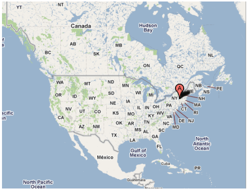
Algorithms for terrain and bathymetric sensor data

W. Randolph Franklin, Zhongyi Xie, Eddie Lau, You Li

Rensselaer Polytechnic Institute
Troy, NY, USA



ICA Workshop on Advances in Sensors and Algorithms for Topographic and Thematic Mapping, 19 Nov 2010

partially supported by NSF grant CMMI-0835762

Intro

Three algorithic advances and a research topic in processing topographic and bathymetric sensor data:

- lossy terrain compression that maintains slope accuracy,
- bathymetric surface fitting to irregular tracklines,
- lossy compression of 5D environmental data, and
- terrain modeling to maintain hydrological validity.

Why? To attack several issues raised by the large amounts of data now available.

Eventual goal: A unified system.

Lossy terrain compression that maintains slope accuracy

- Accurate elevations $\not\Rightarrow$ accurate slopes
- Bad commercial slope representation.
- *Goal:* Compress and reconstruct terrain so that slope derived from reconstructed terrain is good.

Accurate elevations $\not\Rightarrow$ accurate slopes

- Ignoring errors, slope is simply $f'(x)$
- But $\limsup_{i\to\infty} |(f_i(x) - f(x)| \to 0$, gives no guarantees about $\limsup_{i\to\infty} \left|(f_i'(x) - f'(x)\right|$
- Consider two approximations to $y(x) = 0$

- Elevation got better but slope got worse.

Bad commercial slope representation

Commercial SW:

Bad commercial slope representation

Commercial SW: 

Photo:

ODETLAP –Overdetermined Laplacian Method

Fundamental representation for this work

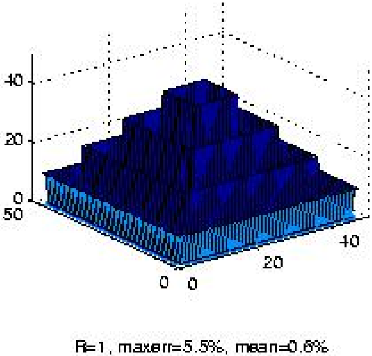
- Small set of posts $\Rightarrow$ complete matrix of posts
- Overdetermined linear system:
  - $z_{ij} = h_{ij}$ for known points,
  - $4z_{ij} = z_{i-1,j} + z_{i+1,j} + z_{i,j-1} + z_{i,j+1}$ for all nonborder points.
  - Emphasize accuracy or smoothness by weighting the two types of equations differently.
- Original goal: fill contours to a grid w/o showing terraces; competing methods have these problems:
  - Information does not flow across contours $\rightarrow$ slopes discontinuous
  - If rays are fired from the test point to the first known point, then method is not conformal etc.

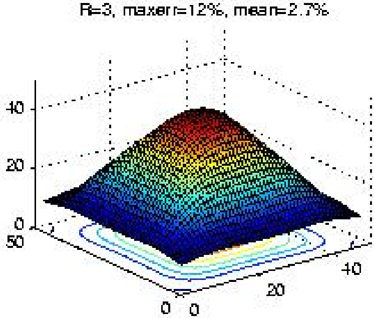
ODETLAP Advantages

Handles

- missing–data holes.
- incomplete contours,
- complete contours,
- kidney–bean contours,
- isolated points,
- inconsistent data.

