A Template Matching Method for Enhancing and Generalizing OpenStreetMap Building Data

Tinghua Ai, Xiongfeng Yan, and Xiang Zhang

School of Resource and Environmental Sciences, Wuhan University, Wuhan, China;

tinghuaai@whu.edu.cn, xiongfeng.yan@whu.edu.cn, xiang.zhang@whu.edu.cn

Abstract:

With the favorable characteristics of wide coverage, free sharing, quasi real-time and so on, Volunteer Geographic Information (VGI) is rapidly becoming a significant supplement of the traditional and professional geographic information, such as the OpenStreetMap (OSM) (Goodchild, 2007; Elwood, 2008). However, this type of data also has some obvious problems because it is gathered by individuals with no professional training, for example, the quality of data is not standardized and the scale is not unified. Therefore, further processes are essential to perform analysis and mapping on VGI, such as the squaring and generalization operations on the OSM building layer (Lokhat & Touya, 2016).

By considering the typical template characteristics of building distributions and representations, this study proposes a template matching method for enhancing and generalizing OSM buildings. Our method first formulates a series of templates to abstract the building shape by generalizing their polygons and analyzing their symbolic meanings, then conducts the generalization by searching and matching the most similar template that can be used later to replace the original building. The propose method includes two stages for shape measurement and template matching, in the first step, we employ the turning function as shape descriptor to measure the similarity to find the best candidate template for original buildings. In the second step, we carry out a geometry transformation to adjust the position of the candidate template based on the least squares model.

To assess the reasonability and feasibility of the proposed method, experiments on the test area that covers a region around **the eastern part of Lund** are carried out. With respect to this dataset, we manually create two levels of template libraries, the first level contains the templates in more detail, but the second more abstract. To ensure the individual geometric accuracy, we define the surface distance by comparing the areas of the intersection and union between building and template, and introduce a rejection threshold for it. In this cases, we set the threshold to 0.3 for the first level templates and the correctness is 95.1%. For the second one, we use a smaller value for the threshold, such as 0.4, to obtain a more abstract generalized result, and the correctness is 92.8%. From these experiments, two conclusions can be drawn: firstly, from the perspective of data processing, the template may make data more standard, to some extent overcome the data quality issues of VGI, for example, keeping the corners as a right angle. Secondly, from the perspective of representations, the same template can be used for these similar buildings in a certain region to make their shapes more regular and their scales more uniform. What's more, for these special shapes, such as alphabetic building, some symbolize templates can be used to enhance the cognitive nature of the map.

Through in-depth analysis we found that there are three main aspects that may cause errors: (1) the shape of building is too complicated to find a suitable template in the library; (2) there is a suitable template but a wrong

template is found when calculating the similarity distance; and (3) a destructive offset of position and orientation occurs when the transformation is performed. For the first type, the completeness of template library has an impact on the quality of generalization, thus, it should be further enriched and extended in the future, for instance, by considering the situation where buildings are complicated with more than one hole and using other sources of information (e.g., street network) to avoid some potential conflicts. The second and third types of error are what might have been expected because of the actuality that there is no method can achieve full correctness for shape representation and similarity measurement, and some high-performance machine learning methods also can be introduced to identify shape structures and community patterns of buildings, then templates may be applied to building communities, which will greatly expand the scale range of generalization.

Keywords: VGI, Template matching, Map generalization, Building data

Mapmaking

We propose a template matching method for enhancing and generalizing OSM buildings. We use two levels of template libraries to generalize buildings, and the comparisons before and after the operation are showed in figure 1 and figure 2, respectively. The processes of map making are as follows.

Step 1. Convert the original OpenStreetMap data to a Gaussian projection.

