Estimation of Road Horizontal Alignment Using Public Bus GPS Data*

Bekir Bartin, Mojibulrahman Jami, and Kaan Ozbay

Abstract -- **This paper introduces an innovative, cost-effective, and efficient method for extracting road horizontal alignment using GPS data collected from public transit buses. An Artificial Neural Network (ANN) model is used to predict whether a GPS point is part of a curved or tangent segment, and to further identify the type of curve (convex or concave). The novelty of this approach lies in the simplicity of the required input to the ANN model, which includes the latitude and longitude of the target GPS vertex and those of its neighboring vertices. Due to the difficulty in obtaining actual road alignment data, the ANN model is trained using synthetically generated GPS data along curved and tangent segments with random horizontal alignments. The model is then independently evaluated using real-world GPS data collected on a freeway located in Istanbul, Turkey. The results demonstrate high accuracy, with the ANN model correctly predicting 84.9 percent of the GPS vertices. The overlap between the estimated and actual section lengths is 95 percent for curved sections and 85 percent for tangent sections. Furthermore, the estimated curve radii show a high degree of similarity to the actual values, ranging from 0.8 to 2.1 percent.**

I. Introduction

The geometry of roadways, particularly horizontal curves, is of interest for traffic safety studies. Previous studies indicate that curved segments are more prone to accidents and account for a larger portion of severe crashes compared to straight road sections [1]. Geometric parameters of these curves, such as their length and radius, directly influence the likelihood of crashes [2], making curve-related crashes a significant component of the American Association of State Highway and Transportation Officials (AASHTO)'s strategic highway safety plan [3]. Identification of these high-risk segments and the determination of their geometric characteristics, inclusive of length, radius, point of curvature, point of tangency, and deflection angle, is not only a crucial task for various safety analyses but also contributes to reducing crash risks through the effective deployment of safety measures.

However, the extraction of roadway horizontal alignment, along with its required geometric properties, is often labor-intensive and costly, especially given the fact that these data are not always readily available in most states' databases [4][5]. This challenge arises primarily because traditional methods of collecting curvature data require considerable cost and time. Consequently, there is a critical need to formulate a new method that can efficiently, accurately, and affordably extract roadway horizontal alignment and identify its geometric characteristics.

Various methods have been employed for the collection of horizontal alignment data, including field surveys, Global Positioning System (GPS) methods, light detection and ranging (LiDAR), high-resolution satellite imagery, manual extraction via online mapping services and Geographic Information System (GIS) methods [6][7]. The widespread use of GPS technology in public transportation systems, such as buses, has introduced new possibilities for large-scale and cost-effective extraction of roadway alignment data. These vehicles, due to their fixed and widespread routes, generate a substantial amount of GPS data which can be used to infer road geometries. However, the interpretation of this data is not straightforward, as buses do not always strictly follow the centerline of roads due to traffic conditions, stops, and diversions, while the accuracy of the data can be influenced by a variety of factors including the quality of GPS receivers, the density of surrounding buildings and irregular bus movement patterns. Despite these challenges, the potential benefits of this approach in terms of scalability, cost-effectiveness, and real-time capability make it a promising area in roadway alignment extraction.

The aim of this paper is to introduce an artificial neural network (ANN) based approach which utilizes GPS trajectory data collected from public transit buses, widely available in public transportation agencies, for the purpose of identifying horizontal curves and determining their geometric properties. The ANN model is initially trained and tested using synthetic GPS data. The model is capable of identifying whether a GPS point is located on a tangent segment or a curved segment, by taking into account its neighboring GPS data points. The model is also able to identify different types of curves, i.e. convex and concave curves. The developed ANN model is independently evaluated using the bus GPS data collected on a freeway located in Istanbul, Turkey.

II. Literature Review

Various approaches have been employed to extract roadway horizontal alignment information, including Field

*The study was supported by the NJDOT (FHWA-NJ-2019-007) and partially by C2SMART, a Tier 1 UTC at New York University funded by the USDOT, and partially by Ozyegin University.

Bekir Bartin is with Civil Engineering Department, Ozyegin University (bekir.bartin@ozyegin.edu.tr)

Mojibulrahman Jami is with Civil Engineering Department, Ozyegin University (mojibulrahman.jami@ozu.edu.tr)

Kaan Ozbay is with Civil and Urban Engineering Department, New York University, (kaan.ozbay@nyu.edu)

surveys, high-resolution satellite imagery, light detection and ranging, GIS maps, and GPS trajectories. Each method has its advantages and limitations in terms of accuracy, efficiency, and cost.

