

# A Multilane Roads Detection Approach for Urban Transport Skeleton Knowledge Discovery

Bisheng Yang<sup>1</sup>, Hongchao Fan<sup>2</sup>, Xuechen Luan<sup>1</sup>

<sup>1</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, China  
<sup>2</sup> Chair of GIScience, University of Heidelberg, 69120 Heidelberg, Germany

Telephone: (+86) 27 68779689

Fax: (+86) 27 68778043

Email: [bsyang@whu.edu.cn](mailto:bsyang@whu.edu.cn), [hongchao.fan@geog.uni-heidelberg.de](mailto:hongchao.fan@geog.uni-heidelberg.de), [xuechen.luan@whu.edu.cn](mailto:xuechen.luan@whu.edu.cn)

## 1 Introduction

Multilane roads in high level-of-detail (LoD) datasets normally represent the constructed high-level roads of urban road network, which have large capacity of transportation and form the pattern of urban movement flows. They are crucial elements for many modern applications such as traffic planning, location-based services (LBS) and emergency evacuation. Nowadays, more and more Volunteered Geographical Information (VGI) with high LoD but low data quality control (DQC) is available, in which the multilane attributes are usually not explicitly stored in data attributes. The existing method (e.g. Yang et al. 2011) is low efficient to extract reasonable multilane roads if the data quality is low. Therefore, a different strategy improving the effectiveness/efficiency needs to be defined.

The algorithm in the presented work is tailored for efficiently extracting multilane roads from OpenStreetMap (OSM) road network, in which there is no attribute indicating the multilane road. Because of lack of data checking process, several duplicated lines representing the same road feature exist independently in the dataset, increasing the difficulties and time consuming of extracting the multilane roads. Moreover, some road lines in the non-professional OSM data may have several errors, such as dangles, broken roads, singular angles, and so on. These issues cause great challenges for extracting multilane roads from the whole networks when using the existing approaches. Instead of road line data analysis, in our research multilane roads are extracted based on polygon analysis. It is developed based on the character that multilane roads are often digitized as multiple parallel lanes, which constitute several long and thin polygons that clearly distinguished from block polygons. Hence, the long and thin polygons can be used as seeds to detect entire multilane road from networks. Besides, the polygon-based approach is probable to overcome the inefficiencies of previous line-based approaches because the number of polygons is much less than the number of corresponding lines.

## 2 Methodology

Our approach solves the problem from both the semantic and spatial two perspectives, which is composed of two key steps as follows.

## 2.1 Semantic detection

To describe the shape characteristic of polygons in a road network, five shape descriptors namely area, perimeter, compactness, parallelism and elongation are used, which are described as follows,

- **Area, Perimeter and Compactness**

The relation among these three descriptors is shown as,

$$compactness = 4\pi \times \frac{area(polygon)}{perimeter^2(polygon)} \quad (1)$$

The value of *compactness* ranges from 0 to 1. The more complex the outline of polygon is, the smaller the compactness value is. Compared with other polygons (e.g. blocks), most multilane polygons usually have small areas, perimeters, and compactness.

- **Parallelism of polygon**

Most multilane polygons also have two primary parallel road lines. The *parallelism* descriptor is proposed in our approach to measure degree of parallel pattern. The formulation of *parallelism* is described as

$$parallelism = \frac{\sum_1^n length(line_i) \cdot x_i}{perimeter(polygon)} \quad (2)$$

Suppose there are  $n$  line segments belonging to the boundary of the polygon. If the angle between  $line_i$  and the polygon's primary direction is smaller than threshold, then the  $x_i$  value is 1; otherwise the value is 0. The polygon's primary direction is defined as the long edge direction of the polygon's minimum boundary rectangle (MBR). This *parallelism* descriptor therefore calculates the ratio between the sum of the lengths of the boundary lines similar to the polygon's primary direction and the polygon's perimeter. The value of *parallelism* also ranges from 0 to 1, and the larger the *parallelism* is, the more likely the polygon is a multilane one.

- **Elongation of polygon**

The *elongation* descriptor is thus defined as the length-area ratio of polygon to indicate the long and narrow shape, which is calculated as,

$$elongation(polygon) = \frac{length(centerline)}{area(polygon)} \quad (3)$$

The central line of a polygon is extracted from Delaunay triangles within the polygon. Based on the triangles inner polygons, the central line can then be extracted according to the central line of triangles (Ai and Guo 2000). On the one hand, the longer the central line and the smaller the area of polygon is, the more likely the polygon is a multilane one. On the other hand, if the central line is short but the area of polygon is also small at the same time, the polygon can also be regarded as multilane polygons with large elongation value. This case occurs most often in the representation of multi-lanes. If the polygon is small, it is unneeded to have central line as long as large polygons in expressway.

Instead of setting threshold of each shape descriptor, these five shape descriptors are defined as the 5-dimensional vector input vector for SVM classification. The SVM

package libsvm (Chang and Lin 2011) is adopted to classify the potential multilane polygons.