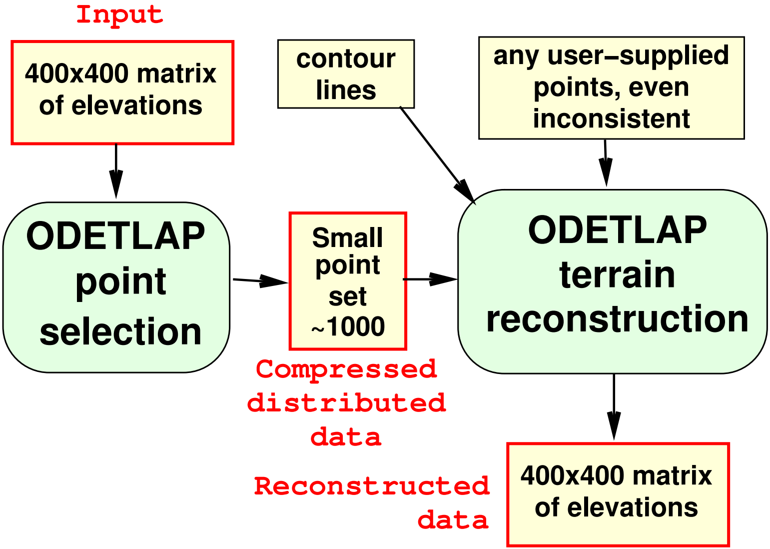
ODETLAP hard example



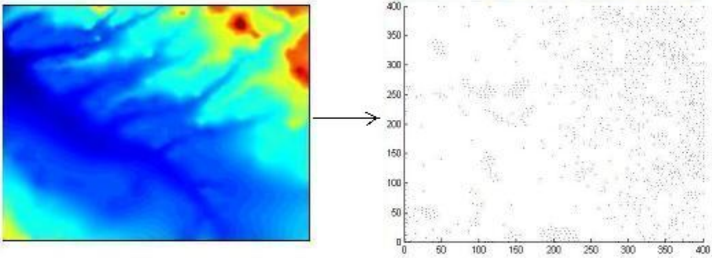


- input: contours with sharp corners
- output: smooth silhouette edges, inferred top

ODETLAP process



ODETLAP summary



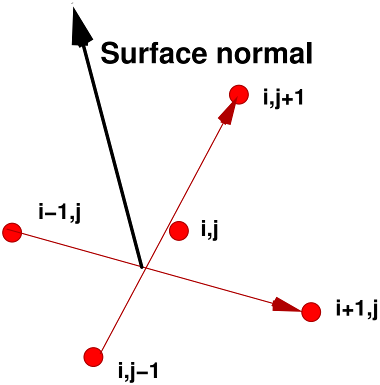
Original Surface
(320 KB)

Compressed Surface
(4071 Bytes)

Average Absolute Error = 2.451
Maximum Absolute Error = 25.822

Slope definition, accuracy

- Zevenbergen-Thorne $\left(\left(p_{i-1,j} - p_{i+1,j}\right) \times \left(p_{i,j-1} - p_{i,j+1}\right)\right)_z$
- p_{ij} not used



Limits of slope accuracy

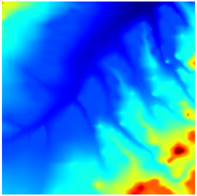
- 1m elevation resolution
- 30m post spacing
- slope precision: $\arctan\left(\frac{1}{30}\right) \approx 3\% \approx 2^\circ$

Info content

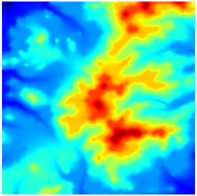
- Slope's autocorrelation distance is smaller than elevation's
- However, slope has less relative precision.

Level-II sample datasets

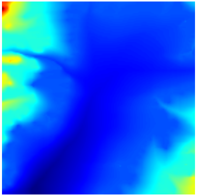
400 $\times$ 400 elevation matrices, *elevation range*



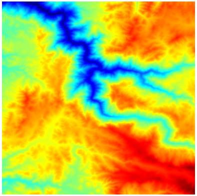
Hill1 *505m*



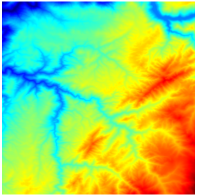
Hill2 *745m*



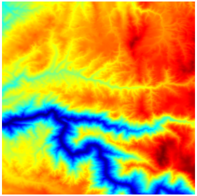
Hill3 *500m*



Mtn1 *1040m*



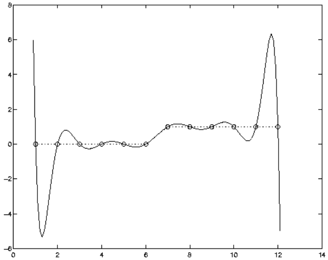
Mtn2 *953m*



Mtn3 *788m*

Idea 1: Pin down the elevation at sets of close points

- When inserting a point into known set, also insert some adjacent points
- *Thesis:* that will force the slope to be accurate there.
- Not really.
- *Analogy:* Lagrangian interpolation.



