

A Generalization Strategy for Discrete Area Feature by Using Stroke Grouping and Polarization Transportation Selection

Xiao Wang,^a Dirk Burghardt^a

^a *Dresden University of Technology, Institute of Cartography, Dresden, Sachsen, Germany; xiao.wang@tu-dresden.de, dirk.burghardt@tu-dresden.de*

Abstract: This paper presents a new strategy for the generalization of discrete area features by using stroke grouping method and polarization transportation selection. The mentioned stroke is constructed on derive of the refined proximity graph of area features, and the refinement is under the control of four constraints to meet different grouping requirements. The area features which belong to the same stroke are detected into the same group. The stroke-based strategy decomposes the generalization process into two sub-processes by judging whether the area features related to strokes or not. For the area features which belong to the same one stroke, they normally present a linear like pattern, and in order to preserve this kind of pattern, typification is chosen as the operator to implement the generalization work. For the remaining area features which are not related by strokes, they are still distributed randomly and discretely, and the selection is chosen to conduct the generalization operation. For the purpose of retaining their original distribution characteristic, a Polarization Transportation (PT) method is introduced to implement the selection operation. Buildings and lakes are selected as the representatives of artificial area feature and natural area feature respectively to take the experiments. The generalized results indicate that by adopting this proposed strategy, the original distribution characteristics of building and lake data can be preserved, and the visual perception is pre-served as before.

Keywords: Generalization, Area feature, Feature group, Stroke, Feature selection

1. Introduction

Since decades it has been a dream of cartographers to imitate the generalization ideas of human brains in computer environment for the derivation of various small-scale target maps or databases from a large-scale source map or database. Automated map generalization is a significant and complex process in the multiple representation of spatial data, which is helpful to reduce data production costs and to improve data maintenance as well as data production efficiency. Area feature is one of the most important features in map, which occupied large proportion of the map load. Due to the complexity of the spatial distribution of area features and for reasons of spatial recognition, area generalization has always been one of the difficult operations in automated map generalization. Regnauld (1996) pointed out that the goal of area feature generalization is to reduce the number of objects while pre-serving the original distribution character.

The objective of this paper is to generalize the discrete area features and preserve the original distribution. To meet this objective, a stroke based strategy is proposed. The overview of the strategy is described as followings: firstly the proximity graph of the discrete area features is generated by the constraint Delaunay triangulation. With refining the original proximity graph by the four constraints, which are location, size, shape and orientation respectively, the strokes are extracted from the refined proximity graph (which can be treated as a line network). Then the strokes are pruned by three rules. Afterwards the area features are divided into two

categories by judging whether they are related with strokes or not. For the stroke related area features, they normally present collinear and curvilinear patterns, while for the non-stroke related area features, they still present an irregular distribution. Therefore different generalization operations are taken into account for the two categories. For the regular stroke related area features, typification is chosen as the operator, while selection is used to generalize the irregular non-stroke related area features. A Polarization Transportation (PT) algorithm which has been used in point feature selection is introduced and modified to do the selection work.

The rest of this paper is organized as followings: Sect.2 briefly summarizes the current existed researches about area feature generalization, especially the grouping process of area features; Sect.3 gives the introduction of the proposed method with the detailed description of the concept and work procedures; then experiments based on the proposed method are shown in Sect.4; finally a conclusion of this paper is given in Sect.5.

2. Related works

The process of area feature generalization is normally separated into two steps (Li et al. 2004), namely the detection of area feature groups and the decision of generalization operator for the detected groups. The detected groups are the basis of generalization. For the group detection or pattern recognition of area features, many algorithms were put forward. Zhang et al. (2013) proposed a framework and several algorithms to recognize collinear and curvilinear building alignments by integrating computational geometry, graph-theoretic

concepts and visual perception theories Christophe and Ruas (2002) presented a method to both detect and characterize building alignments, especially straight line pattern. Regnauld (1996) processed visually identified building clusters together and decided which generalization operation is to be applied by analyzing and comparing these. Yan et al. (2008) adopted three principles of Gestalt theories and six parameters to automated building grouping and generalization. Li et al. (2004) used graph theory, Delaunay triangulation and Voronoi diagram to group buildings and then selected the appropriate operation to generalize the corresponding groups. Through summarizing the current researches briefly, there are still some aspects to be studied or improved for the generalization of area features:

- The current works focus more on the grouping process or the pattern detection, namely the groups are detected, but the further generalization process of the detected area feature groups still wait to be implemented. How to process the groups or which generalization operators should be appropriately chosen aiming at different groups should be studied deeply.
- There are more attention on the features which belong to a certain area group, and some algorithms have proposed to generalize them. However, for the rest of the non-grouped features, they still present the random or discrete distribution, how to design the generalization strategy is still a problem.
- The grouping and generalization process are mostly aiming at building generalization, namely the artificial area features, for the natural area features, such as lakes, islands or vegetation, there are less attention.

