Automated Identification and Extraction of Horizontal Curve Information from Geographic Information System Roadway Maps

Zhixia Li, Madhav V. Chitturi, Andrea R. Bill, and David A. Noyce

Roadway horizontal alignment has long been recognized as one of the most significant contributing factors to lane departure crashes. Knowledge of the location and geometric information of horizontal curves can greatly facilitate the development of appropriate countermeasures. When curve information is unavailable, obtaining curve data in a cost-effective way is of great interest to practitioners and researchers. To date, many approaches have been developed to extract curve information from commercial satellite imagery, Global Positioning System survey data, laser-scanning data, and AutoCAD digital maps. As geographic information system (GIS) roadway maps become more accessible and more widely used, they become another cost-effective source for extraction of curve data. This paper presents a fully automated method for the extraction of horizontal curve data from GIS roadway maps. A specific curve data–extraction algorithm was developed and implemented as a customized add-in tool in ArcMap. With this tool, horizontal curves could be automatically identified from GIS roadway maps. The length, radius, and central angle of the curves were also computed automatically. The only input parameter of the proposed algorithm was calibrated to have the least curve identification errors. Finally, algorithm validation was conducted through a comparison of the algorithm-extracted curve data with the ground truth curve data for 76 curves that were obtained from Bing aerial maps. The validation results indicated that the proposed algorithm was very effective and that it identified completely 96.7% of curves and computed accurately their geometric information.

The horizontal curves of suburban and rural highways have long been recognized as one of the critical locations for roadway departure crashes (*1*). NHTSA indicates that horizontal alignment contributes to 76% of single-vehicle crashes in the United States, based on 2007 Fatality Analysis Reporting System data (*2*). Previous research results have also reported that crash rates on horizontal curves are 1.5 to four times higher than the crash rates on roadway tangents (*3*). Moreover, other researchers have revealed that crash rates increase as the degree of curvature increases and that a curvature of 15° or greater has a high probability of being hazardous (*4, 5*).

Nationwide, a research framework has been formed by research projects that have been conducted to better understand the relationship between the characteristics of horizontal curves and crashes and to explore effective countermeasures to prevent crashes from occurring on horizontal curves (*6–8*). Identifying where horizontal curves are located and what the geometric characteristics of the curves are is an essential and important task in solving the curve safety issue. In the United States, many states have a horizontal curve database for their state routes and Interstate highways; however, most of them do not have such a database for county and local roads, and it is costly and time-consuming to collect this curve information through the traditional approaches. In this context, a high priority should be given to the development of an approach that can identify location and geometric information for state and county or local highways in an accurate, cost-effective, and time-efficient manner.

To date, satellite imagery, Global Positioning System (GPS) survey data, laser-scanning data, and AutoCAD digital maps are four sources that are widely used by researchers to obtain horizontal curve data. Successful applications based on these four sources are well documented in the literature (*9–18*). Geographic information system (GIS) roadway maps are becoming more accessible and widely used by most government agencies and research institutions and provide an alternative source for horizontal curve data extraction. Compared with the traditional data sources, GIS roadway maps have the following potential advantages:

- A zero data cost because of their higher accessibility versus expensive commercial satellite imagery and rare AutoCAD digital maps,
- A zero data collection cost versus expensive and time-consuming GPS surveying and laser scanning,
- A source of complete roadway network data versus the small number of roadways on which GPS surveying and laser scanning have been conducted, and
- A short data preparation time with fewer preprocessing requirements compared with complicated image processing that involves a longer data preparation time and, possibly, more error.

These facts demonstrate the need for a method that can effectively and efficiently detect horizontal curves and extract their information from GIS roadway maps.

Z. Li, Room 1249A Engineering Hall; M. V. Chitturi and A. R. Bill, Room B243 Engineering Hall; and D. A. Noyce, Room 1204 Engineering Hall, Traffic Operations and Safety Laboratory, Department of Civil and Environment Engineering, 1415 Engineering Drive, University of Wisconsin–Madison, Madison, WI 53706. Corresponding author: Z. Li, zli262@wisc.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2291, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 80–92.
DOI: 10.3141/2291-10

Recently, some researchers have started extracting horizontal curve data through the use of ArcGIS roadway Shapefiles in the ArcMap environment (*19*). The ArcGIS coordinate geometry toolbar provides a similar function, through the curve calculator command, for the manual extraction of curve data from GIS roadway maps (*20*). The methods in these applications require the manual identification and construction of tangent points and chord lines on the ArcGIS feature layers, which results in a large work load and lower efficiency. In fact, with the powerful programmable support provided by ArcGIS, the development of an automated curve data–extraction tool that can be embedded into ArcGIS is possible. So far, only one tool, which was developed by the New Hampshire Department of Transportation, has been found in the literature and features a semiautomatic approach. This method extracted roadway alignment information from a geodatabase and output the curve data into a text file (*21*). This method required a special roadway referencing system defined by mileposts and the manual creation of curve feature classes and layers based on the output data. Aside from this semiautomatic approach, no literature that documents a fully automatic method has been found.

This paper presents a fully automated method for horizontal curve identification and curve data extraction from GIS roadway maps. A specific tool, CurveFinder, has been developed based on the ArcGIS programmable package ArcObjects. CurveFinder can be loaded into ArcMap as a customized GIS tool. The internal curve data–extraction algorithm enables the tool to automatically locate simple circular curves and compound curves from a selected county highway layer, compute the length of each curve, as well as the radius and curvature of each simple circular curve, and, finally, create a layer that contains all the identified curve features, along with their geometric information.

LITERATURE REVIEW

High-resolution satellite imagery is one of the sources most widely used in the extraction of elements of highway horizontal alignment. Various approaches that use different image processing techniques have been used by researchers to detect roadway geometry from satellite imagery (*9, 10, 22–25*). Researchers from Ryerson University, Toronto, Ontario, Canada, conducted an insightful investigation into the identification of horizontal curves from IKONOS satellite images and proved the feasibility of deriving the geometric characteristics of simple, compound, and spiral curves through the use of an approximate algorithm based on high-resolution satellite images (*9, 10*). However, the drawbacks of using satellite imagery are obvious: accuracy is reduced when roadway information is extracted from urban roadway images, because the variety of land cover can confuse the target, and accuracy greatly relies on the image resolution, and high-resolution commercial images are relatively expensive.

Since the early 2000s, GPS data have been used by researchers to extract highway horizontal alignment information (*11, 12*). In those approaches, geographic coordinates were recorded by a GPS-equipped vehicle at short time intervals (e.g., 0.5 s) along the roadway. The horizontal curves were then identified, and the radii were computed using a customized GIS program based on the logged GPS data points along the curves. In another study of vehicle paths on horizontal curves in Canada, Imran et al. developed a method to incorporate GPS information into GIS for the calculation of the radius, length, and spiral length of horizontal curves (*17*). Hans et al. used GPS data to develop a statewide curve database for crash analysis in Iowa (*18*). In addition to GPS, Yun and Sung installed multiple sensors, including an inertial measurement unit, a distance measuring instrument, and cameras, on a surveying vehicle to acquire real world highway coordinates at a higher accuracy level (*13*). Other researchers equipped a surveying vehicle with laser-scanning technology, in addition to GPS, to obtain the three-dimensional characteristics of horizontal curves (*14*). The high-density three-dimensional information allowed the faster and easier extraction of cross-section elements. In these methods, GPS data and laser technology both facilitated the collection of high-accuracy curve data. However, the use of these special surveying vehicles for data collection prevented the method from being applied to a larger number of roadways because of the high cost and long data collection time.

Compared with other data sources, a digital map is a more economic and time-saving data source. Researchers from the United Kingdom and Ireland succeeded in their attempts to extract highway geometry from digital maps in an AutoCAD environment (*15, 16*). The researchers' methods used AutoCAD commands to reconstruct the centerline of the roadway based on the digitized roadway map in AutoCAD format. The reconstructed centerline facilitated the location of the start and end points of simple curves. Curve radius, angle, and length were calculated using curve geometry equations. Recently, researchers considered the use of popular GIS digital roadway maps as the source for the identification of horizontal curves, which is considered a step toward higher efficiency and wider applicability. Price introduced a method of manually locating the curves and calculating the turning radii on a digital GIS roadway map in an ArcMap environment (*19*). His method required the manual identification and construction of tangent points and chord lines on the GIS feature layers. ESRI also provides a curve calculator function for the computation of curve information through its coordinate geometry toolbar (*20*). Similarly, the Florida Department of Transportation developed a toolbar in ArcGIS, named curvature extension, which is similar in function and operation to ESRI's curve calculator (*26*). Although in these aforementioned applications the radius and curve length calculation was performed by ArcMap, the full manual identification and construction of tangents was required, which prevented these methods from achieving a high efficiency and a low cost. The New Hampshire Department of Transportation developed a semiautomatic approach for curve data extraction (*21*). The department developed an executable file that retrieved roadway coordinates from a geodatabase and output the curve's starting and ending mileposts, radius, and number of segments into a text file. This method was specifically designed for a GIS roadway map that has a special roadway referencing system defined by mileposts. A manual creation of curve feature classes and layers, based on the output curve data, was also needed.

