ELSEVIER

Computers & Geosciences 32 (2006) 615–623

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

An efficient depression processing algorithm for hydrologic analysis

Qing Zhu, Yixiang Tian*, Jie Zhao

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China

Received 18 October 2004; received in revised form 7 September 2005; accepted 7 September 2005

Abstract

Depression filling and direction assignment over flat areas are critical issues in hydrologic analysis. This paper proposes an efficient approach for the treatment of depressions and flat areas, based on gridded digital elevation models. Being different from the traditional raster neighborhood process which is time consuming, a hybrid method of vector and raster manipulation is designed for depression filling, followed by a neighbor-grouping scan method to assign the flow direction over flat areas. The results from intensive experiments show that there is a linear relationship between time efficiency and data volume, and the extracted hydrologic structures of flat areas are also more reasonable than those proposed by the existing methods.

Keywords: Digital elevation models; Depression; Flat areas; Hydrologic structure; Hydrologic analysis

1. Introduction

Because it is a fundamental problem in digital terrain analysis, the extraction of hydrologic structures plays an important role in applications such as hydrologic analysis, mineral deposition, land erosion, pollution diffusion analysis, etc. (Wolock and McCabe, 1995; Chen, 1991; Freeman, 1991; Moore et al., 1994; Li et al., 2004). Ridges and valleys are the basic features in hydrologic structure information. The most popular application extracts them from gridded digital elevation models (DEMs) and almost all the methods are based on the flow routing model (O'Callaghan and Mark, 1984; Jenson and Domingue, 1988; Tarboton et al., 1991; Moore et al., 1994). In such a model, the main task is to derive three matrices from the original DEM: the depressionless elevation matrix, the flow direction matrix and the flow accumulation matrix.

Depressions and flat areas are common in gridded DEMs; most of them are the result of mistakes, whereas some represent real terrain features, e.g., quarries and grottoes. The majority are spurious features, which arise from interpolation errors during DEM generation, truncation of interpolated values, and the limited spatial resolution of the DEM grid (Martz and Garbrecht, 1993). Depressions and flat areas must be dealt with as a precondition of flow route tracing, but the process is time consuming. So far, a number of methods have been developed for handling the depressions and flat areas of DEMs. Band (1986) simply

*Corresponding author. Tel.: +86 27687780928325; fax: +86 2768778969.

E-mail address: yixiangtian@yahoo.com (Y. Tian).

0098-3004/$ - see front matter
doi:10.1016/j.cageo.2005.09.001

increases the elevation of these cells until a downslope flow path to an adjacent cell becomes available. O'Callaghan and Mark (1984) attempted to treat them by smoothing the data. Jenson and Domingue (1988) and Martz and de Jong (1988) presented a method for filling depressions by increasing the elevation of cells in it to the elevation of the lowest overflow point on the depression boundary. However, these methods are effective only for the simplest cell, they change the nature of the terrain, and they may produce new depressions. Martz and Garbrecht (1995, 1998) proposed algorithms that considered both higher and lower terrain effects in dealing with depressions and flat areas. Thus, they produce more realistic results in applications. However, they still consider each depression separately and thus recursive detecting and filling processes may be required. The inherent problems of the efficiency and accuracy of these approaches have hindered their application in the processing of large-scale DEMs. Aiming at solving these problems, a hybrid method for depression filling using both vector and raster processes is proposed in this paper.

A raster process does not record the relationship of two objects (such as the depression and flat area) directly and searches its adjacent objects only through four or eight neighbors. On the other hand, a vector process considers the vector characteristics of the objects and treats an object as a whole. The hybrid method will be discussed in detail in the next section. Section 3 deals with the direction of flow over flat areas, assigned by applying a neighbor grouping scan method. Intensive experimental analysis is illustrated in Section 4. Finally, a few concluding remarks are given in Section 5.

2. Hybrid method using vector and raster processes for depression filling

2.1. Detecting depressions

There are three kinds of depression in gridded DEMs: single-point depressions, stand-alone depressions and compound depressions. The compound depressions include all complex topographic situations, such as looping depressions (adjacent depressions flowing into each other), depressions within flat areas, and truncation of depressions and flat areas at the edge of the DEM. Compound depressions have been recognized as one of the chief obstacles in the extraction of hydrologic structures (Jenson and Domingue, 1988; Freeman, 1991; Tarboton et al., 1991; Moore et al., 1994).