3. Methodology

Discrete area feature normally presents like linear patterns, such as collinear or curvilinear pattern, this pattern is the main structure of the original data, which could be perceived visually by map users. Therefore the generalization process must preserve this kind of pattern as before. In this paper, the linear patterns of the discrete area features are detected by stroke techniques. After pattern detection, the area features are divided into two categories, and a Polarization Transportation algorithm is introduced to select the remaining non-stroke related area features.

3.1 Area feature grouping by stroke

In road network generalization, stroke techniques are often adopted as the method of selection. The term "stroke" is prompted by the idea of a curvilinear segment that can be drawn in one smooth movement and without a dramatic change in style. The data of area feature generally presents the linear characteristic, therefore it is helpful to introduce the stroke techniques to detect the building groups. There are following steps to construct

stroke, here building data is adopted to illustrate the entire process.

3.1.1 Proximity graph of area feature

The construction of stroke is on the basis of proximity graph which is derived mostly by constrained Delaunay triangulation (CDT). In the proximity graph (Fig. 1), buildings are regarded as vertices and any two buildings that share at least one triangle are regarded as proximal, and an edge forms between the centroids of these two buildings. The detailed instruction and discussion of proximity graph can be referred in references (Zhang et al. 2010, 2013).

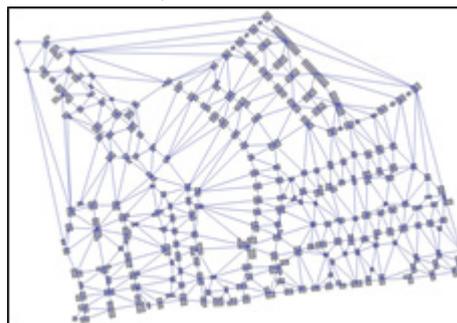


Fig. 1. Proximity graph of area feature

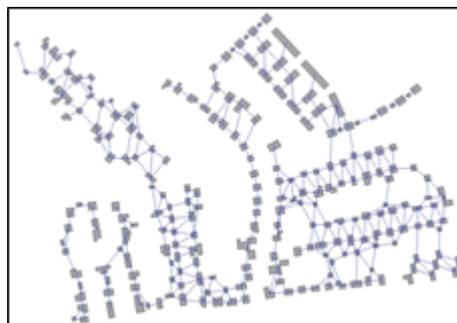


Fig. 2. Refinement of proximity graph

3.1.2 Proximity graph refinement

Proximity graph only reflects the topological proximity of buildings, even two distant buildings may have the proximity relationship, which has less benefits to the further analyzing. Therefore the original proximity graph should be refined. There are many different method to refine proximity graph, such as Nearest Neighbor Graph (NNG), Minimum Spanning Tree (MST), Relative Neighborhood Graph (RNG) (Anders 2003). Here the refinement of the original proximity graph is conducted by calculating the similarity of the two related buildings. The more similar this two buildings are, the more possible they belong to the same group. Four similarity parameters are adopted to measure the similarity of buildings. They are location similarity, size similarity, shape similarity and orientation similarity based on the Gestalt theory (Li 2004). If any one of these four similarities between two buildings is low, their adjacent relationship line will be deleted. By measuring these four similarity, the proximity can be well refined. Figure 2 is the refined proximity graph.

The advantage of this refinement method is that the grouping process can be controlled under different parameters. If it requires that the groups should consider more about the size similarity, the similarity of size must be set in a high threshold. So do the same with the other three parameters. By this way, the degree of grouping can be adjusted and controlled to meet different grouping requirements. For instance, there is one grouping situation which only considers distance between the features, so the thresholds of other three parameters should be set into zero. Or another grouping requires that the size and shape of features should also be considered as well as distance, so only the orientation threshold is set into zero. In a word, by this way the grouping process can be controlled under the four similarity parameters flexibly and is easy to meet different grouping requirements.