In summary, the existing approaches based on different data sources have proved their effectiveness in the extraction of horizontal curve information. However, the previous approaches require costly data collection or extraction or extensive manual labor. There is a need to develop a fully automated approach for curve data extraction from GIS roadway maps, considering the wide use of GIS software and maps. This research aims to develop such a fully automated GIS-based approach.

METHODOLOGY

The research methodology was composed of the following three steps. First, an automated curve data–extraction algorithm for GIS roadway maps was developed. Second, the roadway map was slightly

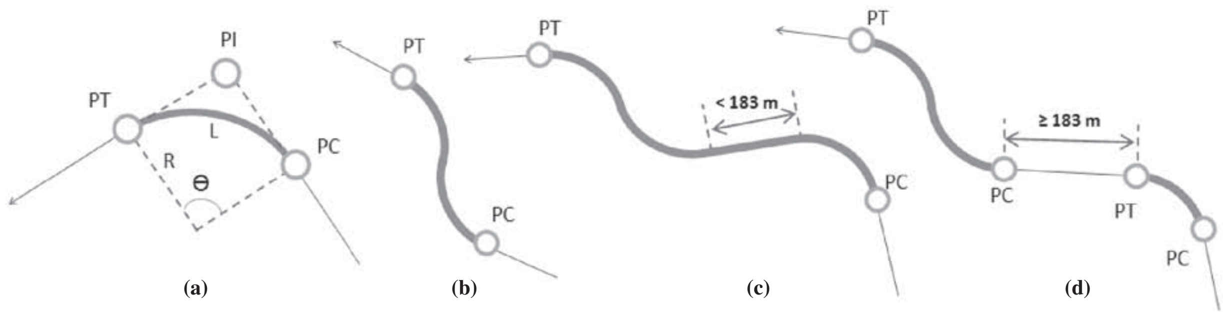


FIGURE 1 Characteristics of horizontal curves.

processed in ArcMap to meet the data source requirement of the algorithm. Third, the algorithm was implemented in ArcMap as an add-in tool by programming through ArcObjects. Before the discussion of the algorithm, the characteristics of horizontal curves and the GIS roadway maps that are used in this research are introduced.

Characteristics of Horizontal Curves

Horizontal curves can be classified into two fundamental types: simple circular curves and compound curves. A simple circular curve is a segment of a circle that is bounded by two tangents, as illustrated in Figure 1*a.* In the figure, the point of curvature, the point of tangency, the point of intersection, the length of the curve, the radius, and the central angle are represented by the symbols PC, PT, PI, *L, R,* and θ, respectively. A larger central angle depicts a sharper or more severe curve, and a larger radius depicts a less severe curve (*27*).

Because most existing GIS roadway maps have been digitized from satellite imagery, with some unavoidable error, the resolution of the existing roadway map data is not high enough to distinguish spiral and compound curves from circular curves. Therefore, all simple curves in this research are assumed to be circular.

The second type of horizontal curve is the compound curve, which is composed of multiple consecutive short curves and inner tangent sections. A typical example of a compound curve is illustrated in Figure 1, *b* and *c.* According to AASHTO, a tangent that separates two consecutive horizontal curves should be at least 183 m (600 ft) in length (*28*). Therefore, in the case illustrated in Figure 1*c,* both curves and the short inner tangent section together form a compound curve, because the length of the inner tangent section is less than 183 m. Figure 1*d* is illustration of a different case, in which there are two separate curves because the tangent that separates them is longer than 183 m.

GIS Roadway Maps and Map Preprocessing

The GIS roadway maps used in this study were the maps from the Wisconsin Information System for Local Roads, which covers all local, county, and state roads for each of the 72 counties in Wisconsin (*29*). In some cases, the roadway maps of different counties in the system had different scales, attributable to the scale of the original map before digitization. The roadways that were targeted for curve data extraction were selected from four counties (Dane, Sheboygan, St. Croix, and Portage), located in different regions of Wisconsin, as noted by the circles in Figure 2*a.* In total, 10 county roads with lengths greater than 10 km were selected for curve extraction. All county road layers had projected georeferencing systems, so the *x* and *y* coordinates could be measured in meters rather than longitudes and latitudes. The only processing required before the algorithm could be run was to dissolve all small polylines with the same road name and direction into a single polyline, which represented a whole county road. In this way, the horizontal alignment of the county road could be continuous, which ensured that no single horizontal curve would be split across different polylines. Figure 2*b* shows an example of the selected county roads in Portage County after the dissolving

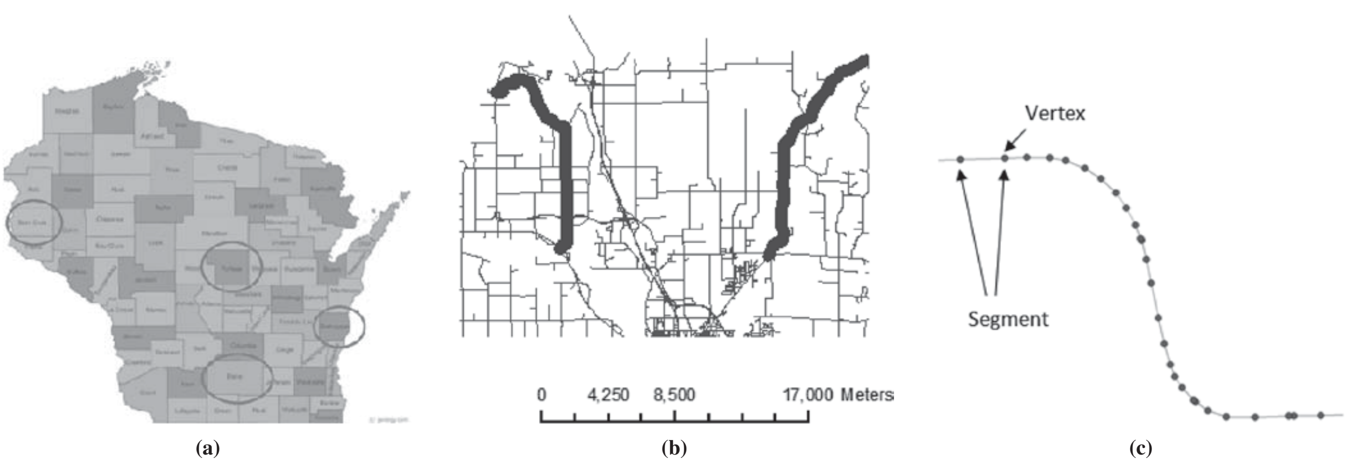


FIGURE 2 Selected counties and Wisconsin Information System for Local Roads GIS roadway map: (*a*) selected counties, (*b*) selected county roads, Portage County, and (*c*) structure of ArcGIS roadway polyline.

treatment. A dissolved road is a polyline composed of multiple connected segments. In GIS, a segment is an undividable feature that is bounded by two vertices, as illustrated in Figure 2*c*.

The rationale of the proposed curve data–extraction algorithm was based on an investigation of the geometric relationship between consecutive segments in the roadway polyline, which is discussed in detail in the following subsection.

Algorithm for Automatic Curve Identification and Curve Information Extraction

The objective of the developed algorithm is fourfold: (*a*) to automatically detect all curves from each road in a selected roadway layer, regardless of the type of curve; (*b*) to automatically classify each curve into one of two categories: simple or compound; (*c*) to automatically compute the radius and degree of curvature for each simple curve, as well as the curve length for simple curves and compound curves; and (*d*) to automatically create curve features and layers for all identified curves in the GIS.

Figure 3 shows a flow chart that explains the algorithm for the identification of curves from a roadway polyline. The algorithm was performed on each road in the selected roadway layer, so that all the curves in the layer could be identified.