In this paper, the lowest cells of each depression (a cell with an elevation lower than all of its eight neighbors) are marked out while calculating the initial flow direction matrix. Starting from these bottom cells, the procedures to detect and fill the depression are carried out by combining vector processes with traditional neighborhood raster processes. Compound depressions have more than one bottom cell and are usually caused by looping depressions.

The single-point depressions can be filled by simply raising the elevation of each bottom cell to the lowest value of its neighbors. After such a step, the steepest descent value of all cells is not less than zero. Since only the cells with a steepest descent value equal to zero can form a mutual-pointing phenomena (adjacent cells with the steepest descent directions point to each other), and mutual pointing can serve as evidence of the existence of a depression, these cells then can be used as clues for detecting depressions. Here, we adopt a stack-based seed-filling algorithm. A flag matrix (a flow accumulation matrix can be utilized for the present) is used to mark the detected depressions. The basic workflow includes the following steps (Fig. 1):

Step 1: An unmarked cell with a steepest descent value equal to zero (used as seed) is pushed into the stack, and its corresponding cell in the flag matrix is marked.

Step 2: If the stack is not empty, a cell at the stack top is taken as the current cell and removed from the stack. Then its corresponding cell in the flag matrix is marked.

Step 3: The cell's eight neighbors are scanned in sequence; if a neighbor's flow direction points to it and the corresponding flag cell of this neighbor is not yet marked, this neighbor is pushed into the stack.

Step 4: Steps (2) and (3) are executed until the stack is detected as being empty at the beginning of step (2).

Step 5: Return to step (1) and deal with another depression until no such seed points remain.

After undergoing such processing, all depressions can be detected. In order to explain our method more clearly, the definitions of some words concerning the vector characteristics of the depressions are now given.

The boundary cells of the depression, which are arranged in a clockwise sequence, compose its rim

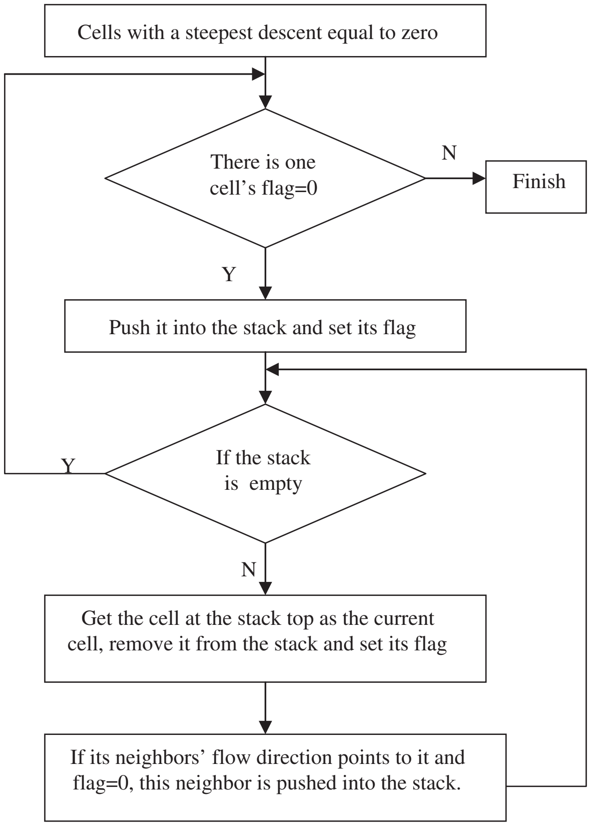

Fig. 1. Workflow of detecting depressions.

and the first point of the rim is its left-top cell. The minimum outer-enclosed rectangle is the minimum rectangular area in which the depression is located and can be determined while detecting depressions. Depressions' relations refer to whether the water of the depression flows to another depression or not, which is determined by its outlet flow direction. A data structure is defined to store the information on the depression, after which the following filling process can be carried out. The C++ language description of the structure is presented below:

```
class CDepression
{
  int nID;                     //Identification of depression
  int nPointedID;              //Identification of spilled depression
  CRect rectExterior;          //Depression's minimum outer-enclosed rectangle
  CpointArray pnts Outlet;     //Depression's flow outlets
  CPolygon* pVergin;           //Depression's rim
}
```

The *nID* is the unique identification of a depression in a flag matrix. The nPointedID of a depression, determined by its outlet flow direction, shows the relationship of adjacent depressions. If the outlet points to a cell belonging to another depression (the water from it flows into another), the nPointedID is the identification number of that depression, which combines the vector character with raster cells. Otherwise, the nPointedID is −1.