Keep trying.

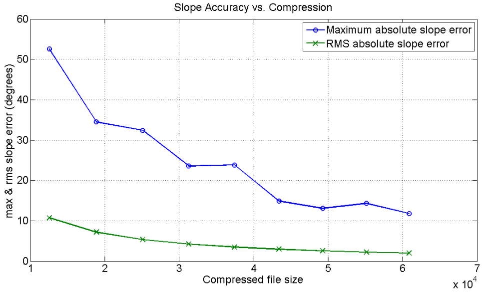
Idea 2: Extend ODETLAP

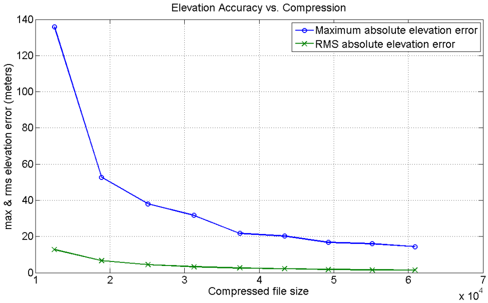
- Explicitly incorporate slope
- New overdetermined linear system:
  - unknowns: z_{ij}
  - known:
    - some h_{ij},
    - some $\Delta_x h_{ij} \triangleq h_{i-1,j} - h_{i+1,j}$,
    - some $\Delta_y h_{ij} \triangleq h_{i,j-1} - h_{i,j+1}$,
  - for all nonborder points:

    $$4z_{ij} = z_{i-1,j} + z_{i+1,j} + z_{i,j-1} + z_{i,j+1}$$
  - for known h_{ij}: $z_{ij} = h_{ij}$
  - for known $\Delta_x h_{ij}$ and $\Delta_y h_{ij}$:

    $$z_{i-1,j} - z_{i+1,j} = \Delta_x h_{ij}$$
    $$z_{i,j-1} - z_{i,j+1} = \Delta_y h_{ij}$$

Preliminary results





Topics

Three algorithic advances and a research topic in processing topographic and bathymetric sensor data:

- lossy terrain compression that maintains slope accuracy,
- **bathymetric surface fitting to irregular tracklines,**
- lossy compression of 5D environmental data, and
- terrain modeling to maintain hydrological validity.

Why? To attack several issues raised by the large amounts of data now available.

Eventual goal: A unified system.

Fitting bathymetric data is hard

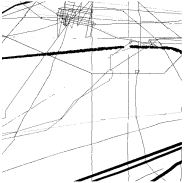
- MBB data is very unevenly spaced (dense in a swath along the ship tracklines, but then nonexistent for a long distance sideways),
- depth accuracy is a few percent, and
- insufficient data to infer features that are probably there.

Current methods often

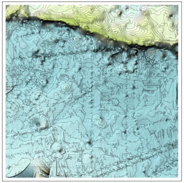
- have a specific distance wired into the formula,
- do not let information flow past data points, and so
- produce artifacts (e.g., abrupt slopes, acquisition footprint);
- show details that aren't justified.

Sea floor bathymetry trackline fitting

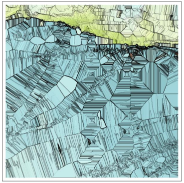
Problem: Trackline data is very unevenly spaced, leading to very bad surface fitting.



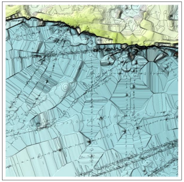
Bathymetry Dataset



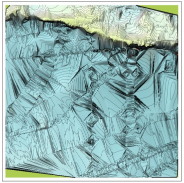
Kriging w. ArcGIS



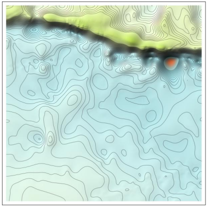
Voronoi Polygons



Inverse Distance



2nd-order Spline Interp



Soln: ODETLAP, $R = 100 \to 10$

Relevant terrain property

- Terrain is unlikely to have created artifacts exactly where the multibeam bathymetry later scanned it.
- How to work this fact into the math?

ODETLAP extension:

- vary R depending on distance to known points

Topics

Three algorithic advances and a research topic in processing topographic and bathymetric sensor data:

- lossy terrain compression that maintains slope accuracy,
- bathymetric surface fitting to irregular tracklines,
- **lossy compression of 5D environmental data,** and
- terrain modeling to maintain hydrological validity.