3.1.3 Constructing stroke of the refined proximity graph

The refined proximity graph can be regarded as a network, thus the frequently-used stroke technology in the road network generalization is came up with. Road network have a natural perceptual grouping characteristic, and "Good Continuation" is the dominant principle when judging the stroke (Thomson and Richardson 1999). By adopting this idea, the edges in the refined proximity graph can be also structured into strokes. Only those edges which satisfy the "Good Continuation" principle can be structured into the same stroke. In the road network, stroke is constructed by geometry information and attribute information. For the stroke construction of the adjacent lines in the proximity graph, here only the geometry information is considered owing to there are no attribute information for the proximity graph edges. Figure 3 presents the strokes of the refined proximity graph.

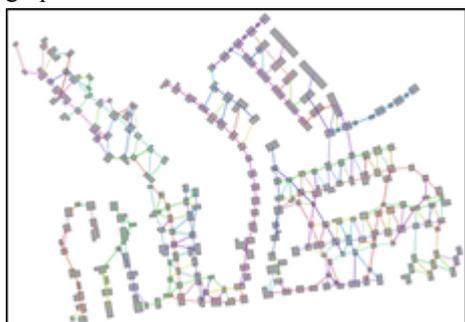


Fig. 3. Constructing strokes of the refined proximity graph

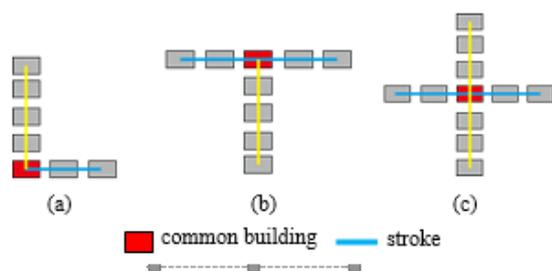


Fig. 4. Different positions of the common building

3.1.4 Stroke pruning

After constructing stroke, a general problem is that several buildings belong to more than one strokes, Figure 4 illustrates three typical position of the common building, namely the L type (Fig. 4-a), T type (Fig. 4-b) and cross type (Fig. 4-c). Therefore the aim of pruning stroke is to decide the belonging of the common feature which locate at the intersecting position of two intersecting strokes, finally getting the non-intersecting strokes. The pruning work should comply with the following three rules:

Rule 1: A building group should consist of at least three buildings, thus those strokes which only relate two buildings should be deleted firstly.

Rule 2: The common building belongs to the stroke which relates more buildings (Fig. 5-(a)). And after judging the common building, if the rest part of stroke who lost the common building relates less than three buildings, the rest part of stroke should be deleted also (Fig.5-(b)).

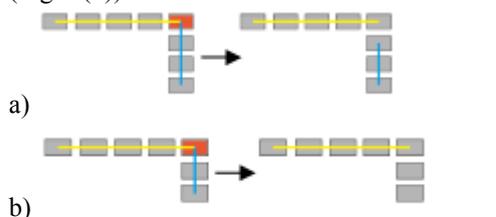


Fig. 5. Pruning rule 2

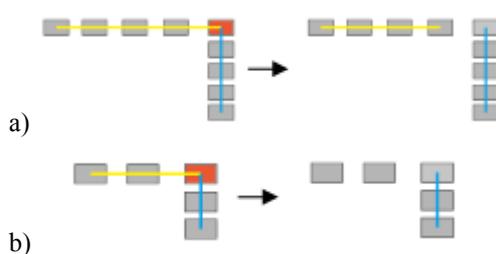


Fig. 6. Pruning rule 3

Rule 3: If two strokes relate the same number of buildings, the common building belongs to the stroke whose length is the shortest (Fig. 6-(a)). And after judging the common building, if the rest part of stroke who lost the common building relates less than three buildings, the rest part of the stroke should be also deleted (Fig. 6-(b)). Figure 7 displays the final pruning results of stroke.

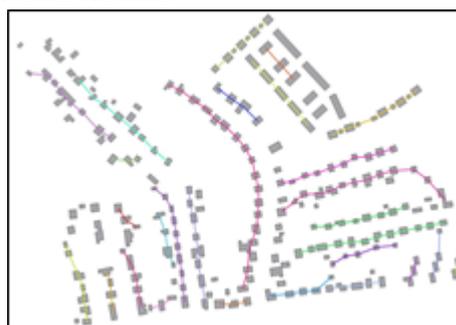


Fig. 7. Pruning results of strokes

3.2 Classification of generalization categories

From chapter 3.1, strokes are constructed, therefore the buildings are divided into two different categories, one category is stroke related buildings (Fig. 8-a), and another one is non-stroke related buildings (Fig. 8-b). Stroke related buildings are normally presented linear like pattern, while non-stroke related buildings are presented like irregular and random distribution. By analyzing different characteristics of these two categories, different generalization operators are adopted to conduct the further generalization process.