Searching along the upper side of the minimum outer-enclosed rectangle, a point in the depression's rim can be found directly. Starting from this point, the depression's rim can be traced out in the flag matrix according to the following steps:

Step 1: Starting from the known rim point, its neighbors are scanned in a clockwise direction until a new rim point is located.

Step 2: Starting from this new rim point, the current scan direction is rotated by 90° counterclockwise. Its neighbors are scanned clockwise until a new rim point is located.

Step 3: Step (2) is repeated until the first known rim point is located a second time.

Step 4: Because any point of the depression boundary may be its outlet, the omission problem should be dealt. When two adjacent rim points form a diagonal, the points of the other diagonal are detected. If there is one point belonging to the depression but is not included in the rim, it will insert into the rim. Any point omission problems arising from a concave corner in the rim are corrected (Fig. 2). Then a more compact and precise rim of the depression is obtained.

The rim of the depression is traced using a typical vector tracing method, which is similar to drawing a line along the depression boundary. By only considering the depression's rim, the lowest cell of the rim is then identified as the outlet of the depression. This method avoids the effects of any cells inside the depression that have lower elevation than that of the outlet. Thus, the problem of

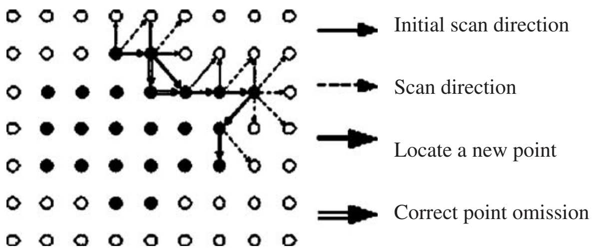

Fig. 2. Tracing rim of depression.

selecting wrong outlets encountered by Martz and de Jong (1988) will not arise here because the searching is based on the vector characteristic of the depression.

All of the outlets can be searched along the rim. Moreover, the outlets can be determined correctly, whether the depression is stand-alone or compound. If its outlet is located in another depression's rim, then it belongs to a compound depression. The filling of these depressions is discussed in the next section.

2.2. Filling depressions

The dominant depressions in a gridded DEM are compound depressions, and quite a lot of them are represented as looping ones (which frequently occur when many depressions are located on a relatively flat area), as shown in Fig. 3(a). In such a case, a new depression may emerge after the adjacent one has been filled; hence, the procedures for detecting and filling need to be performed recursively until there is no new depression. Herein lies the inefficiency of the existing approaches. In this paper, the depressions are filled based on their vector characteristics such as outer-enclosed rectangles, rims, flow outlets and the depressions' relations by adopting vector processes combined with traditional neighborhood raster processes. The filling procedure can be summarized as follows (Fig. 4):

Step 1: Based on the point relations recorded during the detecting of depressions, the looping depressions in compound depressions can be identified (depicted as depressions I, II and III in Fig. 3(a), where the outlets have the same elevation, so the relationship forms a loop).

Step 2: The looping depressions are merged in a new depression (depicted as depression V in Fig. 3(b)), the vector characteristics and point relations of which have been calculated. If the new depression is a stand-alone depression whose outflow does not spill into any other, its elevation will be raised to the value of the outlet to fill it, and step 4 executed. Otherwise, step 3 is executed (depicted as depressions IV and V in Fig. 3(b) that form a new loop).

Step 3: For the new depression, steps (1) and (2) are repeated until all depressions in the DEM have been handled.

Step 4: Another untreated compound depression is chosen and the procedure starts again at step (1).

Because the relationships of adjacent depressions are marked after they are detected, this method estimates whether a new depression would be formed when existing depressions are merged before filling each depression separately, which avoids recursively detecting the whole area and filling process. This is quite different from existing methods, and makes it efficient and effective.

