MTH 312: Project Report

Curve Detection of roads using Random Forest Techniques

Group 4:

Subham Anand (221093), Swapnonil Mondal (231080098),
Samarth Kumar (210911), Sujash Krishna Basak (231080093)

Supervisor: **Dr. Dootika Vats**

Abstract

Roads often have curves to help vehicles change direction smoothly and safely. Understanding whether a road section is straight or curved is crucial for road design and traffic safety. This project aims to classify road points as either part of a curve or a straight section using given coordinates (X, Y). Here two main approaches are taken: creating new features that are related to curves and training machine learning classifiers. The main goal is to develop a function that processes these datasets and classifies each point as either part of a curve (1) or a straight section (0).

1 Introduction

Road alignment classification and curve identification are essential for transportation safety and infrastructure planning. Ai & Tsai (2015) proposed an automatic method using GPS data to detect horizontal curves, improving accuracy in road geometry analysis. Bíl et al. (2018) developed the ROCA toolbox within ArcGIS to automate road alignment identification and compute curve radii, enhancing GIS-based road assessment. More recently, Atif & Sil (2023) evaluated different methodologies for extracting horizontal alignment data, emphasizing cost-effectiveness and efficiency. These studies highlight the growing role of automation and data-driven approaches in road classification, aligning with our project's objective of using machine learning for curve detection. For this study, we use two datasets. The first one is an Indian highway dataset with high-resolution points recorded every 10 meters, along with radius estimates to classify curves. The second dataset, from the Czech Republic, contains road points with variable resolution and labeled curves, though the exact classification method used is unknown.

2 Background

2.1 SVM:

SVM is a supervised learning algorithm that classifies data by finding the optimal hyperplane that maximizes the margin between classes. It handles both linear and non-linear data using kernel functions and is effective in high-dimensional spaces.

2.2 Random Forest:

Random Forest is a machine learning algorithm that works by combining multiple decision trees to improve accuracy and reduce overfitting. Each tree is trained on a random subset of the data and makes its own prediction. The final output is determined by taking a majority vote (for classification). By using multiple trees and random feature selection, random Forest can help us to detect curves more accurately.

3 Methodology

Our dataset consists of only X and Y coordinates, along with labels indicating whether a road section is curved or straight. However, predicting curves using just these features is challenging. To improve accuracy, we created additional features that capture the road's geometric properties. These include angle between three consecutive points, distance between two points, curvature, change in slope. By adding these features, we can increase the model's ability to distinguish between curved and straight sections more accurately.

Now we define our additional features, which are:

- **Angle between three consecutive points** – We take three consecutive points from the road and measured the angle between them.
- **Distance between two points** – We are considering the eucledian distance between two points.
- **Curvature** – Compute curvature as the reciprocal of radius of the circle passing through 3 points.
- **Change in slope** – We took norm of difference of two slopes.
- **Adaptive Sparsity Measure** - Used in predicting road curves by quantifying variations in point distribution; curved sections often show less uniformity and higher sparsity compared to straight, evenly spaced segments. It quantifies the sparsity of a vector $\mathbf{x}$, defined as:

  $$S(\mathbf{x}) = f\left(\frac{\|\mathbf{x}\|_p}{\|\mathbf{x}\|_1}\right)$$

  where $\|\mathbf{x}\|_1$ is the L_1-norm and $\|\mathbf{x}\|_p$ is a generalized L_p-norm.

- **Local Density Variation** - This captures changes in point spacing, as curved sections typically have more variation in distances between consecutive points compared to straight sections. It is defined as the ratio of the standard deviation to the mean of the pairwise Euclidean distances:

  $$\text{LDV} = \frac{\text{std}(d_1, d_2, d_3, d_4)}{\text{mean}(d_1, d_2, d_3, d_4)}$$

  where d_1, d_2, d_3, d_4 are the pairwise distances between consecutive points in the local neighborhood.

- **Local linearity** - Local linearity helps predict road curves by approximating small curve segments as linear, making it easier to detect subtle directional changes.

Since our added features, like curvature and change in slope, likely have a relationship with whether a road is curved or straight, our goal is to identify these patterns. However, a simple linear classifier like SVM may struggle to capture these complex relationships effectively. Curves in a road do not always follow a clear linear boundary, making it difficult for a hyperplane-based model to classify them accurately. To address this, we take a more flexible approach by generalizing the concept of SVM and using Random Forest. This allows us to model nonlinear relationships more effectively and improve the accuracy of curve classification.

After creating the features, we normalized them to bring all values to the same scale for better consistency and model performance.

4 Results

We split the Czech data into train and test with 80% and 20% respectively.

4.1 Model Performance Metrics:

It is shown in Table 1

Table 1: Performance metrics of the random forest model

Class	Precision	Recall	F1-Score	Support
0 (Straight Road)	0.93	0.97	0.95	977
1 (Curve Road)	0.86	0.74	0.80	274
Accuracy	-	-	0.92	1251

4.2 Explanation:

Precision: Measures how many of the predicted curved/straight points are actually correct. A precision of 0.93 for straight roads means that 93% of points predicted as straight are truly straight. Similarly, 86% of points predicted as curves are actually curves.

Recall: Measures how well the model identifies all actual curves and straight sections. The recall for straight roads (0.97) indicates that 97% of all straight points were correctly classified, while for curves (0.74), 74% were correctly identified.

F1-Score: Balances precision and recall. Higher values indicate better performance, with an overall F1-score around 0.92, showing a good balance in classification.

Accuracy (92%): The percentage of total correctly classified points out of all 452 road points.

Overall, the model performs well, achieving an overall F1 score of around 0.99 for the India data set, meaning it effectively classifies road curves and straight sections.

5 Conclusion

In this project, we aimed to classify road sections as curved or straight using X and Y coordinates. To improve the classification, we created additional features like angle between consecutive points, distance, curvature, and change in slope. These features were normalized to bring them to the same scale. We used a Random Forest model, which captured the complex relationships between the features and road curvature better than simpler models like SVM. The model achieved F1-score around 0.82 for our test dataset.

In conclusion, this project demonstrates how Random Forest can effectively classify road curves, providing reliable results for road curve detection.

References

Ai, C. & Tsai, Y. (2015), 'Automatic horizontal curve identification and measurement method using gps data', *Journal of Transportation Engineering* **141**(2), 04014078.

Atif, M. & Sil, G. (2023), Investigation of horizontal alignment data extraction methodologies in terms of cost and time, *in* 'Proceedings of the Institution of Civil Engineers-Transport', Emerald Publishing Limited, pp. 1–23.

Bíl, M., Andrášik, R., Sedoník, J. & Cícha, V. (2018), 'Roca–an arcgis toolbox for road alignment identification and horizontal curve radii computation', *PLoS One* **13**(12), e0208407.