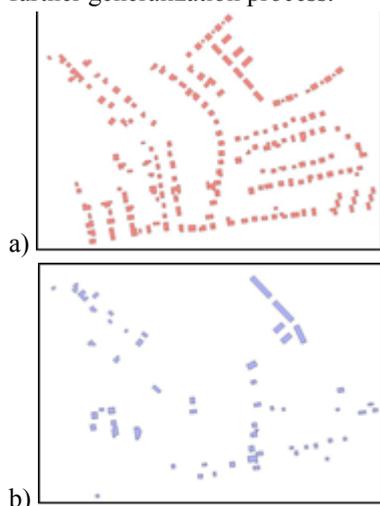


Fig. 8. (a) stroke related buildings and (b) non-stroke related buildings

For the stroke related buildings, they present the linear pattern, namely the regular pattern, so the typification is chosen as the generalization operator. Typification is conducted to each stroke related area feature group. For the non-stroke related area features, they have the random irregular distribution, therefore the selection operator is used for generalization. By considering the irregular distribution of the non-stroke related buildings, and in order to retain their original distribution density, a point selection algorithm named Polarization Transportation (PT) is introduced.

3.3 PT selection algorithm of non-stroke related area features

PT algorithm can preserve the density and distribution of point feature effectively after selection. The original PT algorithm is detailedly introduced and discussed in reference (Qian 2007). In summary, there are mainly five steps to use polarization transportation algorithm to implement the selection operation of point feature cluster in map generalization. Step 1: Determining the origin of the polar coordinate system. Step 2: Converting the coordinate of point features from rectangular coordinate system into polar coordinate system. Step 3: Unfolding the polarized point set by the relative polar angle ranging from 0° to 360° and plotting the sequence on an XY-plane. By connecting each point, a spectrum line is formed. Step 4: Segmentation of the spectrum line in polarization space by angle difference thresholds. Step 5: Simplifying the spectrum line by deleting nodes on the spectrum line

based on circle method which can preserving the local structure.

In some degree, area feature can be treated as point features, because the centroid of the area feature can represent the area feature. But the selection of area feature is also different with point. Point is one-dimensional feature, thus the selection should mainly consider the original distribution density. While for the area feature, it is two-dimensional feature, and its selection process should consider not only the distribution density but also its own size, namely the area. Therefore, the original point based PT algorithm may have some defectiveness when it deals with the areal feature directly. It is necessary to modify the original PT algorithm so that it can adapt to the demands of areal feature selection. By doing the first four steps of PT algorithm, the area features can be divided into different region by its polarization angles.

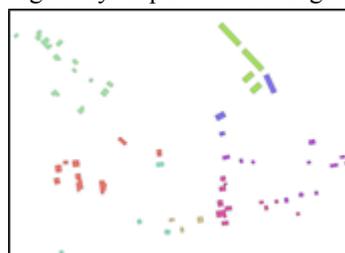


Fig. 9. Area feature categories by angle

After the clustering of area features by angle (Fig. 9), the last step should be to solve the selection of area feature. In the work of area feature selection, it should consider not only the polar coordinate position information, but also the area information of the area features. Normally, the larger an area feature is, the more possible this area feature should be retained after selection.

4. Experiment

Building and lake data are applied as the representatives of artificial and natural areal feature respectively. By these two different type data, the effectiveness and adaptation of the proposed method can be verified. A village around Dresden is selected as the building data, while the lakes around Lyon is selected as the lake data.

4.1 Experiment process and results

Figure 10 and Figure 11 display each step of the proposed generalization strategies for the building and lake data. Here the typification operation is implemented by using the WebGen service provided by the commission of generalization and multiple representation of ICA.