3. Assigning flow direction over flat areas based on a neighbor-group scan

Assigning the direction of flow over flat areas is another stubborn problem in the computation of the flow accumulation matrix from gridded DEMs. The reasons why flat surfaces can arise in a grid DEM include sampling precision, data interpolation, some real terrain features, and the result of filling depressions (O'Callaghan and Mark, 1984; Freeman, 1991). To date, many methods of assigning the direction of the flow over flat areas have been devised. The main disadvantages of existing methods include ineffective parallel flow and saw tooth phenomena, which are usually caused by irrational increases in the elevation of cells (Jenson and Domingue, 1988; Martz and de Jong, 1988; Freeman, 1991; Tribe, 1992; Moore et al., 1994; Garbrecht and Martz, 1997).

Here, as in the detection of depressions, the vector characteristics of a flat area (minimum

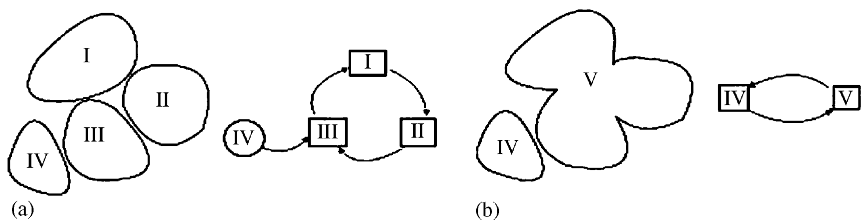

Fig. 3. Processing compound depressions one time (arrow—flow direction of outlet). (a) Compound depressions and their relations. (b) Compound depressions and their relations after filling one time.

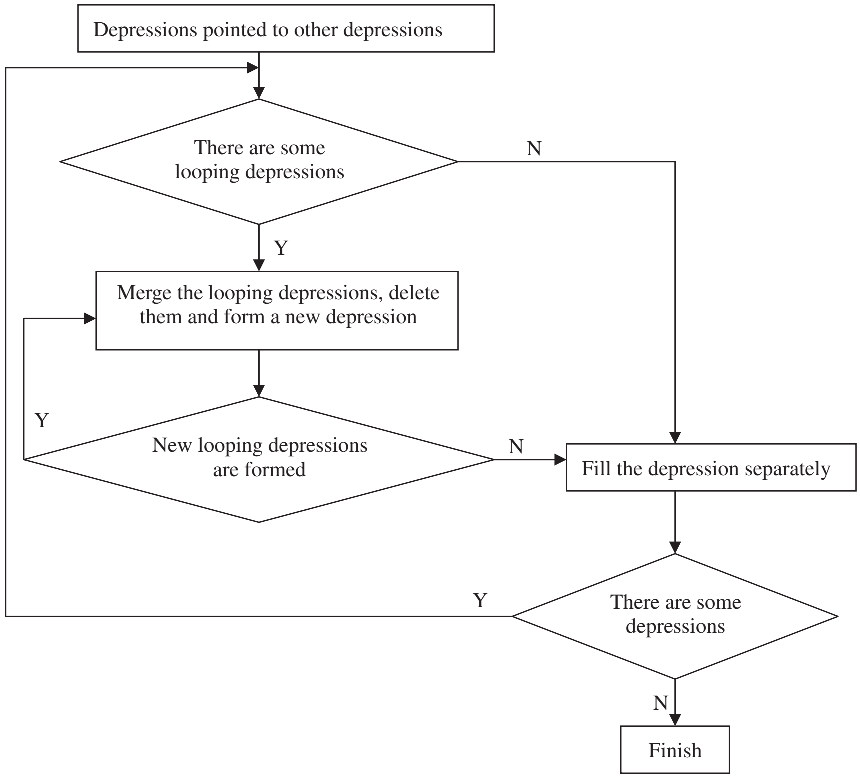

Fig. 4. Workflow of filling depressions.

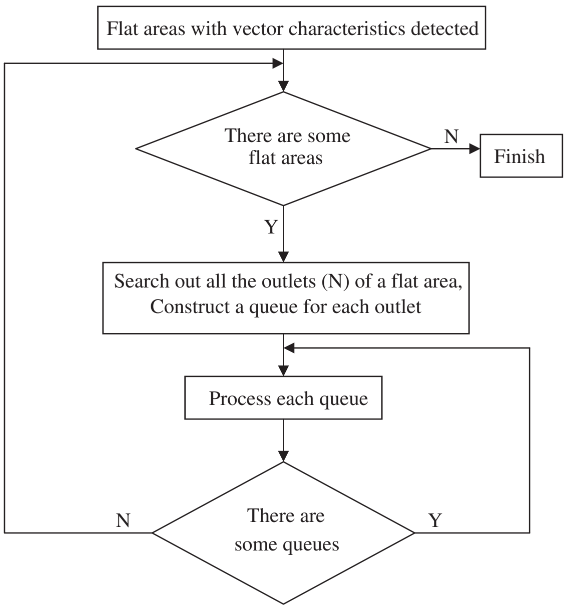

Fig. 5. Workflow of assigning flow direction over flat areas.