Field surveys [8], along with the extraction of information from available as-built plan files, although offering high accuracy, are often time-intensive and expensive. As such, these methods prove impractical for conducting large scope analysis.

High-resolution satellite imagery was also used in extracting roadway horizontal alignment information. Various image processing techniques have been employed by researchers to detect roadway geometry from satellite imagery [9]-[13]. Dong et al. [10][11] and Ease et al. [11] investigated the identification of horizontal curves from IKONOS satellite images, demonstrating the feasibility of deriving geometric characteristics of simple, compound, and spiral curves using an approximate algorithm based on high-resolution satellite images. However, the accuracy of this approach can diminish when extracting information from urban roadway images due to the variety of land cover, and the method relies heavily on image resolution, making high-resolution commercial images relatively expensive.

The extraction of roadway alignment data has also seen advancements with the application of LiDAR technology. Both Mobile LiDAR Mapping (MLM) and Airborne Laser Scanning (ALS) have been utilized to improve accuracy and efficiency in capturing roadway geometry. In this method a survey vehicle is equipped with lidar technology to collect real-world highway coordinates at a higher accuracy level along roadways, which can then be processed to derive highly accurate curvature data [8], [14]-[16]. However, when applied to numerous roads, considerable data post-processing becomes a necessity. On the other hand, ALS, with its ability to rapidly collect dense point clouds over large areas, can provide a high density of the collected data, often reaching 10-50 points per square meter, allowing for precise modeling of the roadway geometry [17]. Zhou et al.[17] developed an efficient framework to identify highway alignments and construct 3D highway models. This method involved using smoothly connected grid cells to pinpoint pavement points from ALS data, extracting pavement boundaries and lane markings with the α-shape algorithm and a tracking strategy, and finally minimizing a complex energy function to extract highway alignments and rebuild 3D models. However, despite its advantages, lidar-based road alignment identification also poses certain challenges. The raw lidar point cloud usually contains noise and outliers that need to be carefully filtered out. Also, the process of distinguishing road surfaces from other types of terrain in the point cloud can be complex, requiring sophisticated algorithms and considerable computational resources.

GIS-based methods appear to be ideal due to their cost and time efficiencies. However, these methods have their challenges. Since the GIS polylines are generated manually along the centerlines on orthophotos during the digitization, this process can compromise the accuracy of geometric properties due to factors like inaccurate digitization or low vertex resolution, leading to potential inaccuracies in geometric determinations. Various GIS-based approaches for the extraction of roadway alignment data were developed by researchers. Curve Calculator extension [18] is a tool for ArcGIS software, which enables curve extraction from GIS road networks. Users manually define the start and end points of a curve, and the tool subsequently computes the radius and curve length. Similarly, Xu and Wei [19] proposed an azimuth computation method, which calculates the directional difference between adjacent vertices at each vertex. A threshold of 2° was recommended on the basis of the bearing change distribution and the horizontal curves were identified if the directional change exceeded the threshold. In a similar vein, Li et al. [20] introduced CurveFinder, a fully automated and network-wide method that calculates bearing angles between successive vertices to automatically detect curves. However, due to the potential inaccuracies in GIS data points, these angle-based approaches may incorrectly classify very short and smooth curves as tangent segments. Bíl et al. [4] developed ROCA (Road Curvature Analyst), an ArcGIS toolbox application that allows users to automatically identify horizontal curve data using the Naive Bayes classification technique. This method considered not only the bearing angle but also five additional explanatory variables of roadway geometry for determining whether a road vertex was part of a curve or tangent. In a distinct approach, Bartin et al. [21] utilized the K-means clustering algorithm to detect horizontal curves using curvature values of GIS vertices. This method was further expanded by Bartin et al. [22] into CurvS, a web-based program capable of extracting, visualizing, and analyzing roadway horizontal alignment data at a network-wide level using GIS data. This method uses a nine-dimensional explanatory variable vector to classify segments as curved or tangent and utilizes a clustering approach to estimate the geometric properties of the segments. In a subsequent study, Bartin et al. [23] developed an ANN model to predict whether a vertex, extracted from GIS road data, was part of a curved or a tangent segment, utilizing seven explanatory variables derived from the latitude and longitude readings of GIS vertices.

