

Article

Automated Residential Area Generalization: Combination of Knowledge-Based Framework and Similarity Measurement

Xiaorong Gao ^{1,2,3}, Haowen Yan ^{1,2,3,*}, Xiaomin Lu ^{1,2,3} and Pengbo Li ^{1,2,3}

¹ Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China; 0119058@stu.lzjtu.edu.cn (X.G.); xiaominlu08@mail.lzjtu.cn (X.L.); 13210073@stu.lzjtu.edu.cn (P.L.)

² National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China

³ Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China

* Correspondence: yanhw@mail.lzjtu.cn; Tel.: +86-1360-931-0452

Abstract: The major reason that the fully automated generalization of residential areas has not been achieved to date is that it is difficult to acquire the knowledge that is required for automated generalization and for the calculation of spatial similarity degrees between map objects at different scales. Furthermore, little attention has been given to generalization methods with a scale reduction that is larger than two-fold. To fill this gap, this article develops a hybrid approach that combines two existing methods to generalize residential areas that range from 1:10,000 to 1:50,000. The two existing methods are Boffet's method for free space acquisition and kernel density analysis for city hotspot detection. Using both methods, the proposed approach follows a knowledge-based framework by implementing map analysis and spatial similarity measurements in a multiscale map space. First, the knowledge required for residential area generalization is obtained by analyzing multiscale residential areas and their corresponding contributions. Second, residential area generalization is divided into two subprocesses: free space acquisition and urban area outer boundary determination. Then, important parameters for the two subprocesses are obtained through map analysis and similarity measurements, reflecting the knowledge that is hidden in the cartographer's mind. Using this acquired knowledge, complete generalization steps are formed. The proposed approach is tested using multiscale datasets from Lanzhou City. The experimental results demonstrate that our method is better than the traditional methods in terms of location precision and actuality. The approach is robust, comparatively insensitive to the noise of the small buildings beyond urban areas, and easy to implement in GIS software.

Keywords: automatic map generalization; knowledge-based framework; geometric similarity quantification; multiscale map space; urban area delineation



Citation: Gao, X.; Yan, H.; Lu, X.; Li, P. Automated Residential Area Generalization: Combination of Knowledge-Based Framework and Similarity Measurement. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 56. <https://doi.org/10.3390/ijgi11010056>

Academic Editor: Wolfgang Kainz

Received: 16 November 2021

Accepted: 5 January 2022

Published: 12 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Automated map generalization has always been both a challenge and a dream for many mapping agencies [1–3]. For example, China's 1:5000 to 1:1,000,000 vector map databases, which consist of the same areas and regions at different levels of detail [4,5], are maintained and updated manually or semi-automatically by cartographers [6]. Current map generalization is undeniably a labor-intensive process that has many disadvantages, such as the repetitive digitization and compilation of data from the same region as well as inconsistent content and relationships between the map databases at different scales [4,6]. Thus, it is of great importance to realize automated map generalization. Indeed, many studies have been conducted to address this issue.

The realization of automated generalization mainly includes four aspects: (1) knowledge acquisition [7–9], i.e., obtaining characteristics, attributes, and relations (i.e., knowledge) from or between map objects; (2) the development of map feature generalization

algorithms (see [10–15] for simplification; see [16–18] for aggregation; see [19–21] for displacement); (3) the design of conceptual frameworks for map generalization [22–25]; and (4) the generation of a cognitive map that establishes the principle foundation for generalization [26–30]. Although many advancements have been made, the full automation of map generalization is still a challenge, as the knowledge acquisition that is required for automatic generalization is still difficult. Brassel and Weibel [22] deemed that generalization is a spatial modeling process that is only simulated by strategies based on understanding and not by a mere sequence of operational processing steps. Mackaness et al. [31] inferred that a cartographer's manual solution reflects a deep knowledge of the map generalization process and the ways in which map features might be illustrated at different scales. Mustiere's [9] proposal of a "cartographic knowledge acquisition bottleneck" also reflects the importance of knowledge. Moreover, most existing approaches focus on the generalization between two adjacent scales, i.e., a two-fold or smaller reduction in map scale, such as generalization from 1:10,000 to 1:20,000. Little attention has been given to generalization with a scale reduction greater than two-fold, e.g., 1:10,000 to 1:50,000.

This is also true for residential area generalization. Residential area generalization is one of the most significant problems in map generalization. On large-scale maps, residential area polygons occupy a large proportion of map loads, whereas on small-scale maps, only big and important settlement areas are retained. They act as indispensable positioning references for map readers. Therefore, this study follows a knowledge-based generalization framework [22] and emphasizes the generalization of residential areas in vector map databases from 1:10,000 to 1:50,000, with the aim of proposing a hybrid approach to automatically generalize residential areas. In this paper, two main factors (i.e., the knowledge-based framework and similarity measurement) are considered. The former is a widely accepted framework that has been proven to be scientific and reasonable [3,23]. It includes five steps: structure recognition, process recognition, process modeling, process execution, and display. More precisely, in this paper, maps in a multiscale map space (1:10,000, 1:50,000, 1:250,000, and 1:1,000,000 maps) are first analyzed (in accordance with the structure recognition). Secondly, the 1:50,000 residential area generalization process is divided into two parts: free space acquisition and urban area outer boundary determination (in accordance with process recognition). Thirdly, approaches to characterize both parts are determined, i.e., (1) Boffet's method for free space acquisition (hereafter referred to as Boffet's method) [32] and (2) a hybrid approach combining Boffet's method and kernel density analysis (KDE) for the outer boundaries of urban areas (in accordance with process modeling); Lastly, a step-by-step method is applied to the original 1:10,000 residential area data (in accordance with process execution). The display step is not discussed, as it was not the focal point in this study.

Specifically, in process modeling, geometric similarity is introduced and quantified to determine the KDE threshold. This is because the essence of map generalization is a kind of spatial similarity transformation in multiscale map spaces [6].

Therefore, the proposed approach has two main features: (1) knowledge is acquired by map analysis and similarity measurements in the multiscale map space, which is then used to automate the map generalization process; (2) two existing methods are combined to realize 1:10,000 to 50,000 residential area generalization. Compared to the reference 1:50,000 data, the results are more precise and are consistent with the latest 1:10,000 data. It should be noted that knowledge here refers specifically to the knowledge required in a certain map generalization task [33,34].

The remainder of the article is organized as follows: Section 2 proposes the hybrid approach, including an introduction to the data, an analysis of existing multiscale vector databases, the knowledge acquisition method from multiscale residential areas, the determination of the threshold values, and the calculation of the geometric similarity of residential areas; Section 3 presents the experimental results and analyzes them; Section 4 investigates and discusses the experiments, the results, and the proposed approaches; Section 5 draws a number of conclusions.

2. Methodology

This section proposes a hybrid method for generalizing residential areas on maps from 1:10K to 1:50K, which is a combination of a knowledge-based framework and geometric similarity measurements. As mentioned before, this approach is a combination of two existing methods: Boffet's method and KDE. However, the main focus of this paper is the knowledge-based framework determined through map analysis and similarity calculations in a multi-scale map space. Therefore, the organization of this part follows the four steps of the knowledge-based framework. However, experimental data were introduced before the steps of the framework.

2.1. Experimental Data

The experimental data are the spatial vector datasets covering the whole area of Lanzhou City, Gansu Province, China, at four different scales: 1:10K; 1:50K; 1:250K; and 1:1M. The 1:10K datasets were provided by the Provincial Geomatics Center of Gansu Province, China; the 1:50K datasets were provided by the National Geomatics Center of China; and the 1:250K and 1:1M datasets are free to download from <http://kmap.ckcest.cn> (accessed on 10 October 2021). The quality of all of the datasets have been thoroughly checked and are accepted by authorized institutions, e.g., the provincial or national quality inspection stations. Therefore, the four scale datasets can be used as standard manual generalization results for the proposed approach and can be used for analysis as well as for reference.

The basic information of the data is listed in Table 1.

Table 1. Experiment data.

Scale	Updating Time	Usage
1:10K	2017–2019	a. Data to validate the similarity measure; b. Data to be aggregated to obtain the target data.
1:50K	2012–2014	a. Data to validate the similarity measure; b. Standard data.
1:250K	2012	a. Data to validate the similarity measure.
1:1M	2015	a. Data to validate the similarity measure.

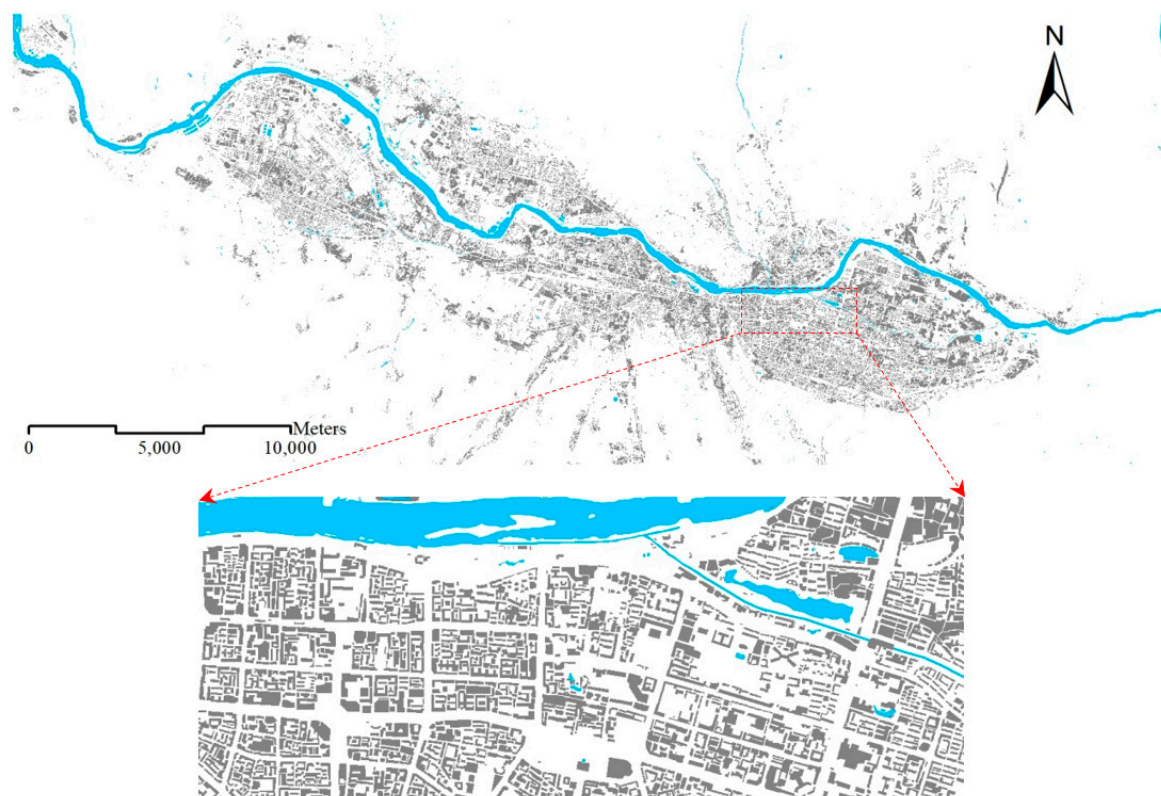
2.2. Residential Areas Generalization Method Follows Knowledge-Based Framework

