COMPLETING RIVER NETWORKS WITH ONLY PARTIAL RIVER OBSERVATIONS VIA HYDROLOGY-AWARE ODETLAP

Tsz-Yam Lau and W. Randolph Franklin

Department of Computer Science
Rensselaer Polytechnic Institute
110 8th Street, Troy NY, USA
laut@cs.rpi.edu, wrf@ecse.rpi.edu

ABSTRACT:

Fragmentary river segments have to be joined appropriately before becoming useful to addressing transportation problems like route planning and pollutant tracking. When height observations are available at the given river locations only, conventional terrain reconstruction techniques often fail to help to classify river/non-river locations at the tributaries. We propose using the hydrology-aware variant of Overdetermined Laplacian Partial Differential Equation (ODETLAP) to improve the situation. Its original version features regularizing every position, known or unknown, with the average of its immediate neighbors. Slope continuity as a result helps infer ridges at regions between the river segments, which is crucial to correct tributary formation. To honor the given river locations as local minima, in this ODETLAP variant we regularize the corresponding heights to values smaller than the respective neighborhood averages. When this average deflation is large enough, the given river locations become local minima in the reconstructed terrain. We foresee application of our solution framework in a few 2D or 3D network tracing problems having similar sample distribution, like dendrite network reconstruction.

1 INTRODUCTION

Knowledge of the complete river network is essential to a number of geographical and environmental applications. For example, we need to know how the river segments are connected to design the shortest route for a ship to travel from one place to another. We have to identify the exact segment connectivity and river locations before determining the areas that are likely to be affected by pollutants or flooding. However, that information is usually not immediately available with conventional surveying techniques. The presence of clouds and canopies often occlude parts of the river network (Asante and Maidment, 1999). We are only able to identify the river segments.

To apply these data to connectivity applications, we need to join these river segments appropriately. A complete river network is usually a fully-connected tree. Branches can be multiple-cell thick, but very often we aim to figure out one-cell thick branches representing the middle lines of the rivers. We usually expect that every river location has a single way for the water to reach some sea shore or terrain edge (Asante and Maidment, 1999). If we have fragmentary river observations only (no height data are available), what we can do is to extend the segments (Asante and Maidment, 1999). Criteria like shortest connection distance or segment curvature preservation may be assumed to guide the process. However, the search space is still huge. Numerous solutions are possible. In contrast, if some other terrain property, which is usually the set of elevations, is also known, we may rule out certain possibilities. For example, it is nearly impossible for a river segment on one side of the hill to meet another river segment on the other side.

In our previous work (Lau and Franklin, 2010), we reported that such height awareness can be achieved by the *induced terrain approach*, which involves first reconstructing a full terrain with a *hydrological corrected terrain reconstruction* algorithm, and then deriving a river network using the *biased river derivation* algorithm that favors given river locations. Figure 1 illustrates the whole workflow. If height samples are available across the terrain, we recommend natural neighbor interpolation and stream burning (NN-SB) as the underlying terrain reconstruction algorithm. This one-pass tactic honors the entire given river locations, and offers reasonably accurate river/non-river classification even when compared with other multi-pass alternatives.

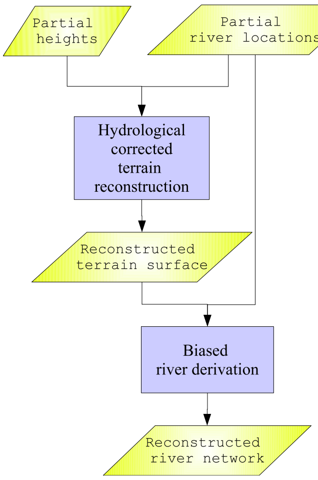

Partial heights
Partial river locations
Hydrological corrected terrain reconstruction
Reconstructed terrain surface
Biased river derivation
Reconstructed river network

Figure 1: Workflow of the induced terrain approach of river network reconstruction.