For extensive analysis on a larger scale, GPS trajectory data also appear to be useful for estimating road horizontal alignment. These methods involve recording geographic coordinates using a GPS-equipped vehicle at frequent time intervals across the roadway. Subsequently, the identification of horizontal curves and computation of their radii are achieved using a specialized GIS program that analyzed the logged GPS data points along the curves. Imran et al. [24] developed a technique to integrate GPS data into GIS to determine the radius, length, and spiral length of horizontal curves. In addition, Yun and Sung [25] implemented various sensors, including an inertial measurement unit, a distance measuring instrument, and cameras, on a survey vehicle to capture real-world highway coordinates at a higher accuracy level[25]. A less expensive approach proposed by [26] involved the identification of horizontal curves using data gathered from sensor-equipped smartphones, which included accelerometers, magnetic sensors, gyroscope readings, and GPS. They used Butterworth low-pass filtering to minimize sensor noise, improved GPS accuracy with extended Kalman filtering, and estimated curve locations using a K-means clustering algorithm. Although the results suggested that this method could measure the radius of sharp curves with reasonable accuracy, it required multiple iterations of data collection on a road.

In summary, previous studies on GPS approaches have largely concentrated on vehicles equipped with a range of advanced technologies. However, the necessity of these specially equipped surveying vehicles for data collection has limited the applicability of the method to a broader set of roadways due to the substantial costs and extended data collection process. Recognizing these constraints, this study aims to develop an ANN approach that utilizes GPS coordinates collected from public transit buses.

III. Problem Formulation

The formulation is adopted from Bartin et al. [23]. A road's horizontal alignment can be represented as a set of K continuous road segments with piecewise curvilinear functions denoted by $\{f_1, f_2, \dots, f_K\}$. Let (x_{k-1}, y_{k-1}) and (x_k, y_k) be the start and end coordinates for each segment k, $1 \le k \le \#K$. Also let C and T denote the set indices of curved and tangent segments, respectively, and $K = C \cup T$. $\#C$ and $\#T$ are the number of curved and tangent segments. Each f_k can then be formulated as:

$$f_k(x) = f_k(x_{k-1}) + m_k(x - x_{k-1}) \text{ for } \boldsymbol{I}_C(k) = 0 \quad (1)$$

$$f_k(x) = y_k^c \mp \left(R_k{}^2 - (x - x_k^c)^2\right)^{1/2} \text{ for } \boldsymbol{I}_C(k) = \pm 1 \quad (2)$$

Where, $\boldsymbol{I}_C$ is an indicator function defined as

$$\boldsymbol{I}_C(k) := \begin{cases} -1 \text{ or } 1, & \text{if } k \in C \\ 0, & \text{if } k \notin C \end{cases} \quad (3)$$

$\boldsymbol{I}_C(k) = 1$ means segment k is convex, and $\boldsymbol{I}_C(k) = -1$ means it is concave. If tangent, then $m_k = (f_k(x_k) - f_k(x_{k-1}))/(x_k - x_{k-1})$ is the slope, otherwise x_k^c, y_k^c are center coordinates and R_k is the radius.

Let us now assume that a road is discretized by a set of n vertices, $P = \{p_i = (u_i, v_i) : u_i \in [x_0, x_K]\}$, $1 \le i \le n$. Further assume that this discretization process is error-prone, such that $f_k(u_i) = v_i + \varepsilon_i$. Let $\varepsilon_i = 0$, for the moment assuming that each vertex is positioned seamlessly along the road centerline. The objective is then, without any prior information on the real road alignment, to estimate $\{\hat{f}_1, \hat{f}_2, \dots, \hat{f}_{\hat{K}}\}$ where $\hat{K}$ is the estimated number of sections and $\hat{f}_{\hat{k}}(u_i) = \hat{v}_i$, $1 \le \hat{k} \le \hat{K}$, using the available set of vertices P that minimizes:

$$E = \sum_{i=1}^{n} \left(v_i - \hat{f}_{\hat{k}}(u_i)\right)^2 \quad (4)$$

The computation of E requires the estimates for the number of curved and tangent segments, i.e. $\hat{C}$, $\hat{T}$, and the segment's boundary points $(\hat{x}_k, \hat{y}_k)$. It also requires continuity constraints $\hat{f}_{k-1} = \hat{f}_k$, and the smoothness constraints $\hat{f}'_k = \hat{f}'_{k-1}$ at each estimated boundary point. The functional form of $\hat{f}_{\hat{k}}$ is also unknown, though it can either be a straight line or a portion of a curve equation. This minimization problem is challenging even with two functional forms, necessitating an iterative solution. However, when n is large, the problem becomes computationally costly, if not impossible, because $\hat{K}$ and $(\hat{x}_k, \hat{y}_k)$ are unknown. The complexity of the problem would be largely reduced if $\hat{K}$ and $(\hat{x}_k, \hat{y}_k)$ and the functional form of $\hat{f}_{\hat{k}}$ were known a priori. A two-step strategy was proposed in [23] as a more efficient resolution.