2.2.1. Structure Recognition and Process Recognition Based on Map Analysis

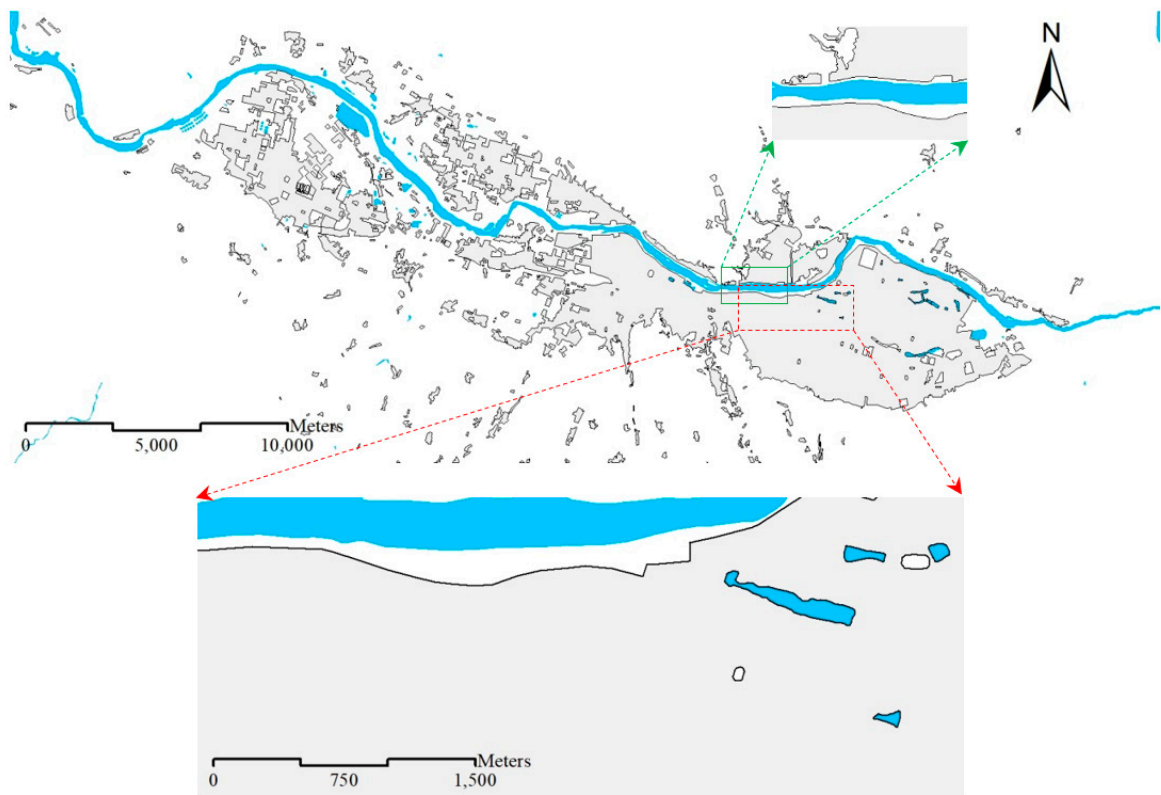
Figures 1 and 2 show the residential areas, roads, and rivers of the region at scales ranging from 1:10K to 1:1M.

1. In the 1:10K dataset, residential areas are represented as buildings (Figure 1a);
2. In the 1:50K dataset, residential areas are represented using building groups and blocks [9] (Figure 1b) At the 1:50K scale, roads are not shown, so the residential areas can be divided into blocks;
3. In the 1:250K dataset, residential areas depict entire urban settlements (Figure 2a);
4. In the 1:1M dataset, only big cities (i.e., residential areas) appear (Figure 2b).

It can be seen from Figure 1b that the residential areas on the 1:50K maps are a compound of aggregated building groups and urban areas, or built-up areas [35,36]. Thus, the generalization process of the residential areas on the 1:50K maps should be considered: (1) the boundaries of the built-up areas should be identified and (2) free space should be acquired within the outer boundaries. The two parts will be analyzed, and a strategy will be made; furthermore, knowledge of how to properly implement the strategy will be obtained.



(a)



(b)

Figure 1. Vector datasets of Lanzhou City at the 1:10K and 1:50K scales: (a) 1:10K data and (b) 1:50K data.

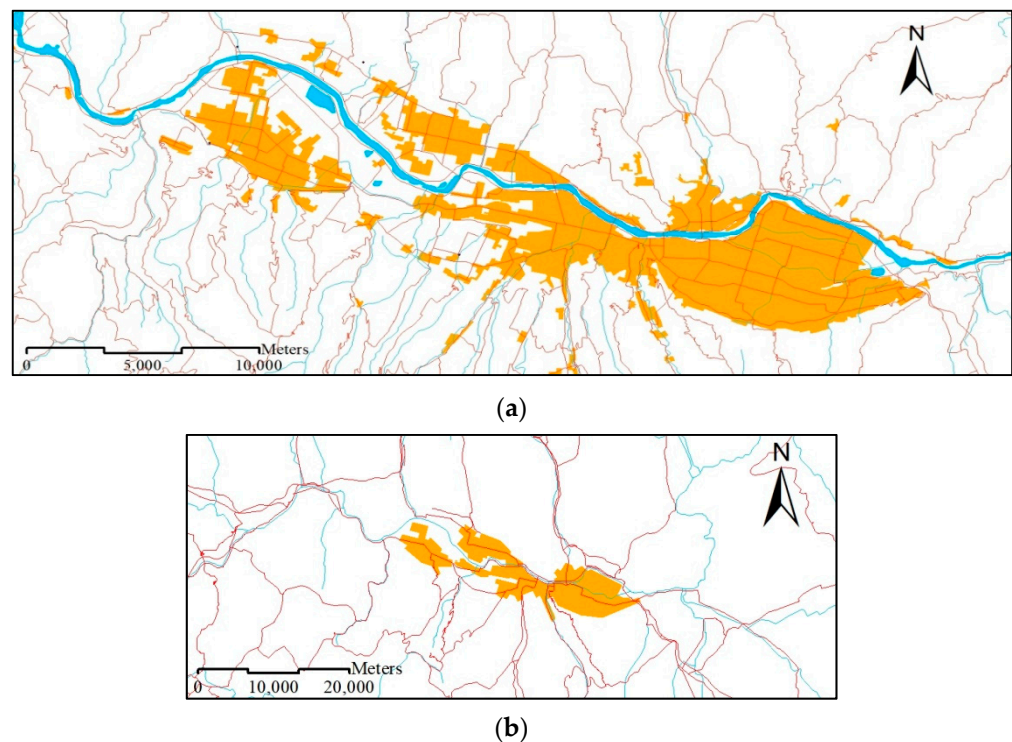


Figure 2. Vector datasets of Lanzhou City at the 1:250K and 1:1M scales: (a) 1:250K data and (b) 1:1M data.

2.2.2. Process Modeling

1. Boffet's method and KDE:

Boffet et al. [32] proposed a complete method for free space identification. It is mainly based on the buffer analysis of buildings and roads. The main idea of this method indicates that the free spaces within urban areas are those spaces are not occupied by buildings and roads. It is not hard to understand and is easy to apply. Through empirical testing, the method was proven to be reasonable [37,38]. The only problem is the buffer distance, which will be discussed later.

To determine the outer boundaries of urban areas, there are many approaches that can be used [33,35,39,40]. Among them, KDE has been found to excel in detecting city hot spots. However, similar to other hot spot maps, the KDE results are shown by contours, which have two main drawbacks: first, contours provide less precise residential area generalization results; and second, the smooth shape of the contours does not match the cognitive habits that map readers have when they are identifying settlements [35].

On the other hand, the experiments have determined that Boffet's method is precise enough to be used for 1:50K residential area generalization. This is because it is a method that is based on buildings and roads dilating and eroding buffer analysis, which is consistent with the city sprawl rule: generally, urban expansion takes place in the form of building and road construction, with a higher density in the hot spots and a lower density near the edges of the city. In addition, the location error for the 1:50K residential areas should be smaller than 0.1 mm on maps, which represents 5 m on the ground (according to Chinese technical rules for quality inspection and acceptance of 1:50K topographical maps). In Boffet's method, the outline of the built-up areas and free spaces was obtained by the dilating and eroding buffers for the 1:10K buildings, and its location error is smaller than 0.1 mm on 1:50K maps. This is because the 1:10K data has a higher location precision than the 1:50K data. However, a problem with Boffet's method is that it aggregates buildings and roads together through buffering, merging, and other operators to form dispersed building groups, but which building groups are included in the urban areas to form the 1:50K residential areas? Which building groups are excluded from the urban areas and

eliminated? As mentioned in the last paragraph, KDE can provide a fuzzy build-up boundary that allows Boffet's method to make judgements. In fact, a city boundary is intrinsically a fuzzy boundary that is cognitively delineated by human beings [33,41]; thus, the results from KDE fit these kinds of boundaries.

Hence, a hybrid method combining KDE and Boffet's method was developed in this study. KDE was used to determine the general extent of built-up areas, while Boffet's method was applied to aggregate buildings and roads together to form a precise boundary, such as the ones seen in 1:50K residential areas. Then, the two results are intersected to determine the belonging relationships between buildings and urban areas. The process is modeled in Figure 3. The area within the blue outlines in (b) and (c) represents the KDE results. This is followed by an illustration of the thresholds of the two methods and detailed generalization steps (Section 2.2.3).

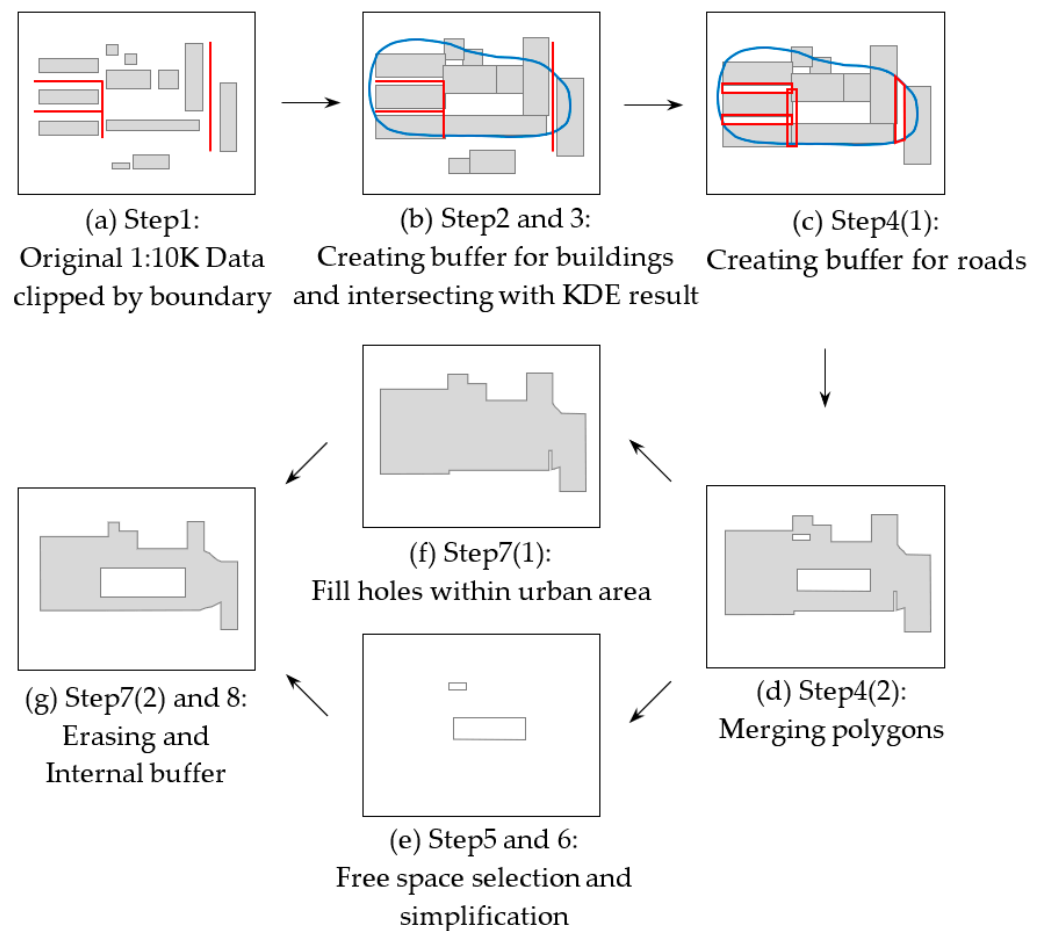


Figure 3. The 1:10K to 1:50K generalization model (it should be noted that the steps in the figure are consistent with those in Figure 7 in Section 2.2.3; a detailed explanation is given in Section 2.2.3): (a) Step 1: The original 1:10K data is clipped by the administrative boundary; (b) Steps 2 and 3: a buffer is created for buildings, and it intersects with the KDE results; (c) Step 4, Part 1: A buffer is created for the roads; (d) Step 4, Part 2: Polygons are merged; (e) Steps 5 and 6: Free spaces are collected, simplified and smaller ones are eliminated; (f) Step 7, Part 1: Holes within urban areas are filled in; (g) Step 7, Part 2 and Step 8: Holes are erased and internal buffer is made.