Why? To attack several issues raised by the large amounts of data now available.

Eventual goal: A unified system.

5D data compression

- Sensors, e.g., in World Ocean Atlas 2005, collecting multiple bands of environmental data –
- temperature, salinity, oxygen concentration,
- producing set of values over 5D grid (x, y, z, t, b).
- *Compress it!*
- little prior art.

Principles:

- Assume one band's large derivative at particular $(x, y, z, t) \Rightarrow$ likely for the other bands,
- Treat the data as one 5-D dataset, and
- Compress lossily since the data is imprecise.

Data compression technique

- extend ODETLAP to 3D, then 4D, 5D.
- *Major challenge:* Everything harder in higher dimensions.
- *To date:* compression ratios of 100:1 (mean error $< 1.5\%$).

	3D-ODETLAP			3D-SPIHT		
Variable	Mean Err(%)	Max Err(%)	Compr Ratio	Mean Err(%)	Max Err(%)	Compr Ratio
Salinity	0.0532	0.2174	77:1	0.0530	0.4946	11:1
Temperature	0.4993	2.0673	98:1	0.50	17.91	135:1
Dissolved O_2	0.9993	4.4145	100:1	1.002	24.9965	71:1
Apparent O_2 util.	0.9999	4.0170	85:1	0.9991	20.3609	81:1
Percent O_2 satur.	0.9985	4.5672	78:1	0.9969	20.3610	65:1
Phosphate	0.9993	4.5241	86:1	0.9978	15.6922	65:1
Nitrate	1.0242	4.6946	66:1	1.0006	18.5360	59:1
Silicate	0.9996	5.1437	91:1	1.0018	21.6457	81:1

ODETLAP's smaller compression error than SPIHT

Topics

Three algorithic advances and a research topic in processing topographic and bathymetric sensor data:

- lossy terrain compression that maintains slope accuracy,
- bathymetric surface fitting to irregular tracklines,
- lossy compression of 5D environmental data, and
- **terrain modeling to maintain hydrological validity**

Why? To attack several issues raised by the large amounts of data now available.

Eventual goal: A unified system.

Maintaining hydrological validity

Given incomplete hydrography; fill in gaps.

- Presented on Wed
- Don’t work directly on partial hydrography.
- Compute deeper representation (terrain) from it.
- Derive hydrography.
- Result is guaranteed internally consistent.

Topics

Three algorithic advances and a research topic in processing topographic and bathymetric sensor data:

- lossy terrain compression that maintains slope accuracy,
- bathymetric surface fitting to irregular tracklines,
- lossy compression of 5D environmental data, and
- terrain modeling to maintain hydrological validity.

Why? To attack several issues raised by the large amounts of data now available.

Eventual goal: A unified system.

Disclaimer and goals

This section is a report on work not yet done.

Goals: Math that

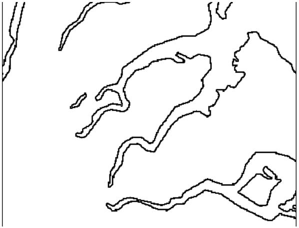
- allows the representation of only legal terrain (= height of land above geoid),
- minimizes what needs to be stated explicitly, and
- enforces global consistencies.

Why? To put compression and other ops on a logical foundation.

Terrain properties

Messy, not theoretically nice.

- Often discontinuous (C^{-1}).
- Many sharp local maxima.
- But very few local minima.
- Lateral symmetry breaking — major river systems.



Peninsulas or fjords?

- Different formation processes in different regions.
- Features do not superimpose linearly; two canyons cannot cross and add their elevations.
- C^{∞} linear systems, e.g,. Fourier series, are wrong.
- Structure that people can recognize even though hard to formalize; see Figure.
- Multiple related layers (elevation, slope, hydrology).