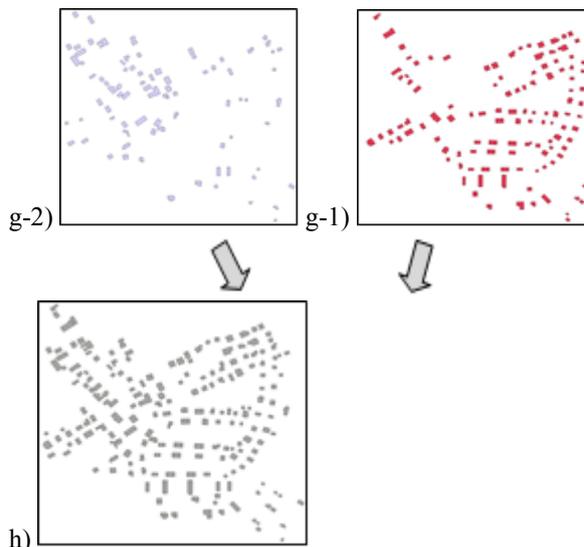
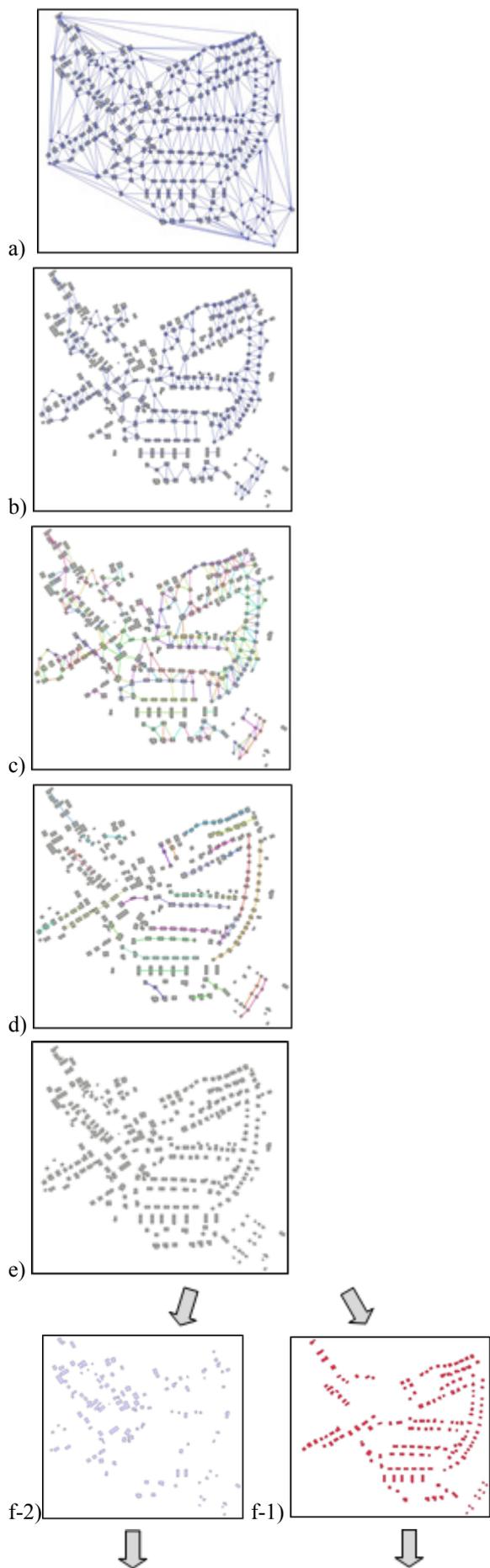
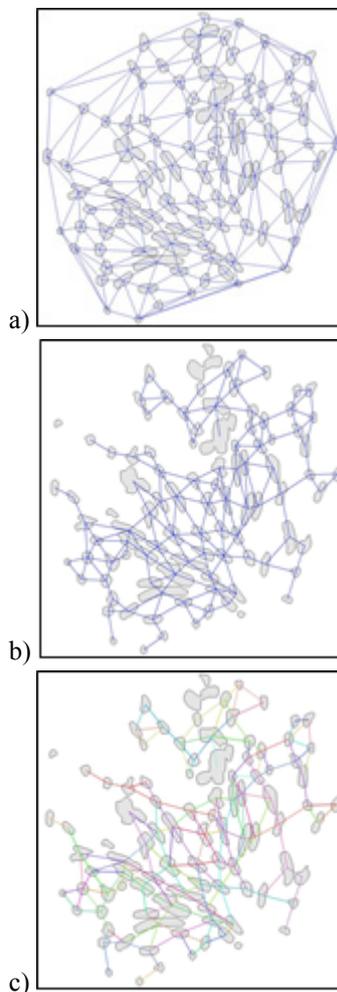


Fig. 10. Group detection and generalization process of building data: (a) proximity graph, (b) refined proximity graph, (c) strokes of refined proximity graph, (d) pruning results of strokes. Original data (e) is classified into two categories stroke related buildings (f-1) and non-stroke related building (f-2), (g-1) typification results of stroke related buildings and (g-2) selection results of non-stroke related buildings, (h) is the final generalized results.