## 2.2 Spatial detection

These polygons discriminated by SVM need to be complemented by connecting operations according to the spatial relationship. In the presented work, a regional growing strategy is proposed regarding these polygons detected by SVM as seed polygons to complement the whole multilane roads. It iteratively compares the spatial similarity of adjacent polygons, which includes the arrangement of polygons and the new width and length of connected polygons. For each adjacent two polygons  $P_i$  and  $P_j$ , it calculates the two acute angles  $A_{ij}$  and  $A_{iB}$  to determine the arrangement of polygons, in which the angle  $A_{ij}$  is the acute angle of the major orientations of  $P_i$  and  $P_j$ , and the angle  $A_{iB}$  is the acute angle between the common boundary  $B_{ij}$  and the major orientations of polygon  $P_i$ . Base on that, the arrangement between two polygons can be categorized into three cases. For each group, road length and width are updated respectively. Suppose the length of the common boundary  $B_{ij}$  is represented as  $L_{boundary}$ , and the lengths and widths of polygon  $P_i$  and  $P_j$  are represented as  $L_i$ ,  $W_i$  and  $L_j$ ,  $W_j$  respectively. The three cases in updating process are thus described as follows.

**Case 1:** If the both the angles  $A_{ij}$  and  $A_{iB}$  are smaller than 45 degree, as illustrated in figure 6-1, they will be regarded as the side-to-side pair. The new length of these two polygons is calculated as

$$L_{new} = L_i + L_j - L_{boundary} \quad (4)$$

**Case 2:** If only the angle  $A_{ij}$  is smaller than 45 degree, as illustrated in figure 6-2, they will be regarded as the end-to-end pair. The new length of these two polygons is calculated as

$$L_{new} = L_i + L_j \quad (5)$$

**Case 3:** If the angle  $A_{ij}$  is larger than 45 degree, they will be regarded as the side-to-end or end-to side pair, where if  $A_{iB}$  is smaller than 45 degree, as illustrated in figure 6-3, the new length of these two polygons is calculated as

$$L_{new} = L_i + W_i + L_j - L_{boundary} \quad (6)$$

Else if  $A_{iB}$  is larger than 45 degree, as illustrated in figure 6-4, the new length of these two polygons is calculated as

$$L_{new} = L_i + W_j + L_j - L_{boundary} \quad (7)$$

For all the three cases, the new width of these two polygons is calculated as

$$W_{new} = (area_i + area_j) / L_{new} \quad (8)$$

**Step 3:** If the new width  $W_{new}$  of two polygons is smaller than the width threshold, the polygon  $P_i$  and  $P_j$  are connected as the new detected multilane polygons and pushed into seed polygons set. Otherwise, the connecting algorithm will judge the other adjacent candidate polygons of polygon  $P_i$ .

**Step 4:** The connecting algorithm will retrieve all the seed polygons and stop when no more seeds are left.

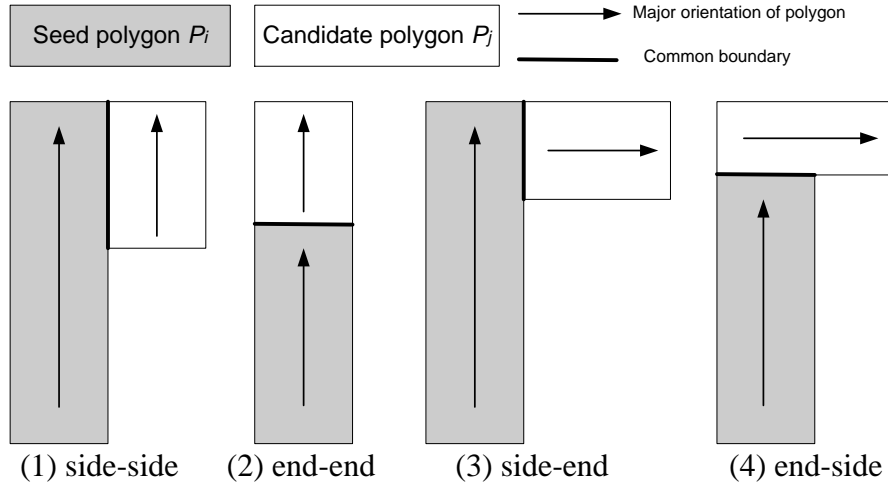


Figure 1 Arrangement cases of two polygons

### 3 Experiments

To verify the validity of the proposed approach, road networks of Munich from the OSM datasets was tested. It is originated from precisely captured real-world roads and the data volume is quite large. There are 240321 roads and 84247 polygons in the Munich OSM data. All these multilane roads in urban networks cannot be totally identified by the attributes in OSM. To apply the SVM method, training areas were firstly selected from road networks, which is about  $1500 \times 2500 m^2$  including 1180 polygons. There are 613 multilane polygons artificial selected as training sample to get the SVM classification model. After the training process, the accuracy ratio of this model is 77.16% for this training data. The final extractions are illustrated in figure 2, the cyan color indicates the roads that do not belong to multilane road, and the dark color indicates the finally extracted multilane roads.



Figure 2 Extraction results

## 4 Acknowledgements

This research is supported by the 863 project (No. 2012AA12A211), the National Nature Science Foundation (41101443), the Fundamental Research Funds for the Central Universities (No. 3103005, No. 201161902020015), the special funding for Wuhan University doctoral candidates short term studying abroad, and the academic award for excellent Ph.D. Candidates funded by Ministry of Education of China (No. 5052011619019, No. 5052012619001).

## References

- Ai, T.H. and Guo, R.Z., 2000, Extracting Center-lines and Building Street Network Based on Constrained Delaunay Triangulation. *Acta Geodaetica et Cartographica Sinica*, 29(4): 348-354. [In Chinese]
- Chang, C.-C. and Lin, C.-J., 2011, LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), Article No. 27.
- Yang, B.S., Luan, X.C., Li, Q.Q., 2011, Generating hierarchical strokes from urban street networks based on spatial pattern recognition. *International Journal of Geographical Information Science*. 25(12): 2025-2050.