This work describes a strategy for some even more hostile situations in which height samples are only available at certain given river locations. In such cases, we find that replacing NN-SB with the hydrology-aware variant of Overdetermined Laplacian Partial Differential Equation (HA-ODETLAP) gives much more accurate river/non-river classifications at the tributaries. These types of locations are where other conventional schemes depend heavily on given height samples to infer ridges and valleys crucial to correct water distribution and river formation. In Section 2 we will explain why this is the case through a review of the basic ODETLAP. In Section 3 we will describe how the basic ODETLAP framework is adapted to incorporate given river locations in order to honor them as local minima as we expect for real river locations. In Section 4, we will suggest what this finding implies for future work, like potential applications to problems of a similar nature.

2 OVERDETERMINED LAPLACIAN PARTIAL DIFFERENTIAL EQUATION

2.1 Basic version

ODETLAP stands for *Overdetermined Laplacian Partial Differential Equation* (Gousie and Franklin, 2005, Xie et al., 2007) (Some refer to it as *Inverse interpolation* (Claerbout and Fomel, 2008)). This approach solves for an $n \times n$ elevation grid $Z = \{z_{i,j}\}$ where $1 \leq i, j \leq n$ through setting up an overdetermined system. The unknowns are all the n^2 points in the elevation grid, whether their height values are already known or not. Each position whose height is known induces two equations, an averaging equation and an exact equation. The averaging equation is in general

$$z_{i,j} = (z_{i-1,j} + z_{i+1,j} + z_{i,j-1} + z_{i,j+1})/4 \qquad (1)$$

which is the discretized version of the following Laplacian Partial Differential Equation.

$$\frac{\delta^2 z}{\delta x^2} + \frac{\delta^2 z}{\delta y^2} = 0 \qquad (2)$$

(N. B.: We will have slightly different treatments if the point is on an edge or at a corner of the grid.)

The exact equation is

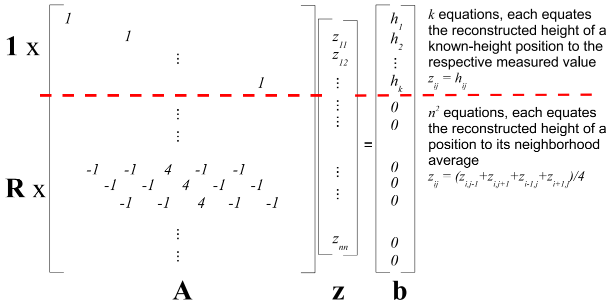
$$z_{i,j} = h_{i,j} \qquad (3)$$

where $h_{i,j}$ is the known height at that position.

For each unknown-height position, only an averaging equation is created.

As a result, if k of the n^2 positions have known heights, there are $n^2 + k$ equations for the n^2 unknown variables. k of them attempts to set the heights of the known-height positions to the respective values, while the remaining n^2 equations impose a regularization constraint, which in this case smoothes the overall surface by trying to set the height value of each position to the average of its immediate neighbors.

When there are more equations than the unknowns and the system of equations is inconsistent, we optimize for a least squares solution. We can bias the solution to particular equations by multiplying a factor $R > 1$ to both sides of the respective equations. In the original implementation, we can increase our bias to all the averaging equations for a smoother surface but reduced accuracy and vice versa. Figure 2 presents an overview of the system.



k equations, each equates the reconstructed height of a known-height position to the respective measured value $z_{ij} = h_{ij}$

n^2 equations, each equates the reconstructed height of a position to its neighborhood average $z_{ij} = (z_{i,j-1}+z_{i,j+1}+z_{i-1,j}+z_{i+1,j})/4$

Figure 2: Original ODETLAP

The superiority of the ODETLAP approach lies in its ability to handle continuous contour lines of elevations and isolated points at the same time. Also, while producing a surface that infers mountain tops inside innermost contours, it enforces continuity of slope across contours. As a result, it shows no visible indication of the input contours, i.e., no generated terraces (Franklin et al., 2006).