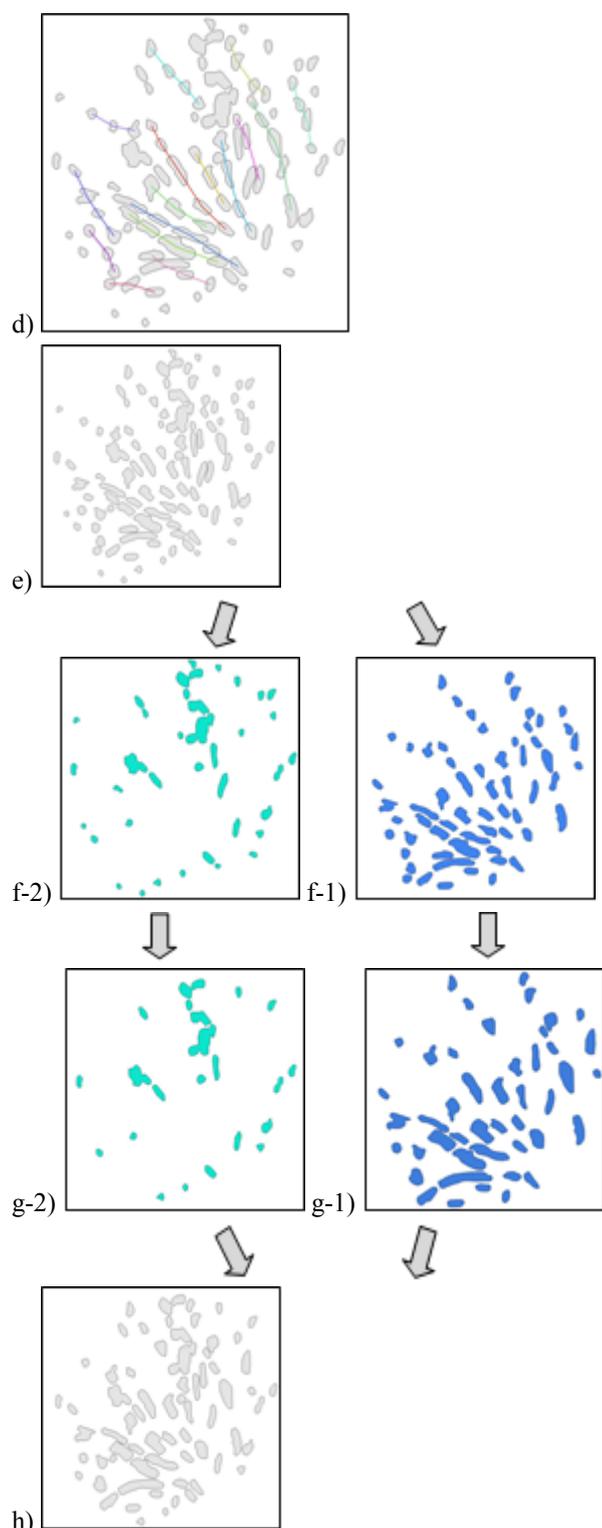


Fig. 11. Group detection and generalization process of lake data: (a) proximity graph, (b) refined proximity graph, (c) strokes of refined proximity graph, (d) pruning results of strokes. Original data (e) is classified into two categories stroke related lakes (f-1) and non-stroke related lakes (f-2), (g-1) typification results of stroke related lakes and (g-2) selection results of non-stroke related lake s, (h) is the final generalized results.

4.2 Discussion

The discussion consist of three parts, section 4.2.1 discusses the satisfied and unsatisfied parts of the

proposed stroke based grouping method and section 4.2.2 argues about the selection results by comparing with the selection method which only considers the area factor, and section 4.2.3 evaluates the proposed methods generally.

4.2.1 Discussion of the stroke based grouping method

From the results of the experiments in section 4.1, the linear like patterns are well detected by the introduced stroke based method, and the linear patterns of buildings and lakes construct the main structures of this region. For the typification results of the stroke related area features, the results preserves the original linear like patterns.

Meanwhile, there are also some objects get unsatisfied results. The limitation of the proposed stroke grouping method mainly reflects in two aspects. One is that the detection effectiveness of linear patterns is affected by the parameters in some degree. Currently the value of parameters is set mainly by the experience of experts, and if the parameter value is set not appropriately, the linear pattern detection process may get unsatisfied results. Another one is that the pruning process of the original strokes may cause two types unsatisfied results. One is that the more likeable linear patterns may be neglected by the current pruning rules, shown like examples A and B with the black dashes line in Figure 12. Visually, the black dashes line related buildings are more like a linear group than the detected ones. Another one is that the less likeable linear patterns may be wrongly detected, shown like example C in Figure 12, these buildings do not have obvious linear pattern, but they are detected as the linear pattern.

