Automatic Horizontal Road Design Information
Extraction from Georeferenced Polygonals: A Brazilian Federal Highway Network Study

Alexandre H. Coelho¹, Nataniel P. Borges Jr.², Nicolas P. Borges², Marcos D. Gallo² and Amir M. Valente¹
1. Department of Civil Engineering, Federal University of Santa Catarina (UFSC), Florianópolis SC 88040-970, Brazil
2. Transportation and Logistics Laboratory (LabTrans), Federal University of Santa Catarina (UFSC), Florianópolis SC 88040-970, Brazil

Abstract: Road geometric design data are a vital input for diverse transportation studies. This information is usually obtained from the road design project. However, these are not always available and the as-built course of the road may diverge considerably from its projected one, rendering subsequent studies inaccurate or impossible. Moreover, the systematic acquisition of this data for the entire road network of a country or even a state represents a very challenging and laborious task. This study’s goal was the extraction of geometric design data for the paved segments of the Brazilian federal highway network, containing more than 47,000 km of highways. It presents the details of the method’s adoption process, the particularities of its application to the dataset and the obtained geometric design information. Additionally, it provides a first overview of the Brazilian federal highway network composition (curves and tangents) and geometry.

Key words: Curve identification, information extraction, geometric design, polygonal segmentation, road.

1. Introduction

Geometric design data are a vital input for several transportation studies. Some operating speed prediction models [1-4] require curvature degree, curve radius and curve length information. Safety studies [5, 6] require the knowledge about segments’ location and length. Design guides [7] requires design information to evaluate vehicle skidding. These data are also required for level of service estimation [8, 9]; The method presented in the highway capacity model, for example, recommends sharp curves to be separately analyzed, because they have lower free-flow speeds. Moreover, geometric design data are also necessary for railway [10] and waterway [11] transportation studies, since curve radii are used to estimate transportation costs at railways and to determine maximum size of vessels and cargoes.

However, the acquisition of geometric information from design plans is a laborious task, since many design plans, when available, do not provide tools for automatic information extraction. It is also not guaranteed that design plans represent the as-built road situation, while differences may emerge from layout modifications, due to corrective actions or faulty project execution. Traditionally, curve radii and length could only be extracted during field visits which make the development and maintenance of a nationwide or, even statewide, database prohibitive.

Today, the availability and precision of GPS (global positioning system) equipment allows highway layout elements to be easily measured by vehicles travelling at highway speed and leading to a reduction of time and resources during the data acquisition step [12]. Raw GPS data that represent highway axis are also available from volunteered geographic information, like OpenStreetMap, and from private companies [13].

Corresponding author: Alexandre H. Coelho, Dr.-Ing., research fields: traffic engineering, geographic information systems, simulation and road design. E-mail: alexandre@labtrans.ufsc.br.
However, differently from the measurement performed during field visits, the geometric information is not immediately available after the acquisition; it needs to be extracted from the Raw GPS data.

This paper describes the adaptation of a process for generating digitized maps, using GPS data, into a procedure to systematically identify horizontal segments, using georeferenced polygonals that represent the road axis and the possibility of geometric design data extraction from the identified tangents and circular curves. The presented process is then applied to more than 47,000 km of paved roads in the Brazilian federal highway network, providing a first look on its geometric characteristics.

2. Related Work

Literature presents several applications of GPS and GIS (geographic information systems) for transportation studies. Applications supported by GPS and GIS have been used for on-road traffic data acquisition [14], transportation network monitoring [15], driver behavior analysis [16], digitalized maps development [17, 18] and highway safety analysis [19, 20].

Models for the extraction of geometric design information from measured points were also developed. The method of least squares was used to deduce design elements of horizontal alignment [21], $\psi$-$S$ curves were used in models to determinate line segments, circular arcs and clothoidal arcs that form a complex [22], clothoids and splines where used to represent transition curves [23, 24].

Algorithms to segment and classify data into tangents and circular curves were also proposed [25]. An algorithm, to automatically extract the curve length and the degree of curvature using the information provided by a gyro compass attached to a vehicle driving at highways, was proposed. The New Hampshire Department of Transportation developed an application, called “curve finder” [26], to implement its own method to automatically identify horizontal curves based exclusively on GPS points. Their method, however, does not distinguish between circular and transition curves. Other segmentation model that considers only tangents and circular curves is presented in Ref. [27], where a cloud of GPS points, acquired by several vehicles driving on a mine, is clustered and segmented to extract the curve radii.

Studies were also developed with algorithms that identify transition curves, such as Refs. [28, 29], respectively used to analyze driver behavior and provide data for advanced driver assistance systems.