A critical step in the curve identification algorithm is the computation of the bearing angle between two consecutive segments. Figure 4*a* gives an example that facilitates understanding of the bearing angle.

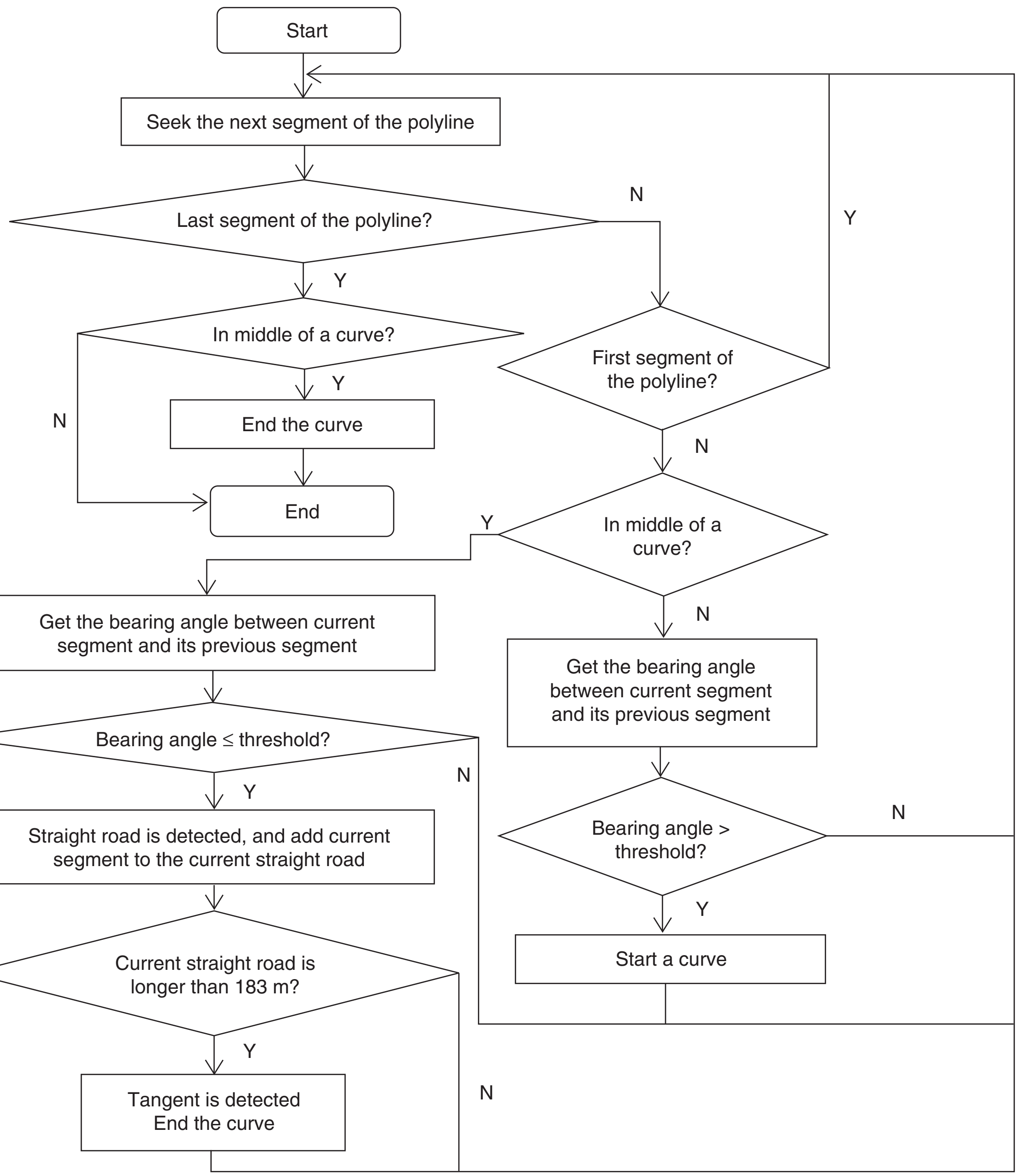


FIGURE 3 Curve identification algorithm (Y = yes; N = no).

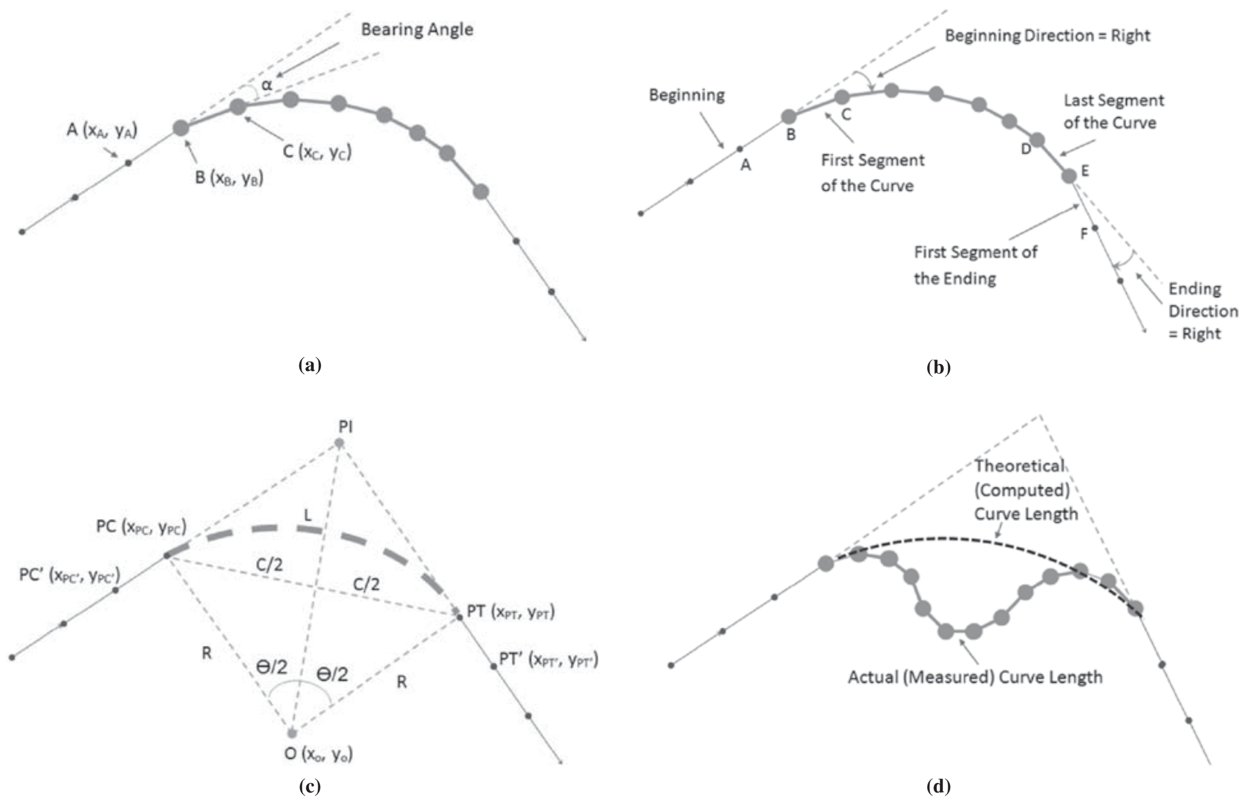


FIGURE 4 Critical steps in curve identification and classification.

In Figure 4*a,* the bearing angle α between Segment AB and Segment BC can be computed based on the geographic coordinates of vertices A, B, and C, through the use of the following equations:

$$\cos\left(\frac{\alpha\cdot\pi}{180}\right)=\frac{\overrightarrow{AB}\cdot\overrightarrow{BC}}{\left|\overrightarrow{AB}\right|\cdot\left|\overrightarrow{BC}\right|}=\frac{(x_B-x_A)(x_C-x_B)+(y_B-y_A)(y_C-y_B)}{\sqrt{(x_B-x_A)^2+(y_B-y_A)^2}\times\sqrt{(x_C-x_B)^2+(y_C-y_B)^2}} \tag{1}$$

$$\alpha=\cos^{-1}\left(\frac{(x_B-x_A)(x_C-x_B)+(y_B-y_A)(y_C-y_B)}{\sqrt{(x_B-x_A)^2+(y_B-y_A)^2}\times\sqrt{(x_C-x_B)^2+(y_C-y_B)^2}}\right)\times\frac{180}{\pi} \tag{2}$$

where x_A, x_B, and x_C are the x coordinates of Points A, B, and C, respectively, and y_A, y_B, and y_C are the y coordinates of Points A, B, and C, respectively.