Note that any equation in the system cannot have more than five variables. As a result, the coefficient matrix $\mathbf{A}$ is *sparse*. Sparse linear systems can be solved rather quickly using sparse QR factorization. We can speed up the computation even more by multiplying the adjoint of $\mathbf{A}$ to both sides of the linear system (from $\mathbf{Az} = \mathbf{b}$ to $\mathbf{A}^T\mathbf{Az} = \mathbf{A}^T\mathbf{b}$) before solving for $\mathbf{z}$. The new coefficient matrix $\mathbf{A}^T\mathbf{A}$ is *symmetric positive definite*. Much more efficient computation using Cholesky factorization is available (Li et al., 2010). Solving a 400×400 dataset takes about 30 CPU-seconds on a Lenovo Thinkpad X61 with one 2.1GHz dual-core Centrino vPro processor and 3GB RAM running Windows Vista. For larger datasets, we can partition the problem, solve the pieces in parallel, and smoothly merge them together (Stookey et al., 2008). Such a divide-and-conquer approach means we are required to recompute a local subblock rather than the whole grid when there is a local data update. Such cost saving is critical for dealing with terrain data which are increasingly massive.

2.2 Extension to include known river locations

River locations are expected to be local minima relative to surrounding non-river locations. To honor this expectation in HA-ODETLAP which is the hydrology-aware version of ODETLAP, we modify the averaging equation at each known river location: we regularize its height to a value smaller than the respective neighborhood average. For example,

$$z_{i,j} = (z_{i-1,j} + z_{i+1,j} + z_{i,j-1} + z_{i,j+1})/4f \qquad (4)$$

where f is the average deflation factor: if we increase its value from 1, the value imposed to $z_{i,j}$ by this averaging equation decreases from the neighborhood average. With sufficiently large f, that value will be smaller than any of its non-river immediate neighbors. The computed surface will have a local minimum at that location. An illustration is shown in Figure 3. Another possible realization of the formula is available (Muckell, 2008).

			10				
			10				
			10				
	10				10	10	10
10							

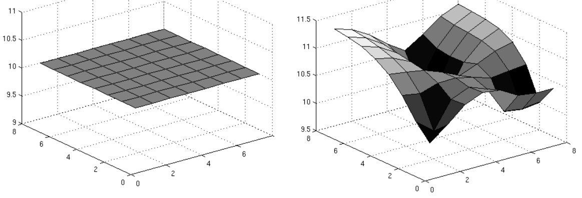


Figure 3: Test case for HA-ODETLAP. Known river locations with respective heights (top). Terrain recovered using basic ODETLAP (bottom left). Terrain recovered using HA-ODETLAP $f = 1.04$ (bottom right).

Figure 4 illustrates the overall system of our hydrology-aware extension. Note that we just change the values of some non-zero entries in the $\mathbf{A}$ matrix. Zero locations in the original system remain zeros. The sparsity structure of the $\mathbf{A}$ matrix is not changed.

Thus the modified system can be solved within a time similar to the original system.

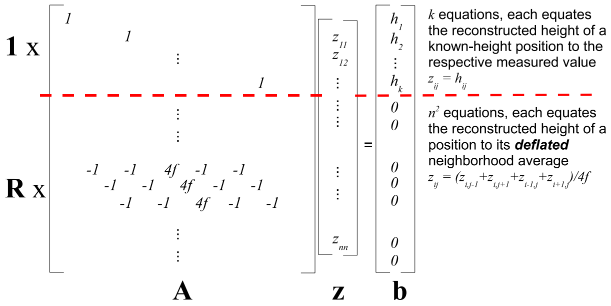

1 x
R x
1
1
1
-1 -1 4f -1 -1
-1 -1 4f -1 -1
-1 -1 4f -1 -1
z_{11}
z_{12}
z_{nn}
=
h_1
h_2
h_k
0
A
z
b
k equations, each equates the reconstructed height of a known-height position to the respective measured value
$z_{ij} = h_{ij}$
n^2 equations, each equates the reconstructed height of a position to its ***deflated*** neighborhood average
$z_{ij} = (z_{i,j-1}+z_{i,j+1}+z_{i-1,j}+z_{i+1,j})/4f$