2. Threshold determination by map analysis and similarity measurement:

To describe the generalization process clearly, some parameters need to be clarified, i.e., buffer distance, selection threshold, and KDE threshold. Map analysis and similarity measurements are employed to find the most reasonable values for these parameters.

- Buffer Distance

In Boffet's method, the buffer distance around buildings is 20 m, and the internal buffer distance of the blocks is 20 m. This is because 20 m is the standard distance of "ownership". However, in Lanzhou City, the widest road is 52 m, which is obtained through measurements taken on the 1:10K road dataset. Therefore, in our experiment, a 20 m buffer around buildings and an buffer of 26 m (half of 52 m) around roads were determined to identify the free space in Lanzhou City.

- Selection Threshold

After the free spaces are obtained, elimination should be applied to select those large and informative areas so that they may be retained on a 1:50K residential area map. This decision is made through measuring the areas of the holes that are present on the 1:50K map. The determined threshold is 5000 square meters. Thus, free spaces smaller than 5000 square meters may be eliminated.

- KDE Threshold

Here, city hot spot detection (i.e., the KDE) of 1:10K buildings is critical. This is because the results achieved using Boffet's method remain unchanged under the definite buffer distance. Conversely, the KDE result, namely the hot spot detection result, varies according to its threshold values. Additionally, only the regions that were determined using Boffet's method that intersect with the hot spots will be categorized into the generalization result. On the other hand, others will be classified as field areas and will be deleted. For example, in Figure 4, which shows KDE result 1, built-up areas include region A and region B; however, when considering KDE result 2, built-up areas consist of region A, and region B is only seen as noise and will be eliminated.

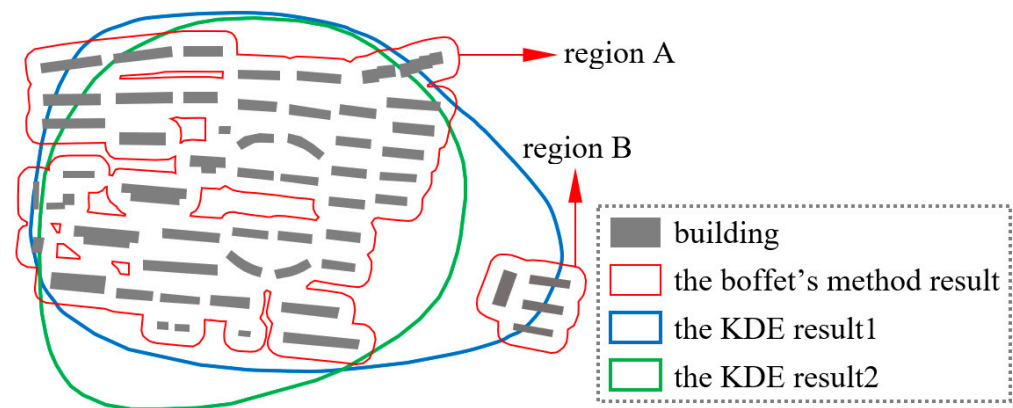


Figure 4. Relationship between buildings, Boffet's method result, different KDE results, and obtained urban areas.

Therefore, delineating an objective and scientific urban area is crucial to achieve a reasonable generalization result. In order to achieve this aim, the KDE and the principles of its threshold determination can be analyzed by following three steps:

(1) KDE is used to detect the distribution hot spots of 1:10K buildings. The Kernel Density tool in ArcGIS (V10.2) is used to calculate the feature density in the neighborhood around those features. The formula for calculating the kernel density is [42]:

$$Density(x, y) = \frac{1}{(radius)^2} \sum_{i=1}^n \left[\frac{3}{\pi} \cdot pop_i \left(1 - \left(\frac{dist_i}{radius} \right)^2 \right)^2 \right] \text{ For } dist_i < radius \quad (1)$$

where $Density(x, y)$ is the predicted density at a new location (x, y) ; $radius$ is the search radius from the point (x, y) ; $i = 1, \dots, n$ are the points within the (x, y) radius distance; and pop_i is the population field value; and $dist_i$ denotes the distance between the cell and the i th point in the circular neighborhood.

The search radius and output cell size are two parameters that affect analysis results, and previous studies have suggested different thresholds for the two parameters, with both achieving different results. Through empirical testing, it has been found that the search radius has more influence than the cell size, as the cell size only determines the level of detail on the output grid. To determine the search radius, previous research used different values ranging from 125 m to 500 m [43–46]. To fully automate the process, this study uses Tversky’s similarity ratio model to determine the optimal threshold [47].

(2) Why is similarity modeling necessary? The reason is because the target is to generalize 1:10K buildings to achieve a 1:50K residential areas; thus, the existing 1:50K residential area dataset can act as a reference for KDE to make a comparison in order to obtain the most similar results. Implicit knowledge can be gained through this process. The similarity is context-dependent [48], i.e., the similarity between two things is always relative to the certain context in which they are compared, such as from a certain perspective or based on a specific interest [48,49]. This is also true for spatial similarity. In this study, the geometric similarity relationship is considered. To be specific, the similarities determined based on the KDE results and the 1:50K residential overlapping areas are considered. These area similarities are compared for two reasons: first, shape similarity is controlled by Boffet’s method, which is precise enough for a 1:50K result; and secondly, for polygonal features, area maintenance is another important aspect in generalization [50]. Thus, area is the main factor that is considering for geometric similarity measurements in this study. Its function is to automate KDE threshold determination. To quantify and formally express it, Tversky’s ratio model is used, and a detailed explanation is given below.

Firstly, a good similarity measure should match intuition [51,52]. However, for residential areas in multi-scale map spaces, aggregation is a common generalization operator that causes that the correspondence relationships between residential areas at different scales become complicated. Figure 5 shows a portion of the example data. In the 1:10K dataset, 18 buildings represent field objects; while in the 1:50K dataset, two polygons are obtained after applying aggregation and division operators based on these 18 buildings; furthermore, in the 1:250K dataset, all of the buildings are represented by only one polygon. As a result, the common measurements for geometric similarity, e.g., the turning function is not suitable for features in multi-scale map spaces [53]. Instead, using Tversky’s ratio model to measure similarity is a sound assumption because all of the maps used in the study are geographically rectified [54], so overlaying areas between different scale residential areas can reflect their commonality to some extent; additionally, it is not too sensitive to the shape details and is robust enough as a geometric similarity measurement. This assumption is tested by experiments in Section 3.1.1.



Figure 5. An example of a correspondence relationship between residential areas at different scales.

Secondly, cartography and geographical information science is a discipline that is closely related to cognition science, and decades of human behavior research have demonstrated that psychological similarity is not equivalent to mathematical similarity as it is typically defined [47,55]. Tversky's ratio model is a psychological similarity model that can be used in map generalization. However, it must be validated before it is applied in KDE; therefore, an experiment is designed below.

The matching function of Tversky's ratio model is:

$$GeoSim(ObjD, RefD) = \frac{f(ObjD \cap RefD)}{f(ObjD \cap RefD) + \alpha f(ObjD - RefD) + \beta f(RefD - ObjD)}, \alpha, \beta \geq 0 \quad (2)$$

where $GeoSim(ObjD, RefD)$ is the similarity degree between the object data and reference data; the function f measures the area of certain regions; $ObjD \cap RefD$ denotes the common regions belonging to both the object data and reference data; $ObjD - RefD$ represent the regions that belong to the object data but not to the reference data; and $RefD - ObjD$ are the regions that belong to the reference data but not to the object data.

Here, $\alpha = \beta = \frac{1}{2}$ because at present, data at different scales are maintained and updated separately, and their update time are independent of each other (Table 1). $ObjD - RefD$ and $RefD - ObjD$ are equally important, and as a result, the weights are set to be equal, i.e., α equals β .

$GeoSim(ObjD, RefD)$ is calculated between two different-scale experimental residential areas, including $GeoSim(50K, 10K), \dots, GeoSim(1M, 250K)$, etc. There are six $GeoSim$ values. The values are then analyzed; if Tversky's ratio model is proven to be effective, namely, its results match the intuition of human beings, then it can be used to determine the KDE threshold. Theoretically, the reference data should always be large-scale data, and the object data should be small-scale data because small-scale data should be gained through generalization from large-scale data, and large-scale data should be used as the reference. However, in practice, different data are maintained and updated separately; therefore, in this study, for the convenience of presentation, in $GeoSim(S_i, S_j)$, S_i can be greater than S_j . Here, S_i and S_j are data scales. However, in most cases, S_j is a larger scale, and S_i is a smaller scale.

(3) The KDE threshold is determined through the following steps: Firstly, $GeoSim(KDE_{different\ threshold}, 50K)$ is computed, and the search radius that maximizes $GeoSim(KDE, 50K)$ is chosen as the most appropriate threshold, i.e., the most similar KDE result to the existing 1:50K residential areas is acquired with the determined search radius.

The resulting raster is then reclassified and converted to vector data, which are denoted by $KDEPolygons$. The polygons in this dataset act as a "container" for Lanzhou City. The "container" changes according to different density values (Figure 6). Areas with lower density values can be used to extract the city's overall sprawl extent, and areas with higher density values represent the central urban areas with higher concentrations. For example, in Figure 6, areas with density values of 1 or greater include the entire region (Figure 6b), while areas with a density value of 2 or greater include two smaller individual regions (Figure 6c). Through experiments, this study used the "Equal Interval" classification method (9 classes) in ArcGIS and chose the areas with density values bigger than or equal to classes 2–9 as urban city regions because the KDE result under this classification method was the most similar to the 1:50K reference residential areas.

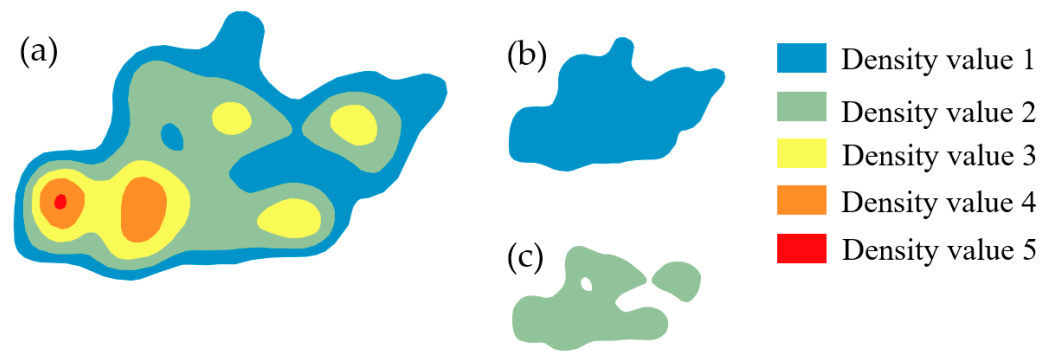


Figure 6. Illustration of different density values and corresponding urban sprawl extents: (a) is the KDE result, (b) is the obtained urban area with density values of 1 or greater, and (c) is the obtained urban areas with density values of 2 or greater.