The first step is segmentation, where $\hat{K}$ and $(\hat{x}_k, \hat{y}_k)$ are estimated. The second step is curve fitting, where the piecewise least squares estimation of circular curve segments is performed.

Algebraic or geometric methods can be used to estimate the curve parameters $\hat{x}_k^c$, $\hat{y}_k^c$ and $\hat{R}_k$ in the second step [27]. Taubin circle fit, an algebraic method, is selected for estimating these curve parameters. The focus of this study is the segmentation step. The proposed methodology for the segmentation step is presented next.

IV. Methodology

This work stems from ANN-based classification method for extraction of roadway alignment data presented in [23], in which an ANN approach was used for the segmentation step of the minimization problem stated in Equation (4). Bartin et al. [23] applied this approach to GIS data to identify curved and tangent segments. This current study applies the same approach to GPS data, and extends it by identifying the curved segments by direction, i.e. convex and concave. The ANN model is expected to predict whether the GPS vertex belongs to a curved (convex or concave) or a tangent segment.

The novelty in the proposed segmentation process in this current paper is the simplicity of the required input to the ANN model. The input vector contains the latitude and longitude of the target GPS vertex and those of its neighboring vertices. The coordinates of one GPS vertex irrespective of the positions of its surrounding vertices are not sufficient for predicting whether it is part of a curved or a tangent segment. Therefore, the input includes the coordinates of the vertices within 300 meters before and after the target GPS vertex.

The caveat of the proposed methodology lies in the requirement for extensive real-world alignment data to sufficiently ensure the applicability of the ANN model for segmenting various types of roadway alignments. However, obtaining actual road alignment data proves challenging, as mentioned earlier. Therefore, this study employed synthetically generated road data with curved and tangent segments with random horizontal alignments. The accurately labeled vertices along these roads were utilized for training the proposed ANN model.

A. Synthetic Data Generation

Figure 1 presents the pseudocode for generating the synthetic roadway alignment dataset, performed using the C programming language. The process can be described briefly as follows. Following the same notation used earlier, let K denote the number of road segments to be generated. Each segment, k, is randomly assigned a label, 0, -1 or 1, indicating whether it is concave or convex, i.e. $I_C(k) = \pm 1$, or tangent segment, i.e. $I_C(k) = 0$. The first vertex starts at (0, 0). Then a chord length, LC_k is assigned randomly between LC^{min} and LC^{max}, with a deflection angle, θ_k randomly assigned between θ^{min} and θ^{max}, using a uniform distribution. End point (x_k, y_k) of the segment is calculated by using θ_k and

LC_k. If the segment is tangent, then the chord becomes the roadway alignment itself. Otherwise, using θ_k and LC_k, the radius of the curved segment is computed by $R_k = LC_k/2sin(\theta_k)$, and the curve direction is randomly assigned e.g. concave or convex curve, only when $k = 1$ or $I_C(k-1) = 0$. The center coordinates of the curve (x_k^c, y_k^c) are computed, and the circular curve function, f_k is determined. The consecutive segments are then generated by imposing an appropriate degree of angle at the start of next segment to ensure smoothness. To prevent excessive number of tangent or curve segments in sequence $cons\_lim$ threshold is imposed. Vertices are then generated at randomly generated lengths, Δ. Each vertex is rotated by a selected angle increment ω to populate training data. For output, vertices are chosen at random to achieve a balanced distribution of curved and tangent vertices. Also, the coordinates are distorted randomly by 4 meters, corresponding to roughly one lane width around the actual alignment, to replicate the noise in GPS data.

Repeat for each roadway
Set $(x_0, y_0) = (0,0)$
Set $\alpha = 0$ *#rotation angle,* $\omega \in (0,360)$ *#rotation increment in degrees*
Set h *#number of neighboring vertices based on buffer distance of ±300 m.*
Set $cons\_lim$, and set $\_cons[0] = 0$, $\_cons[1] = 0$
for each segment $k = 1 \dots K$ **do**
    $\boldsymbol{I}_C(k) \leftarrow \{0,-1,1\}$ *#randomly assign to 0, -1 or 1*
    $\_cons[\boldsymbol{I}(k)] \leftarrow \_cons[\boldsymbol{I}_C(k)] + 1$ *#increase the consecutive number of curved or tangent segments*
    if $\_cons\ [\boldsymbol{I}_C(k)] > cons\_lim$ **then**
        $\boldsymbol{I}_C(k) \leftarrow \{0,-1,1\} - \boldsymbol{I}_C(k)$ *#randomly assign that excludes the previously assigned label*
    end if
    Assign $LC_k \in [LC^{min}, LC^{max}]$
    Assign $\theta_k \in [\theta^{min}, \theta^{max}]$
    Compute (x_k, y_k) *#using* θ_k *and* LC_k
    if $\boldsymbol{I}(k) = 0$ and $k > 1$ and $\boldsymbol{I}_C(k-1) = 0$ **then** $\theta_k = \theta_{k-1}$

