Research Article

Land Use Patch Generalization Based on Semantic Priority

Jun Yang,[1,2] Fanqiang Kong,[2] Jianchao Xi,[1] Quansheng Ge,[1] Xueming Li,[2] and Peng Xie[2]

[1] *Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China*
[2] *Liaoning Key Laboratory of Physical Geography and Geomatics, Dalian 116029, China*

Correspondence should be addressed to Fanqiang Kong; kongfankongfan@163.com

Land use patch generalization is the key technology to achieve multiscale representation. We research patches and achieve the following. (1) We establish a neighborhood analysis model by taking semantic similarity between features as the prerequisite and accounting for spatial topological relationships, retrieve the most neighboring patches of a feature using the model for data combination, and thus guarantee the area of various land types in patch combination. (2) We establish patch features using nodes at the intersection of separate feature buffers to fill the bridge area to achieve feature aggregation and effectively control nonbridge area deformation during feature aggregation. (3) We simplify the narrow zones by dividing them from the adjacent feature buffer area and then amalgamating them into the surrounding features. This effectively deletes narrow features and meets the area requirements, better generalizes land use features, and guarantees simple and attractive maps with appropriate loads. (4) We simplify the feature sidelines using the Douglas-Peucker algorithm to effectively eliminate nodes having little impact on overall shapes and characteristics. Here, we discuss the model and algorithm process in detail and provide experimental results of the actual data.

1. Introduction

Land use generalization is a complicated process involving complex spatial and semantic relationships between land use features, and thus it is very difficult to satisfy such conditions concurrently. A significant amount of research has been conducted in this area: for example, Zongbo [1] discussed proportion image generalization, purpose image generalization, and visual image generalization in image map compilation and elaborated the compilation process on the basis of practice; Chithambaram et al. [2] integrated the data based on extracting feature skeletons; that is, secondary patches were compressed into lines or points, secondary lines were compressed into points, and evaluations were given; Ai and Wu [3] conducted neighborhood analysis using the Delaunay triangulation network and carried out a consistency correction for the shared boundary of vector patches after simplification; Ai et al. [4] applied the Delaunay triangulation network executing neighborhood analysis to retrieve neighbor patches in patch aggregation and subdivided, merged, and simplified secondary patches by generating skeleton lines using the Delaunay triangulation

network; Harrie [5] established appropriate weights for various generalization constraints to solve the balance between constraining conditions and map qualification; Kulik et al. [6] proposed an ontology-oriented cartographic generalization and matched the appropriate needs for different users; Zhao et al. [7] studied the consistent update system of geospatial databases based on digital map generalization; Li et al. [8] and Huang et al. [9] discussed the area proportion of each patch after generalization and investigated patch boundary simplification, achieving constraints in the balanced area of various features in boundary simplification and attaining good adaptability; Stoter et al. [10] discussed the noncustomized automated cartographic generalization of commercial software, comprehensively considered the elevation results of man and machine, and revealed the possible differences; Qiao and Zhang [11] studied cartographic generalization in a distributed environment, which could be adapted to large quantity spatial data; Dilo et al. [12] proposed tGAP to achieve map generalization between two scales in a certain area, with large-scale maps used for generalization and small-scale maps used for constraint; Stanislawski [13] achieved automated generalization in U.S.

national hydrological datasets by deleting the corresponding features based on upstream drainage areas; Foerster et al. [14] studied the feasibility of geospatial data integration in a network service environment; Ai et al. [15] and Liu et al. [16], respectively, provided a detailed analysis and calculation models for the semantic similarity of land use data; Zhu et al. [17] applied a curve fit algorithm to line generalization and compared it with traditional algorithms.

The above studies comprehensively considered the semantic and spatial neighborhood of features when establishing an integrated model and obtained quantitative results through the corresponding weights of various parts. However, the requirements for total area of each land use type before and after land use generalization are strict, and the total area of each feature must fluctuate within a certain range. Thus, this paper prioritized the semantic neighborhood when establishing the model and took the spatial topology relationship as an auxiliary factor to determine final results relating to the same semantic neighborhoods and thus ensured the total area of each land use type optimally.

2. Analysis Model of Feature Neighborhood

2.1. Semantic Neighborhood of Features. Land use data is completely encompassed, seamless, and nonoverlapping in space, has hierarchical semantic divisions [18], and generalizes the feature set in the above premise. Land use data is divided into three-level land types as shown in Figure 1 (each layer is one level from top to bottom). Integration is difficult due to semantic diversity, so a clear generalization rule can only be developed after defining the relationship between semantics and determining the semantic neighborhood.

Land use data is often concerned with the total amount of first and second land use and is only interested in urban and rural construction land subclasses for third land use. Accordingly, this paper argues that semantic neighborhoods exist only among features at the same first land use type or that semantics are unrelated. We developed a land type sequence of semantic neighborhoods at the same level for each second and third land use. Taking arid land of the third land type, we first considered the lands with the same parent type and obtained the following sequence: arid land, irrigated land, and paddy field (see Figure 1). We then considered the relationship between the same first land types and arid land; that is, paddy field was followed by garden plot, woodland, grassland, raised path, irrigation and water conservancy land, agricultural land, and rural road (building land and other first land use types were not related to arid land semantics).

