as identification of new homes or structures lost during storms and the estimation of time of the observed change. Massive time series of sub-meter resolution DEMs have been processed for this type of analysis using the workflow developed for computation of grid-based DEMs (Figure 6). The approach is now being applied to large section of NC Outer banks and it is evaluated for identification of coastal hazards using leveraging support from the Sea Grant. See Figure 7.

(a) (b) (c)

Figure 6. Increasing lidar point density is improving the representation of subtle but important natural terrain features: (a) buildings and foredunes in 1999 data with average densities 1pt/2m grid; (b), (c) Similar features represented in 0.5m DEM computed from 2004 data with density 15pt/2m grid.

Multivariate functions. Building upon our previous research on modeling spatially and temporally distributed phenomena and analysis of topographic change we are exploring new concepts and measures for characterization of terrain dynamics based on multivariate function representation. The analysis based on time series of DEMs outlined above handles evolution over time as discrete events. To apply the full power of analysis based on multivariate differential geometry we represent land surface evolution as a trivariate function where the third dimension is time and elevation is the modelled variable (attribute). Then we can extract evolution of contours (e.g. shorelines) as isosurfaces and derive spatio-temporal gradients using partial derivatives of trivariate interpolation function and estimate the direction and rate of fastest elevation change in both space and time. This approach will be the focus of the project continuation and offers the most innovative approach to surface evolution analysis.

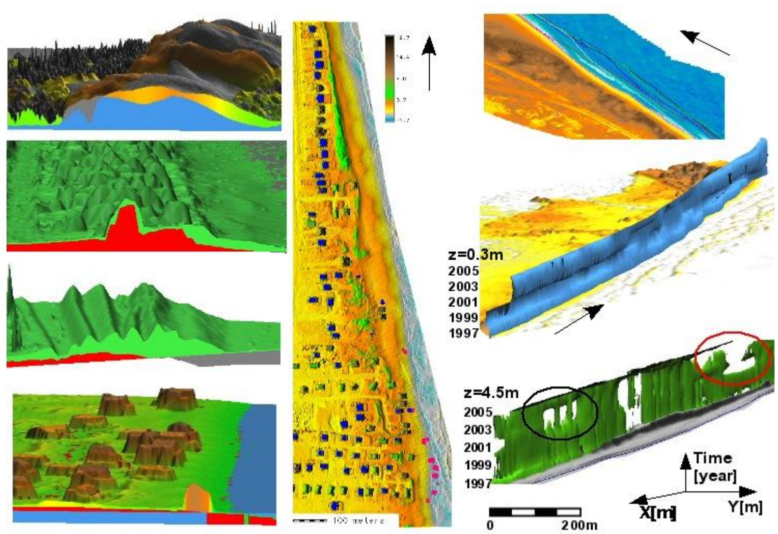
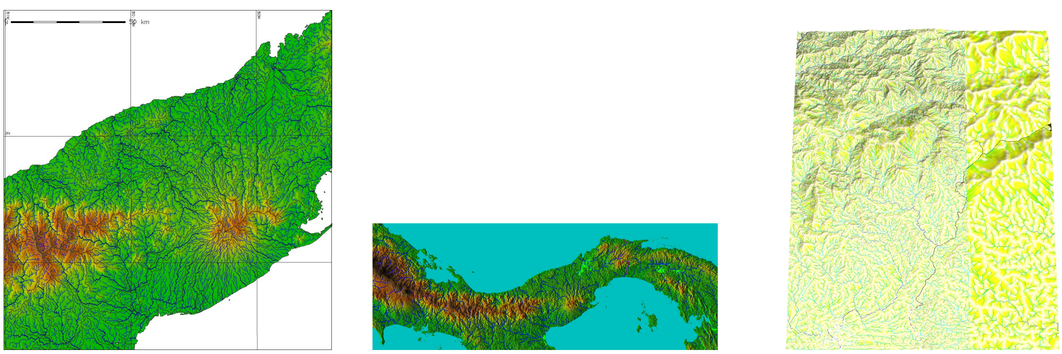


Figure 7. Coastal application: (a) core surface and envelope crossections, (b) identification of new (blue) and lost (red) homes from core and envelope surfaces, (c) isosurfaces representing evolution of 0.3m and 4.5m contours over time, top image shows time series of 0.3m contours for study period displayed in 2D image.

5.5 Watershed analysis

Extraction of streams from lidar-based DEMs in the US and SRTM globally offers automated, highly efficient approach for stream mapping. We have tested the efficiency of deriving statewide stream network representation at two levels of detail using combination of SRTM 90m and IFSARE 10m resolution data for the entire state of Panama with the higher resolution data available for the Panama canal region. Stream extraction was performed on each of the DEMs separately at their original resolutions and then on a seamless 30m resolution DEM with two levels of detail created by resampling and merging the IFSARE and SRTM using regularized spline with tension to ensure adequate routing for rivers flowing along the borders of the two DEMs. Combination of the new algorithms and increased memory reduced the computational time for this task from days to hours for the first year of the project, and to 15 minutes at the end of the project. See Figure 8. Despite the triple canopy tropical forest captured by the IFSARE and SRTM data the improved algorithms that avoid depression filling lead to the horizontal accuracy that is better than 1.5 grid cell resolution (15m for IFSARE, 45m for SRTM) in mountains and hilly terrain, however, errors are still high in coastal plains where combination with imagery is needed.



(a) (b) (c)

Figure 8. (a) Section of Panama SRTM DEM and streams with watershed boundaries extracted using existing GIS tools requiring tile-based processing; (b) SRTM DEM and streams extracted for entire Panama in a single run using r.terraflow processed faster than the much smaller section shown in (a). (c) section of Panama stream network derived from combined SRTM and IFSARE data reinterpolated and smoothed by RST to a common 30m resolution. SRTM was used where IFSARE was not available. Both r.terraflow and standard tools were tested for the combined DEM.

