15th International Symposium on Spatial Data Handling
Bonn, Germany, 22-24 August 2012

Better completion of fragmentary river networks with the induced terrain approach by using known non-river locations

Tsz-Yam Lau and W. Randolph Franklin

Rensselaer Polytechnic Institute,
110 8th Street, Troy NY 12180, USA
laut@cs.rpi.edu, mail@wrfranklin.org

Abstract. We improve the precision of our induced terrain approach for completing fragmentary river networks by incorporating known non-river location data. First, we reconstruct a terrain with known terrain attributes, and then we derive a network passing through all the segments using a river derivation scheme. The original implementation honors height samples and hydrological consistency constraints, which raises the reliability of the reconnected networks in addressing transportation problems. However, we still observe a few wrong reconnections, which pass through areas which are clearly known to have no water flow. To force the river to avoid those places, we opt for raising their heights to a level unreachable by any water in the known river locations instead of modifying the river derivation scheme. While effective in boosting accuracy by around 5 percentage points, this alternative makes it simple to switch between river derivation schemes with different execution speeds and hydrographical consistency constraints.

Keywords: river reconnections, induced terrain, terrain modeling

1 Motivation

The hydrography network is essential for cross-terrain transportation issues such as ship routing, pollutant monitoring and flood plain control. Since they can cover a wide area within a short time, aerial surveys are now the main source of hydrography network data on Earth, and can be expected to remain so for the respective data on other planets. Very often, the extraction of river networks from such survey data involves a classification of terrain features according to multispectral data collected from the spatial locations. In a few locations, due to view obstacles such as canopies and clouds, we fail to detect any signal. Even at other locations, we may not know a river's presence for certain [17]. To avoid false positives, the classification result is usually a disconnected network containing gaps. Such an outcome is not a problem if other surveys are conducted to obtain the missing data, or ample time is provided to complete the reconnections manually. Otherwise (such as the real-time monitoring of a large quantity of data for potentially changing hydrography in a tropical rain forest), automatic bridging of those fragmentary river segments is needed.

This paper extends our automatic reconnection solution called the *induced terrain approach* [7]. As the overview in Section 2 will reveal, this approach reconnects the segments in a way that honors both given height samples and hydrological consistency. Despite their great influence on the network topology, they are not fully considered by conventional geometry-based approaches [11,12,15, 17]. We observe that our original induced terrain approach has yet made complete use of the terrain feature classification data: The reconstructed river network passes through a few locations which have been identified as non-river features. This is obviously not right. Eliminating those situations raises the probability that the correct reconnections will be made and thus improves accuracy. In Section 3, we will illustrate how this problem can be fixed with our induced terrain approach, before we conclude our paper in Section 4.

2 Induced terrain approach

2.1 Structure reconstruction

Our induced terrain approach starts with reconstructing a full height grid $\mathbf{Z}$ according to the given river locations and other additional terrain attributes. We emphasize the use of available height samples because the topological relationship among neighbors determines how water is routed: water flows to its lower neighbors. A hill sitting between two river segments could act as an obstacle to block water from one segment to flow to another segment. These height samples are often available in airborne surveys using the Light Detection and Ranging (LiDAR) technology. We previously found that with sparsely and evenly distributed height samples, such as those from closed-canopy forests [4], we may first reconstruct a preliminary terrain from the given partial heights with natural neighbor interpolation (NN) [14]. We then lower the heights of given river locations by a trench amount, a technique known as stream burning (SB) [5], to help those river locations trap water there.

2.2 Information derivation

To reconnect the segments, we compute a river network from the above reconstructed terrain using a *river derivation algorithm*. Such a river derivation algorithm first computes the water flow directions of all the terrain locations based on their heights, and then finds the amount of water passing through each cell from those directions. In our induced terrain approach, this step is *biased*: Instead of giving all locations with the same amount of initial water, we offer each given river location an initial water amount that is exactly the critical amount (accumulation cutoff) to make sure that each location has sufficient water to be identified as river locations. Meanwhile, all other locations are allocated zero initial water. They have to receive water from the given river locations for river flow. Furthermore, we protect the given river locations from being trimmed away in the final river thinning process. This ensures that the resulting river network passes through, and hence reconnects, the entire given river locations.

The use of a river derivation algorithm helps enforce hydrological consistency in the reconnected river network. Hydrological consistency refers to the set of global constraints we expect on the complete river network. Similar to the completion of road networks, these global restrictions help eliminate invalid reconnections and thus give way to the correct ones [16]. One common constraint governs how water is routed away from sinks. A class of river derivation algorithms, such as `r.watershed` [2] available in GRASS GIS, assign flow directions such that water flows "uphill" and escapes the sink by following a least-cost path [10]. Some others, such as `Terraflow` [1], choose to fill all sinks by flooding so that every cell can then be assigned flow directions without routing water uphill [6]. Another common constraint determines whether the network should consist of a number of tree branching structures. For example, the single-flow direction (SFD) version of `r.watershed` guarantees every location is assigned a single direction to route its water to a terrain edge or a sink in a loop-free manner. If one wishes to have a river network with loop, he may opt for the multiple-flow direction (MFD) version of `r.watershed`, which allows distribution of water from a river location to two or more neighbors and hence river branching. It is still an active topic to develop faster river derivation schemes that satisfy different sets of hydrological consistency constraints [8, 9].