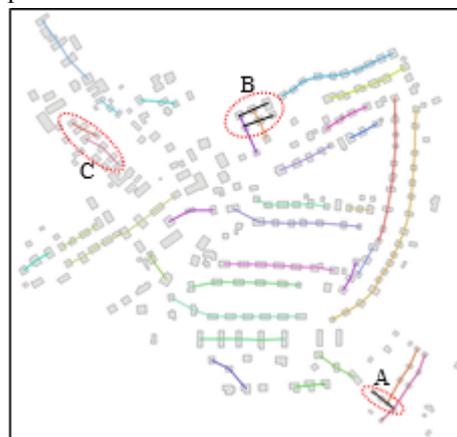


Fig. 12. Drawbacks of the stroke based grouping method

4.2.2 Comparison analysis of PT selection algorithm

In general, the selection method of area features considers only about the area of object, which may result in the density variation in some region, and here it is called as “area selection algorithm”. The proposed PT selection algorithm divides the region into different parts and the selection is conducted under these different region, which can well preserve the original density of data. Figure 13 describes the selection results of these two algorithms. By comparing with the original data distribution, the circled A, B, C region are the regions where the density changes significantly by using area selection algorithm (Fig. 13-

b), while the proposed PT algorithm do not appear this problems in these three regions, which changes more homogeneous. Therefore the original distribution density is preserved (Fig. 13-c).

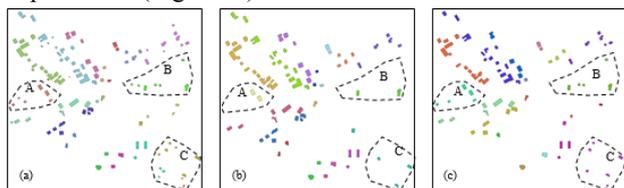


Fig. 13. Selection results of the proposed method (c) and area selection algorithm (b)

For the selection results of the non-stroke related area feature, by the PT selection algorithm, the selected results preserves the original distribution. Table 1 displays the statistic comparison of PT selection algorithm and area selection algorithm. The standard selection rate is set into 0.7. The PT Zone denotes the regions that are divided by the polarization angles of features. In this building experiment, 17 zones are divided in total. In each PT Zone, the selected objects number and selection rate are calculated. In order to reflect the change range of selection rate in each PT Zone, the absolute difference between real selection rate with standard selection rate (0.7) is also calculated.

Zone No.	Original No.	Selected No. (Area al.)	Select ion Rate (Area al.)	Differ ence with standa rd rate (Area al.)	Selec ted No. (PT al.)	Select ion Rate (PT al.)	Differ ence with standa rd rate (PT al.)
1	7	2	0.286	0.414	5	0.714	0.014
2	9	5	0.556	0.144	6	0.667	0.033
3	5	4	0.800	0.100	3	0.600	0.100
4	3	3	1.000	0.300	2	0.667	0.033
5	2	2	1.000	0.300	1	0.500	0.200
6	6	6	1.000	0.300	4	0.667	0.033
7	19	15	0.789	0.089	14	0.737	0.037
8	17	14	0.824	0.124	12	0.706	0.006
9	7	2	0.286	0.414	5	0.714	0.014
10	8	6	0.75	0.050	6	0.750	0.050
11	7	7	1.000	0.300	5	0.714	0.014
12	2	2	1.000	0.300	1	0.500	0.200
13	2	2	1.000	0.300	1	0.500	0.200
14	2	2	1.000	0.300	1	0.500	0.200
15	6	3	0.500	0.200	4	0.667	0.033
16	9	4	0.444	0.256	6	0.667	0.033
17	3	0	0.000	0.700	2	0.667	0.033
	114 (sum)	79(s um)	0.693(avg.)	0.270(avg.)	78(s um)	0.684(avg.)	0.073(avg.)

Table 1. Statistics comparison of PT selection algorithm and area selection algorithm

In table 1, it can be found that the selection rate of area algorithm in each PT Zone changes hugely and has very obvious difference with the standard rate, especially in Zone 17, the selection rate even reach 0.0, which means in this zone all the features are deleted and the original density is absolutely destroyed. On the contrary, the PT algorithm has a stable selection rate which is close to the standard rate in each zone, and the largest difference with standard rate is only 0.2, which can be explained by the less quantity of the original features in this zone (the number is 2). Figure 14 and Figure 15 are the curve graphs of the selection rate and selection rate difference in different zones which are generated from the above table. Figure 14 shows that the selection rate of PT algorithm is more stable than the area algorithm and the difference with standard selection rate of PT algorithm is also much smaller than the area algorithm, which means that the original distribution density is better preserved.

