Development of map-matching algorithm using AI-ML techniques distinguish vehicular movement on highway and service road

Chintada Tarun
Chitrada Sravani
Chitturi Pavani satya sai
Duba Jaya shree

Author Affiliations
JNTU-GV COLLEGE OF ENGINEERING, VIZIANAGARAM
JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY- GURAJADA VIZIANAGARAM
DWARAPUDI, VIZIANAGARAM, ANDHRA PRADESH – 535003.
(A constituent college of JNTU-GV & Approved by AICTE ,New Delhi)(Recognised by UGC under section 2(f)&12(B) of UGC Act 1956)

Author Emails
tarunchintada2004@gmail.com
sravanichitrada03@gmail.com
pavanichitturi24@gmail.com
dubajayashree@gmail.com

Abstract: Efficient and accurate mapping of vehicular movement on road networks is essential for traffic management, urban planning, and infrastructure development. This project develops a data-driven solution that leverages geographic information and vehicle tracking data to determine the positioning of vehicles on highways or service roads using clustering and map-matching techniques. The system integrates two datasets: a geo dataset containing latitude, longitude, and address details, and a vehicle path dataset with attributes such as GPS coordinates, speed, and direction.

The core process involves creating a graph-based model from the geo dataset, where each node represents a road segment with geodesic distances as edge weights, and matching vehicle GPS points to the closest nodes in the road network using geodesic distance calculations. This approach provides accurate road segmentation by aligning vehicle trajectories with predefined road structures, distinguishing between highways and service roads in real-time. For visualization, the graph-based road network and matched vehicle points are plotted, showing the association of each GPS trajectory with its closest road segment.

Python libraries such as NetworkX and Geopy enable graph construction and distance measurement, while Matplotlib is used for graphical representation. The system's potential applications span from real-time traffic monitoring to automated route optimization, supporting smart city initiatives and enhancing data-driven decision-making for urban infrastructure. This document details the project's design, functionality, and implementation approach, offering insights into potential enhancements such as scaling the model to larger geographic areas and integrating additional real-time data sources.

1.**Introduction**: In recent years, the demand for automated traffic management and enhanced urban planning has surged as cities seek to accommodate increasing vehicle numbers and improve traffic flow. This project focuses on the development of a solution that can accurately determine vehicular positioning on highways or service roads, utilizing a combination of GPS data and map-matching techniques. By applying clustering algorithms and geospatial data analysis, the system processes complex movement patterns from raw GPS data, providing insights into road usage and aiding in traffic flow optimization.

The project integrates two key datasets: a geo dataset containing latitude, longitude, and address data to define road networks, and a vehicle path dataset that includes GPS coordinates, speed, and direction of vehicles. By creating a graph from the geo dataset, where each node represents a location on the road network and each edge is weighted by geodesic distance, the system accurately models road connectivity. Each vehicle GPS point is then matched to the closest road segment using geodesic distance calculations, distinguishing whether the vehicle is on a highway or a service road.

Python libraries such as NetworkX, Geopy, and Matplotlib enable efficient graph construction, distance calculations, and visualization. The system's visualization plots the road network and GPS points, visually confirming the vehicle-road association, enhancing its applicability to real-time monitoring and automated traffic systems.

Beyond simple vehicle-road association, this solution has potential to support adaptive traffic management systems, enabling cities to respond dynamically to real-time traffic conditions. With further refinement, this technology could also be integrated into broader smart city frameworks, where data from various sources (e.g., weather, events) could help predict traffic surges or identify alternate routes. This capability would aid in both short-term adjustments and long-term urban planning, facilitating smoother commutes, reducing emissions from idling vehicles, and contributing to sustainable infrastructure development.

Each vehicle GPS point is then matched to the closest road segment using geodesic distance calculations, which ensures a high degree of accuracy in determining whether a vehicle is traveling on a highway or service road. This map-matching technique, which considers the geographic relationship between GPS data points and the road network, allows for effective clustering of vehicular movements. By identifying specific vehicle flow patterns and road type usage, the system can provide valuable insights into traffic density, peak hour congestion, and road wear-and-tear trends.