2.2.3. Detailed Generalization Process

Figure 7 shows the procedures of the proposed method. The results obtained from city hot spot detection are denoted as *KDEPolygons* and have been described in detail in Section 2.2.2, where the threshold for city hot spot detection is closely related to geometric similarity measurements. The detailed generalization process is as follows:

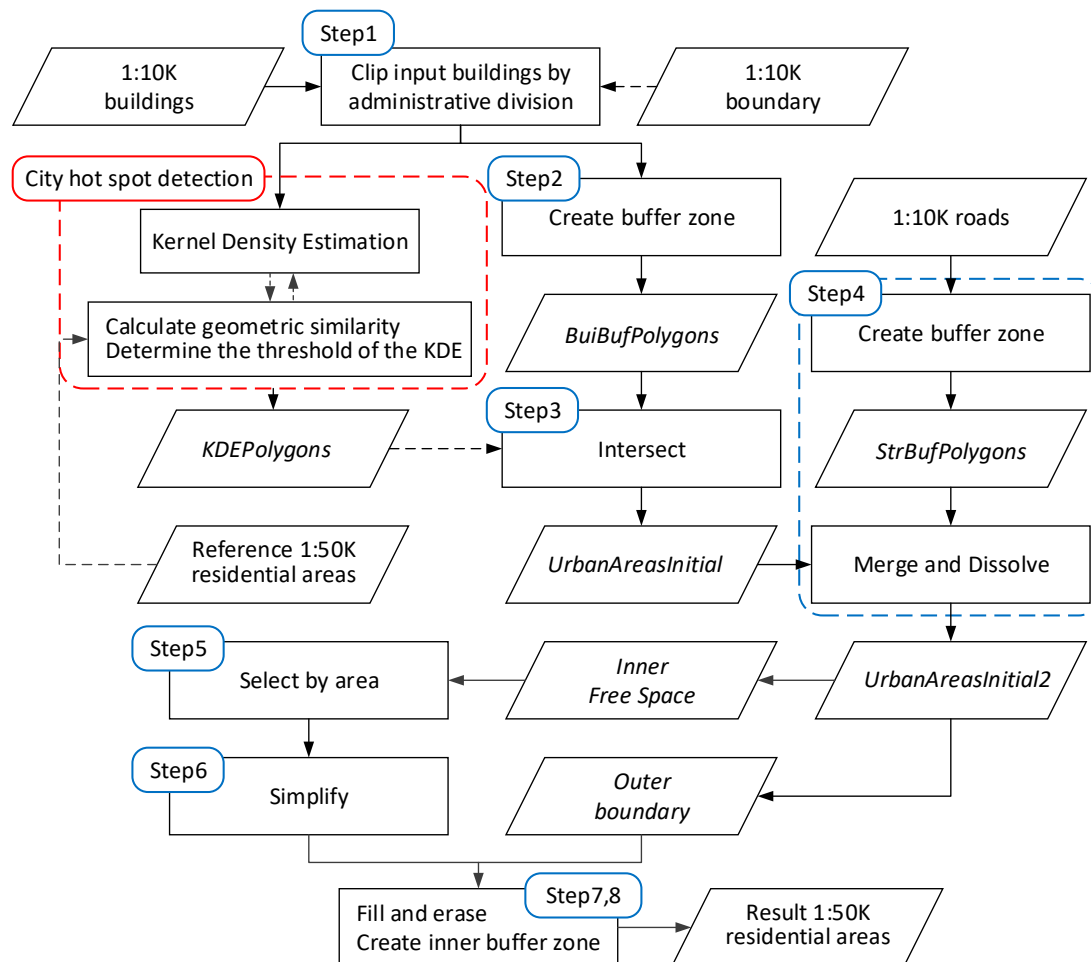


Figure 7. Procedures of the proposed method.

Step 1: The 1:10K buildings are clipped by the administrative district polygons to minimize the study area.

Step 2: According to the knowledge acquired in Section 2.2.2, a 20 m buffer around the buildings is computed to aggregate neighbor buildings, and the result is denoted by *BuiBufPolygons*.

Step 3: The *BuiBufPolygons* that intersect with the *KDEPolygons* are selected, as denoted by *UrbanAreasInitial*, representing the initial built-up areas. Due to the wide streets and free spaces in cities, there are many small holes or narrow gaps. The narrow gaps will be filled in the next step, while free spaces will be extracted as input data in Step 5.

Step 4: According to the knowledge acquired in Section 2.2.2, a buffer of 26 m around 1:10K road lines is made, and the result is clipped by the *KDEPolygons*, as denoted by *StrBufPolygons*. The *UrbanAreasInitial* are “glued” together by the *StrBufPolygons* to form a more complete built-up area, as denoted by *UrbanAreasInitial2*.

Step 5 and Step 6 are processes for acquiring free spaces.

Step 5: Extract all holes in *UrbanAreasInitial2*, denoted by *FreeAreaPolygons*. According to the knowledge acquired in Section 2.2.2, the polygons in *FreeAreaPolygons* whose field areas are smaller than 5000 square meters are deleted.

Step 6: Simplify *FreeAreaPolygons* using the Bend Simplify Algorithm [56] to obtain simpler polygons depicting open spaces within *UrbanAreasInitial2*.

Based on *UrbanAreasInitial2* and the inner free spaces, Step 7 and Step 8 are applied to obtain the 1:50K residential areas.

Step 7: Fill all the holes within the outer boundaries of *UrbanAreasInitial2* o, and then, erase *UrbanAreasInitial2* using the *FreeAreaPolygons* to achieve the final result, denoted by *UrbanAreasFinal*.

Step 8: An internal buffer of 20 m is adopted to contract the dilated boundaries to obtain higher position accuracy.

Note that in KDE, 1:10K buildings must be turned into points because the KDE tool in ArcGIS only can use building points as input features.

3. Experiments and Results

3.1. Results and Analysis

3.1.1. Geometric Similarity Results and Analysis

Table 2 shows similarity values between the residential areas at four scales. They can also be expressed as a matrix (1). Figure 8 evaluates the results and the effectiveness of the geometric similarity measures. Many insights can be gained when considering the three forms of similarity representation.

Table 2. Similarity values between residential areas at four scales.

Object Data and Reference Data	Common Area (m ²)	Symmetrical Difference Area (m ²)	Similarity Value
1:50K, 1:10K	43,071,658	74,941,399	0.53
1:250K, 1:10K	40,623,846	85,319,088	0.49
1:1M, 1:10K	39,926,929	98,906,243	0.45
1:250K, 1:50K	92,497,605	28,961,830	0.86
1:1M, 1:50K	88,426,174	49,297,954	0.78
1:1M, 1:250K	96,648,146	38,336,073	0.83

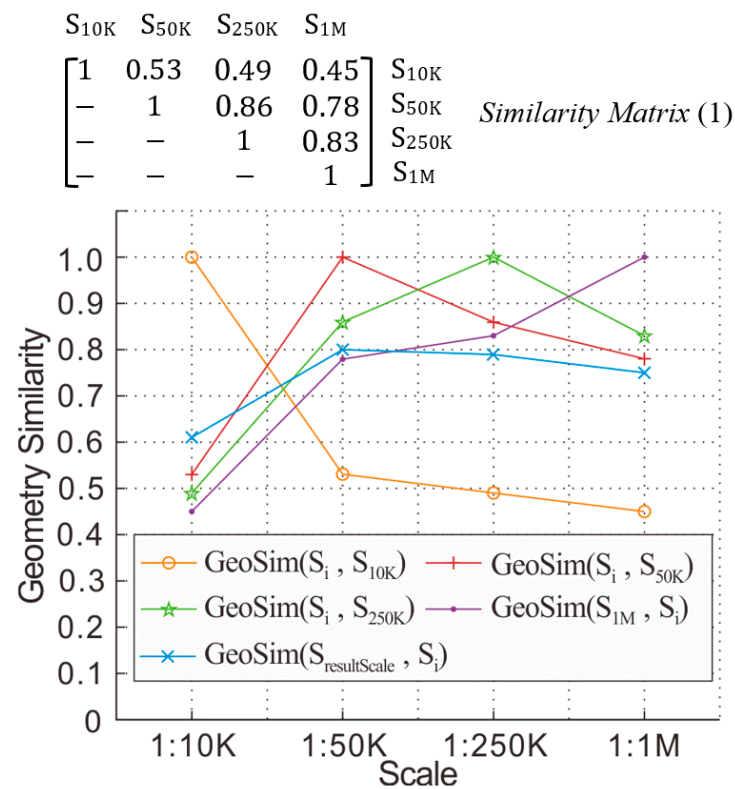


Figure 8. Geometry similarity values between residential areas at four scales.

Firstly, in Matrix (1), where the row number equals the column number, all of the similarity values equal 1, indicating that the same dataset has the greatest similarity value with itself. This conclusion can also be drawn from Figure 8. For example, the orange line represents the similarity between residential areas at all four scales and residential areas at 1:10K, and it reaches its peak value when the abscissa is 1:10K. In addition, it decreases monotonically from the 1:10K abscissa on both sides (1:10K has the lowest abscissa value; therefore, it only has one side).

Secondly, Matrix (1) is an upper triangular matrix because in multi-scale map space, similarities between different scales are asymmetrical. For example, people usually say “the 1:50K data are 53 percent similar to the 1:10K data” but seldom say “the 1:10K data are 53 percent similar to the 1:50K data” because smaller scale data are an abstraction of larger scale data, and a metaphor that is able to connect two concepts with a “similar relationship” involves a selective rather than an unconstrained comparison process [57]. The same rule applies to Figure 8.

For each row of Matrix (1), the similarity values decrease as the map scale decreases. On the whole, the similarity values between any other scales and the 10K data are the smallest (except for 10K data, which are 100% similar to themselves). This should be readily interpretable, for the 1:10K data provide a detailed description of individual buildings, but the 1:50K scale is a turning point where the data form a block view. However, in the 1:50K to 1:1M data, the similarity value decreases at a relatively slow speed, but the overall trend remains unchanged. An example of this would be $c_{23} = 0.86 > c_{24} = 0.78$, which means that the 1:250K data are more similar to the 1:50K data than the 1:1M data are.