The only adjustable parameter of the curve identification algorithm is the threshold of the bearing angle. When the bearing angle between the current segment and the preceding segment is greater than the threshold, a new curve should begin or the existing curve should continue. On the contrary, when the bearing angle between the current segment and the preceding segment is equal to or less than the threshold, the segments are considered to be a part of a tangent section. Therefore, the selection of a reasonable threshold value is of great importance for the accurate identification of curves.

The curve classification algorithm is designed to classify all identified curves into one of two types: simple or compound. A curve would be classified as a simple curve if the curve satisfied both of the following criteria: (*a*) it had the same beginning and ending directions, and (*b*) the difference between the theoretical (computed) curve length and the actual (measured) curve length was within a certain percentage. If either condition was not met, the curve would be classified as a compound curve.

Figure 4*b* shows the curve's beginning and ending directions. The curve's beginning direction is defined as the bending direction of the first segment of the curve against the beginning tangent. The curve's ending direction is defined as the bending direction of the ending tangent against the last segment of the curve. In the case shown in Figure 4*b,* the beginning direction is the same as the ending direction, an indication that the curve satisfies the first criterion for being classified as a simple circular curve.

In the second criterion, the difference between the theoretical and actual curve lengths should be no more than a certain threshold percentage. After a simple sensitivity analysis, 2.5% was used for the threshold in this paper. The theoretical (computed) curve length is the length of the ideal simple circular curve that can be estimated based on the detected tangents, as shown in Figure 4*c.* According to Figure 4*c*, the theoretical curve length can be computed when the coordinates of PC, PC′, PT, and PT′ are known, through the use of the following equations:

$$k_{O-\mathrm{PC}}=\frac{x_{\mathrm{PC}'}-x_{\mathrm{PC}}}{y_{\mathrm{PC}}-y_{\mathrm{PC}'}} \tag{3}$$

$$b_{O\text{-PC}} = y_{PC} - x_{PC} \cdot \frac{x_{PC'} - x_{PC}}{y_{PC} - y_{PC'}} \tag{4}$$

$$k_{O\text{-PT}} = \frac{x_{PT'} - x_{PT}}{y_{PT} - y_{PT'}} \tag{5}$$

$$b_{O\text{-PT}} = y_{PT} - x_{PT} \cdot \frac{x_{PT'} - x_{PT}}{y_{PT} - y_{PT'}} \tag{6}$$

$$x_O = \frac{b_{O\text{-PT}} - b_{O\text{-PC}}}{k_{O\text{-PC}} - k_{O\text{-PT}}} \tag{7}$$

$$y_O = k_{O\text{-PC}} \cdot \frac{b_{O\text{-PT}} - b_{O\text{-PC}}}{k_{O\text{-PC}} - k_{O\text{-PT}}} + b_{O\text{-PC}} \tag{8}$$

$$R = \sqrt{\left(x_{PC} - x_O\right)^2 + \left(y_{PC} - y_O\right)^2} \tag{9}$$

$$C = \sqrt{\left(x_{PT} - x_{PC}\right)^2 + \left(y_{PT} - y_{PC}\right)^2} \tag{10}$$

$$\theta = 2 \times \sin^{-1}\left(\frac{C}{2R}\right) \times \frac{180}{\pi} \tag{11}$$

$$L = \frac{\theta \cdot \pi}{180} \times R \tag{12}$$

where

O = center point of curve,
$k_{O\text{-PC}}$ = slope of line equation for Line O–PC,
$b_{O\text{-PC}}$ = intercept of line equation for Line O–PC,
$k_{O\text{-PT}}$ = slope of line equation for Line O–PT,
$b_{O\text{-PT}}$ = intercept of line equation for Line O–PT,
x_O = x coordinate of curve's center point,
y_O = y coordinate of curve's center point,
R = curve's radius (m),
C = length of curve's long chord (m),
θ = curve's central angle (degrees), and
L = theoretical (computed) length of curve (m).

The actual (measured) length of the curve can be measured by summing up the length of each segment of the curve. Figure 4*d* shows a curve that does not satisfy the second criterion for classification as a simple curve. In Figure 4*d*, the difference between the theoretical curve length and the actual curve length is greater than the threshold of 2.5%, an indication that the curve would not be classified as a simple curve but as a compound curve.

Finally, the geometric information of each curve will also be automatically extracted after all the curves are classified. The geometric information includes the actual length of each curve, regardless of the type of curve, as well as the radius, central angle, and theoretical length of the simple curves. The algorithm computes the radius, central angle, and theoretical curve length based on Equations 9, 11, and 12, respectively.

Software Implementation of Algorithm

The proposed curve data–extraction algorithm was implemented in ArcMap as a customized add-in tool (CurveFinder). The development of CurveFinder was based on the ArcGIS programming package ArcObjects. Microsoft Visual Studio 2010 was selected as the software development interface, and the programming language was C#. Figure 5*a* shows the user interface of CurveFinder, as well as a snapshot of the identified curves that is overlaid on a Bing aerial map layer for a better presentation.

CurveFinder allows users to select a roadway layer from the drop-down list as the source layer for curve identification. Users can specify the bearing angle threshold before running the algorithm. CurveFinder also features a query-based result filter for simple curves. Only those simple curves that satisfy the user-designated minimum central angle and maximum radius will be extracted from the selected roadway layer. Clicking the Extract Curves button will trigger the algorithm, and a layer that contains all the identified curves will be created automatically. Figure 5*b* shows an example of the identified simple curves with their geometric information stored in the property table; Figure 5, *c* and *d,* shows examples of different compound curves identified by the algorithm.

SENSITIVITY ANALYSIS AND ALGORITHM CALIBRATION

As discussed in the section on methodology, the only adjustable parameter in the curve identification algorithm is the threshold of the bearing angle. Users can specify the threshold value through the CurveFinder interface before performing the curve data extraction. An inappropriate threshold value can negatively affect the curve identification results and generate many identification errors. This section discusses the sensitivity analysis of the bearing angle threshold. The sensitivity analysis enabled the algorithm to be calibrated using the optimal bearing angle threshold (the threshold that produces the fewest identification errors).

For the sensitivity analysis to be prepared, four county roads were randomly selected from four counties in Wisconsin as the source roadway map. The ground truth curve data were obtained by manually and carefully identifying the horizontal curves from the selected county roads based on the Bing aerial map. In total, 51 ground truth curves were identified and created as features in the ground truth curve layer in ArcMap.

The sensitivity analysis tested 13 bearing angle thresholds, ranging from 0.5° to 5°, in terms of the accuracy of curve identification. For each threshold, the curves identified by the algorithm were compared with the ground truth curves in ArcMap. There are two possible types of identification error: Type 1 errors and Type 2 errors. A Type 1 error is a failure to detect an existing curve. A Type 2 error is the misidentification of a tangent section as a curve. After a comparison of the identified curves with the ground truth curves, all Type 1 and Type 2 errors were recorded.

Figure 6, *a* through *i,* shows examples of the comparison, as well as different scenarios of Type 1 and Type 2 errors. Figure 6, *a* through *c*, depicts three scenarios of complete identification (the curve is 100% identified). In this research, a complete identification was reached if the difference between the identified curve and the ground truth curve was no more than one segment. Figure 6, *d* and *e*, shows different scenarios of Type 2 errors. During the sensitivity analysis, if part of the identified curve had not overlapped with the ground truth curve, that part was recorded as a Type 2 error. Figure 6, *f* through *i*, shows four scenarios of Type 1 errors. In these scenarios, the curve was only partially identified, which meant that