3. Data Segmentation Process

The segmentation algorithm presented in this paper is a modified version of the algorithm presented in Ref. [27]. This algorithm was originally developed to create digitized road maps, using a cloud of GPS points collected by vehicles driving on a mine. A comparison between the algorithms is presented in Fig. 1; The main differences between the algorithms are the removal of the clustering step and the addition of two filters to remove speed variation, GPS noise and vehicle maneuvers from the data.

3.1 Uniform Spacing

Previous analysis performed over the database [30] established an average of 59 points per kilometer, analyzing some segments with a GIS application. It was possible to see differences between the point spacing on curves and tangents, as exemplified with a highway segment in Fig. 2.

To remove this variation, equally-spaced points were interpolated on the polygonal by the use of the following algorithm:

1. Considering a polygonal $P$, whose geodesic length is $c$, and $S$ is treated as the desired point spacing, calculate $r$ as the relation between $c$ and $S$;
2. Generate the interpolating function, defined as $f:R \rightarrow R$;
3. Create an empty point set;
4. While the geodesic length of the new point set is
smaller, the ε adds a new point to the new point set. The new point coordinates are defined by Eqs. (1) and (2):
\[ x_i = x_i + (i - 1)\varepsilon \]  
\[ y_i = f(x_i) \]

This step is necessary because the segmentation mechanism presented in the next sections is affected by the point spacing and a single parameter set is used to process the whole polygonal. Fig. 3a presents the original highway segment, and Fig. 3b presents the same segment after the uniform spacing procedure.

3.2 Heading Calculation

To allow an algebraic data segmentation, the method proposed by Ref. [27] establishes a relation between the geographic distance (d) of consecutive points pairs and the angle (θ) formed by the line segment connecting them and the north azimuth. From the starting point of an n point polygonal, it is possible to determinate both d and θ for each point i, where \( 1 \leq i \leq n \), as presented in Fig. 4.

Once d and θ are calculated for each polygonal point, it is possible to build a \( d \times \theta \) chart, where tangent segments generate almost horizontal lines with some noise, and curves generates steep lines with steeper slopes corresponding to a sharper road curve. Fig. 5 presents a highway segment and its corresponding \( d \times \theta \) chart. In this figure, d is shown as the highway mileage.

3.3 Smoothing

This work considers both vehicle maneuvers and GPS precision loss as noise since they can mislead the segmentation results by affecting the correct location of the segments start and end points or identifying non-existent curves.

To reduce the noise influence, a smoothing procedure using a moving average filter was adopted. In this procedure, each θ value is compared to its k closest neighbors and k is a predefined value. Fig. 6

---

1 The last polygonal point receives the value of its predecessor.
2 Speed variation noise was already removed in the uniform spacing phase.
Fig. 3  Highway segment before and after uniform spacing: (a) original highway segment; (b) highway segment after uniform spacing.

Fig. 4  Points over a highway axis and their relation according to \( d \) and \( \theta \) [27].

Fig. 5  Highway segment and its equivalent \( d \times \theta \) chart: (a) highway segment; (b) highway \( d \times \theta \) chart.

Fig. 6  Smoothing of a noisy segment represented by its \( d \times \theta \) chart.

presents a \( d \times \theta \) chart of a noisy highway segment and the results of the filter. In this figure, the original data are represented by a dashed line and the smoothed data is represented by a black one.

3.4 Segmentation

According to Ref. [27], to segment the data is to identify the start and end points of almost all the horizontal lines, ignoring noise, on the \( d \times \theta \) chart. The further classification of each segment, as a tangent or circular curve, is done by checking the line slope. The segmentation mechanism proposed by the authors consisted of a two-stage threshold process.

At the first stage, the two initial polygonal points are respectively defined as the start (\( P_S \)) and end (\( P_E \)) of a tangent segment. \( P_E \) is then iteratively moved forward. During each iteration, a line segment (\( \overline{P_S P_E} \)) is created

\[ (x, y) \]

\[ (x, y)_{i-1} \]

\[ (x, y)_i \]

\[ (x, y)_{i+1} \]

\[ d_i \]

\[ \theta_i \]

\[ \theta_{i+1} \]

\[ \text{Mileage (km)} \]

\[ \theta (\text{rad}) \]

\[ \text{Mileage (km)} \]

\[ \theta (\text{rad}) \]

\[ 1 \]

\[ 2 \]

\[ 0.5 \]

\[ 1.0 \]

\[ 1.5 \]

\[ 2.0 \]

\[ 0.5 \]

\[ 1.0 \]

\[ 1.5 \]

\[ 2.0 \]

\[ 0.5 \]

\[ 1.0 \]

\[ 1.5 \]

\[ 2.0 \]

\[ 0.5 \]

\[ 1.0 \]

\[ 1.5 \]

\[ 2.0 \]
between $P_S$ and $P_E$, and the sum of the squared error of each point belonging to $\overline{P_S P_E}$ is calculated. When the total error exceeds a coarse threshold and the iterations are stopped, this stage aims to provide the noise tolerance.