Figure 4: HA-ODETLAP

As in the original system, the accuracy-smoothness parameter R determines our trade-off between smoothness of the reconstructed terrain and accuracy of the known heights. Here since all the height samples are concentrated at the river locations, terrain surface smoothness is not that relevant, and this justifies a high accuracy setting, say $R = 20$. The new parameter f is much more interesting and thus worth more attention. To look into how this parameter affects river location production and recovery, we conduct the following experiment.

Our test dataset includes some six 400 × 400 digital elevation models (DEMs) shown in Figure 5. A bigger plot of *mtn1* is available in Figure 6 top left.

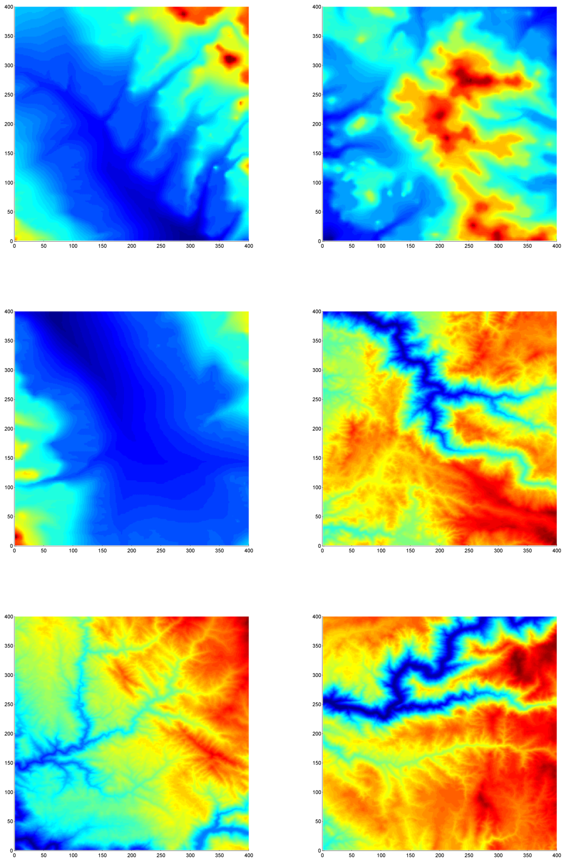


Figure 5: Test 400×400 DEMs: (first row) *hill1*, *hill2*, (second row) *hill3*, *mtn1*, (third row) *mtn2*, *mtn3*.

We first run `r.watershed` in GRASS GIS (Ehlschlaeger, 2008) with accumulation cutoff threshold = 200, initial water amount at each location = 1 over these six DEMs to obtain the respective theoretical four-connected river networks. (Figure 6 top middle shows the full river network of *mtn1*.) Next, we sample for the partial heights and river locations as follows: For river locations, first we divide the whole grid into 20 × 20 subgrids. In each subgrid, we randomly pick a point and mask an area of 12 × 12 around it. Heights are provided at given river position only. (Figure 6 top right shows the resulting partial river network of *mtn1*, while its bottom left shows the locations where heights are available.) After that, we pass the partial heights and river locations data to our algorithm at five different f settings: 1.00, 1.01, 1.02, 1.03 and 1.04, and obtain the respective reconstructed terrain surfaces. Finally, we bias `r.watershed` as described in our previous work (Lau and Franklin, 2010) to figure out the respective river networks that honor the entire given river locations. We compute the river/non-river correct classification rates to quantify how well the respective settings perform.