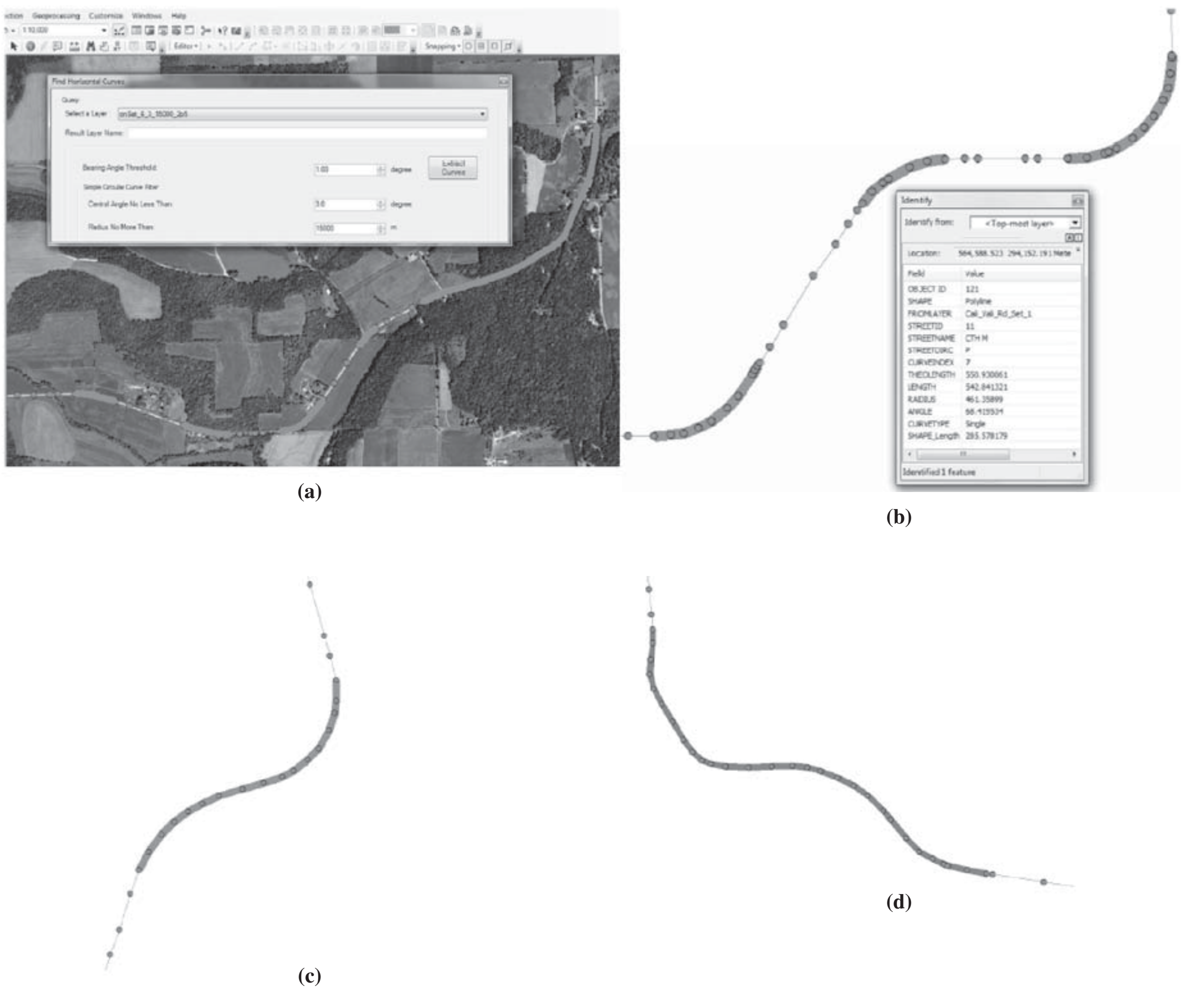


FIGURE 5 CurveFinder interface and curve identification results.

some parts of the ground truth curve were not detected as curves by the algorithm. Figure 6, *f* through *i*, shows the scenarios of 25% missed, 50% missed, 75% missed, and 100% missed, respectively. During the sensitivity analysis, the percentage missed was recorded for each Type 1 error.

In the sensitivity analysis, the identification rate and the Type 2 error ratio were used as measures of effectiveness. The identification rate is a reflection of the Type 1 errors and is defined as the percentage of curves that are successfully identified by the algorithm. A higher identification rate reflects a lower number of Type 1 errors. The Type 2 error ratio is defined as the ratio between the number of Type 2 errors and the number of curves. A higher ratio reflects a higher number of Type 2 errors. The following equations mathematically explain how the identification rate and the Type 2 error ratio are computed:

$$\mathrm{IR} = \frac{\sum_{i}^{n}\left(1 - P\mathrm{miss}_i\right)}{n} \tag{13}$$

$$\mathrm{TIIR} = \frac{m}{n} \tag{14}$$

where

IR = identification rate,
$P\mathrm{miss}_i$ = percentage of the ground truth curve i that was missed,
n = number of ground truth curves,
TIIR = Type 2 error ratio, and
m = number of Type 2 errors.

Figure 7, *a* and *b*, depicts the Type 2 error ratio and the identification rate for each bearing angle threshold.

According to Figure 7*a*, the Type 2 error ratio decreases as the bearing angle threshold increases. This trend indicates that the use of a larger bearing angle threshold can reduce the number of Type 2 errors. However, according to Figure 7*b*, the identification rate decreases as the bearing angle threshold increases, which means that the use of a larger bearing angle threshold will greatly increase the number of Type 1 errors. From the perspective of crash analysis, a Type 1 error is more serious than a Type 2 error, because curves will not be included in the analysis if Type 1 errors occur. Therefore, the optimal threshold of the bearing angle was determined using the following rules: (*a*) the optimal threshold is the one that generates the fewest Type 1 errors (i.e., the one that has the highest identification rate), and (*b*) if multiple thresholds generate similarly low numbers of Type 1 errors, the one that generates the fewest

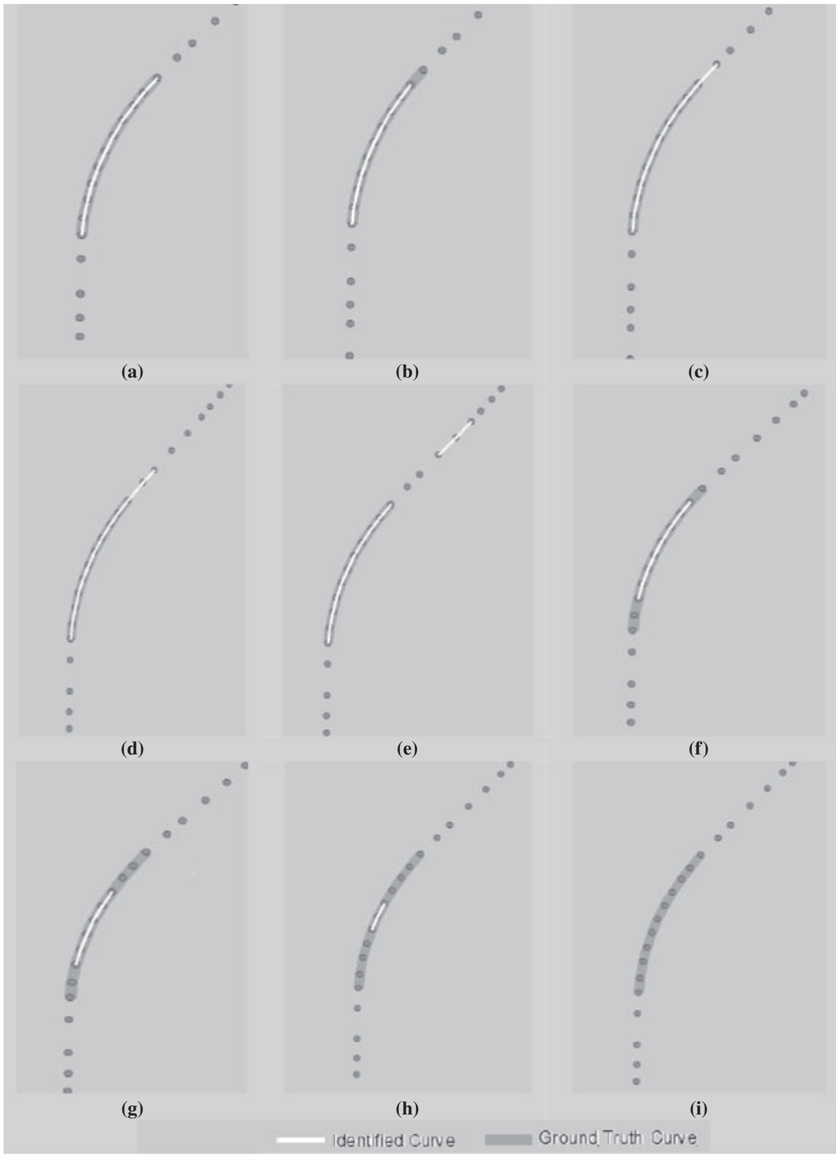


FIGURE 6 Comparison between ground truth and identified curves.

Type 2 errors (i.e., the one that has the lowest Type 2 error ratio) will be selected as the optimal threshold value.