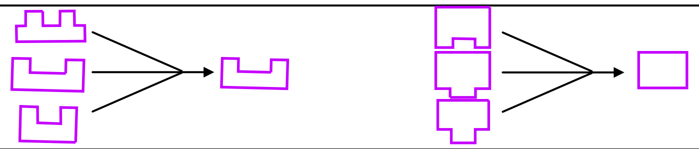
Step 2. With respect to the characteristics of this dataset, we create two levels of template libraries manually. The first level template library contains 91 templates, including standard geometries, symbolic symbols with rich details (e.g., L-, T-, U-shaped), and some special templates derived from the typical building shapes in the dataset. In the template library at this level, templates are detailed and diversified, and there may be several templates that are pertinent to the same symbolic meaning, twelve examples are shown in table 1. The second level template library contains 23 templates, mainly including some standard geometries, such as square, rectangle and circle-shaped. At this level, templates will be further abstracted, and these templates with slightly different shapes will be represented as one, two examples are shown in table 2.

Table 1 Examples of the first-level template library for the test dataset.

Categories	**Template**			
Simple template	Square-shaped	Rectangle-shaped	Hexagon-shaped	Circle-shaped
Symbolic template	L-shaped	T-shaped	U-shaped	E-shaped
Composite shape				

Table 2 Examples of the second-level templates abstracted from the first-level templates.

Level-1	Level-2	Level-1	Level-2



Step 3. Then, we can obtain two different levels generalized map by using these two template libraries. For each library, perform the following steps for each building in the dataset:

Step 3.1 Calculate the similar distances with all template in this library, and find the 'most similar' one as a candidate generalized version for the building. In this step, the turning function that describes shape by measuring the changes of the tangent-angle as a function f of the arc-length s is employed to obtain this similarity

(Arkin,1991).

Step 3.2 Find the matching correspondence between the outline of this building and its candidate template, a dynamic programming algorithm proposed by Nöllenburg et al. (2008) is used in this step.

Step 3.3 For the matching correspondence, solve the four-parameters (the scale component k, the rotation component θ, and the translation components Δx and Δy) of a 2D geometry transformation model based on the least squares adjustment (Yan et al. 2017).

Step 3.4 Scale, rotate and translate the candidate template to match this building at an appropriate location, and calculate the surface distance between the adjusted template and building by comparing the areas of the intersection and union between them.

Step 3.5 If the surface distance is smaller than the given threshold (in our case, it was set as 0.3 and 0.4 in the two-levels libraries, respectively), the matching operator is available, and original building will be replaced by the adjusted template. Otherwise, the operator will be rejected.

Step 4. Render and export the map.

Implementation

The code will be released as an open-source software soon.

References

Goodchild, M.F. (2007). Citizens as Sensors: The World of Volunteered Geography. *GeoJournal.* 69(4): 211-221.

Elwood, S. (2008). Volunteered Geographic Information: Future Research Directions Motivated by Critical, Participatory, and Feminist GIS. *GeoJournal*. 72(3&4): 173-183.

Lokhat, I. & Touya, G. (2016). Enhancing Building Footprints with Squaring Operations. *Journal of Spatial Information Science*, 13: 1-28.

Arkin, E.M. et al. (1991). An Efficiently Computable Metric for Comparing Polygonal Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3): 209–216.

Yan X. et al. (2017). Template Matching and Simplification Method for Building Features Based on Shape Cognition. *ISPRS International Journal of Geo-Information*, 6(8): 250.

Nöllenburg, M. et al. (2008). Morphing Polylines: A Step towards Continuous Generalization. *Computer Environment and Urban System*, 32(4): 248–260.