2.2. Definition of Feature Relationship in the Model. We supposed land use data as LandUseSet = $\{F1, F2, F3 \ldots Fn\}$, and SArea and DFeature respectively represented the minimum area of features in the map and the minimum distance between the features; land type name was represented by Land Name (Fn); the parent land type of feature land type (e.g., the parent land type of farmland and garden plot was agricultural land) was represented by Father[LandName (Fi)]; feature area was represented by Area (Fi); Dis (Fi, Fj)

represented the minimum distance between features Fi and Fj; the spatial topology relationship between features Fi and Fj was represented by TopoRel (Fi, Fj); the semantic similarity was represented by SemRel (Fi, Fj). The values of TopoRel (Fi, Fj) and SemRel (Fi, Fj) are as follows.

(1) The values of TopoRel (Fi, Fj) were −1, 0, and 1. We first determined ColLine (Fi, Fj) (whether two features are collinear), with spaces of features Fi and Fj being adjacent if they were collinear, and thus TopoRel (Fi, Fj) = 0; otherwise we determined the relationship between Dis (Fi, Fj) and DFeature; if Dis (Fi, Fj) < DFeature, the $F1$ and $F2$ spaces were adjacent, and TopoRel (Fi, Fj) = 1; otherwise TopoRel (Fi, Fj) = −1, and the $F1$ and $F2$ spaces were unrelated.

(2) The range of SemRel (Fi, Fj) was determined by the number of land types close to Fi. As mentioned before, there were 10 land types with similar semantics (including itself); when Fi was arid land, the values of SemRel (Fi, Fj) were $0, 1, 2 \ldots 9$ in order based on the semantic neighborhood of dry land; when the semantics of features Fi and Fj were unrelated, SemRel (Fi, Fj) = −1.

2.3. Model Rules. Land type area in each administrative region should be counted before and after land use integration, so the administrative region is an independent integrated unit. The following rules were formulated under this precondition. The secondary feature dataset FeaSet{Fi} (Area (Fi) < SArea) should