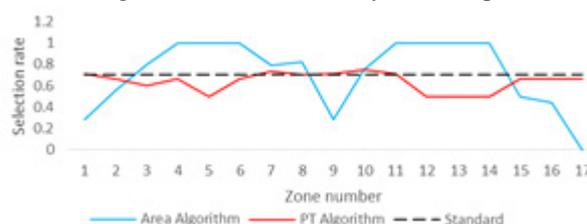


Fig. 14. Curve graph of selection rate in different PT zone



Fig. 15. Curve graph of selection rate difference in different PT zone

4.2.3 Evaluation of the proposed method

In general, by this new generalization strategy the original distribution character of the area features can be pre-served. The generalized results can well reflect the original characteristic of the data, the linear like patterns are pre-served and the rest of the non-linear pattern objects also retain their original density and distribution character. The introduced stroke techniques can distinguish area features in the dataset with different distribution characters, thereby different generalization operations are more targeted implement to the different areal features.

The stroke related area features present some certain patterns which construct the framework structure of the whole dataset, and this framework structure can be easily recognized or perceived visually by human. Thus after generalization, this kind of framework should also be recognized obviously in vision, and the targeted typification operation can preserve it. The non-stroke

related area features can be regarded as the supplement for the frame-work structure ones. They are distributed among the framework areal features randomly, but they have different density in different regions, which should also be retained. By introducing and modifying the polarization transformation method which is used in point feature selection, the density of the areal feature distribution can also be pre-served.

In a word, by this strategy the original distribution character of the area features can be preserved, and the generalized results can keep the framework structure and density of the original data. The proposed method are more effective on the data which have more linear like patterns, and for the data which present more like a cluster or grid like patterns, the method should be improved in order to detect better.

5. Conclusions

In this paper, a strategy for the generalization of discrete area feature has been presented. This strategy divides the area features into two categories by using stroke technology. The area features which are related by strokes are presented linear pattern and the area features which are not related by strokes presented a randomly pattern. By analyzing the character of these two different parts, different generalization operators are chosen for the mentioned two categories. Experiments are conducted aiming at two different types (artificial and natural) of area features to verify the effectiveness of the proposed method, and by analyzing the generalization results, the advantages of the proposed method and some limitations are discussed.

The future work should focuses on the evaluation of the generalization results. Up to now the generalization results is evaluated by the visual perception, which is very subjective and non-precisely, how to design a system to evaluate the generalization results should be considered. And for the area features which has linear patterns, how to control the degree of generalization is still to be researched, because the typification operator is chosen in this situation, and the aggregation may be also suitable in some other situations. Thus the generalization implement of linear patterns should still be paid more attention.

6. References

- Thompson R, Richardson D. (1999). The 'Good Continuity' principle of perceptual organization applied to the generalization of road networks. In: Proceedings of the 19th international cartographic conference, Ottawa, 1215-1225.
- Zhang X, Ai T, Stoter J. (2010). Characterization and detection of building patterns in cartographic data: two algorithms. In: Joint in-ternational conference on theory, data handling and modelling in geospatial information science (SDH' 2010), vol XXXVIII, part 2, pp 261–266.
- Zhang X, Ai T, Stoter J, Kraak M.J., Molenaar M. (2013). Building pattern recognition in topographic

data: examples on collinear and curvilinear alignments. *Geoinformatica*, 17:1-33.

Christophe S, Ruas A. (2002). Detecting building alignments for generalisation purposes. In: Richardson DE, van Oosterom P (eds) *Advances in spatial data handling (SDH 2002)*. Springer, Ottawa, pp 419–432.

Regnauld N. (1996). Recognition of Building Clustering for Generalization. In: Kraak MJ, Molenaar M (eds) *Advances in GIS research (Proceedings of 6th SDH'96, Delft)*. Taylor & Francis, London, pp 4B. 1-4B.14.

Yan, H., Weibel, R., Yang, B. (2008). A multi-parameter approach to automated building grouping and generalization. *GeoInformatica* 12, 73–89.

Li, Z., Yan, H., Ai, T., Chen, J. (2004). Automated building generalization based on urban morphology and gestalt theory. *International Journal of Geographical Information Science*, 18 (5), 513–534.

Qian, H.Z., and Meng, L.Q. (2007). Polarization transformation as an algorithm for automatic generalization and quality assessment. In *Proceedings of the Fifteenth International Conference on Geoinformatics*, Nanjing, China: 1–12.