Current representations

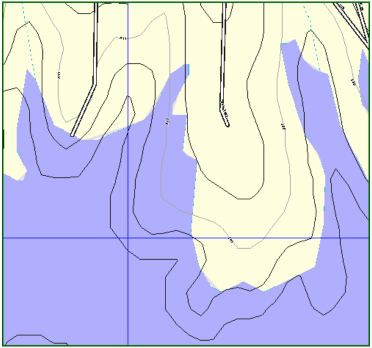
- Array of elevation posts.
- Triangular splines, linear or higher.
- Fourier series.
- Wavelets

Theory vs practice:

- Slope is derivative of elevation, but
- that amplifies errors, and
- lossy compression has errors, so
- maybe we want to store it explicitly.

Also, shoreline is a level set, but...

Inconsistencies between layers



Elevation contours crossing shoreline

Slope is important

- mobility
- erosion
- aircraft
- visibility
- recognition

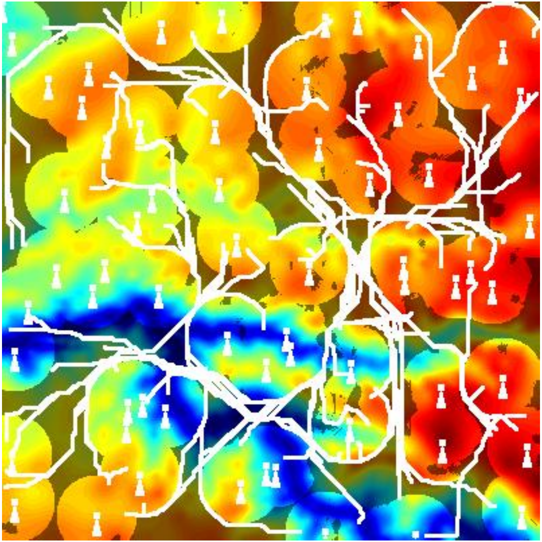
Path planning

Example of a common terrain operation. Cost depends on

- distance
- uphill climb
- being seen
- not a *metric:* $d(a,b) \neq d(b,a)$
- not a *scalar field difference:* $d(a,b) \neq -d(b,a)$

$$\mathcal{C} = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \cdot \left(1 + \max\left(0, \frac{\Delta z}{\sqrt{\Delta x^2 + \Delta y^2}}\right)\right) \cdot (1 + 100v)$$

Smugglers and Border Guards

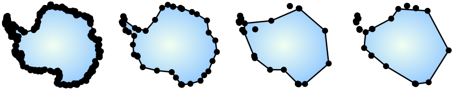


Math should match physics

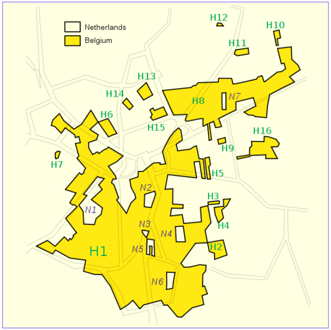
- Fourier series appropriate for small vibrations, not terrain.
- Truncating a series produces really bad terrain.
- Anything, like Morse complexes, assuming continuity is irrelevant.
- Fractal terrain is not terrain.
- Wavelets: how to enforce long-range consistency?
- Topology, by itself, is too weak.
- Terrain is not linear, not a sum of multiples of basis function.

Examples of rich structure

- group theory
- trig
- constructive solid geometry
- line generalization from level sets
- hydrology from terrain
- polygon properties from local geometry and topology.



Line generalization



Baarle-Nassau - Baarle-Hertog border

Terrain formation by scooping

- Problem: Determine the appropriate operators, somewhere inside the range from conceptually shallow (ignoring all the geology) to deep (simulating every molecule).
- One solution: **Scooping.** Carve terrain from a block using a scoop that starts at some point, and following some trajectory, digs ever deeper until falling off the edge of the earth.
- Properties: Creates natural river systems w cliffs w/o local minima.
- Every sequence of scoops forms a legal terrain.
- Progressive transmission is easy.

Terrain formation by features

- Represent terrain as a sequence of features — hills, rivers, etc ..
- plus a combining rule.
- This matches how people describe terrain.
- Progressive transmission.
- The intelligence is in the combining rule.

How compact is this rep? How to evaluate it?

Implications of a better rep

- Put earlier empirical work on a proper foundation.
- Formal analysis and design of compression.
- Maximum likelihood interpolation, w/o artifacts.
- Treat more sophisticated metrics, like suitability for operations like path planning, or recognizability.
- Close the loop to pre-computer descriptive geometry.