be obtained before integration. According to 2.2, when TopoRel (Fi, Fj) = −1, no relationship existed between Fj and Fi due to the too long distance; when SemRel (Fi, Fj) = −1, the semantics of the two features were unrelated, so aggregation treatment cannot be conducted. The model process was as follows. Step 1: retrieve feature dataset FeaSet{} based on condition (1) SemRel (Fi, Fj) = 0 and TopoRel (Fi, Fj) = 0, and the feature that had the longest shared boundary with Feature Fi was the desired one in the dataset. For example, Feature $F7$ in Figure 2(a) was a secondary feature, the dataset meeting condition (1) should be FeaSet{$F3, F4$}, and the feature with the longest boundary with $F7$ was the desired one, which was the nearest feature in the dataset ($F4$). If the dataset meeting condition (1) was empty, Step 2 was conducted: retrieve feature dataset FeaSet{} based on condition (2) SemRel (Fi, Fj) = 0 and TopoRel (Fi, Fj) = 1; the feature with the largest area in the buffer of the DFeature radius of Feature Fj was the desired one. Taking $F7$ in Figure 2(b) as an example, when the dataset meeting condition (1) was empty, the dataset meeting condition (2) was FeaSet{$F4, F3$} and consisted of two features, and the buffer of the $F7$ Buffer ($F7$) was made by taking DFeature as the radius; attention should be paid to $F3$ and $F4$ in Buffer ($F7$), with $F3$ as being the desired feature because its area was larger than that of $F4$ in Buffer ($F7$). If the nearest feature was not retrieved after the aforementioned two steps, 1 was added to the value of SemRel (Fi, Fj) for recycling, until the most neighboring feature was retrieved. If the aforementioned features were not found when the maximum of SemRel (Fi, Fj) was achieved, Feature Fi was integrated into the neighboring feature with the largest area. For example, Feature $F7$ in Figure 2(c) was finally merged
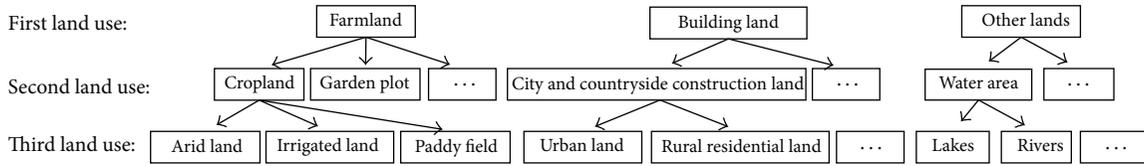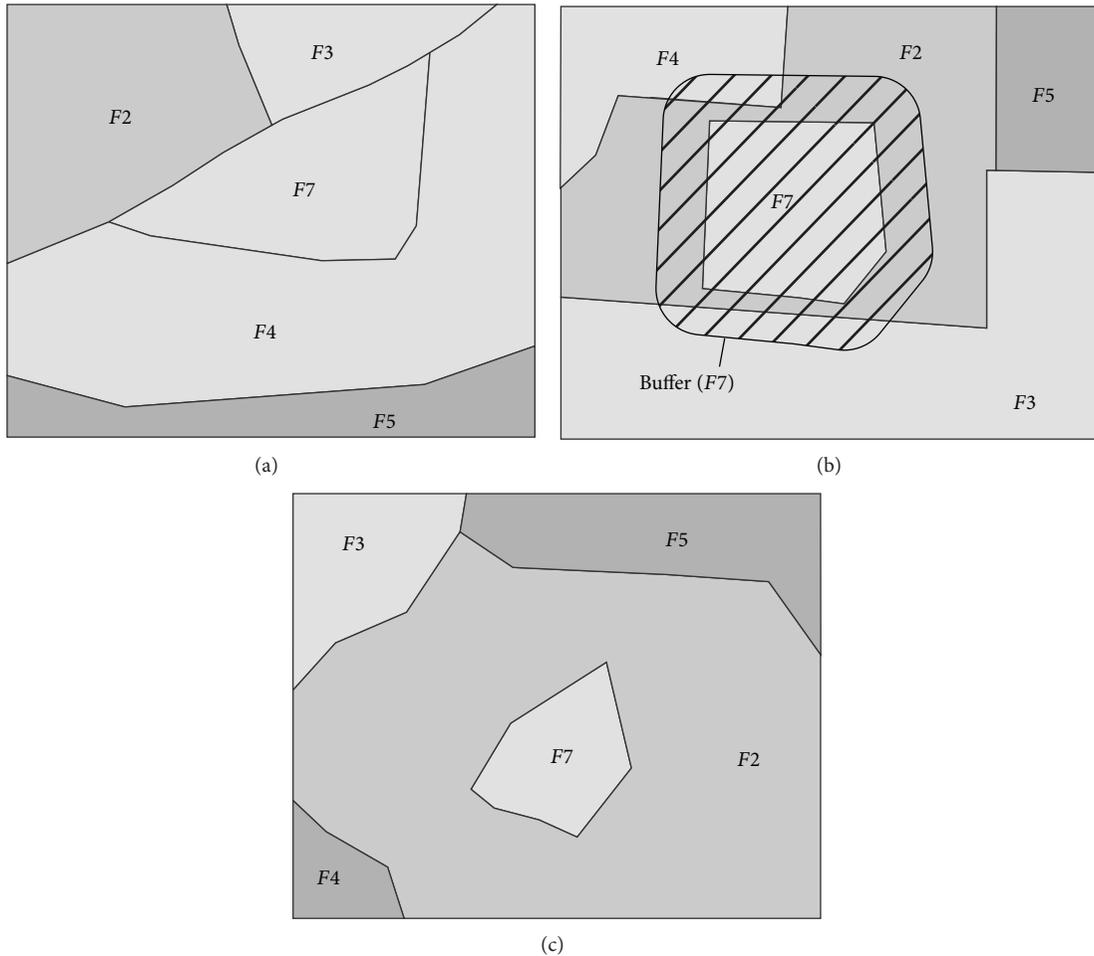
First land use:

| Farmland | Building land | Other lands |

Second land use:

| Cropland | Garden plot | $\cdots$ | City and countryside construction land | $\cdots$ | Water area | $\cdots$ |

Third land use:

| Arid land | Irrigated land | Paddy field | Urban land | Rural residential land | $\cdots$ | Lakes | Rivers | $\cdots$ |

FIGURE 1: Hierarchical tree of land use type.

(a)

(b)

(c)

FIGURE 2: Neighborhood degree.

into $F2$. In the previous process, if the Fi and Fj spaces were adjacent, the amalgamation method was taken; if the Fi and Fj spaces were neighboring, the aggregation method was taken. This model determines the nearest feature of secondary features by focusing on the semantic neighborhood of features with spatial topology relationships. This modeling process was simple and the changes in each land type area were minimized during integration, and the requirements of land use integration were met. The workflow of the neighborhood analysis model is shown in Figure 3.

3. Feature Processing Algorithms

3.1. Aggregation Processing. Feature aggregation is the merging of separate features in space, and it can prevent the same

type of features with short distance from being removed and avoid large changes in total land type area after integration [19]. The specific aggregation algorithm steps in buffer were as follows (taking $F1$ in Figure 4(a) as an example): (1) create the buffer of the $F1$ Buffer ($F1$) using DFeature (the minimum distance between features); (2) look for neighborhood patch $F2$ intersecting Buffer ($F1$); (3) create the buffer of the $F2$ Buffer ($F2$) by taking DFeature as the buffer radius, as shown in Figure 4(a); (4) calculate Buffer ($F1$) ∩ Buffer ($F2$) of the two buffers, and the buffer intersection of the two features (Figure 4(a)) was the grid region in the middle part; (5) calculate NodeSet{$N1, N2, N3 \ldots Ni$}, the node set of features $F1$ and $F2$ in the buffer intersection (black boundary in Figure 4(b)); (6) establishe patch Feature Fn using the nodes in the NodeSet, that is, the dark brown
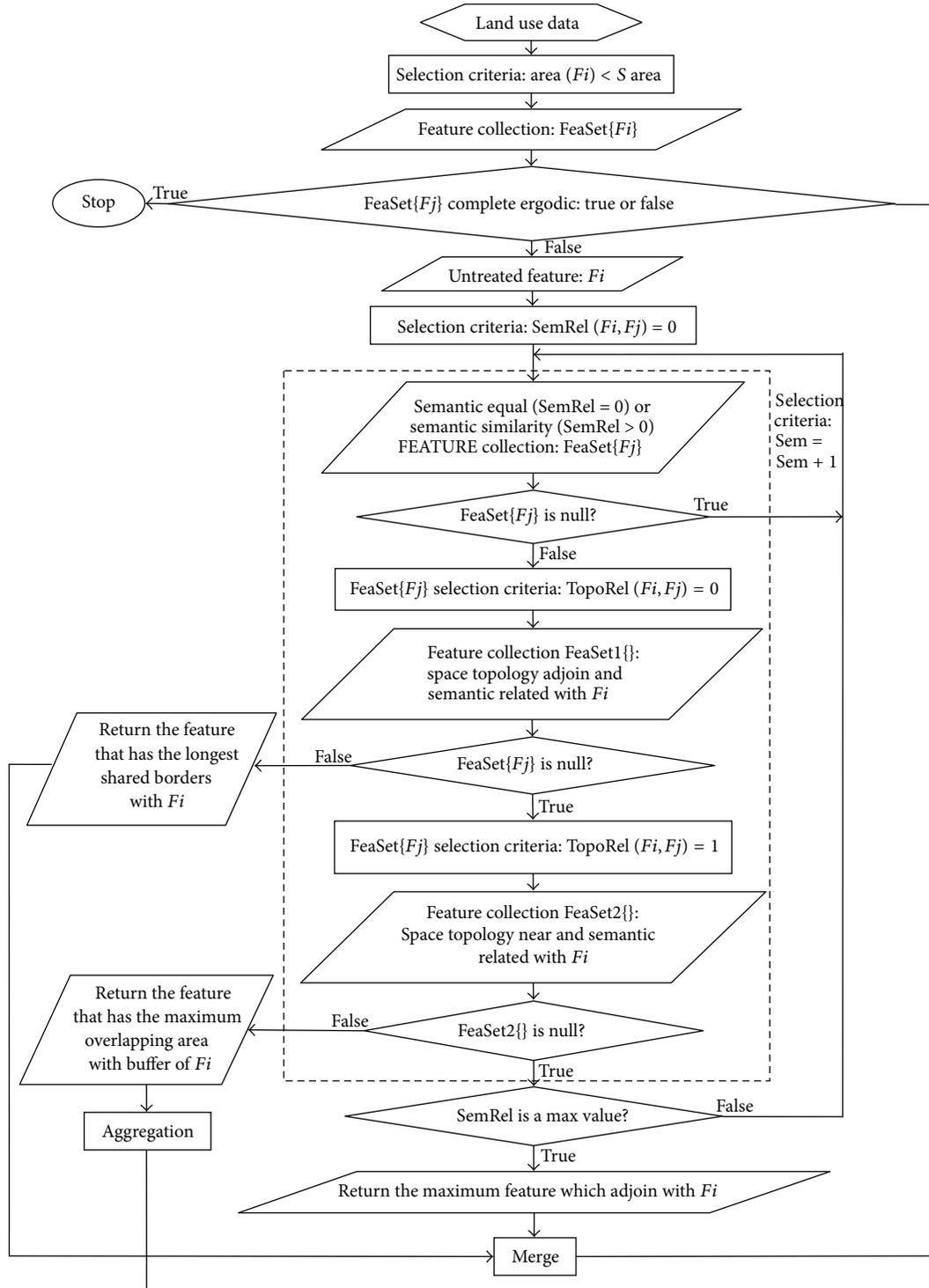
Figure 3: Model workflow.

region in the middle part of Figure 4(b); (7) merge $F1$, $F2$, and Fn to generate the feature after aggregation, as shown in Figure 4(c).

As seen from Figures 4(b) and 4(c), the feature using the buffer intersection nodes was the bridge area of separate features, which was effectively eliminated after the separate features were merged, effectively maintained the original shapes and characteristics of features, and met the requirements of integration. Attention should be paid to the feature overlapping when conducting aggregation processing by this method, and the bridge area can be directly excised for newly added features and overlapping in the bridge area.
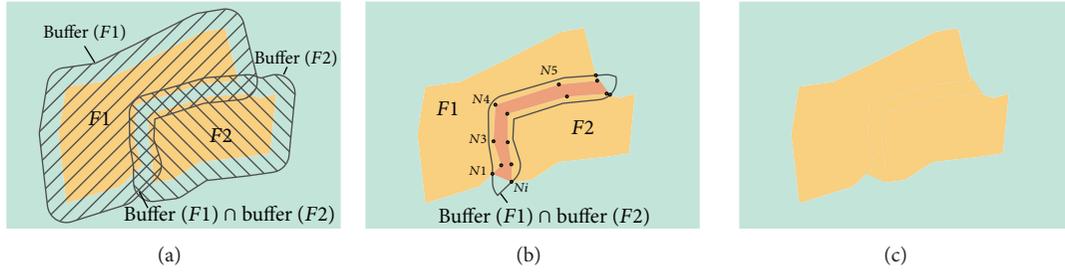
Figure 4: Process of aggregation.