At the second stage, $P_E$ is iteratively moved backwards, and the sum of the squared error of each point between $P_S$ and $P_E$ is calculated, when the total error is smaller than a fine threshold and the iterations are stopped. If the new segment is larger than a minimum length, the segment is identified; Otherwise, it will be merged to the next segment. This minimum length constraint is used to prevent a single curve with radius variation from being identified as multiple curves.

Once the segments are identified, they are classified as tangents or curves based on their line slope at the $d \times \theta$ chart. If their slope is larger than a threshold, they are classified as curves, otherwise, they are classified as tangents. After this first classification, curve segments with a radius bigger than a maximum value are reclassified as tangents. The segmentation of the polygonal presented in Fig. 5 can be seen in Fig. 7. The vertical lines in Fig. 7a represent the identified segments boundaries, while in Fig. 7b, the darker parts represent segments classified as tangents.

### 3.5 Curve Information Extraction

The tangent segments geometry is described once their start and end points are located. The extraction of radius and central angle from circular curves requires the use of a nonlinear least squared fitting technique. This work uses the Levenberg-Marquardt method to fit the circles [31]. Once the method is applied to each located curve, their radii and center points are identified. With these values and the curvature and tangency points’ location, the internal angle and the curve direction (right or left) are calculated. Table 1 shows the information extracted from the polygonal shown in Fig. 8.

![Fig. 7 Highway segment and its equivalent $d \times \theta$ chart: (a) segmented $d \times \theta$ chart; (b) segmented highway segment.](image)

![Table 1 Extracted geometric information.](image)

<table>
<thead>
<tr>
<th>Number</th>
<th>Type</th>
<th>Points</th>
<th>Length (m)</th>
<th>Radius (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Tangent</td>
<td>45</td>
<td>620.57</td>
<td>-</td>
</tr>
<tr>
<td>S2</td>
<td>Right curve</td>
<td>30</td>
<td>384.96</td>
<td>220.00</td>
</tr>
<tr>
<td>S3</td>
<td>Tangent</td>
<td>22</td>
<td>298.14</td>
<td>-</td>
</tr>
<tr>
<td>S4</td>
<td>Left curve</td>
<td>43</td>
<td>553.71</td>
<td>349.38</td>
</tr>
<tr>
<td>S5</td>
<td>Tangent</td>
<td>35</td>
<td>478.76</td>
<td>-</td>
</tr>
<tr>
<td>S6</td>
<td>Right curve</td>
<td>18</td>
<td>226.19</td>
<td>127.80</td>
</tr>
<tr>
<td>S7</td>
<td>Tangent</td>
<td>23</td>
<td>293.14</td>
<td>-</td>
</tr>
</tbody>
</table>

![Fig. 8 Segmentation results with extraction of curve radii and center.](image)

Finally, after all the segments are identified and the geometric information is extracted, the average horizontal curvature [32] is calculated using the following Eq. (3):

$$ AHCS = \frac{\sum_{i=1}^{n} \alpha_i}{l} $$

where, $S$ is a segment, $l$ is the segment’s length and $\alpha$ is the internal angle of the curves belonging to the segment.
4. Brazilian Federal Highway Network Segmentation

The Brazilian DNIT (National Department of Transportation Infrastructure) performs periodical surveys on the network. One of these surveys was made available to LabTrans as a geospatial database. This database contains 118,623.67 km of georeferenced polygons representing the axis of the whole Brazilian federal highway network, as presented in Fig. 9. The data, however, were not completely acquired by GPS. Some segments were digitalized over satellite images and others over road maps. Furthermore, several segments do not possess information regarding their origin. Table 2 presents the amount of segments and its source.

According to the SNV (National Highway System), there are 66,712.20 km of paved federal highways. To reduce the gap between the published amount of paved federal highway segments and the number of segments whose source information was correctly assigned, the dataset was filtered based on its geometry, not based on its source information. The following filters were applied to extract only valid segments surveyed by GPS:

- only paved segments;
- average points spacing below 25 m with standard deviation below 10 m;
- geometry with more than 20 and less than 5,000 points.