On the basis of the previously discussed rules, 1.25° was determined to be the optimal threshold of the bearing angle. The threshold of 1.25° shares the highest identification rate (98%), with thresholds of 0.5°, 0.75°, and 1°, and has the lowest Type 2 error ratio of these four threshold values. The curve identification algorithm was therefore calibrated by using 1.25° as the optimal bearing angle threshold value in the algorithm.

ALGORITHM VALIDATION AND RESULTS

Although the curve extraction algorithm is presumed to operate optimally with the application of the calibrated optimal bearing angle threshold, the algorithm's performance still needs to be further validated by using different sets of ground truth data. The section is therefore dedicated to discussing the validation of the calibrated algorithm from three perspectives: curve identification, curve classification, and the extraction of the curve's geometric information.

In the validation process, six county roads from four counties were used as the input roadways for the algorithm. To ensure the credibility of the validation results, these six county roads were different from the roads that were used to calibrate the algorithm. Based on the same approach described in the previous section, 76 curves were identified from Bing aerial maps as the ground truth curves. In addition, the type of each ground truth curve was also recorded as either a simple circular curve or a compound curve. Moreover, 10 simple curves were randomly selected from the ground truth curves for the extraction of their ground truth geometric information. To do this, the aerial maps of these simple curves were imported into AutoCAD. Each curve's length, radius, and degree of curvature were then computed by AutoCAD as the ground truth data for the curves' geometric information.

Figure 8, *a* through *e*, depicts the results of the algorithm validation from different perspectives.

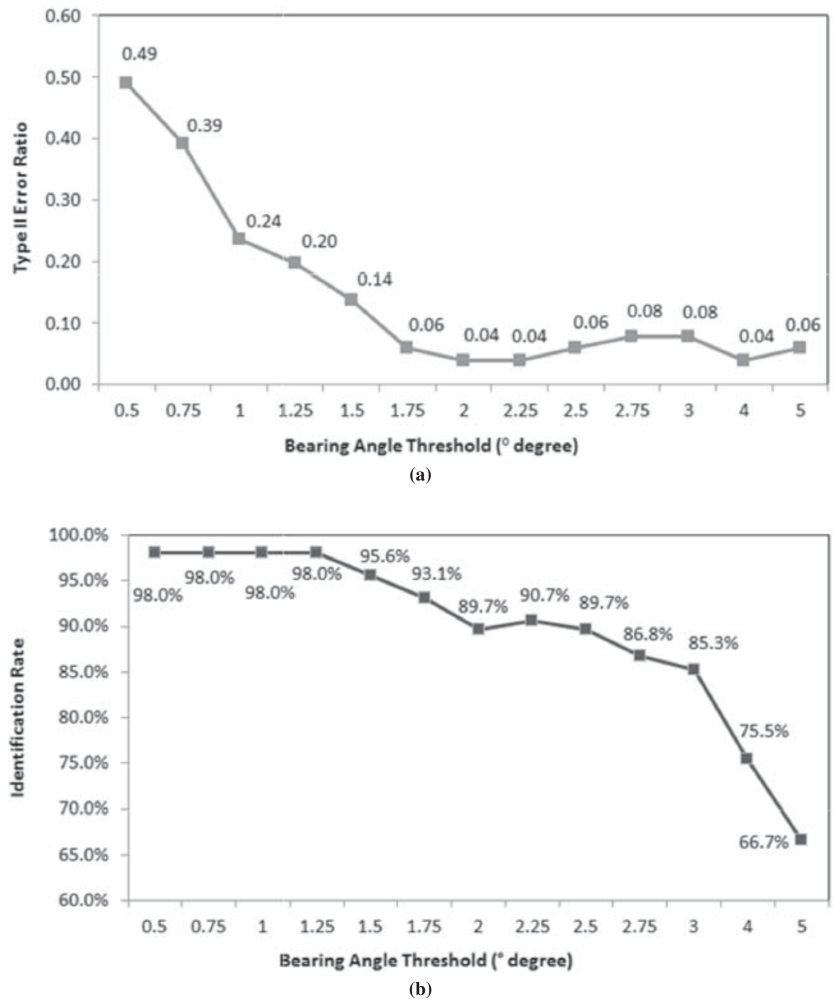


FIGURE 7 Algorithm sensitivity analysis results.

Performance of Curve Identification

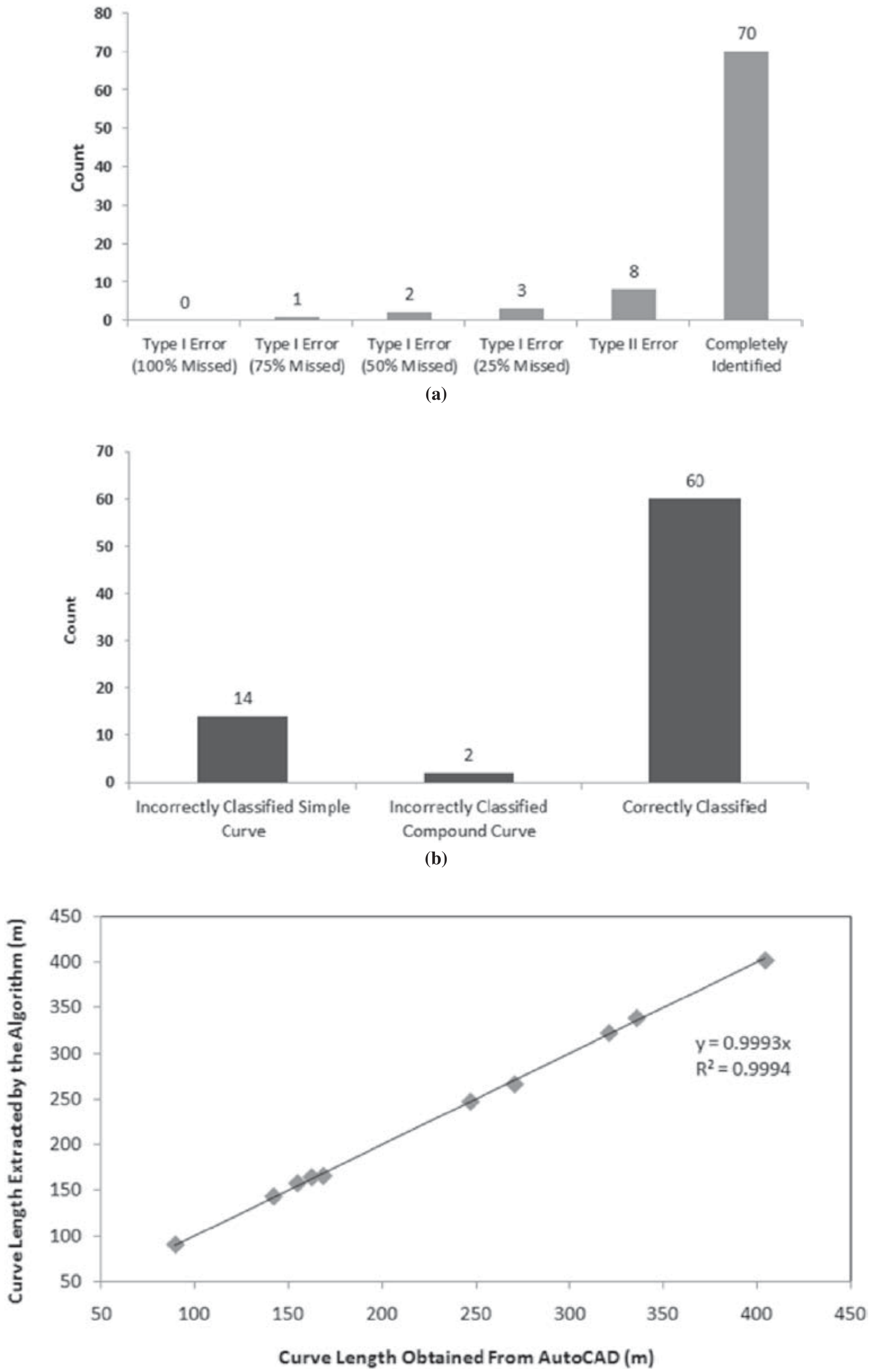
The algorithm identified 84 curves on the six county roads. Seventy of the 76 curves were identified completely; the other six were identified to varying degrees, as shown in Figure 8*a.* None of the curves were completely missed by the algorithm. Eight tangent sections were misidentified as curves. On the basis of Equation 13, the overall identification rate was calculated to be 96.7%, which was very close to the highest value (98.0%) that was obtained in the sensitivity analysis. The algorithm also produced a relatively low number of Type 2 errors. Computed on the basis of Equation 14, the Type 2 error ratio was as low as 0.11, which was also close to the lowest value (0.04) that was obtained in the calibration step. Both the high identification rate and the low Type 2 error ratio indicated that the algorithm had almost reached its optimal performance for curve identification by using the calibrated parameter value.