3.2. Processing of Narrow Features. Narrow features in land use data mainly include railways, roads, rivers, and ditches. Simple integration or aggregation with the surrounding features is not enough because the data is long and narrow and the influence of the feature on the data cannot be eliminated by simple merger processing. We propose a buffer-based method to subdivide the narrow features according to semantic similarity and integrate the divided features into the surrounding features. The algorithm is simple and easy to implement with high efficiency.

Taking into account the semantic similarity of narrow features with spatial adjoining features at both sides, we first extracted the centerline of the narrow surface feature (such as the crimson line centerline in Figure 5(a)) and then divided the narrow feature River into upper and lower parts using the centerline (Upriver and Downriver in Figure 5(a)). Upriver was divided by $F1$, which adjoined it in space; although $F2$ and $F3$ adjoined River, in space they did not directly contact Upriver or participate in the division; while Downriver adjoined $F2$ and $F3$ in space, so it can be divided by Feature $F2$ and $F3$. We took the division of Downriver as an example to describe the processing steps of narrow features. (1) Establish buffer ($F2$) and buffer ($F3$), the buffers of features $F2$ and $F3$ adjoining Down_River in space (buffer distance was half the widest length of the narrow surface feature), and the buffers were overlapping, as shown in Figure 5(b). (2) Judge the features with neighboring semantics based on SemRel ($F2$, River), SemRel ($F3$, River), and the semantic similarity of $F2$ and $F3$ with the River. The semantics of $F2$ were more neighboring with those of the River. (3) Cut the buffers of the other features with the buffer of the feature that had neighboring semantics with the narrow feature; that is, cut buffer ($F3$) with buffer ($F2$), as shown in Figure 5(c). At this stage, there was no overlapping in the buffer. When $F2$ and $F3$ belonged to the same type, we cut the buffer with a small area with the one with a large area. (4) Divid Downriver with the buffer after processing. Downriver was divided into River 1 and River 2, as shown in Figure 5(d). (5) Respectively, merge River 1 and River 2 into the corresponding features and merge River 1 into $F2$ and River 2 into $F3$. The final processing results are shown in Figure 5(e). For land use integration, dimension-reduction treatment should be conducted for narrow surface features to compress the strip surface into the line feature with partial proportional scale. As for this example, the centerline extracted by strip feature could be used as its line feature, and

this line feature did not run through the strip feature, so the topological location of the feature was expressed clearly.

3.3. Sideline Simplification Algorithm. Line feature simplification algorithms consist of some classic algorithms, such as the Douglas-Peucker algorithm [9, 20], progressive approach simplification algorithm [21], oblique dividing curve algorithm [22], and Li-Openshaw algorithm. The Douglas-Peucker algorithm was used in this paper. Commonly used in global line simplification, this algorithm not only maintains the shape characteristics of vector lines but also determines the simplification tolerance based on mapping requirements and effectively removes nodes that have small influence on the overall shape of features. Its principle is to first connect two line endpoints into a straight line, measure the vertical distance from each node between the two endpoints to the straight line, remove all nodes between the two endpoints if the maximum distance is within the specified tolerance limit, make two straight lines, respectively, from the node to the two endpoints if the distance from a certain node to the straight line is greater than the tolerance limit, and then, respectively, compare them, until the line cannot be divided (see Figure 6).

When conducting sideline simplification for land use data, we note that consistent simplification should be conducted for important lines of administrative boundaries, roads, and rivers, and independent simplification should be avoided because it will result in inconsistent administrative boundaries or changes in topological relationships between rivers, roads, and other surrounding features.