Figures

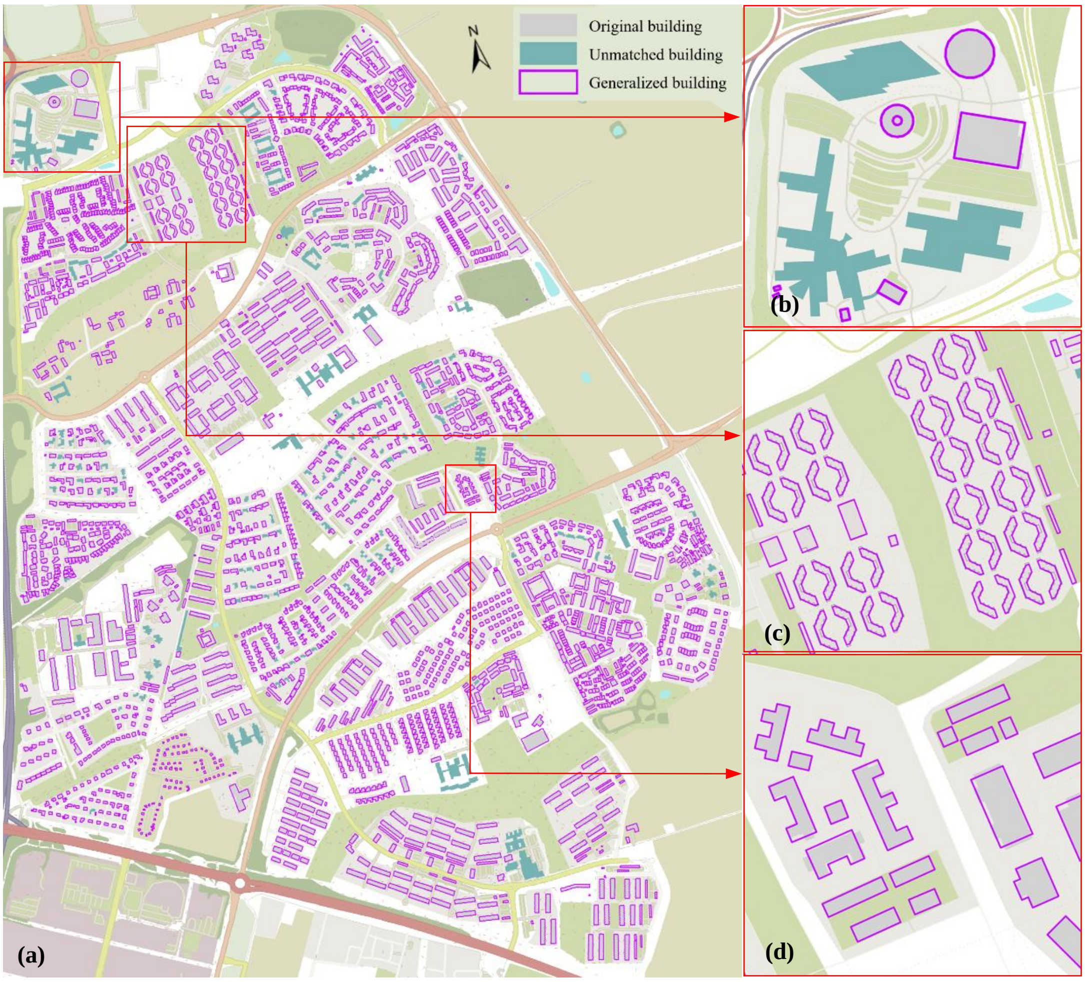


Figure 1. The generalized map by using the first level template library: (**a**) The comparison of buildings before and after generalization; (**b**-**d**) showing the local details. As we can see, a majority of the buildings (95.1%) in the dataset are generalized by the predefined templates, and their shapes are well-preserved, but there are still some buildings that are far from satisfactory (green polygons, as seen in figure 1b). Some building polygons that have similar characteristics are generalized successfully by the same template, as seen in Figure 1c, which will make their shapes more regular and uniform. For these buildings have an auxiliary cognizance, for example, the shapes are similar to letters, they can also be generalized by the symbolic templates, and their main tendencies are maintained after generalization, as seen in Figure 1d.