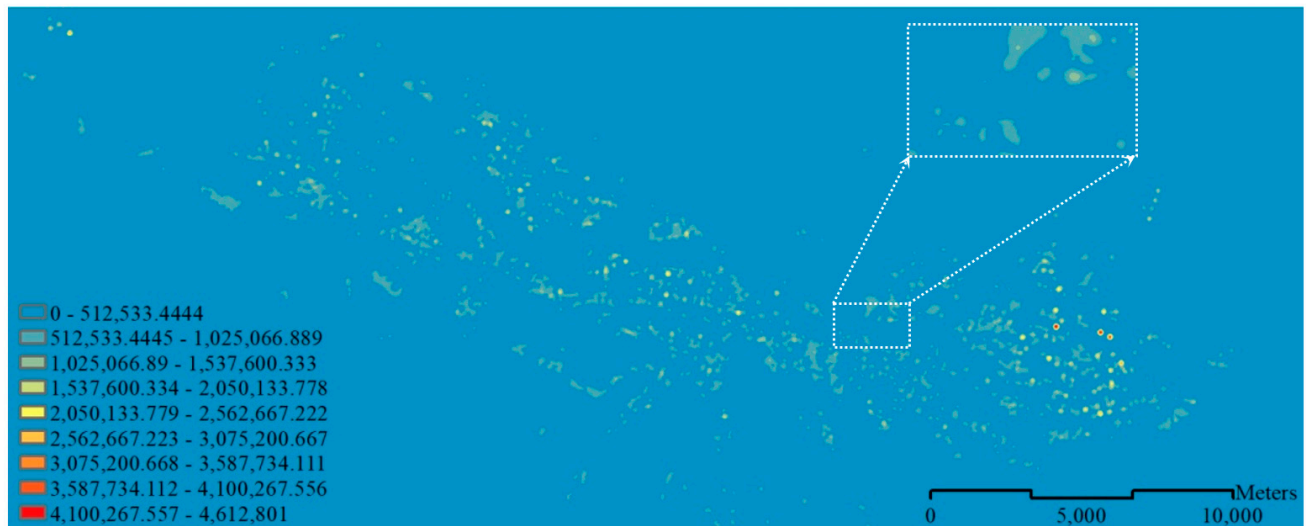
The ground truth time described by the data also plays an important role in similarity measurements, which is called “time similarity”. The results also indicate that 1:50K is a turning point scale for map generalization from which the main task for mapping residential areas becomes a more generalized “block view” (compared with individual buildings of 1:10K). For building generalization from 1:10K to 1:50K, the most commonly used algorithm is aggregation. While for 1:50K to 1:250K and even to 1:1M generalization,

elimination and simplification are more used to delete smaller areas and details more often [17].

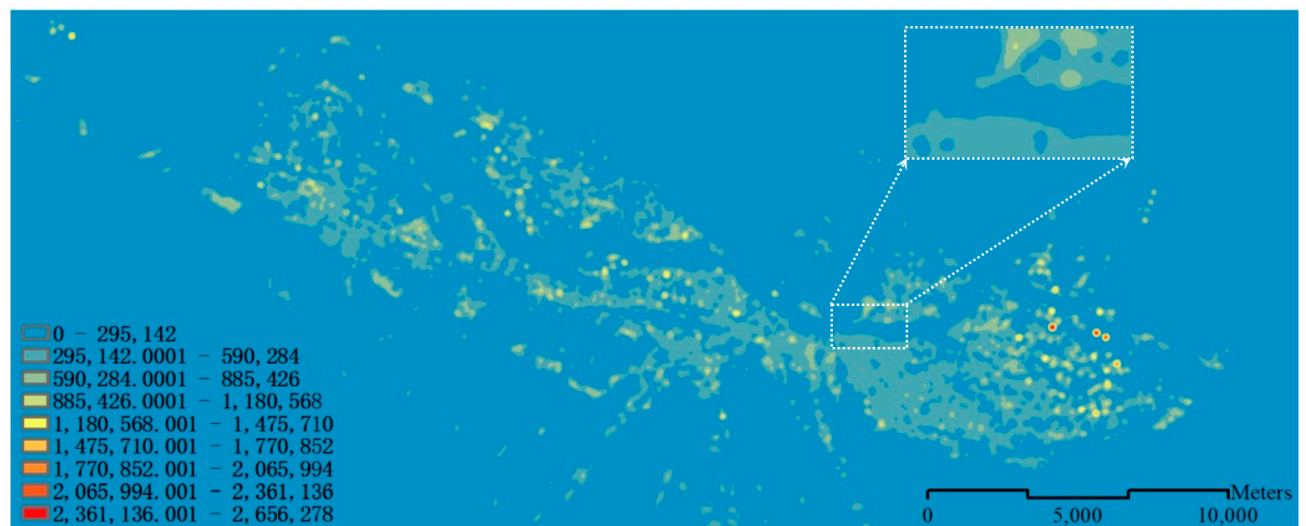
It can be seen from the above analysis that the method used to measure geometric similarity is effective. Thus, it can be used to fully automate KDE and to determine its threshold values.

3.1.2. The KDE Results with Different Threshold Values and Manual Inspection

Figure 9 shows the urban hot spots delineated by the kernel density analysis with different threshold values.

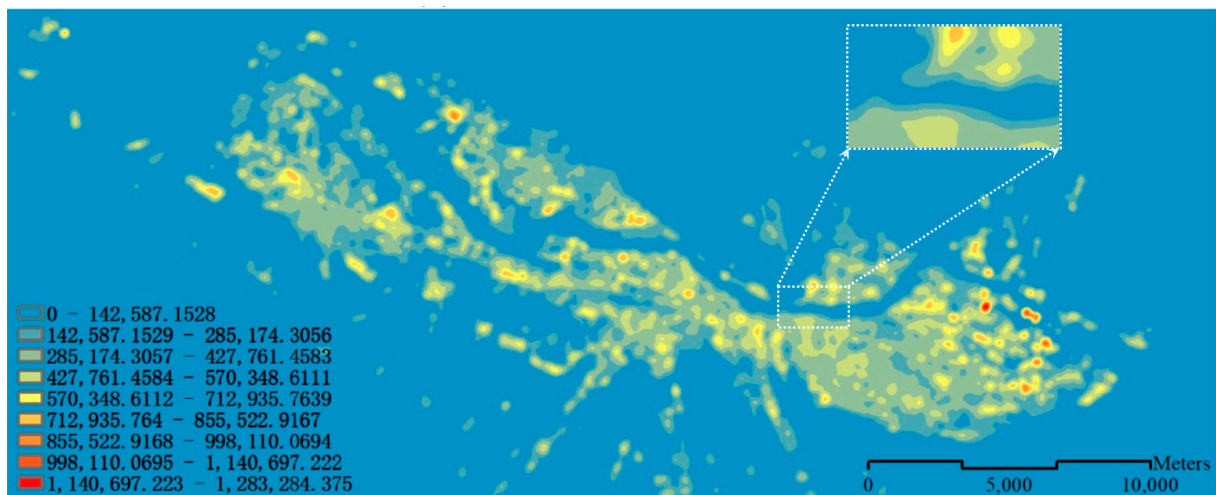


(a)

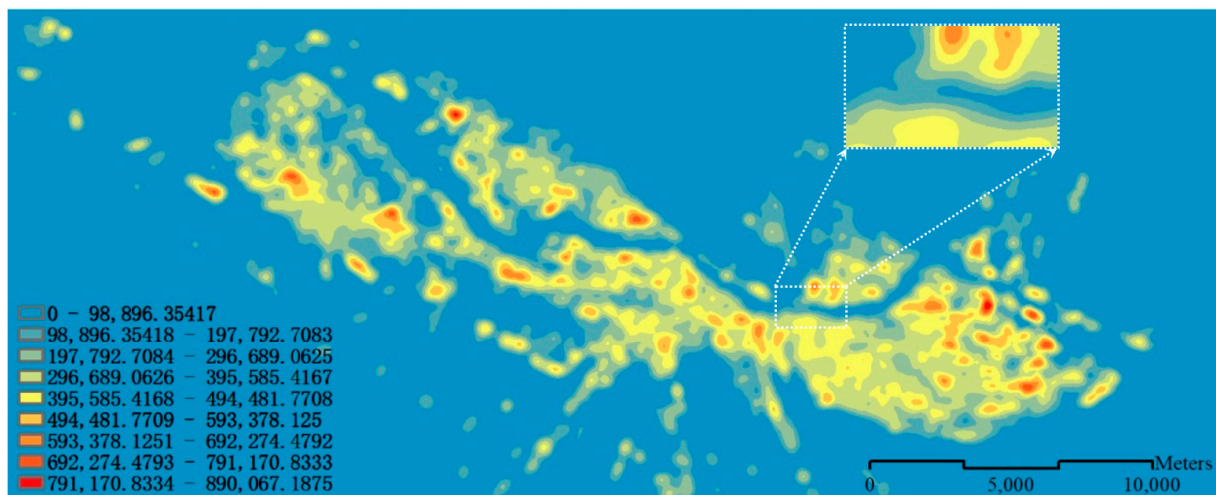


(b)

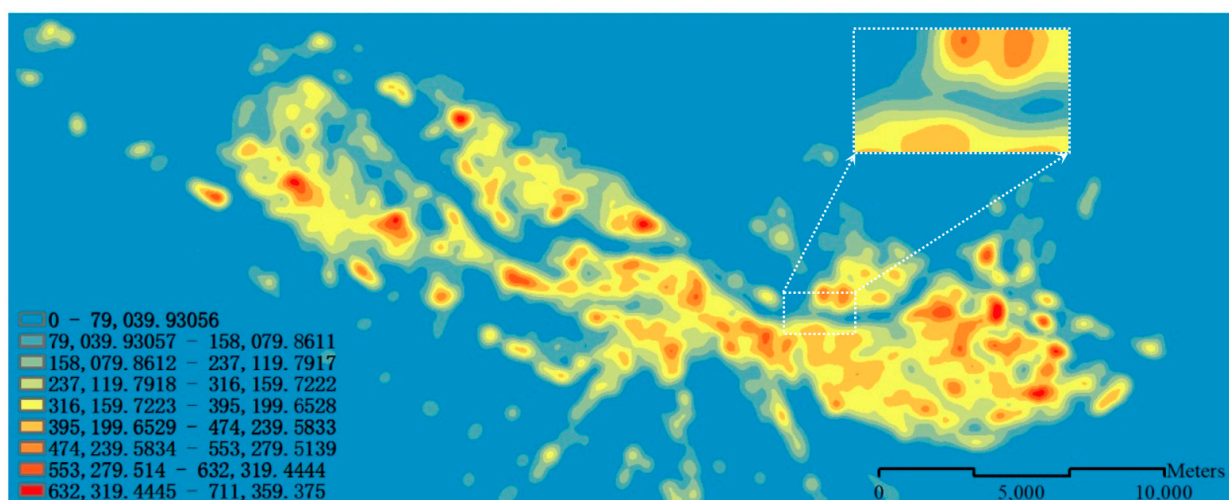
Figure 9. Cont.



(c)



(d)



(e)

Figure 9. Urban hot spots detected by KDE with different thresholds but with the same refined cell size (10 m): (a) is the result when the search radius is 150 m; (b) is the result when the search radius is 200 m; (c) is the result when the search radius is 300 m; (d) is the result when the search radius is 400 m; and (e) is the result when the search radius is 500 m.

From Figure 9, it can be seen that the detected urban hot spots sprawl as the search radius increases. A manual comparison between Figure 9 with Figure 1b (1:50K residential area data) shows that:

- (1) The detected hot spots are mostly isolated contours when the search radius is small, e.g., 150 m (Figure 9a). The topological relationships of the contours are disjointed, which makes them less suitable as containers for aggregated buildings;
- (2) The detected hot spots are almost connected when the search radius is large, e.g., 500 m (Figure 9e). The topological relationships of the contours are contained, resulting in an urban area that is larger than it actually is;
- (3) The white rectangles in Figure 9 show a partial enlargement of the KDE results. Compared to the same region in the 1:50K data (green rectangle in Figure 1b), the residential area is divided by the Yellow River into two parts. Intuitively, from the perspective of the topological relationship, the best KDE result is the urban hot spots that were detected with the 300 m search radius.

3.1.3. Similarity Measure and Threshold Determination

Table 3 lists the quantitative similarity results that were computed using Equation (2) (Section 2.2.2). It can be seen that when the search radius is 300 m, the maximum similarity it obtained (0.77). This result also matches the initiation of human beings. Consequently, the hot spots result in Figure 9c is used as the building aggregation container (*KDEPolygons*).

Table 3. The similarity of different kernel density analysis results and the standard 1:50K residential areas.