    if $\boldsymbol{I}(k) = -1$ or $\boldsymbol{I}(k) = 1$ **then**
        Compute R_k *#using* θ_k *and* LC_k
        if $\boldsymbol{I}(k-1) = 0$ or $k = 1$ **then**
            Randomly assign curve direction from $\{-1,1\}$
        else find curve direction *#using* f'_{k-1}, θ_k *and* R_k
        Compute center coordinates (x_k^c, y_k^c) and f_k
    end if
    Discretize segment k by vertices (u, v) at random intervals, Δ
    Insert random error to each (u_i, v_i)
    while $\alpha < 360$ **do**
        for each vertex i **do**
            Rotate vertex by α
            Update label of i if the curve direction is changed by rotation
            Output vector (label $i, u_{i-h}, v_{i-h}, \dots, u_i, v_i, \dots, u_{i+h}, v_{i+h}$)
        end for
        $\alpha = \alpha + \omega$
    end while
end for

Figure 1. Generating synthetic GPS data

The following values were used when generating synthetic data: $K = 5$ segments, $\omega = 18^o$, $cons\_lim = 2$ segments, $\theta^{min} = -30^o$, $\theta^{max} = 30^o$, $LC^{min} = 200$ meters, $LC^{max} = 1000$ meters. Using this process, a total of roughly 1.05 million synthetic vertices were generated to be used as a training dataset for the proposed ANN model.

B. ANN Model

The training of the ANN model was performed using the synthetically generated GPS dataset, as presented above. 80 percent of the generated dataset was used for training and 20 percent for tuning the model hyperparameters and testing the model.

The model was generated using the Keras library available in the TensorFlow package in Python. The ANN model was of a 512-256 structure, consisting of two hidden layers, and the last node is a softmax function. The model was trained for 100 epochs. The accuracies achieved were 92.9 percent with the training dataset and 90.1 percent with the test dataset.

V. Data Description

A. Available GPS Data

The public bus GPS dataset from March 8, 2022 was provided by the Istanbul Electric Tram and Tunnel Company (IETT), the public transport provider in the city of Istanbul, among other modes of public transportation, operates nearly 6,000 buses on 814 bus routes with a daily ridership of 4 million [29]. The dataset includes the following fields: Bus ID number, date, time, latitude and longitude readings in WGS 84 coordinate system. The dataset was in the format of comma separated values (*.csv) with a file size of 2.1 Gb. A C programming code was used to parse the dataset. Figure 2 shows the coverage of the GPS.

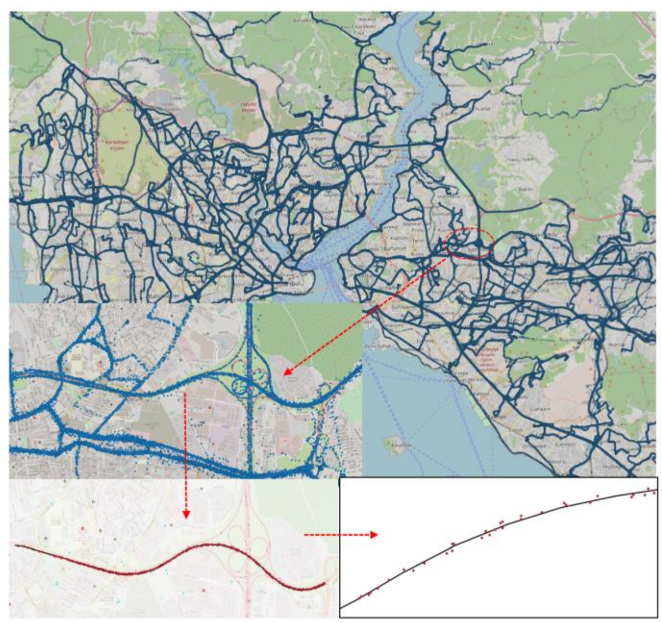

Figure 2. Bus GPS data of Istanbul

B. Study Roadway

The selected roadway is an old freeway located in the Asian Side of Istanbul. The roadway, although referred to as a freeway, has a traffic control that is neither of a freeway nor a principal arterial, with a combination of ramp access and uncontrolled at-grade access points. The portion of the roadway included for the analysis starts at (41.031206, 29.09403) and ends at (41.02786, 29.12992), and is 3.22

kilometers. It consists of three curved and three tangent segments. The selection of this specific portion was due to the availability of the horizontal alignment data, provided by the Istanbul Metropolitan Municipality (IMM).