We have investigated standard and enhanced flow routing methods that allow us to take advantage of high resolution lidar data and extract stream networks and watershed hierarchies with accuracy and efficiency that exceeds the currently used practice. Using leveraging funding from related projects we have acquired field measured data for location of streams that provided us with unique opportunity to evaluate the algorithms in different environments, from tropical forest (Panama), undulating piedmont topography, urban terrain, to coastal plain. The gained knowledge was used to improve the algorithms that were developed for massive data sets. New research will be needed for development of algorithms that will perform flow routing and watershed analysis in developed (e.g., urban) areas, using combination of elevation surfaces and augmented data that take into account uncertainty in multiple data sources. Significance and Army Value The new algorithms provide Army with new capabilities to extract critical information such as stream networks and watershed hierarchies from large DEMs at significantly higher level of efficiency, accuracy and flexibility, and reduce the time between the data acquisition and delivery of core derived information for planning and decision making.

The multidimensional analysis of elevation data time series provides new approach to identification locations with highly unstable topography that may influence military activities. Classification of topography based on its sensitivity to change also provides important information on locations that need to be mapped with increased accuracy and shorter time interval.

5.6 Geometric data structures

Point location. Planar point location is a fundamental problem in computational geometry with important applications in GIS (as well as in many other application areas): Store a planar subdivision, i.e. a decomposition of the plane into polygonal regions induced by a straight-line planar graph, such that the region containing a query point can be found efficiently. We developed the first linear-space data structure for the dynamic version of the problem (where edges can be inserted or removed into/from the subdivision) that achieves logarithmic query time and polylogarithmic update time. We also developed a simpler optimal structure for the incremental problem; it supports queries and insertions in logarithmic time.

Range searching. Planar orthogonal range searching is a fundamental problem in computational geometry with important applications in GIS (as well as in many other application areas): Store a set of points in the plane such that the points in a query rectangle can be found efficiently. We developed a new cache-oblivious structure for the simpler two- and three-sided versions of the problem (where the rectangle is unbounded in two or one directions) with bounds matching the previous best known structure; a cache-oblivious structure is a structure that is efficient on all levels of a multi-level (possibly unknown) memory hierarchy. Our new structures a much simpler than the previously known structures and can be made semi-dynamic using standard techniques.

Handling mobile data. We study the problem of designing geometric data structures for moving objects under the so-called kinetic data structure (KDS) framework. As the objects move the data structure has to be updated when certain *events* occur. Because of uncertainty in data and numerical errors the event times cannot be computed exactly and events may be processed in a wrong order. In traditional KDS's this can lead to major inconsistencies from which the data structure cannot recover. We present robust data structures for the maintenance of several fundamental structures, which overcome the difficulty by employing a refined event scheduling and processing technique. We prove that our data structures maintain the correct information except possibly at some finite number of short time intervals (namely the uncertainty intervals of the event-time computations). Even in these uncertainty intervals, we can bound the error in the data structure.

Continuous queries. A *continuous query* is a standing query over a data stream that, once issued by the user, needs to keep generating query results or changes to old results subject to the same query condition, as new data continue to arrive in a stream. This is in contrast to traditional queries, where each query is executed only once on a snapshot of the data set. The main challenge in continuous query processing is how to organize a large number of continuous queries in a scalable manner, so that for each incoming data update, one can quickly identify the subset of continuous queries whose query results are affected by the update, and compute changes to the results of these affected queries.

We developed fast algorithms for processing continuous *band-join* queries, an important type of queries that naturally arises in many continuous query applications and forms the basis of more complex join queries. We further developed simple and practical data structures for processing continuous band-join queries as well as other join queries. Our key idea is to exploit clustering patterns in the query ranges so that the performance of the data structure depends on the degree of clusteredness. We expect that our technique will be useful in other areas of spatial database research besides continuous queries.

6 Technology Transfer

- TerraStream software package is publicly available at http://www.madalgo.au.dk/Trac-TerraSTREAM in standard installation packages for Windows, Linux, and Macintosh. It has been downloaded by more than twenty-five institutions and companies worldwide, including The World Bank Group, Risk Management Solutions, Jones Edmunds & Associates in USA; Ministry of Natural Resources and CARIS in Canada; and SURDEX, Kaya Consulting Limited, University of Weizburg, ATKINS Danmark, Hvidorre Municipality, and NIRAS in Europe. We are currently discussing software licensing to companies in US, Canada, and Denmark
- Code enhancements were implemented in Open source GIS - the GRASS6.3 and the most recent GRASS6.4 release
- Implementation of the developed workflows as GIS-based modules in Python is planned for GRASS7 as well as an on-line tutorial following the format of successful on-line GIS-based erosion modeling tutorial
- The results of this research were also used for the development of high resolution (10m) seamless topobathy for entire North Carolina coast prepared with RENCI for NC DENR as input for coastal NC flooding maps and the project has been extended up to New Jersey coast.

7 Visits to Army Labs

- Agarwal has visited the Engineering Research and Development Center (ERDC), Vicksburg, and the Topographic Engineering Center (TEC), Alexanderia multiple times; presented the groups work there. These visits have led to direct contacts with the researchers Mr. James Rogers, Dr. James Shine, and Mr. Harry Puffenberger at TEC.
- Arge has also visited ERDC and TEC and presented his work there.
- Mitasova has been coordinating the research with Mr. James Rogers and Michael Campbell from TEC and Dr. Ehlschlaeger and S. Tweddale from CERL