Dataset	HA-ODETLAP with f					NN-SB
	1.00	1.01	1.02	1.03	1.04	
$mtn1$	6.89	2.59	2.50	2.45	2.45	4.20
$mtn2$	6.86	2.73	2.71	2.64	2.62	4.42
$mtn3$	6.17	2.77	2.64	2.69	2.71	4.14
$hill1$	6.15	2.60	2.67	2.73	2.72	3.18
$hill2$	6.75	2.41	2.51	2.56	2.58	3.78
$hill3$	6.52	3.50	3.50	3.51	3.54	3.43

Table 1: Error rates of HA-ODETLAP under different average deflation factors f, together with natural neighbor with stream burning (trench amount=30). Inside each cell is the classification error with respect to the ground truth.

Table 1 shows how f matters. f = 1.00 is equivalent to the original implementation. At that setting, HA-ODETLAP performs worse than NN-SB because the algorithm does not model the given river locations as local minima properly. For all other settings, we see significant improvement over NN-SB. Indeed, we find that for this particular set of terrains, $f = 1.02$ gives consistently satisfactory results in general.

Figure 6 bottom middle and right show the recovered river connections using these two terrain reconstruction schemes. One may immediately realize that the tributaries cannot be correctly reproduced when NN-SB is used: in this case we do not have any height sample in the non-river area. As a result, NN-SB fails to reconstruct proper V-shape centered at the river lines across the non-river regions, which is important for deducing proper tributaries. In contrast, the hydrology-aware adaptation in ODETLAP infers local minima as long as the corresponding locations are defined to be river locations, regardless of whether heights are available.

3 CONCLUSION AND FUTURE WORK

We have proposed the hydrology-aware Overdetermined Laplacian Partial Differential Equation (HA-ODETLAP) as a neat solution to resolving for tributaries when height sample is only available at the given river segment locations. By regularizing all given segment locations as local minima and other locations as the immediate neighborhood average, ridges and valleys are properly induced not only at given river locations but also areas without any height samples. We believe that is the reason why HA-ODETLAP induced-terrains offer superior tributaries and hence smaller river/non-river classification errors because of their huge contribution to the complete river network locations.