outer-enclosed rectangle and rim) are also calculated. Fig. 5 is the workflow of assigning flow direction over flat areas. Fig. 6 shows the process of assigning flow direction over one flat area. All of the outlets, which are the lowest ones of the boundary, can be searched out along the rim (see Fig. 6(a)). As the basic method, starting from the outlets, the flow direction of their neighbor cells is assigned first. Then, the flow direction of the cells adjacent to these cells is assigned. From such an iterative processing, all the cells in the flat area can be dealt. The point to mention is that the elevation of inside cells of the flat area is not increased so as to avoid irrational elevation increases and the cells contributing water to every outlet are searched collaterally. It makes sure every cell contributes water to the nearest outlet, which helps to eliminate parallel flow.

To avoid the parallel flows and saw tooth phenomena caused by the fixed scanning direction, the direction of the flow is assigned over the flat area by adopting a neighbor-grouping scan method, which is the most important difference between the proposed method and existing methods: the neighbor cells without flow directions are grouped according to their locality (cardinal and diagonal), and scanned in sequence (cardinal prior to diagonal) to assign flow directions, and the scan sequence within the same group is determined randomly. In

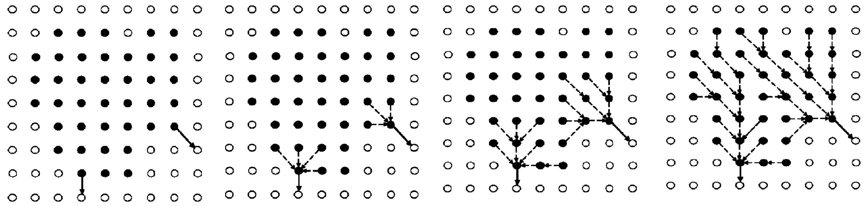

Fig. 6. Assigning direction of flow over one flat area. (a) Flat and its outlets. (b) Execute Step 2 one time. (c) Execute Step 2 several times. (d) Final flow direction over flat areas.

Table 1
Basic descriptive information of various sized DEMs

Number of rows and columns in DEMs	250 × 250	500 × 500	1000 × 1000	1500 × 1500	2000 × 2000	2500 × 2500	3000 × 3000
Min elevation (m)	595.4	602.6	1345.0	1076.7	1319.5	1256.2	1162.9
Max elevation (m)	721.8	792.4	1640.2	1580.2	1640.2	1619.7	1619.7
Mean elevation (m)	634.1	690.3	1484.1	1236.2	1458.5	1395.9	1423.2
Standard deviation	27.9	41.6	76.2	82.8	73.9	93.4	107.3

other words, it partitions the neighbors into two groups. One group contains the cells of the horizontal or vertical directions, the other contains the cells of diagonal directions. Different from the fixed scanning direction, the cells in the first group are scanned prior to those in the second one, since the cardinal cells are nearer to the center cell than the diagonal ones. And the scanning order of the cells in each group is random.

Assigning flow direction for one flat area is as follows:

Step 1: Supposing there are N outlets, N first-in-first-out queues are constructed, and each outlet is put in a separate queue. The N queues are sorted by the steepest descent of the outlet placed in them.

Step 2: If one queue is empty, it is deleted and the next queue is processed. Otherwise, the head of the queue is taken as the current cell and removed from the queue. Then its eight neighbors (cardinal ones prior to diagonal ones) are scanned. If one of the neighbor is in the flat area and has not been assigned a flow direction, then its direction is set to the current cell (depicted in Fig. 6(b) and (c)), and this neighbor is added to the tail of the queue.

Step 3: For each queue, step (2) is executed once until all the queues are deleted.

Thus, the approach can handle flat surfaces with multiple outlets and avoids parallel flows and saw tooth phenomena caused by a fixed scanning direction.