Threshold (m)	Common Area (m ²)	Symmetrical Difference Area (m ²)	Similarity
150	20,470,299	87,143,961	0.32
200	66,457,811	52,045,137	0.72
300	98,036,747	59,650,663	0.77
400	101,323,067	83,361,424	0.71
500	102,154,882	102,910,295	0.67
600	102,457,973	116,361,813	0.64

3.1.4. Results of Generalization

(1) Figure 10 shows the 20 m buffer results of the 1:10K buildings. Two layers, grey areas (*BuiBufPolygons*) and colored areas (*UrbanAreasInitial*), are shown. *BuiBufPolygons* is the original buffer result, while *UrbanAreasInitial* is a subset of *BuiBufPolygons* that intersects with *KDEPolygons*. *BuiBufPolygons* that are covered by *UrbanAreasInitial* are identical to the *UrbanAreasInitial* on the top, and the *BuiBufPolygons* that are not covered by *UrbanAreasInitial* are excluded from the urban areas (e.g., red circles in Figure 10).

It can be seen that the results are separated parts, also, interior of each parts of *UrbanAreasInitial* has many narrow areas and holes.

(2) Figure 11 shows the polygons merged by *UrbanAreasInitial* and *StrBufPolygons* (a), they are dissolved to obtain (b). Figure 11c shows *UrbanAreasInitial2*.

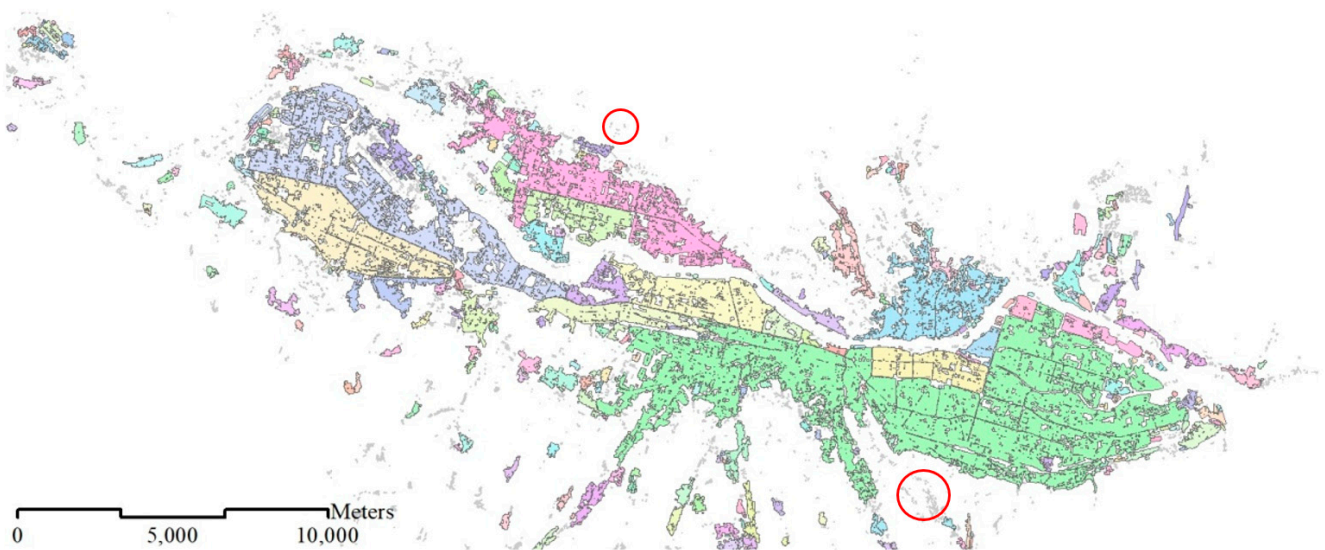
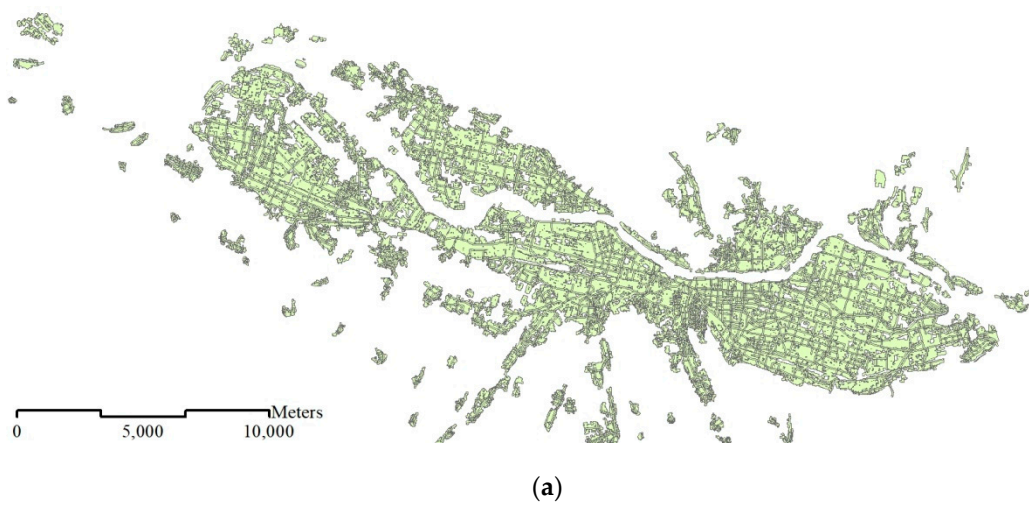
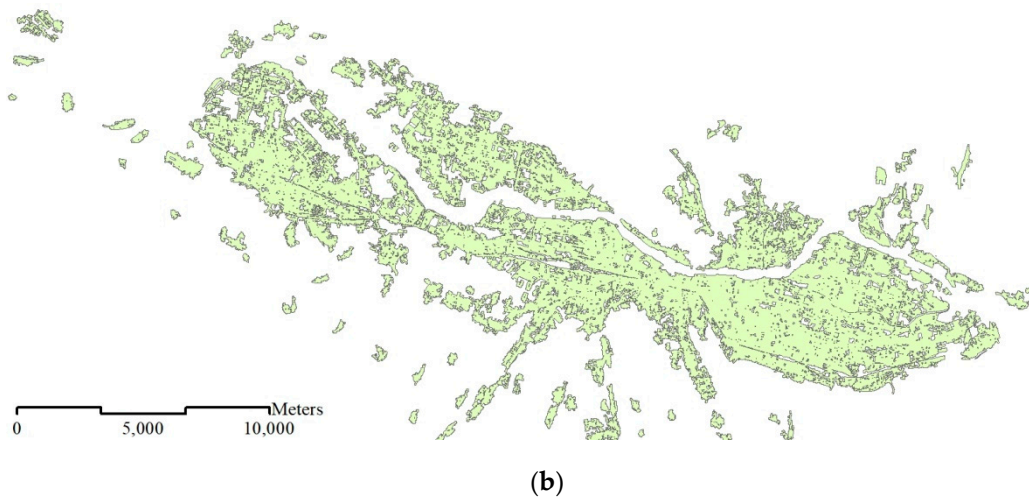


Figure 10. Buffering results (20 m) of 1:10K buildings.



(a)



(b)

Figure 11. *Cont.*

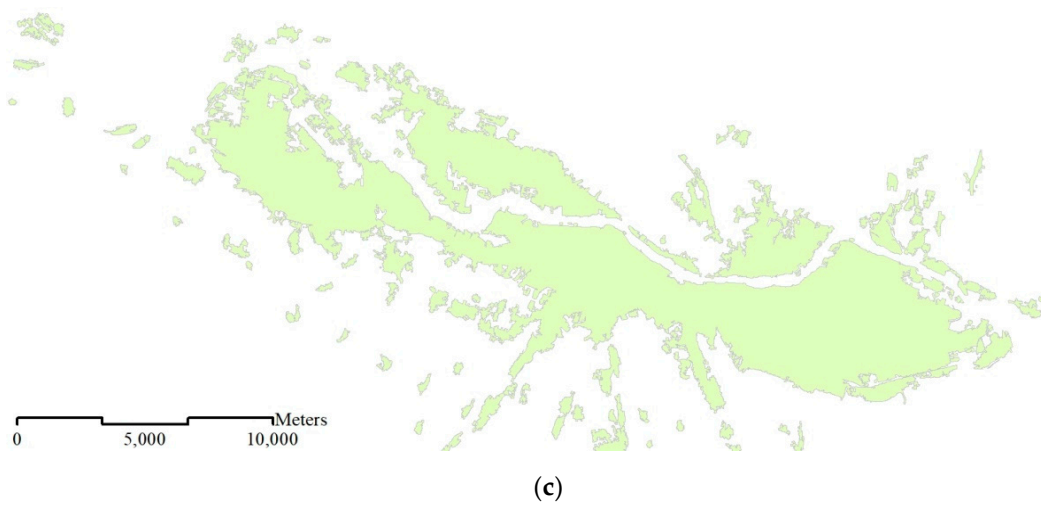


Figure 11. Results of experiments: (a) *UrbanAreasInitial*, (b) *UrbanAreasInitial_StrBufPolygons_Merge*, and (c) *UrbanAreasInitial2*.

(3) Figure 12 shows *FreeAreaPolygons*, which represents the free spaces within urban areas larger than 5000 square meters.

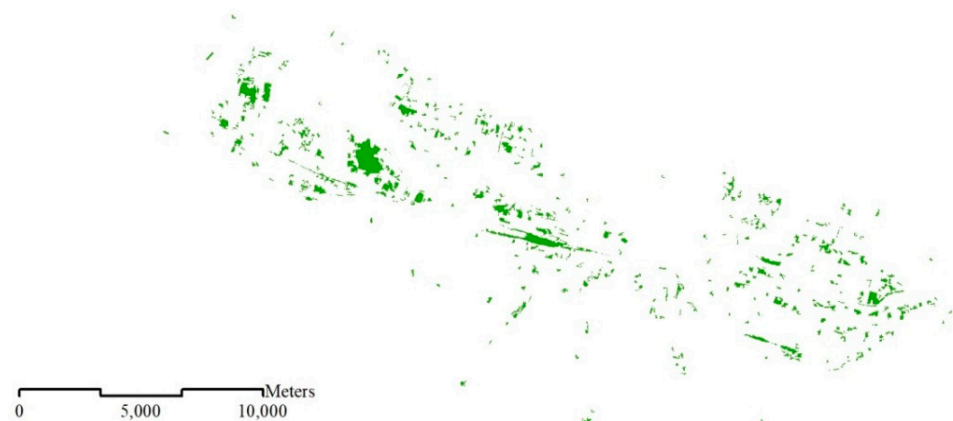


Figure 12. Obtained free spaces within urban areas.

(4) After Step 7 and Step 8, *UrbanAreasFinal* can be obtained, as shown in Figure 13.

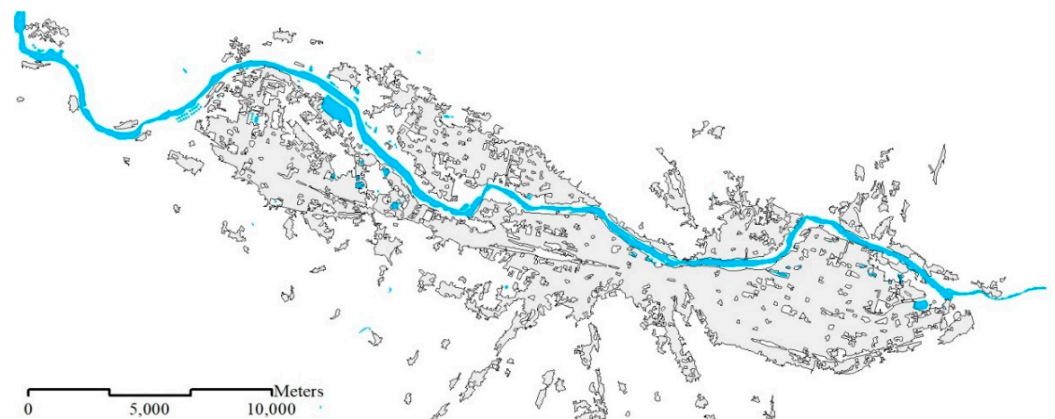


Figure 13. Urban areas obtained from the proposed approach.