Performance of Curve Classification

According to Figure 8*b,* 60 of the 76 curves were correctly classified; in other words, the classification success rate was around 79%. Of the 16 incorrectly classified curves, 14 were simple curves that had been misclassified as compound curves. Only two were compound curves that had been misclassified as simple curves. This fact reveals that the major error of the curve classification comes from the misclassification of simple curves. The specific reasons are to be discussed in the section on the algorithm's performance. Overall, the 79% classification success rate is acceptable. This rate also indicates that the classification algorithm still has room for improvement.

Performance of Curve Information Extraction

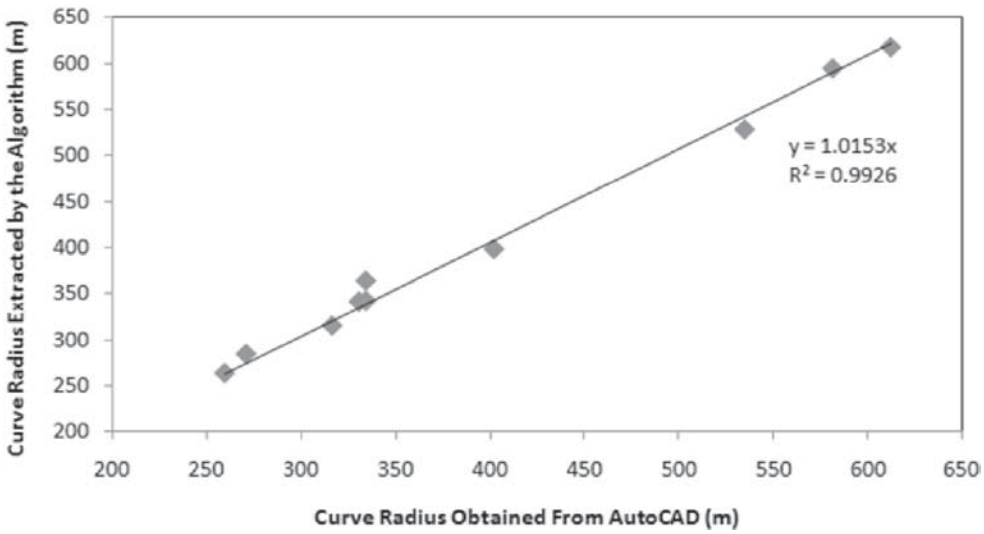
Figure 8, *c* through *e,* compares the algorithm-extracted curve length, radius, and degree of curvature with the ground truth geometric information obtained from AutoCAD for 10 simple curves. In each of the three figures, the horizontal axis represents the ground truth data, and the vertical axis represents the geometric information extracted by the algorithm. The linear regression analysis was performed with a fixed intercept of zero. The algorithm's output was considered accurate if the slope of the regression line was near one, which meant that the algorithm-extracted geometric information was expected to be identical to the ground truth geometric information. According to Figure 8, *c* through *e,* the slopes of the regression lines of curve length, radius, and degree of curvature are 0.9993, 1.0153, and 0.9789, respectively. All of the slopes are very close to one. In addition, all the regressions



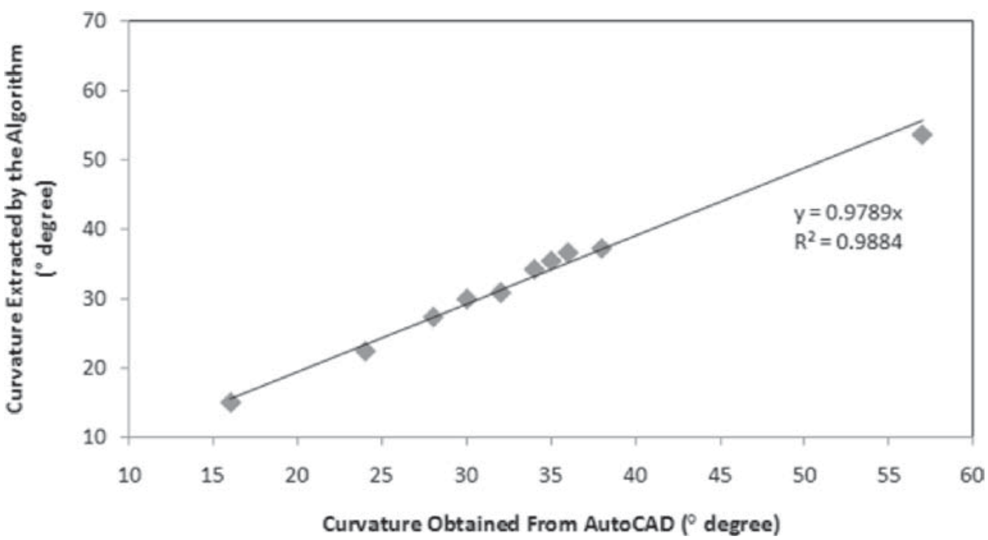
(c)

FIGURE 8 Algorithm validation results: (*a*) curve identification, (*b*) curve classification, and (*c*) curve length.

(continued on next page)



(d)



(e)

FIGURE 8 *(continued)* **Algorithm validation results: (*d*) curve radius and (*e*) degree of curvature.**

have an R^2 that is close to one. Both facts strongly suggest that the curve information extracted by the algorithm is very accurate.

DISCUSSION OF ALGORITHM PERFORMANCE

The validation results presented in the previous section have proved the effectiveness of the proposed fully automated algorithm for the extraction of horizontal curve information from GIS roadway maps. The overall curve identification rate is as high as 96.7%, which means that 96.7% of all curves can be completely detected by the proposed algorithm. Through the validation by ground truth geometric data, the algorithm was also tested to be reliable in extracting curves' geometric information, including the curve length, radius, and degree of curvature, with a high accuracy.

However, identification and classification errors were also found during the validation process. Each error was investigated using the aerial map, and the major reasons for the errors are summarized below.

The typical reason for Type 1 errors is the use of an improper bearing angle threshold. The sensitivity analysis found that the selection of the bearing angle is critical to the accuracy of the curve identification. A larger threshold increases the possibility of identifying curve sections with large radii as tangents, a Type 1 error. For example, in some long and smooth curves that are composed of many small segments, the bearing angle between consecutive segments is sometimes smaller than 1°. Therefore, if the optimal threshold of 1.25° were used, these curve segments would be identified as tangents. Another example of a Type 1 error is that the middle section of a long and smooth reverse curve is very likely to be misidentified as a tangent, because the bearing angles in the middle section of this type of curve are typically smaller than 1°. However, simply reducing the bearing angle threshold is not a solution, as a threshold that is below 1° can significantly increase the number of Type 2 errors.

In addition, the inconsistency in the roadway alignment between the GIS map and the aerial map is a major contributor to Type 2 and curve classification errors. This inconsistency is a result of the existing alignment bias of the GIS roadway map (i.e., an error of the GIS roadway map). The bias is mostly caused by the map's mapping and digitizing error. For example, some tangents are not properly mapped as a straight line in GIS but are a combination of sawtooth-like segments whose vertices are distributed on both sides of the tangent. In this case, the tangent segment may be misidentified as a compound curve by the algorithm, thereby causing a Type 2 error. Similarly, this type of sawtooth-like segment sometimes occurs in the middle of a long simple circular curve, and a classification error may therefore occur because of the misclassification of the simple circular curve as a compound curve. Based on observation, about 10% of all the curves used in the algorithm calibration and validation process contain intrinsic GIS map error.

Moreover, the scale of GIS roadway maps also contributes to the errors in the identification of horizontal curves. This scale can be reflected by the distance between successive GIS roadway vertices. The literature indicated that the distance between consecutive GIS points can impact the error occurrence (*30*). A similar finding was also observed in this research: that longer distances may increase the occurrence of Type 1 errors.