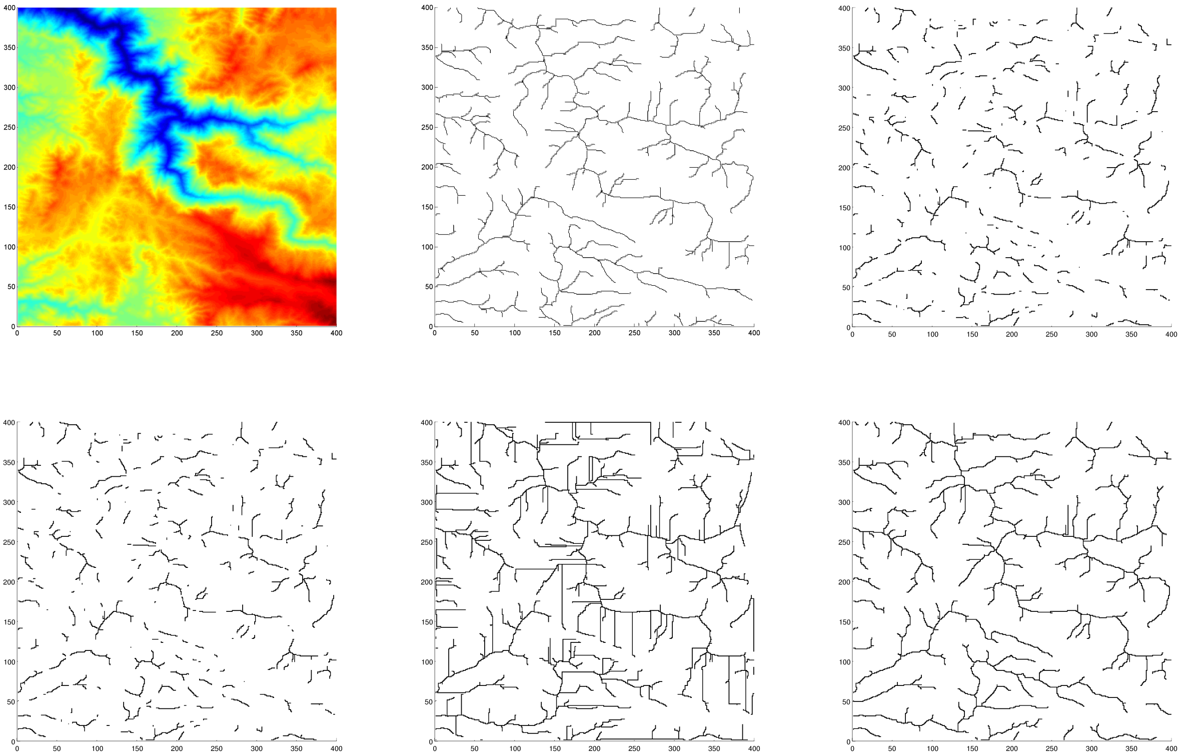


Figure 6: Full elevation (top left). Full river network (top middle). Available partial river segments(top right). Locations where heights are available (bottom left). NN-SB river reconstruction result (bottom middle). Hydrology-aware ODETLAP result with $f = 1.02$ (bottom right).

Having seen the success of this set of techniques with completing hydrology networks, we are eager to port the same solution framework to complete some other 2D or 3D networks with similar sample data distribution. For example, in 3D dendrite network recovery problems, we are trying to connect the dendrite segments, each of which is represented by a tubular structure with decreasing gray-scale intensity away from the axis, in such a way that they are all eventually linked to the neuron nucleus. If we model the dendrite pieces as river segments, the inverted gray-scale intensities as heights, and the nucleus location as the only place for all rivers to end (instead of any edges of the terrain grid), we essentially transform this problem to a 3D river segment connection problem. We are investigating whether our solution framework can give an elegant and more accurate solution than the existing approaches.

4 ACKNOWLEDGMENTS

This research was partially supported by grant CMMI-0835762 from the National Science Foundation.

REFERENCES

Asante, K. and Maidment, D., 1999. Creatng a river network from the arcs in the digital chart of the world. `http://www.ce.utexas.edu/prof/maidment/grad/asante/dcw/rivernet.htm`, (retrieved Jun 23, 2010).

Claerbout, J. F. and Fomel, S., 2008. Image estimation by example: geophysical soundings image construction. `http://sepwww.stanford.edu/sep/prof/gee2.2010.pdf`, (retrieved Jun 23, 2010), chapter 3, pp. 85–108.

Ehlschlaeger, C., 2008. GRASS GIS: r.watershed. `http://grass.itc.it/gdp/html_grass63/r.watershed.html`, (retrieved Jun 23, 2010).

Franklin, W. R., Inanc, M., and Xie, Z., 2006. Two novel surface representation techniques. In: Autocarto.

Gousie, M. B. and Franklin, W. R., 2005. Augmenting grid-based contours to improve thin plate DEM generation. Photogrammetric Engineering & Remote Sensing 71(1), pp. 69–79.

Lau, T. Y. and Franklin, W. R., 2010. Completing fragmentary networks via induced terrain. In: Autocarto.

Li, Y., Lau, T. Y., Stuetzle, C. S., Fox, P. and Franklin, W. R., 2010. 3D oceanographic data compression using 3D-ODETLAP. In: 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.

Muckell, J., 2008. Evaluating and compressing hydrology on simplified terrain. Master's thesis, Rensselaer Polytechnic Institute.

Stookey, J., Xie, Z., Cutler, B., Franklin, W. R., Tracy, D. and Andrade, M. V., 2008. Parallel ODETLAP for terrain compression and reconstruction. In: 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.

Xie, Z., Andrade, M. A., Franklin, W. R., Cutler, B., Inanc, M., Tracy, D. M. and Muckell, J., 2007. Approximating terrain with over-determined Laplacian PDEs. In: 17th Fall Workshop on Computational Geometry, IBM TJ Watson Research Center, Hawthorne NY. poster session, no formal proceedings.