3.2. Evaluation

3.2.1. Visual Comparison with Reference 1:50K Data

Through a manual comparison of the generalization results (Figure 13) and the standard data (Figure 1b), the differences between the two datasets can be classified into two types: new free spaces within urban areas and extension regions on the edges of the cities or towns. The results show more free spaces and extension areas than the standard data do. This indicates that from 2012 to 2019 (Table 1), many leisure and entertainment places and green land parks have been built in the city, allowing the city to become more developed and increasing livability. Additionally, the overall land areas have increased, most of which are new residential communities.

Figure 14 shows examples of free space from Figure 12. They are compared with the latest web satellite map <https://map.qq.com/> (accessed on 10 October 2021). It can be seen that the free spaces within urban areas generally fall into the following categories: (a) fields; (b) open-air playgrounds; (c) railway stations; (d) squares; (e) parks; and (f) space yet to be built. Therefore, compared to the 1:50K reference residential areas, the obtained free spaces are accurate and are in accordance with the ground conditions.

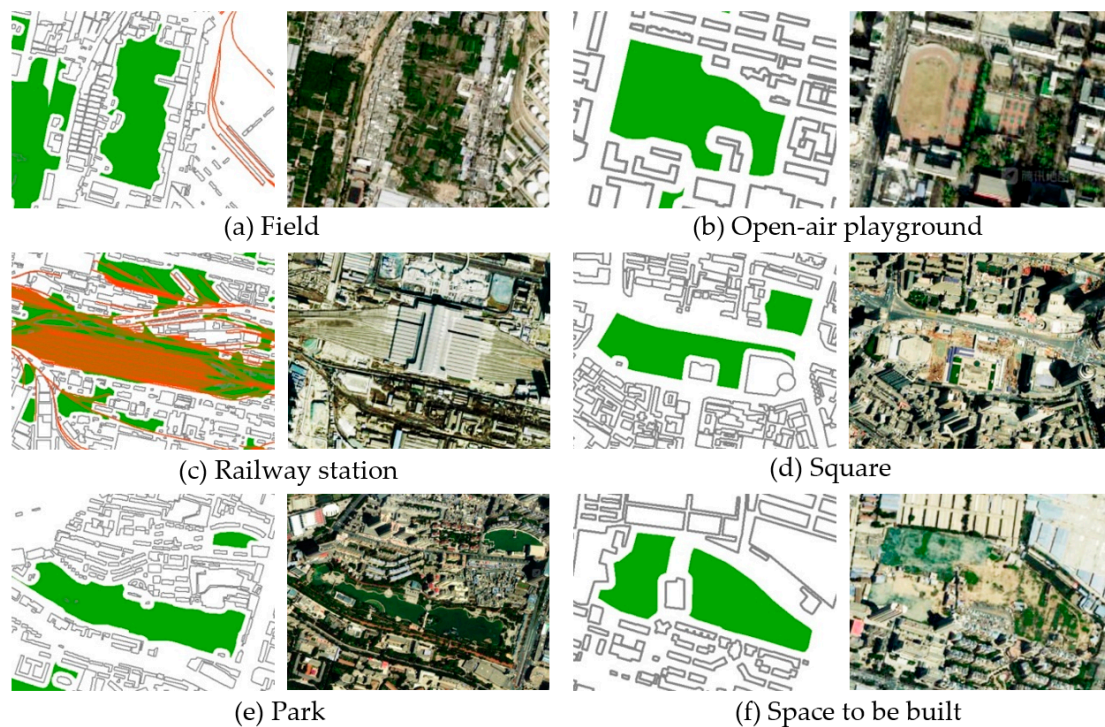


Figure 14. Examples of free spaces obtained from 1:10K buildings (left) and image map (right): (a) field; (b) open-air playground; (c) railway station; (d) square; (e) park; and (f) space to be built.

3.2.2. Quantitative Evaluation

Equation (2) is used to calculate the similarity between the result and the standard data. Moreover, the similarity between the results and the 1:250K residential areas and the results and the 1:1M data are measured (Table 4).

Table 4. The similarity between the resulting urban areas and the standard residential areas at varying scales.

Object Data (Resulting Data) and Reference Data	Common Area (m ²)	Symmetrical Difference Area (m ²)	Similarity
Result, 1:10K	53,260,312	68,916,946	0.61
Result, 1:50K	88,326,204	44,336,786	0.80
Result, 1:250K	89,024,948	48,421,362	0.79
Result, 1:1M	89,229,289	60,206,001	0.75

Table 4 shows that similarity varies with the reference data. A general trend is that the similarity between the obtained data (equivalent to 1:50K data) and the reference data decreases as the scale difference between the reference data and the result data increases, but the rate of the decrease is relatively slow, especially for the 1:250K data, whose similarity value (0.79) is almost the same as that between *UrbanAreasFinal* and the 1:50K data. This is because the actuality (i.e., the time the data is updated) of the 1:50K and 1:250K data are relatively consistent (Table 1). Additionally, the similarity value between the obtained data and the 1:1M data reaches 0.75, which is also not small, due to the fact that the actuality of the 1:1M data is close to that of the original 1:10K data. Both of them are relatively new among the four scales.

3.2.3. Satisfaction of Requirements in Practice

Table 5 provides a qualitative comparison of the traditional approaches with the proposed method based on six criteria. These six criteria are commonly used standards in practice and include position correctness, updating cycle, data consistency, etc. It can be seen that the traditional method has a long data updating cycle due to the repetitive manual updating mode. As a result, before the small-scale data are updated, the data consistency between the large-scale (1:10K) buildings, roads, and small-scale (1:50K) residential areas is not good. Moreover, the correctness of the proposed method is ensured by Boffet's method as well as by KDE. It can delineate urban areas objectively and automatically, while the data quality of the traditional method mainly depends on the cartographer's experience and skill, and it is subjective.

Table 5. Satisfaction of requirements for ideal data generalization method.

Criterion	Traditional Method	Proposed Method
1:10K and 1:50K data updating mode	Manual updating	1:10K data updating + 1:50K data generalization from the newest 1:10K data
Data updating cycle	Longer (1:10K data manual updating cycle + 1:50K data manual updating cycle)	Shorter (1:10K manual updating cycle + automatic 1:50K data generalization time)
Consistency between 1:10K data and 1:50K data	Not good before updating is finished	Good
Correctness of outer boundaries and free spaces	Depends on the cartographer's experience and skill	Correctness can be ensured due to the dilating and eroding buffers of updated buildings and roads
Threshold determination automation	/	Yes
Method is objective or Subjective	Subjective	Objective

4. Discussion

This study provides an example of residential area generalization using a knowledge-based framework with a scale reduction of five times. In the map generalization process, geometric similarity measures are used to determine the KDE threshold. A number of insights can be gained from this study.

(1) In practice, the delineation of urban areas is mainly finished manually by cartographers using big scale data, e.g., 1:10K buildings. The artificial determination of urban boundaries is inevitably subjective. In addition, the biggest drawback of the manual method is that the multi-scale databases are maintained and updated separately, which not only leads to repeated work, but also makes it difficult to maintain consistency among multi-scale datasets. The result of this study is a 1:50K residential area dataset in multiple representation databases. The proposed method can delineate the urban areas objectively, which is in accordance with the “natural city” proposed by Jiang [40].

(2) Through geometric similarity calculations of the residential areas from 1:10K to 1:1M, it was found that the similarity values decreased when the map scale reduced. In addition, 1:50K is an essential scale from which residential areas are represented from individual buildings to city blocks. From this insight, it can be inferred that different scale maps should focus on different tasks: a 1:10K map is mainly for careful observation and measurement, whereas a 1:50K map should mainly be used to express the overall rule of a whole mapping area. This is because it is consistent with human beings’ intuition, which is also an important standard for a similarity measure.

(3) Table 4 and Figure 8 show that the similarity values between the result data and the reference data from 1:10K to 1:1M reaches its peak (0.80) when the scale is 1:50K. This result demonstrates that compared to other data, the result is the most similar to the 1:50K reference data. Additionally, the differences between the two lines reflect the differences in the field conditions. This is because the result is abstracted from the newest 1:10K data (2017–2019), but the actuality of the existing 1:50K data is 2012–2014, and during this period, the field situation has changed greatly.

(4) It can be inferred from (3) that, for a city that has developed to a relatively mature stage or for a city whose sprawl is limited by specific factors, such as Lanzhou, there is less and less land for it to expand to due to the fact that its land in the suitable terrain slope range on which people can live is becoming less and less, making this method applicable. This is because the changes that took place between 2012 and 2019 were in an acceptable range. On the contrary, for a city that sprawls quickly, e.g., a newly built city, geometric similarity cannot be measured by the proposed method.

(5) The analysis of the experiment process and the results shows that: knowledge plays a fundamental role in intelligent generalization. Some knowledge can be formalized; however, it is not easy to adequately obtain and record other types of knowledge. Most of the knowledge in the experiment was acquired by map analysis and spatial reasoning. Additionally, from another point of view, similarity quantification is a type of knowledge that is hidden in the minds of cartographers.

5. Conclusions and Future Work

This article presents a hybrid method that uses a knowledge-based framework and similarity measurements to generalize residential area maps in the scale ranging from 1:10K to 1:50K. The generalization process is divided into two parts: (1) identifying the boundaries of built-up areas; and (2) acquiring the free space within those boundaries. For the latter, Boffet’s method was used; for the former, a hybrid method using Boffet’s method and kernel density analysis was utilized. Map analysis and similarity measurements were applied to acquire knowledge about parameters in the two methods. Generalization steps were created based on the knowledge acquired.

The proposed approach was tested using multi-scale datasets of Lanzhou City. A manual comparison between the results and the existing 1:50K residential areas show that the results are reasonable and in accordance with the actual field situation; in addition, the similarity values between the results and the reference 1:50K data are the largest among the calculated similarity values. However, there are several limitations that must be considered when using this model. The similarity calculation is based on the overlay of multi-scale maps; its result depends on the coordinate consistency and the time span between data sources. For example, as described in the Discussion (4), if the 1:50K data are very old and

the 1:10K data are updated within a short cycle, then the 1:50K data cannot be employed to act as a reference that the kernel density analysis can be compared to. Moreover, the obtained result still can be improved through the use of a specific algorithm to avoid unpleasant curves caused by buffer-internal buffer operation.

Overall, the approach is robust, comparatively insensitive to the noise of the small buildings beyond the urban areas, and easy to implement in the GIS software. It should be noted that data model harmonization and topological relationship correctness after generalization are not considered, which are critical for achieving scientific automatic generalization. Our future work will concentrate on two aspects, including the generalization of residential areas by taking into account how to maintain their topological relationships with city roads.