CONCLUSIONS AND RECOMMENDATIONS

The original efforts that have been carried out in this research are summarized as follows:

- The development of a fully automated algorithm and the implementation of the algorithm in ArcMap for the identification of curve locations. The algorithm also classifies curves as simple or compound and computes the curve radius, the central angle for every simple curve, and the length of each curve. In addition, the curve features and layers are automatically created in ArcMap.
- The calibration of the only input parameter for the algorithm and the identification of an optimal value for the bearing angle threshold: 1.25°.
- The comparison of advantages over existing GIS-based approaches. The advantages include full automation, only GIS roadway maps required, and no additional data collection needed.
- The validation of the algorithm. The algorithm successfully identifies curves at an identification rate of 96.7%. The algorithm also accurately extracts geometric information of simple curves.

Future research will focus on improving the algorithm by reducing the Type 2 errors and by increasing the overall identification rate, as well as the classification success rate. The algorithm will also be improved to accommodate more existing alignment errors that are present in most GIS maps. In addition, methods for the extraction of the geometric information of spiral and reverse curves will be investigated.

ACKNOWLEDGMENTS

This project was funded by the Wisconsin Department of Transportation (DOT). The authors thank John Corbin and Rebecca Szymkowski from the Wisconsin DOT for their support of this project. The authors thank also Wilson O. Vega of the University of Puerto Rico and Kelvin R. Santiago-Chaparro, Lang Yu, and Michael DeAmico of the University of Wisconsin–Madison for their help during the algorithm calibration and validation process.

REFERENCES

1. National Center for Statistics and Analysis. *Traffic Safety Facts, 2007 Data: Rural/Urban Comparison.* Publication DOT HS 810 996. NHTSA, Washington, D.C., 2007.
2. NHTSA, U.S. Department of Transportation. Fatality Analysis Reporting System (FARS) Encyclopedia. 2007. http://www-fars.nhtsa.dot.gov/Crashes/.
3. Zegeer, C. V., J. R. Stewart, F. M. Council, D. W. Reinfurt, and E. Hamilton. Safety Effects of Geometric Improvements on Horizontal Curves. In *Transportation Research Record 1356,* TRB, National Research Council, Washington, D.C., 1992, pp. 11–19.
4. Lamm, R., E. M. Choueiri, J. C. Hayward, and A. Paluri. Possible Design Procedure to Promote Design Consistency in Highway Geometric Design on Two-Lane Rural Roads. In *Transportation Research Record 1195,* TRB, National Research Council, Washington, D.C., 1988, pp. 111–122.
5. Lin, F. B. Flattening of Horizontal Curves on Rural Two-Lane Highways. *Journal of Transportation Engineering,* Vol. 116, No. 2, 1990, pp. 181–193.
6. Torbic, D. J., D. W. Harwood, D. K. Gilmore, R. Pfefer, T. R. Neuman, K. L. Slack, and K. K. Hardy. *NCHRP Report 500: Guidance for Implementation of the AASHTO Strategic Highway Safety Zone. Volume 7: A Guide for Reducing Collisions on Horizontal Curves.* Transportation Research Board of the National Academies, Washington, D.C., 2004.
7. Schneider, W. H., IV, K. H. Zimmerman, D. Van Boxel, and S. Vavilikolanu. Bayesian Analysis of the Effect of Horizontal Curvature on Truck Crashes Using Training and Validation Data Sets. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2096,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 41–46.
8. Schneider, W. H., IV, P. T. Savolainen, and D. N. Moore. Effects of Horizontal Curvature on Single-Vehicle Motorcycle Crashes Along Rural Two-Lane Highways. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2194,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 91–98.
9. Dong, H., S. M. Easa, and J. Li. Approximate Extraction of Spiralled Horizontal Curves from Satellite Imagery. *Journal of Surveying Engineering,* Vol. 133, No. 1, 2007, pp. 36–40.
10. Easa, S. M., H. Dong, and J. Li. Use of Satellite Imagery for Establishing Road Horizontal Alignments. *Journal of Surveying Engineering,* Vol. 133, No. 1, 2007, pp. 29–35.
11. Breyer, J. P. Tools to Identify Safety Issues for a Corridor Safety-Improvement Program. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1717,* TRB, National Research Council, Washington, D.C., 2000, pp. 19–27.
12. Souleyrette, R. R., A. R. Kamyab, Z. Hans, A. J. Khattak, and K. K. Knapp. Systematic Identification of High Crash Locations. Presented at 81st Annual Meeting of the Transportation Research Board, Washington D.C., 2002.
13. Yun, D., and J. Sung. Development of Highway Horizontal Alignment Analysis Algorithm Applicable to the Road Safety Survey and Analysis Vehicle, ROSSAV. *Proceedings of the Eastern Asia Society for Transportation Studies,* Vol. 5, 2005, pp. 1909–1922.
14. Kim, J. S., J. C. Lee, I. J. Kang, S. Y. Cha, H. Choi, and T. G. Lee. Extraction of Geometric Information on Highway Using Terrestrial Laser Scanning Technology. *Proc., International Society for Photogrammetry and Remote Sensing Congress,* Vol. XXXVII, Beijing, China, 2008, pp. 539–544.
15. Hashim, I. H., and R. N. Bird. Exploring the Relationships Between the Geometric Design Consistency and Safety in Rural Single Carriageways in the UK. *Proc., 36th Annual Conference of the Universities' Transport Study Group,* Newcastle, United Kingdom, 2004.
16. Watters, P., and M. O'Mahony. The Relationship Between Geometric Design Consistency and Safety on Rural Single Carriageways in Ireland. *Proc., European Transport Conference,* Leiden, Netherlands, Association for European Transport, London, 2007.

17. Imran, M., Y. Hassan, and D. Patterson. GPS–GIS-Based Procedure for Tracking Vehicle Path on Horizontal Alignments. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 21, No. 5, 2006, pp. 383–394.
18. Hans, Z., T. Jantscher, R. Souleyrette, and R. Larkin. Development of a Statewide Horizontal Curve Database for Crash Analysis. *Proc., Mid-Continent Transportation Research Symposium,* Ames, Iowa, Institute for Transportation, University of Iowa, 2009.
19. Price, M. Under Construction: Building and Calculating Turn Radii. *ArcUser,* winter 2010, pp. 50–56.
20. *ArcGIS Desktop 9.3.1.* Environmental Systems Research Institute (ESRI), Inc., Redlands, Calif., 2009.
21. Rasdorf, W., D. J. Findley, C. V. Zegeer, C. A. Sundstrom, and J. E. Hummer. Evaluation of GIS Applications for Horizontal Curve Data Collection. *Journal of Computing in Civil Engineering,* Vol. 25, No. 2, 2012, pp. 191–203.
22. Dial, G., L. Gibson, and R. Poulsen. IKONOS Satellite Imagery and Its Use in Automated Road Extraction. In *Automatic Extraction of Man-made Objects from Aerial and Space Images,* Vol. 3, (E. P. Baltsavias, A. Grün, and L. V. Gool, eds.), A. A. Balkema, Lisse, Netherlands, 2001, pp. 357–367.
23. Zhao, H., J. Kumagai, M. Nakagawa, and R. Shibasaki. Semiautomatic Road Extraction from High-Resolution Satellite Image. *International Archives of Photogrammetry and Remote Sensing,* Vol. 34, No. 3B, 2002, pp. 406–411.
24. Anil, P. N., and S. Natarajan. Automatic Road Extraction from High Resolution Imagery Based on Statistical Region Merging and Skeletonization. *International Journal of Engineering Science and Technology,* Vol. 2, No. 3, 2010, pp. 165–171.
25. Keaton, T., and J. Brokish. Evolving Roads in IKONOS Multispectral Imagery. *Proc., International Conference on Image Processing,* Barcelona, Spain, IEEE, New York, 2003, pp. 1001–1004.
26. Transportation Statistics Office. *Geographic Information System (GIS): Curvature Extension for ArcMap 9.x.* Florida Department of Transportation, Tallahassee, 2010.
27. Roess, R. P., E. S. Prassas, and W. R. McShane. *Traffic Engineering,* 4th ed. Prentice Hall, Upper Saddle River, N.J., 2010.
28. *A Policy on Geometric Design of Highways and Streets,* 5th ed. AASHTO, Washington, D.C., 2004.
29. Wisconsin Department of Transportation. Wisconsin Information System for Local Roads. http://www.dot.wisconsin.gov/localgov/w-islr/.
30. Findley, D. J. *A Comprehensive Two-Lane, Rural Road Horizontal Curve Study Procedure.* PhD dissertation. North Carolina State University, Raleigh, 2011.

The Geographic Information Science and Applications Committee peer-reviewed this paper.