C. Data Filtering

To identify and filter out GPS vertices along the selected portion of the study roadway, the GIS road centerline data were extracted as polylines from the OpenStreet Maps database, available online [29]. With the same C code used to parse the dataset, each vertex in the close vicinity of the roadway polylines was filtered out by removing any vertex whose projected distance from the polyline was greater than 4 meters. The filtered vertices are shown in the bottom left portion of Figure 2, and a close-up view of how the vertices are scattered around the polyline are shown in the bottom right corner. After filtering, a total of 385 GPS points were detected along the selected portion of the study road. Each GPS vertex's label was assigned based on the actual horizontal alignment, provided by the IMM.

It is worth mentioning that despite the smoother trajectory formed by the GPS vertices of the same bus along the road, the relatively longer time intervals between each GPS reading, approximately 30 seconds, resulted in significant spatial gaps between consecutive vertices, as shown in Fig 3.

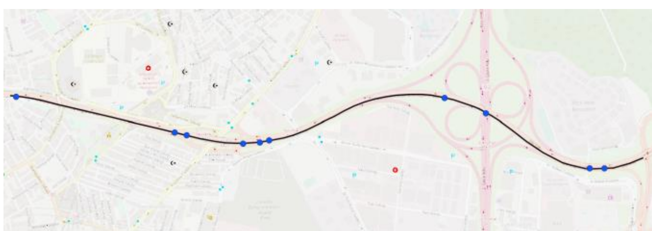

Figure 3. GPS vertices obtained from one bus

This spatial gap poses a challenge in accurately capturing the complete alignment information. Furthermore, it is important to note that buses do not traverse the entire length of the chosen road. However, by incorporating the GPS vertices obtained from multiple buses, a more comprehensive horizontal geometry emerges, albeit a noisy one, as shown in Figure 2. The proposed ANN model, trained with noisy synthetic GPS data, therefore, is valuable in identifying distinct segment types.

VI. Analysis Results

The ANN model was evaluated independently by using the filtered GPS vertices along the selected roadway. Note that ANN model was used for the segmentation process, in which $\hat{K}$ and $(\hat{x}_k, \hat{y}_k)$ were estimated. In the subsequent curve fitting step, these outputs of the segmentation step were used in the estimation of curve parameters, i.e. $\hat{x}_k^c$, $\hat{y}_k^c$ and $\hat{R}_k$. As mentioned earlier, Taubin circle fit method was implemented for the curve fitting step, coded in the C programming language.

The results show that the ANN model correctly predicted 84.9 percent of the GPS vertices. Further investigation revealed that 74.6 percent of the incorrectly predicted vertices are on tangent segments, and the remainder on curved segments. Also, 80.9 percent of the incorrectly labeled vertices are located at the transition regions between curves and tangent segments. In addition, none of the vertices on curved segments incorrectly predicted the curve direction, i.e. convex instead of concave curve, vice versa.

Using the predicted labels of the GPS vertices, segments on the selected road were estimated. These estimated segments, along with the start and end are presented in Table 1. The actual total curved and tangent section lengths are 1,938 meters and 1,281 meters, respectively. As shown in Table 1, all curved and tangent segments were detected by the ANN method. In addition, the estimated curve directions also matched the actual ones, i.e. two convex curves and one concave curve.

As seen in Table 1, the start and end points of sections do not match precisely, which is expected due to the noise in GPS data. Therefore, the ratio of overlap of the estimated tangent segments with the actual ones in length was 0.85. The total undetected tangent length was 196.8 meters. The overlap ratio of the estimated curved segments with the actual ones in length was 0.95, and the total undetected curve length was 94.9 meters.

Table 1. Analysis Results

Actual Alignment Data				**ANN Method**			
k	***Start***	***End*** *(m)*	R_k*(m)*	$\hat{k}$	***Start***	***End*** *(m)*	$\hat{R}_{\hat{k}}$ *(m)*
1	0	967.3	-	1	0	799.3	-
2*	967.3	1,479.6	800	2	799.3	1,551.7	782.9
3	1,479.6	1,673.0	-	3	1,551.7	1,714.6	-
4*	1,673.0	2,514.3	750	4	1,714.6	2,471.0	743.6
5	2,514.3	2,634.5	-	5	2,471.0	2,644.5	-
6*	2,634.5	3,218.9	515	6	2,644.5	3,218.9	496.3

*Segments 2 and 6 are convex curves and segment 4 is a concave curve.