Author Contributions: Xiaorong Gao and Haowen Yan conceived the knowledge-based and similarity measurement approach together; Xiaorong Gao collected the data, designed the experiments, and wrote the manuscript; Xiaomin Lu and Pengbo Li performed the analysis; Haowen Yan modified the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant No. 41930101, 42161066 and 41801395), Department of Education of Gansu Province: The Excellent Postgraduate Student “Innovation Star” Project (Grant No. 2021CXZX-549), and the LZJTU EP (Grant No. 201806).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ruas, A. Automating the generalisation of geographical data: The age of maturity? In Proceedings of the 20th International Cartographic Conference, Beijing, China, 6 August 2001.
2. Lee, D.; Hardy, P. Automating Generalization—Tools and Models. In Proceedings of the XXII International Cartographic Conference, La Coruna, Spain, 11–16 July 2005.
3. Duchêne, C.; Ruas, A.; Cambier, C. The CartACom model: Transforming cartographic features into communicating agents for cartographic generalization. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1533–1562. [\[CrossRef\]](#)
4. Yang, M.; Ai, T.; Yan, X.; Chen, Y.; Zhang, X. A map-algebra-based method for automatic change detection and spatial data updating across multiple scales. *Trans. GIS* **2018**, *22*, 435–454. [\[CrossRef\]](#)
5. Yu, W.; Zhang, Y.; Chen, Z. Automated generalization of facility points-of-interest with service area delimitation. *IEEE Access* **2019**, *7*, 63921–63935. [\[CrossRef\]](#)
6. Yan, H.; Li, J. *Spatial Similarity Relations in Multi-Scale Map Spaces*; Springer International Publishing: Cham, Switzerland, 2015.
7. Weibel, R.; Keller, S.; Reichenbacher, T. Overcoming the knowledge acquisition bottleneck in map generalization: The role of interactive systems and computational intelligence. In Proceedings of the 2nd International Conference on Spatial Information Theory (COSIT 95), Semmering, Austria, 21 September 1995.
8. Kilpeläinen, T. Knowledge Acquisition for Generalization Rules. *Cartogr. Geogr. Inf. Sci.* **2000**, *27*, 41–50. [\[CrossRef\]](#)
9. Mustiere, S. Cartographic generalization of roads in a local and adaptive approach: A knowledge acquisition problem. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 937–955. [\[CrossRef\]](#)
10. Douglas, D.H.; Peucker, T.K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Can. Cartogr. J.* **1973**, *10*, 112–122. [\[CrossRef\]](#)
11. Regnauld, N.; Edwardes, A.; Barrault, M. Strategies in building generalization: Modelling the sequence, constraining the choice. In Proceedings of the 19th ICC Workshop on Progress and Developments in Automated Map Generalization, Ottawa, ON, Canada, 14–21 August 1999.
12. Sester, M. Optimizing approaches for generalization and data abstraction. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 871–897. [\[CrossRef\]](#)
13. Bayer, T. Automated building simplification using a recursive approach. In *Cartography in Central and Eastern Europe*; Gartner, G., Ortog, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 121–146.
14. Yan, X.; Ai, T.; Zhang, X. Template matching and simplification method for building features based on shape cognition. *Int. J. Geo-Inf.* **2017**, *6*, 250. [\[CrossRef\]](#)
15. Yang, M.; Yuan, T.; Yan, X.; Ai, T.; Jiang, C. A hybrid approach to building simplification with an evaluator from a backpropagation neural network. *Int. J. Geogr. Inf. Sci.* **2021**, *5*, 1–30. [\[CrossRef\]](#)
16. Su, B.; Li, Z.; Lodwick, G. Jean-Claude Muller Algebraic models for the aggregation of area features based upon morphological operators. *Int. J. Geogr. Inf. Sci.* **1997**, *11*, 233–246. [\[CrossRef\]](#)

17. Li, Z.; Yan, H.; Ai, T.; Chen, J. Automated building generalization based on urban morphology and Gestalt theory. *Int. J. Geogr. Inf. Sci.* **2004**, *18*, 513–534. [\[CrossRef\]](#)
18. QI, H.; Li, Z. An Approach to Building Grouping Based on Hierarchical Constraints. In the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B2 (pp. 449–454). Available online: http://www.isprs.org/proceedings/XXXVII/congress/2_pdf/3_WG-II-3/13.pdf (accessed on 10 October 2021).
19. Bader, M. Energy Minimization Methods for Feature Displacement in Map Generalization. Ph.D. Thesis, Universität Zürich, Zürich, Switzerland, 2001.
20. Ai, T.; Zhang, X.; Zhou, Q.; Yang, M. A vector field model to handle the displacement of multiple conflicts in building generalization. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 1310–1331. [\[CrossRef\]](#)
21. Pilehforooshha, P.; Karimi, M.; Mansourian, A. A new model combining building block displacement and building block area reduction for resolving spatial conflicts. *Trans. GIS* **2021**, *25*, 1366–1395. [\[CrossRef\]](#)
22. Brassel, K.; Weibel, R. A Review and Conceptual Framework of Automated Map Generalization. *Int. J. Geogr. Inf. Syst.* **1988**, *2*, 229–244. [\[CrossRef\]](#)
23. McMaster, R.B.; Shea, K.S. *Generalization in Digital Cartography*; Association of American Geographers: Washington, DC, USA, 1992.
24. Galanda, M. Automated Polygon Generalization in a Multi Agent System. Ph.D. Thesis, Universität Zürich, Zürich, Switzerland, 2003.
25. Sarjakoski, L.T. Conceptual Models of Generalisation and Multiple Representation. In *Generalisation of Geographic Information: Cartographic Modelling and Applications*; Mackaness, W., Ruas, A., Sarjakoski, L.T., Eds.; Elsevier: Oxford, UK, 2007; pp. 11–36.
26. O’Keefe, J.; Dostrovsky, J. The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **1971**, *34*, 171–175. [\[CrossRef\]](#)
27. O’Keefe, J.; Nadel, L. *The Hippocampus as a Cognitive Map*; Clarendon: Oxford, UK, 1978.
28. Hafting, T.; Fyhn, M.; Molden, S.; Moser, M.; Moser, E. Microstructure of a spatial map in the entorhinal cortex. *Nature* **2005**, *436*, 801–806. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Yan, X.; Ai, T.; Yang, M.; Tong, X.; Liu, Q. A graph deep learning approach for urban building grouping. *Geocarto Int.* **2020**, *8*, 1–24. [\[CrossRef\]](#)
30. Li, Z.; Huang, P. Quantitative measures for spatial information of maps. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 699–709. [\[CrossRef\]](#)
31. Mackaness, W.A.; Burghardt, D.; Duchêne, C. Map Generalization. In *International Encyclopedia of Geography: People, the Earth, Environment and Technology*; Richardson, D., Castree, N., Goodchild, M.F., Kobayashi, A., Liu, W., Marston, R.A., Eds.; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2017. [\[CrossRef\]](#)
32. Boffet, A.; Serra, S.R. Identification of spatial structures within urban blocks for town characterization. In Proceedings of the 20th International Cartographic Conference, Beijing, China, 6 August 2001; pp. 1974–1983.
33. Chaudhry, O.; Mackaness, W. Automatic identification of urban settlement boundaries for multiple representation databases. *Comput. Environ. Urban Syst.* **2008**, *32*, 95–109. [\[CrossRef\]](#)
34. Liu, Y.; Martin, M.; Menno-Jan, K. Semantic Similarity Evaluation Model in Categorical Database Generalization. In Proceedings of the Symposium on Geospatial Theory, Processing and Applications, Ottawa, ON, Canada, 9–12 July 2002.
35. Zhou, Q. Comparative Study of Approaches to Delineating Built-Up Areas Using Road Network Data. *Trans. GIS* **2015**, *19*, 848–876. [\[CrossRef\]](#)
36. Li, Y.; Sun, Q.; Ji, X.; Xu, L.; Lu, C.; Zhao, Y. Defining the Boundaries of Urban Built-up Area Based on Taxi Trajectories: A Case Study of Beijing. *J. Geovisualization Spat. Anal.* **2020**, *4*, 8. [\[CrossRef\]](#)
37. Burghardt, D.; Steiniger, S. Usage of principal component analysis in the process of automated generalisation. In Proceedings of the 22nd International Cartographic Conference, La Coruna, Spain, 9–16 July 2005; International Cartographic Association (ICA): La Coruna, Spain, 2005.
38. Du, S.; Luo, L.; Cao, K.; Shu, M. Extracting building patterns with multilevel graph partition and building grouping. *Isprs J. Photogramm. Remote Sens.* **2016**, *122*, 81–96. [\[CrossRef\]](#)
39. Chaudhry, O.; Mackaness, W. Visualisation of Settlements over Large Changes in Scale. In Proceedings of the 8th ICA Workshop on Generalisation and Multiple Representation, La Coruna, Spain, 9–16 July 2005.
40. Jiang, B.; Liu, X. Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 215–229. [\[CrossRef\]](#)
41. Smith, B. On drawing lines on a map. In *Spatial Information Theory, Proceedings of COSIT ’95 Berlin/Heidelberg/Vienna/New York/London/Tokyo, Vienna, Austria, 21–23 September 1995*; Frank, A.U., Kuhn, W., Mark, D., Eds.; Springer: Berlin/Heidelberg, Germany, 1995; pp. 475–484.
42. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: New York, NY, USA, 1986.
43. Thurstain-Goodwin, M.T.; Unwin, D. Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations. *Trans. GIS* **2000**, *4*, 305–317. [\[CrossRef\]](#)
44. Borruso, G. Network Density and the Delimitation of Urban Areas. *Trans. GIS* **2003**, *7*, 177–191. [\[CrossRef\]](#)
45. Borruso, G. Network Density Estimation: A GIS Approach for Analysing Point Patterns in a Network Space. *Trans. GIS* **2010**, *12*, 377–402. [\[CrossRef\]](#)

46. Jia, T.; Jiang, B. Measuring Urban Sprawl Based on Massive Street Nodes and the Novel Concept of Natural Cities. *arXiv* **2011**, arXiv:1010.0541. WWW Document. Available online: <https://arxiv.org/ftp/arxiv/papers/1010/1010.0541.pdf> (accessed on 10 October 2021).
47. Tversky, A. Features of similarity. *Psychol. Rev.* **1977**, *84*, 327–352. [[CrossRef](#)]
48. Holt, A. Spatial Similarity and GIS: The Grouping of Spatial Kinds. In Proceedings of the 11th Annual Colloquium of the Spatial Information Research Centre, Dunedin, New Zealand, 13–15 December 1999.
49. Popper, K.R. *The Logic of Scientific Discovery*; Hutchinson: London, UK, 1972; 480p.
50. Ai, T.; Ke, S.; Yang, M.; Li, J. Envelope generation and simplification of polylines using Delaunay triangulation. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 297–319. [[CrossRef](#)]
51. Arkin, E.M.; Chew, L.P.; Huttenlocher, D.P.; Kedem, K.; Mitchell, J.S.B. An Efficiently Computable Metric for Comparing Polygonal Shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 209–216. [[CrossRef](#)]
52. Samal, A.; Seth, S.; Cueto, K. A feature-based approach to conflation of geospatial source. *Int. J. Geogr. Inf. Sci.* **2004**, *18*, 459–489. [[CrossRef](#)]
53. Frank, R.; Ester, M. A quantitative similarity measure for maps. In *Progress in Spatial Data Handling*; Riedl, A., Kainz, W., Elmes, G.A., Eds.; Springer: Berlin, Germany, 2006; pp. 435–450.
54. Goodchild, M.F.; Hunter, G.J. A simple positional accuracy measure for linear features. *Int. J. Geogr. Inf. Sci.* **1997**, *11*, 299–306. [[CrossRef](#)]
55. Winecoff, A.A.; Brasoveanu, F.; Casavant, B. Users in the loop: A psychologically-informed approach to similar item retrieval. In Proceedings of the 13th ACM Conference on Recommender Systems, Copenhagen, Denmark, 16–20 September 2019.
56. Wang, Z.; Müller, J.-C. Line Generalization Based on Analysis of Shape Characteristics. *Cartogr. Geogr. Inf. Syst.* **1998**, *25*, 3–15. [[CrossRef](#)]
57. Rodríguez, M.A. Assessing Semantic Similarity among Spatial Entity Classes. Ph.D. Thesis, University of Maine, Orono, ME, USA, 2000.