3 Known non-river locations

In the terrain feature classification step, we may have identified locations that are clearly land features. We then do not expect a river to flow through them. If any reconnection segment occupies those regions, we can tell immediately that there is something wrong with it. Consider the terrain and its hydrography shown in Figure 1, left. In the middle of the same figure, we have a snapshot of the given river locations and the respective reconnections using our original induced terrain approach. Note the reconnection segment at (30, 35). We can say it is incorrect as it passes through a region that has been identified to have no river flow. We should figure out some way to avoid those situations, in order to increase the chance for the actual reconnection to be picked up and hence raise overall accuracy.

3.1 Method 1: modify the river derivation scheme

Our ultimate goal is to block water from flowing into those known non-river locations. The most direct way to achieve that purpose is to ask the river derivation scheme to ignore those non-river locations: They should not accept any water inflow and should not give water away to the others. A few river derivation schemes have already provided such options for users. For instance, in `r.watershed`, one can specify those locations with its `blocking` option.

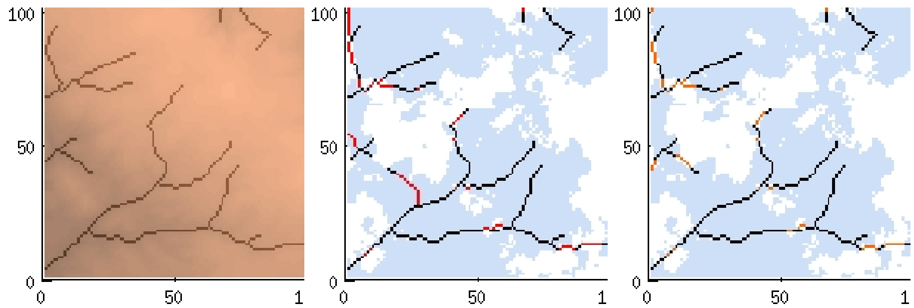

Fig. 1. Left: Full terrain (copper grayscale) with hydrography (black). Middle: With view blocked by obstacles (white), the given river locations (black) and reconnections with the original induced terrain approach (red). Note some reconnections pass through known non-river locations in blue. Right: Reconnections with our improved induced terrain approach (orange).

3.2 Method 2: modifying the induced terrain

However, there can be a few problems if the above method is not readily available in the implementations provided: The source code may not be available, and even if one has the code, understanding and modifying the source code could take considerable time. With the induced terrain approach, we have the alternative of modifying the induced terrain: we raise the heights of those known non-river locations by the maximum height found in the reconstructed terrain before feeding it to the information derivation step. Suppose that the original reconstructed height of the location (i, j) is $z_{i,j}$, its height after the process, $z'_{i,j}$, is given by the following formula.

$$z'_{i,j} = \begin{cases} z_{i,j} + MAXZ + \Delta & \text{if it is a known non-river location} \\ z_{i,j} & \text{otherwise} \end{cases} \tag{1}$$

In the above, $MAXZ$ is the maximum height in the original elevation grid $\mathbf{Z}$, and Δ is a small positive value to avoid ties. Since now those given non-river locations are extra high, water at the given river locations will find a way that passes through the other locations instead to drain their water away in order to avoid the large cost.

3.3 Evaluation

Our test datasets consist of twelve 400 $\times$ 400 DEMs extracted from two SRTM1 cells and two SRTM3 cells. We obtain the ground-truth eight-connected river network by running run `r.watershed` with accumulation cutoff threshold = 200 and initial water amount at each location = 1 over these twelve DEMs. We derive the full river networks with complete elevation data since we need the accurate ground-truth river networks for comparison with the reconnection results. Real partial observations rely on humans to complete the missing parts, which may mean errors.

We prepare two sets of observed river locations. Set 1 attempts to mimic the situations with rather regularly distributed obstacles (such as cumulus clouds [13]). We first 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. Set 2 simulates clouds that exhibit spatial autocorrelation. We use the diamond square algorithm [3] to generate fractal cloud patterns that cover at least 30% of the total area.

Correct adjacent segment reconnection is chosen as the evaluation criterion. It calculates the proportion of river segments that are connected back to their respective adjacent downstream segments. Table 1 shows the segment reconnection results. No degradation is observed. The mean improvements to the regularly-spaced obstacles and spatially autocorrelated obstacles are 6.01% and 3.84% respectively, and they are all statistically significant.