The radii of curved segments were estimated using Taubin curve fitting method, and the results are shown in Table 1. As seen, the estimated curve radii demonstrate a high degree of similarity to the corresponding actual values. The percentage differences are 2.1, 0.8 and 3.6 percent for segments 2, 4 and 6, respectively.

VII. Conclusions & Future Work

This paper introduced an innovative approach to estimating horizontal alignment data of roadways using GPS data. The extraction of horizontal alignment is known to be time-consuming and expensive, thus necessitating an efficient and rapid method due to its vital role in various traffic safety-related analyses. Prior research endeavors regarding the use of GPS data predominantly focused on vehicles equipped with a diverse array of advanced technologies. In contrast, the proposed approach capitalizes on the available GPS data obtained from public buses, taxi cabs available in many countries, which often cover a vast majority of roadways in any study area.

An ANN-based segmentation approach was used to handle a large volume of GPS data collected from numerous vehicles, enabling the construction of complete yet noisy roadway trajectories. Each GPS vertex's classification, indicating whether it corresponds to a tangent or curved segment, was predicted through the proposed ANN model.

These predicted classifications were then utilized to estimate the number and characteristics (tangent or curved) of the road segments.

To train the ANN model, synthetically generated GPS data were employed. The results of the segmentation process were subsequently utilized in circular curve fitting to determine essential geometric parameters such as curve boundaries and radii. Notably, the developed ANN model was designed not only to identify curve segments but also to ascertain their direction, distinguishing between concave and convex curves. Consequently, the developed ANN model possessed the capability to detect reverse S-curves.

Public bus dataset along a selected route in Istanbul was used to independently evaluate the proposed approach. The findings of the study showed that the overlap between the estimated and actual section lengths was 95 percent for curved sections and 85 percent for tangent sections, indicating strong agreement. In addition, the percentage differences between the actual radii of the three curves along the selected route and the estimated ones by the proposed approach were 2.1, 0.8 and 3.6 percent.

At present, the proposed approach executes the segmentation and circular curve fitting steps independently. A complete analysis would require imposing both continuity and smoothness constraints in Equation (4) for a global minimum E, which would require simultaneous fitting of all segments. Finally, further evaluation of the proposed ANN model will be conducted once more actual horizontal alignment information becomes available.

References

[1] Gibreel, G. M., Easa, S. M., & El-Dimeery, I. A. (2001). Prediction of operating speed on three-dimensional highway alignments. Journal of Transportation Engineering, 127(1), 21–30

[2] AASHTO, *Highway Safety Manual.* Washington, DC, 2011.

[3] Torbic, D. J., Harwood, D. W., Gilmore, D. K., Pfefer, R., Neuman, T. R., Slack, K. L., & Hardy, K. K. (2004). Guidance for implementation of the AASHTO strategic highway safety plan. In Volume 7: A guide for reducing collisions on horizontal curves (No. Project G17-18 (3) FY'00).

[4] Bíl, M., Andrášik, R., Sedoník, J., and Cícha, V. (2018). ROCA – An ArcGIS toolbox for road alignment identification and horizontal curve radii computation. *PLoS ONE.* 2018.*13*(12).

[5] Ma, Q., Yang, H., Wang, Z., Xie, K. and Yang, D. (2020). Modeling crash risk of horizontal curves using large-scale auto-extracted roadway geometry data. *Accident Analysis & Prevention*. Vol. 144.

[6] Carlson, P. J., Burris, M., Black, K., and Rose, E. (2005). Comparison of Radius-Estimating Techniques for Horizontal Curves *Transportation Research Record: Journal of the Transportation Research Board.* 1918 (1), 76–83.

[7] Ibraheem, A.Th., and Janan, F. (2011). Developing a Computer Program for the Methods of Radius-Estimating Techniques for Horizontal Curves. *American Journal of Engineering and Applied Sciences.* 4 (2), 276-287.

[8] Findley, D.J., Hummer, J.E., Rasdorf, W. and Laton, B.T. (2012). Collecting Horizontal Curve Data: Mobile Asset Vehicles and Other Techniques. Journal of Infrastructure Systems. 19 (1), 74-84.

[9] Anil, P. N., & Natarajan, S. (2010). Automatic Road Extraction from High Resolution Imagery Based on Statistical Region Merging and Skeletoniz. In International Journal of Engineering Science and Technology (Vol. 2, Issue 3).