After the filter, 47,767.10 km where identified as suitable for the extraction of geometric information, representing approximately 72% of the published number of paved segments. The filtered dataset is presented in Fig. 10 and its point spacing histogram is presented in Fig. 11.

Each polygon on the filtered dataset was then segmented using the parameters specified in Table 3. These parameters were manually adjusted plotting a sample dataset over satellite images and visually analyzing the results.

---

1DNIT surveys only paved roads with GPS as part of a pavement survey project called “Continuous Visual Survey” [33] where a single vehicle drives at each segment a single time.

2According to the SNV (National Highway System) [34], the paved segments length vary from 0.1 m to 96.6 km. However, after verification in the geometry database, no curves were identified in paved segments smaller than 500 m.

---

Table 2 Database composition.

<table>
<thead>
<tr>
<th>Source</th>
<th>Length (km)</th>
<th>Length (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS</td>
<td>25,435.59</td>
<td>21.44</td>
</tr>
<tr>
<td>Maps</td>
<td>31,490.59</td>
<td>26.55</td>
</tr>
<tr>
<td>Satellite images</td>
<td>283.20</td>
<td>0.24</td>
</tr>
<tr>
<td>Unknown origin</td>
<td>61,414.19</td>
<td>51.77</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>118,623.67</strong></td>
<td><strong>100.00</strong></td>
</tr>
</tbody>
</table>

---

3The used polygons are described in geographic coordinates system. Different coordinate systems would impact the parameters values.
Table 3  Segmentation parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum segment length</td>
<td>60 m</td>
</tr>
<tr>
<td>Maximum curve radius</td>
<td>2,500 m</td>
</tr>
<tr>
<td>Coarse threshold</td>
<td>0.048</td>
</tr>
<tr>
<td>Fine threshold</td>
<td>0.022</td>
</tr>
<tr>
<td>Slope limit</td>
<td>18°</td>
</tr>
<tr>
<td>Number of neighbors for smoothing</td>
<td>3</td>
</tr>
<tr>
<td>Smoothing average threshold</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The segmentation process resulted in 135,567 horizontal segments, of which 70,718 (52.16%) were classified as curves. In extent, 37,215.34 km (77.91%) were classified as tangents and 10,551.75 km (22.09%) were classified as curves. The identified curve radii and internal angle histograms are summarized on Fig. 12.

To evaluate the segmentation mechanism accuracy, the resulting dataset was drawn over satellite images and visually analyzed, because design plans for Brazilian federal highways are not available. This manual evaluation, although very superficial, showed several problems on curves with radius below 100 m. Those curves do not represent real curves but GPS noise that was not removed by the filters. An example of this problem is presented in Fig. 13. After the identification of this trend, a manual analysis of the 9,918 curves with radius below 100 m identified only 547 valid curves (5.52%) and a single valid curve with radius below 40 m.

5. Conclusions

This paper presents a procedure to automatically extract highway horizontal geometric information on a nationwide scale. It adapted a mechanism previously used to build digitalized maps and converted it into a segmentation mechanism that processes georeferenced polygons. This mechanism was then used to analyze over 47,000 km of paved highways in the Brazilian federal highway network, acquired as part of their Continuous Visual Survey project, providing a first
overview of its geometric design and an input for some transportation studies.

This overview showed that most of the Brazilian federal highway network segments, 77.91% in extent, are tangents and that most of the curve segments have radius between 100 m and 400 m and internal angle lower than 30°. It also demonstrated that DNIT’s information is unreliable for curves with radius below 100 m, with 94.48% of invalid segments. It is possible to replace the invalid network parts with data obtained from other sources, like volunteered geographic information projects. However, the automatic identification and replacement of invalid segments require further studies.

Possible extensions to this work are the inclusion of transition curves in the mechanism’s model and the development of a procedure to automatically identify an optimal parameter set for each polygonal. Both researches are currently under development.

Acknowledgments

This work was supported by the LabTrans/UFSC (Transportation and Logistics Laboratory of the Federal University of Santa Catarina) and by the DNIT (National Department of Transportation Infrastructure) under Cooperation Agreement 497/2012.

References

Automatic Horizontal Road Design Information Extraction from Georeferenced Polygonals: A Brazilian Federal Highway Network Study

Transportation Research Record: Journal of the Transportation Research Board 2194 (1): 91-8.


[31] DNIT (National Department of Transportation...
Automatic Horizontal Road Design Information Extraction from Georeferenced Polygonals: A Brazilian Federal Highway Network Study