	Set 1		**Set 2**	
	Regularly spaced		**Spatially autocorrelated**	
Non-river locations used?	No	Yes	No	Yes
hill1	83.98%	89.00%	80.38%	84.69%
hill2	83.98%	88.31%	79.22%	80.52%
hill3	58.45%	76.10%	58.82%	64.71%
hill4	78.10%	86.13%	78.10%	84.31%
hill5	87.04%	89.88%	80.97%	83.81%
hill6	85.16%	94.92%	81.64%	85.16%
mtn1	91.25%	94.17%	77.08%	80.00%
mtn2	95.13%	98.67%	88.05%	91.15%
mtn3	91.96%	95.54%	82.59%	84.82%
mtn4	79.59%	88.10%	73.81%	79.25%
mtn5	93.44%	96.53%	80.70%	83.40%
mtn6	92.77%	95.58%	78.71%	84.34%
mean improvement		6.01%		3.84%
95% confidence interval min		3.51%		2.44%
95% confidence interval max		8.50%		5.23%

Table 1. Correct adjacent segment reconnection rates of different DEMs, with and without the use of known non-river location data.

4 Conclusion

This paper presents one of our approaches for extending the induced terrain approach for completing fragmentary river networks. In particular, we have discussed how known non-river locations may be used. The crux is to block any water inflow in those locations, so they will never become parts of the river network. New river derivation schemes are coming out from time to time for improved speed and different hydrological consistency constraint enforcement, and it could be hard to modify the respective implementations. In case the scheme so picked does not allow such modifications to be done readily, in the induced terrain approach one can alternatively raise the heights of known non-river locations above the maximum height found in the terrain. This simple yet effective practice achieves around 5 percentage points of improvement. (This research was partially supported by NSF grants CMMI-0835762 and IIS-1117277)

References

1. L. Arge, H. Mitasova, and L. Toma. r.terraflow. http://www.cs.duke.edu/geo*/terraflow/r.terraflow.html, (retrieved Feb 18, 2011), 2002.
2. C. Ehlschlaeger. GRASS GIS: r.watershed. http://grass.itc.it/gdp/html_grass63/r.watershed.html, (retrieved Jun 23, 2010), February 2008.
3. A. Fournier, D. Fussell, and L. Carpenter. Computer rendering of stochastic models. *Communications of the ACM*, 25(6):371–384, 6 1982.
4. D. Gatziolis, J. S. Fried, and V. S. Monleon. Challenges to estimating tree height via LiDAR in closed-canopy forests: A parable from western Oregon. *Forest Science*, 56(2):139–155, April 2010.
5. M. F. Hutchinson. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology*, 106:211–232, 1989.
6. S. K. Jenson and J. O. Domingue. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing*, 54(11), 1988.
7. T.-Y. Lau and W. R. Franklin. Completing fragmentary river networks via induced terrain. *Cartography and Geographic Information Science*, 38(2):162–174, Apr. 2011.
8. S. V. G. Magalhães, M. V. A. Andrade, W. R. Franklin, and G. C. Pena. A new method for computing the drainage network based on raising the level of an ocean surrounding the terrain. In *AGILE'2012 15th AGILE international conference on geographic information science*, Avignon, 24–27 April 2012.
9. M. Metz, H. Mitasova, and R. S. Harmon. Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search. *Hydrology and Earth System Sciences*, 15(2):667–678, 2011.
10. D. Morris and R. Heerdegen. Automatically derived catchment boundary and channel networks and their hydrological applications. *Geomorphology*, 1:131–141, 1988.
11. J. A. Noble. The effect of morphological filters on texture boundary localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:554–561, 1996.
12. Paul V. C. Hough. Method and Means for Recognizing Complex Patterns. US Patent 3069654, 1962.
13. J. A. Ramirez and R. L. Bras. Clustered or regular cumulus cloud fields: The statistical character of observed and simulated cloud fields. *Journal of Geophysical Research*, 95(3):2035–2045, 2 1990.
14. R. Sibson. A brief description of natural neighbor interpolation. In V. Barnett, editor, *Interpreting Multivariate Data*, chapter 2, pages 21–36. John Wiley & Sons, New York, 1981.
15. P. Soille and J. Grazzini. Extraction of river networks from satellite images by combining mathematical morphology and hydrology. In W. Kropatsch, M. Kampel, and A. Hanbury, editors, *Computer Analysis of Images and Patterns*, volume 4673 of *Lecture Notes in Computer Science*, pages 636–644. Springer Berlin / Heidelberg, 2007.
16. C. Steger, H. Mayer, and B. Radig. The role of grouping for road extraction. In *In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkh auser Verlag Basel*, pages 245–256. Verlag, 1997.
17. Y. Zhang. A Method for Continuous Extraction of Multispectrally Classified Urban Rivers. *Photogrammetric Engineering and Remote Sensing*, 66(8):991–1000, August 2000.