4. Discussion and Conclusions

Data from the second national land survey of Longtou Sub-district of Dalian Lushun Port of Liaoning province was used in this study. We unified the land use type of the data into type division of Appendix B in the People's Republic of China land management industry standard TD/T 1027–2010 file (Figure 7(a)). The minimum patch area of research data is 400 m^2, and the scale is 1 : 10,000. According to the 1 : 10 land use data requirements of the 2006–2020 overall plan for land utilization, the minimum patch area of a map is 10,000 m^2, and 30 m is the furthest aggregation process distance. We used the previous algorithm to generalize the data the results of which are shown in Figure 7(b). The number of patches in the data decreased from 1007 to 428, and the compression
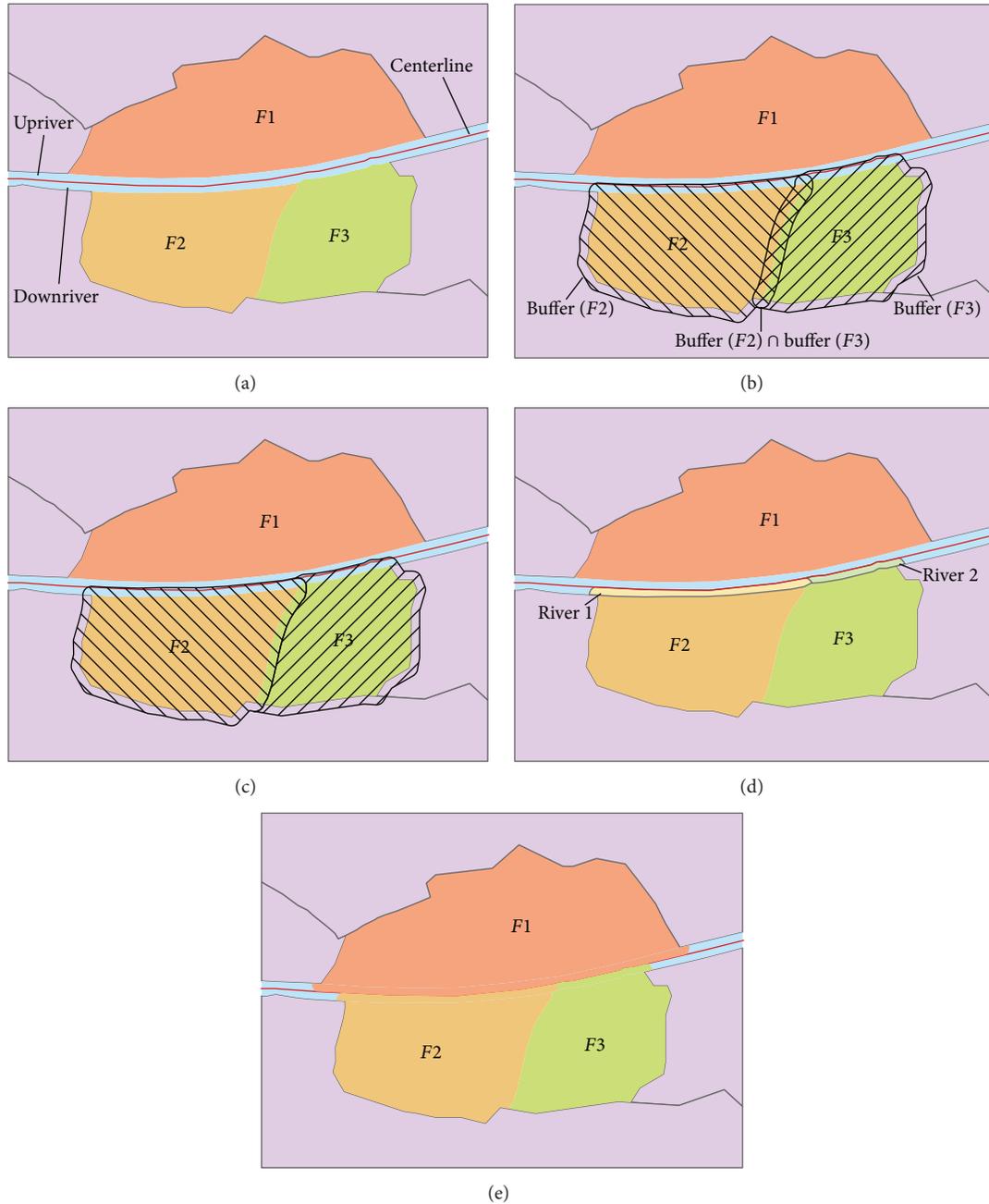
Figure 5: Process of handling strip feature.

ratio was 52.1%. The change rate of the total area of the important city and countryside construction land was 0.72%. The change rates of all land use types were less than 4%, except for agroland (19.36%) due to its scarcity and being highly dispersed. After generalization, some agro-land was integrated into other classes, and therefore changes in its area were larger than appropriate limits, which were considered as special circumstances. These generalization methods above must cause information loss as follows: the amalgamation of adjacent small area patches did not cause information loss;