4. Experimental analysis

The approach proposed above was implemented using Microsoft Visual C++ and tested with real DEMs of various sizes. A comparison of the results was carried out between the proposed method and one of the existing methods (the Spatial Analysis Tools of ArcGIS V9.0 developed by ESRI Inc.). The main purpose of the experiment was to investigate the efficiency and accuracy of the approach. The data sets used in the experiment were the square grid DEMs (with a 5 m grid cell size) obtained from aerial photogrammetry using a digital photogrammetric workstation. A detailed description of the various DEMs is given in Table 1. Most of these DEMs are of mountainous regions. There are also a few rivers along which are scattered some large flood plains.

The test environment was as follows: the CPU P4 at 2.4 GHz, 512 M of Memory, and the operating system was Windows 2000 Professional. Hydrologic structure extraction is a time-consuming process in hydrologic analysis and thus the algorithm efficiency, especially when dealing with large-scale DEMs, is the most significant evaluating factor. Table 2 shows that the proposed approach has a significantly improved level of efficiency compared to existing approaches and the time required for

Table 2
Comparison of computing times for extraction of valleys

Number of rows and columns in DEMs	250 × 250	500 × 500	1000 × 1000	1500 × 1500	2000 × 2000	2500 × 2500	3000 × 3000
Number of depressions and flats	96	176	796	1490	3264	5348	6692
CPU time of proposed approach (s)	1	2	4	8	18	37	62
CPU time of ArcGIS (s)	11	13	28	53	96	121	167

Table 3
Computing time required for extraction of valleys from various depressionless DEMs

Number of rows and columns in DEMs	250 × 250	500 × 500	1000 × 1000	1500 × 1500	2000 × 2000	2500 × 2500	3000 × 3000
CPU time (s)	1	1	2	2	5	11	18

extraction increases in a roughly linear manner when applied on DEMs of various sizes.

In order to investigate the efficiency of the proposed algorithm related to the data volume and number of depressions in a gridded DEM, a set of various depressionless DEMs and a set of size-equivalent DEMs with different numbers of depressions were used separately for the experiment. A comparison of the time taken to extract valley lines from various depressionless DEMs is shown in Table 3. These depressionless DEMs were generated by using the depression filling method proposed in this paper on the DEMs described in Table 1. The results show that the time required for extraction has a roughly linear dependency on the size of the DEM. Table 4 depicts the comparison of the time required to extract valleys from size-equivalent DEMs (1000 × 1000) with different numbers of depressions. The results show that the time needed to execute the extraction also has a roughly linear dependency on the number of depressions.

A comparison of the accuracy of the results was also carried out between the proposed method and ArcGIS V9.0. Fig. 7 shows the results derived using these two approaches. In mountainous regions, there is no apparent difference, but in the region dominated by flood plains (Fig. 8), the proposed approach has a more proper convergence of flow and results in many fewer parallel lines than with ArcGIS (see the flat regions marked by circles). The two methods also produce different network topology. Analyzing from the original DEM data value

Table 4
Time required to compute extraction of valleys from size-equivalent DEMs with different numbers of depressions

Number of depressions and flats	265	436	634	796	823	1230
CPU time (s)	1	2	3	4	4	5

and considering that water of a cell in the flat area flows to the nearest lower cell, the network topology produced by our method is more reasonable.

5. Conclusion

This paper presents a new method to fill depressions by adopting vector processes combined with traditional eight neighborhood raster processes and assigns the direction of the flow over flat surfaces by applying a neighbor-grouping scan method. The approach has been tested with real DEMs, and the results show that it is efficient and its accuracy is better than existing methods. It is therefore more appropriate for the processing of large-scale DEMs, which can be used for spatial decision-making with regard to large regional sustainable development projects, such as the planning of drainage areas, formulating responses to flood disasters, determining borderlines and other applications.

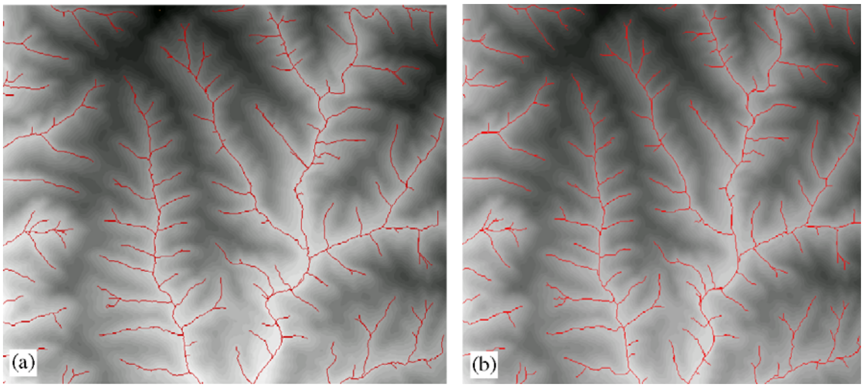

Fig. 7. Comparison of extraction on a mountainous region. (a) Extraction result of proposed method. (b) Extraction result of method implemented in ArcGIS.