[10] Dong, H., Easa, S. M., & Li, J. (2007). Approximate Extraction of Spiralled Horizontal Curves from Satellite Imagery. Journal of Surveying Engineering, 133(1), 36–40. https://doi.org/10.1061/(ASCE)0733-9453(2007)133:1(36)

[11] Easa, S. M., Dong, H., & Li, J. (2007). Use of Satellite Imagery for Establishing Road Horizontal Alignments. Journal of Surveying Engineering, 133(1), 29–35. https://doi.org/10.1061/(ASCE)0733-9453(2007)133:1(29)

[12] Keaton, T., & Brokish, J. (2003). Evolving roads in IKONOS multispectral imagery. IEEE International Conference on Image Processing, 3, 1001–1004. https://doi.org/10.1109/ICIP.2003.1247416

[13] Zhao, H., Kumagai, J., Nakagawa, M., & Shibasaki, R. (2002). Semi-automatic road extraction from high-resolution satellite image. International Archives of Photogrammetry and Remote Sensing, Vol.34((Part3B)), 406–411.

[14] Ai, C., & Tsai, Y. (2015). Automatic horizontal curve identification and measurement method using GPS data. Journal of Transportation Engineering, 141(2). https://doi.org/10.1061/(ASCE)TE.1943-5436.0000740

[15] Gargoum, S., El-Basyouny, K., & Sabbagh, J. (2018). Automated Extraction of Horizontal Curve Attributes using LiDAR Data. Transportation Research Record, 2672(39), 98–106. https://doi.org/10.1177/0361198118758685

[16] Luo, W., Li, L., & Wang, K. C. P. (2018). Automatic Horizontal Curve Identification and Measurement Using Mobile Mapping System. Journal of Surveying Engineering, 144(4). https://doi.org/10.1061/(ASCE)SU.1943-5428.0000257

[17] Zhou, Y., Huang, R., Jiang, T., Dong, Z., & Yang, B. (2021). Highway alignments extraction and 3D modeling from airborne laser scanning point clouds. International Journal of Applied Earth Observation and Geoinformation, 102. https://doi.org/10.1016/j.jag.2021.102429

[18] ArcGIS Desktop. Environmental Systems Research Website. https://www.esri.com/en-us/arcgis/products/arcgis-desktop/Accessed on May 15 2023.

[19] Xu, H., & Wei, D. (2016). Improved identification and calculation of horizontal curves with geographic information system road layers. Transportation Research Record, 2595, 50–58. https://doi.org/10.3141/2595-06

[20] Li, Z., Chitturi, M., Bill, A., & Noyce, D. (2012). Automated identification and extraction of horizontal curve information from geographic information system roadway maps. Transportation Research Record, 2291, 80–92. https://doi.org/10.3141/2291-10

[21] Bartin, B., Ozbay, K., & Xu, C. (2019). Extracting Horizontal Curvature Data from GIS Maps: Clustering Method. Transportation Research Record, 2673(11), 264–275. https://doi.org/10.1177/0361198119850789

[22] Bartin, B., Demiroluk, S., Ozbay, K., & Jami, M. (2022). Automatic Identification of Roadway Horizontal Alignment Information Using Geographic Information System Data: CurvS Tool. Transportation Research Record, 2676(1), 532–543. https://doi.org/10.1177/03611981211036364

[23] Bartin, B., Jami, M., & Ozbay, K. (2021). Estimating Roadway Horizontal Alignment using Artificial Neural Network. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2021-September, 2245–2250. https://doi.org/10.1109/ITSC48978.2021.9565062

[24] Imran, M., Hassan, Y., & Patterson, D. (2006). GPS-GIS-based procedure for tracking vehicle path on horizontal alignments. Computer-Aided Civil and Infrastructure Engineering, 21(5). https://doi.org/10.1111/j.1467-8667.2006.00444.x

[25] Yun, D., & Sung, J. (2005). 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, 5.

[26] Wood, J. S., & Zhang, S. (2018). Identification and Calculation of Horizontal Curves for Low-Volume Roadways Using Smartphone Sensors. Transportation Research Record, 2672(39), 1–10. https://doi.org/10.1177/0361198118759005.

[27] Chernov, N. (2010). Circular and linear regression : fitting circles and lines by least squares. 286.

[28] Song, Z., Yang, F., Schonfeld, P., Liu, H., & Li, J. (2021). Integrating segmentation and parameter estimation for recreating vertical alignments. Computer-Aided Civil and Infrastructure Engineering, 36(4), 472–488.

[29] Istanbul Electric Tram and Tunnel Company Website (2023). www.iett.istanbul. Accessed on May 17, 2023.

[30] Open Street Map Website. (2023). www.openstreetmap.org. Accessed on May 2, 2023.