the aggregation of separate small patches with neighboring semantics caused area information loss, but its attribute and location information was preserved. Long and narrow terrain was simplified into lines, which maintained information and resulted in a very little loss. The most serious loss of information comes from the merging of isolated patches into other land types. In conclusion, methods based on semantic priority maintained the general characteristics of the original data, and thus the change in total area of each land type was very small. Microelements and the narrow area were
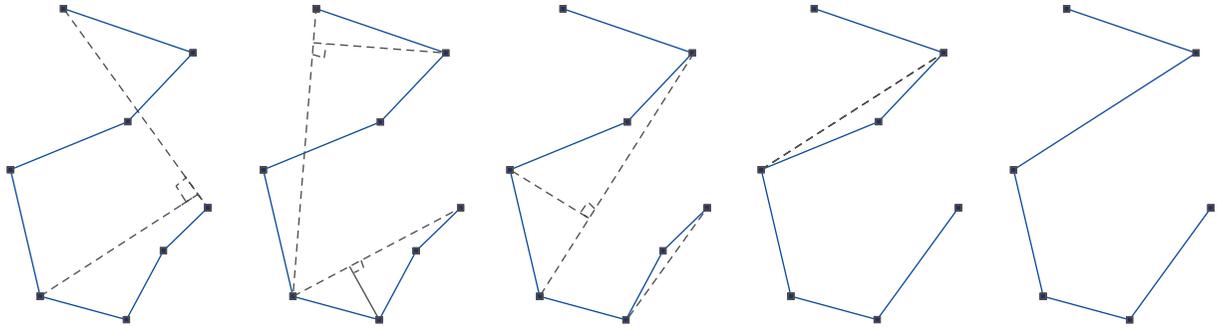
FIGURE 6: Douglas-Peucker algorithm procedure.
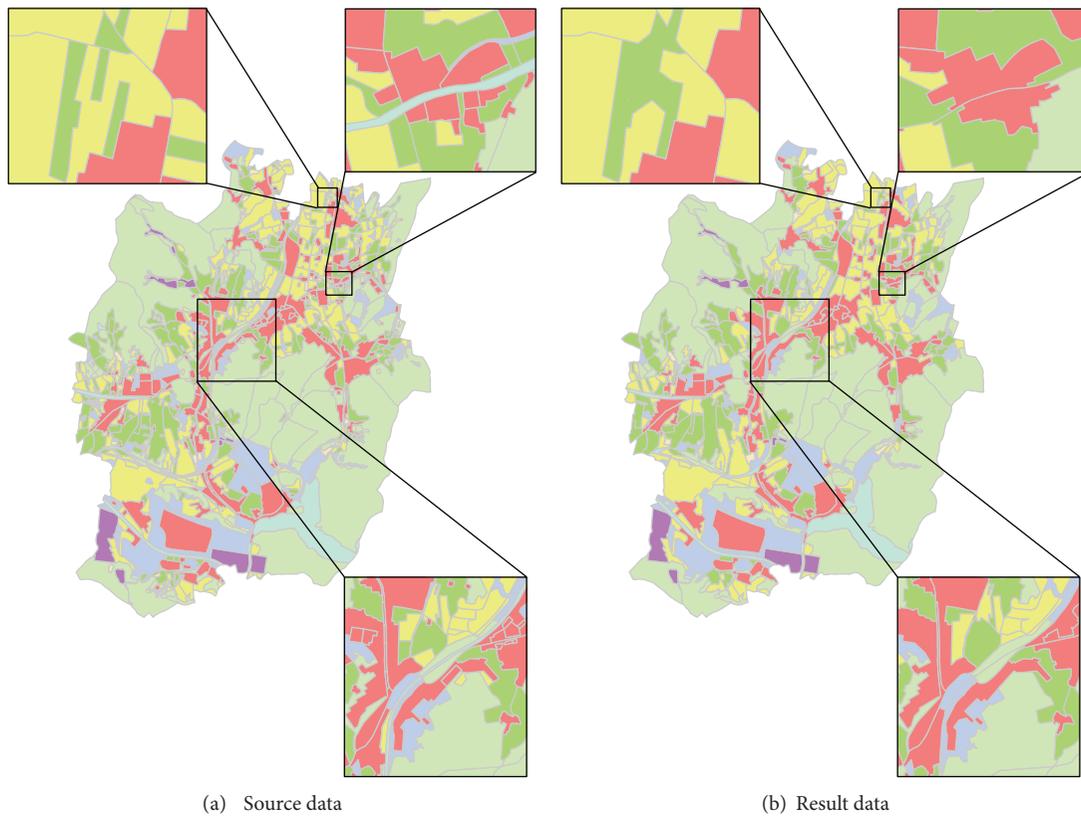

(a) Source data

(b) Result data

FIGURE 7: Comparison of generalization.

managed effectively and reasonably. In addition, the buffer algorithm was simple and fast. However, because the division was not smooth when dividing narrow features using the buffer (see Figure 5(c)), there was a small raised area where the narrow feature absorbed by Feature $F3$ contacted $F2$, which will be the focus of future research.
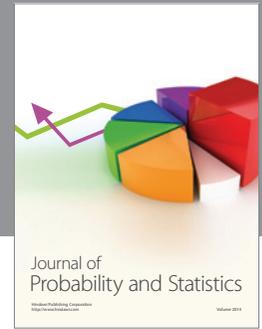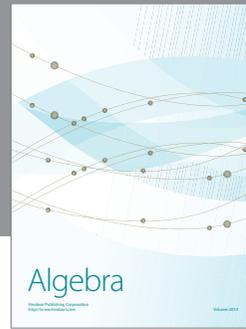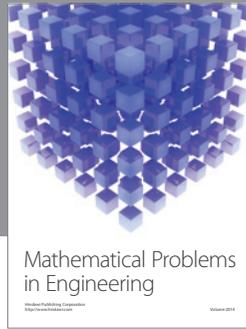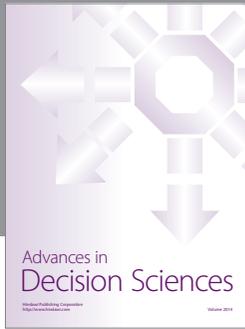
References

[1] Y. Zongbo, "On mapping integration and compiling technique of image maps," *Scientia Geographica Sinica*, vol. 8, no. 1, pp. 87–93, 1988.