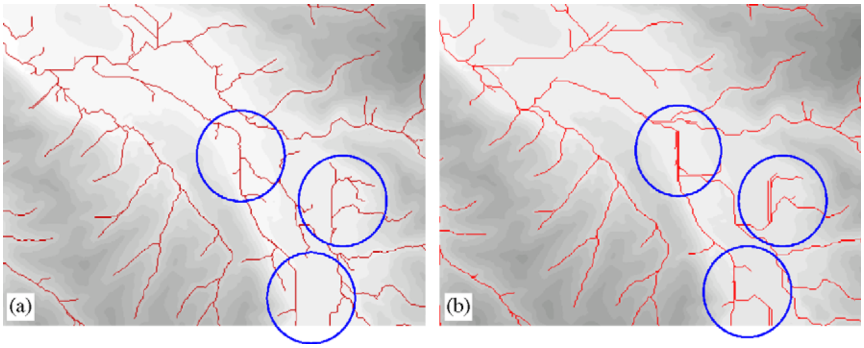

Fig. 8. Comparison of extraction on a plain region. (a) Extraction result of proposed method. (b) Extraction result of method implemented in ArcGIS.

Acknowledgments

This paper is supported by the National Basic Research Program of China (2004CB318206 and 2002CB312101), and the Hubei Outstanding Young Researchers Foundation of China (2004ABB018).

References

Band, L.E., 1986. Topographic partition of watersheds with digital elevation models. Water Research 22 (1), 15–24.

Chen, X., 1991. Mathematical Morphology and Image Analysis. Survey & Mapping Press, Beijing, China (in Chinese).

Freeman, T.G., 1991. Calculating catchment area with divergent flow based on a regular grid. Computers & Geosciences 17 (3), 413–422.

Garbrecht, J., Martz, L.W., 1997. Assignment of drainage direction over flat surfaces in raster digital elevation models. Journal of Hydrology 193 (1–4), 204–213.

Jenson, S.K., Domingue, J.O., 1988. Extraction topographic structure from digital elevation data for geographic information system analysis. Photogrammetric Engineering and Remote Sensing 54 (11), 1593–1600.

Li, Z., Zhu, Q., Gold, C., 2004. Digital Terrain Modeling: Principles and Methodology. CRC Press, Boca Raton, FL 323pp.

Martz, L.W., de Jong, E., 1988. CATCH: a FORTRAN program for measuring catchment area from digital elevation models. Computers & Geosciences 14 (5), 627–640.

Martz, L.W., Garbrecht, J., 1993. Automated extraction of drainage network and watershed data from digital elevation models. Water Resources Bulletin 39 (6), 901–908.

Martz, L.W., Garbrecht, J., 1995. Automated recognition of valley lines and drainage networks from grid digital elevation models: a review and a new method–comment. Journal of Hydrology 167 (5), 393–396.

Martz, L.W., Garbrecht, J., 1998. Treatment of flat areas and depressions in automated drainage analysis of raster digital elevation models. Hydrological Processes 12 (6), 843–855.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1994. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. In: Beven, K.J., Moore, I.D. (Eds.), Terrain Analysis and Distributed Modelling in Hydrology. Wiley, Chichester, UK, pp. 7–34.

O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Computer Vision, Graphics, and Image Processing 28 (4), 323–344.

Tarboton, D.G., Bras, R.L., Rodriquez, I.I., 1991. On the extraction of channel networks from digital elevation data. Hydrologic Processes 5 (1), 81–100.

Tribe, A., 1992. Automated recognition of valley lines and drainage networks from grid digital elevation models: a review and a new method. Journal of Hydrology 139 (3), 263–293.

Wolock, D.M., McCabe, G.J., 1995. Comparison of single and multi-flow direction algorithms for computing topographic parameters in TOPMODEL. Water Resources Research 31 (5), 1315–1324.