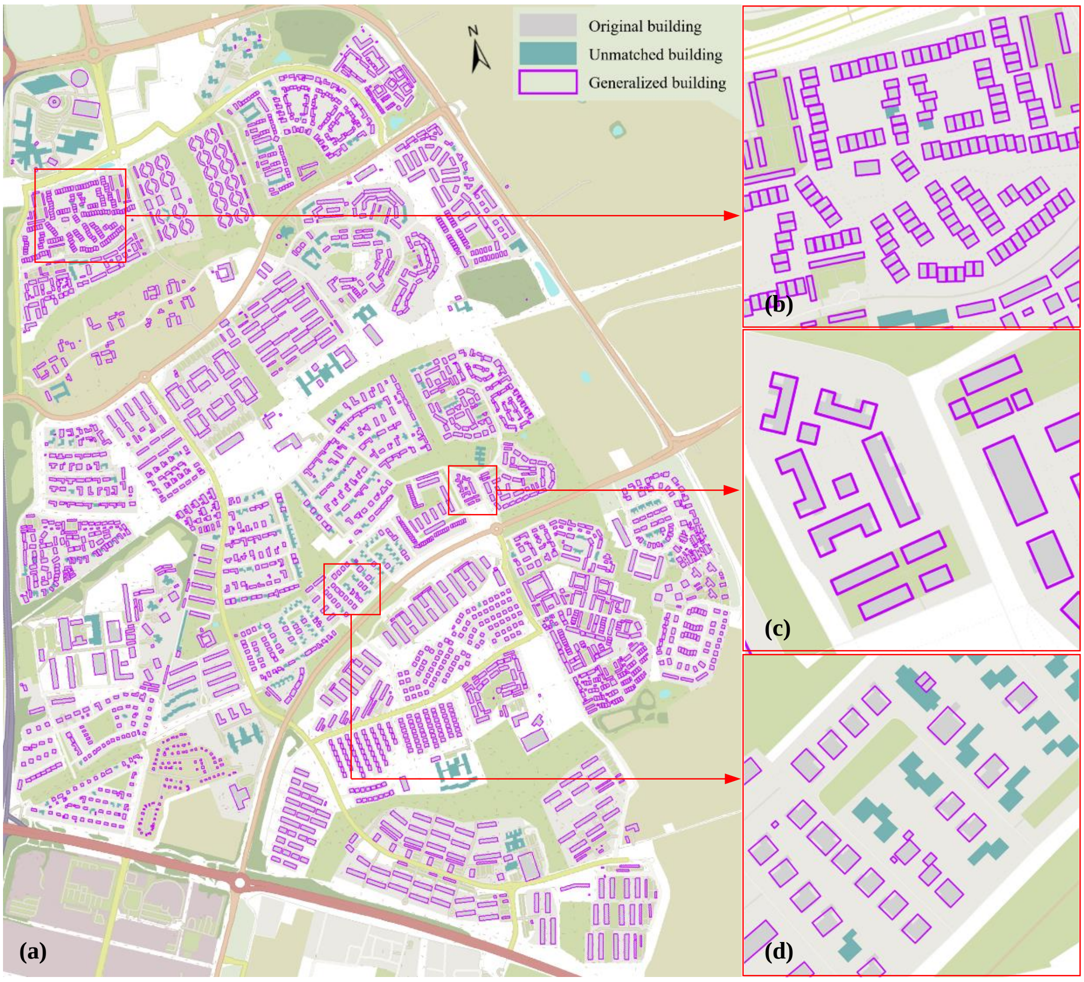


Figure 2. The generalized map by using the second level template library: (**a**) The comparison of buildings before and after generalization; (**b-d**) showing the local details. As we can see, if we use a more abstract template library for generalization, the generalized map will be more succinct and concise, but the geometric accuracy will be lower. We use 0.4 for the surface distance, and the correctness is 92.8%. Detailed comparisons and analyses are shown in figure 3.

Figure 3. The comparison of these two levels generalized maps: (**a**)(**c**)(**e**) showing the local details from the first level generalized map, and (**b**)(**d**)(**f**) showing the local details from the second level generalized map. The main templates used in each figure are shown in the upper-right corner. From the figure 3a and 3b we can see, these buildings generalized by different templates at the first level are eventually generalized by the same one at the second level, for example, the building A, B, C, and D. In addition, we see some buildings are generalized by a simpler template, for example, the building E, but the geometric accuracy will be lower. This situation can also be seen from the figure 3c and 3d. However, as the templates become simpler and more abstract, some buildings will not find a suitable template, as we can see in figure 3e and 3f, this is the reason for the lower correctness in second level generalized map.