[2] R. Chithambaram, K. Beard, and R. Barrera, "Skeletonizing polygons for map generalization," *Technical Papers ACSM. Baltimore*, vol. 2, pp. 44–55, 1991.

[3] T. Ai and H. Wu, "Consistency correction of shared boundary between adjacent polygons," *Geomatics and Information Science of Wuhan University*, no. 5, pp. 426–442, 2000.

[4] T. H. Ai, R. Z. Guo, and X. D. Chen, "Simplification and aggregation of polygon object supported by delaunay triangulation structure," *Journal of Image and Graphics*, no. 7, pp. 93–99, 2001.

[5] L. Harrie, "Weight-setting and quality assessment in simultaneous graphic generalization," *The Cartographic Journal*, vol. 40, no. 3, pp. 221–233, 2003.

[6] L. Kulik, M. Duckham, and M. Egenhofer, "Ontology-driven map generalization," *Journal of Visual Languages & Computing*, vol. 16, no. 3, pp. 245–267, 2005.

[7] R. Zhao, J. Chen, D. Wang, Y. Shang, and T. Ai, "The design and implementation of geo-spatial database updating system based on digital map generalization," in *2nd International Conference on Space Information Technology*, Proceedings of SPIE, Huazhong University of Science and Technology; The Second Academy of China Aerospace Science and Industry Corporation; The National Natural Science Foundation of China; Chinese Academy of Space Technology; China Aerospace Science and Industry Corporation, Wuhan, China, November 2007.

[8] J. Li, D. Zhu, X. Song, Y. Chen, and Y. Yang, "A Polygon simplification algorithm with area-balance consideration," *Geography and Geo-Information Science*, no. 1, pp. 103–106, 2009.

[9] W. Huang, W. Dai, and S. Yu, "Using modified Douglas-Peucher algorithm based on area preservation to simplify polygons," *Science Technology and Engineering*, no. 24, pp. 7325–7328, 2009.

[10] J. Stoter, D. Burghardt, C. Duchêne et al., "Methodology for evaluating automated map generalization in commercial software," *Computers, Environment and Urban Systems*, vol. 33, no. 5, pp. 311–324, 2009.

[11] Q. Qiao and T. Zhang, "Automated map generalization in distributed environments," in *Proceedings of International Joint Conference on Computational Sciences and Optimization (CSO '09)*, pp. 181–183, IEEE Computer Society, Hainan, China, April 2009.

[12] A. Dilo, P. van Oosterom, and A. Hofman, "Constrained tGAP for generalization between scales: the case of Dutch topographic data," *Computers, Environment and Urban Systems*, vol. 33, no. 5, pp. 388–402, 2009.

[13] L. V. Stanislawski, "Feature pruning by upstream drainage area to support automated generalization of the United States National Hydrography Dataset," *Computers, Environment and Urban Systems*, vol. 33, no. 5, pp. 325–333, 2009.

[14] T. Foerster, L. Lehto, T. Sarjakoski, L. T. Sarjakoski, and J. Stoter, "Map generalization and schema transformation of geospatial data combined in a Web Service context," *Computers, Environment and Urban Systems*, vol. 34, no. 1, pp. 79–88, 2010.

[15] T. Ai, F. Yang, and J. Li, "Land-use data generalization for the database construction of the second land resource survey," *Geomatics and Information Science of Wuhan University*, vol. 35, no. 8, pp. 887–891, 2010.

[16] Y. Liu, H. Li, and C. Yang, "Ontology based land use data generalization," *Geomatics and Information Science of Wuhan University*, vol. 35, no. 8, pp. 883–886, 2010.

[17] Y. Zhu, S. Zhou, and T. Lu, "Rearch on spatial data line generalization algorithm in map generalization," *Journal of Software*, vol. 6, no. 2, pp. 241–248, 2011.

[18] T. Ai and Y. Liu, "Aggregation and amalgamation in land-use data generalization," *Geomatics and Information Science of Wuhan University*, vol. 27, no. 5, p. 486, 2002.

[19] J. Weng, Q. Guo, X. Wang, and P. Liu, "An improved algorithm for combination of land-use data," *Geomatics and Information Science of Wunan University*, no. 9, pp. 1116–1118, 2012.

[20] X. Liu, S. Li, and W. Huang, "Study of Douglas-Peucker algorithm controlling by the goniometry in generalization," *Geomatics & Spatial Information Technology*, vol. 29, no. 1, pp. 59–60, 2006.

[21] Q. Guo, "Study on progressive approach to graphic generalization of linear feature," *Geomatics and Information Science of Wuhan University*, no. 1, pp. 54–58, 1998.

[22] H. Z. Qian, F. Wu, B. Chen, J. H. Zhang, and J. Y. Wang, "Simplifying line with oblique dividing curve method," *Acta Geodaetica et Cartographica Sinica*, vol. 36, no. 4, pp. 443–456, 2007.