This document provides an overview of the project's technical approach, design, and visualization outputs, demonstrating its value for smart city initiatives. Further potential improvements, such as real-time data integration and expanded geographic coverage, are also discussed, showcasing the system's adaptability to a range of urban planning and traffic management applications.

3. System Architecture:

The architecture of the map-matching and vehicle tracking solution is built on a modular design that facilitates efficient processing and easy integration into existing traffic management systems. Key architectural components include:

Machine Learning and Data Processing Foundation:

The system leverages advanced data processing and clustering techniques for identifying vehicular movement patterns. Using geodesic calculations, the architecture can accurately map GPS points to specific road segments, distinguishing vehicles traveling on highways versus service roads. This foundation ensures high precision in vehicle-road associations, which is critical for generating accurate traffic insights and flow analysis.

Core Technologies

- **Programming Languages and Libraries:** Python, with libraries such as Pandas, NetworkX, and Geopy, for data manipulation, graph-based road network modeling, and distance calculations.
- **Data Visualization:** Matplotlib for visualizing the road network and matched GPS points, assisting in the validation and presentation of results.
- **Geospatial Tools:** Geopy for precise geodesic calculations to ensure reliable road-vehicle mapping.

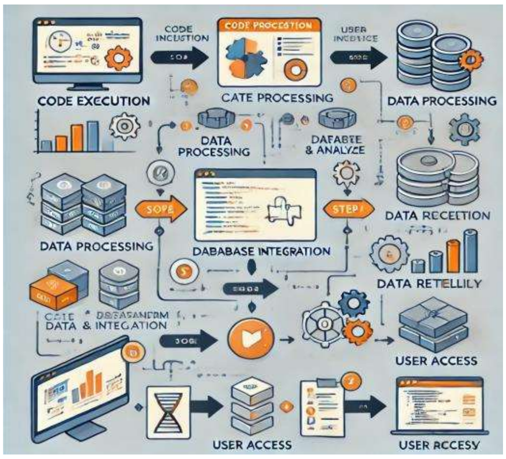
System Integration:

The system is designed with a microservices-inspired structure, where data processing, map-matching, and clustering components interact seamlessly. Each module performs specific tasks, from GPS filtering to graph creation and visualization, enhancing the system's adaptability. The structure is ideal for scaling to meet increasing demands, allowing seamless integration with traffic monitoring systems or as part of a larger smart city framework.

Scalability and Cloud Deployment:

The architecture is built to be cloud-deployable, making it capable of handling large GPS datasets in real-time or batch mode. With further development, it can be hosted on cloud platforms for scalable storage and processing, enabling broader accessibility for cities or organizations with high data throughput needs.

Scalable API Integration: The architecture can be expanded with RESTful APIs to allow seamless integration with external traffic management systems, navigation applications, and city infrastructure platforms. This will enable real-time data sharing and enhance interoperability across various smart city initiatives, making it easier for organizations to incorporate traffic insights into broader decision-making and operational workflows.



4. Data Processing and Feature Engineering:

Effective data processing and feature engineering are essential for accurately mapping vehicle GPS data to road network nodes and distinguishing between highway and service road traffic. This project employs a comprehensive data preprocessing pipeline that optimizes raw input data, calculates geodesic distances, and extracts features relevant to vehicular movement patterns. The approach enhances the system's ability to predict and differentiate between various road types, providing valuable insights for traffic analysis and road usage.

4.1 Data Collection and Preprocessing:

Data collection is a crucial step for building an accurate map-matching system, as it determines the quality and diversity of input data. The project incorporates two datasets: a geo dataset representing road nodes with latitude and longitude points, and a vehicle path dataset containing GPS points along with speed, direction, and vehicle type attributes. This combination covers a broad range of data needed for effective road network mapping and vehicle tracking.

After data collection, a series of preprocessing steps are applied to ensure the datasets are compatible for feature extraction and subsequent analysis:

- **Geodesic Distance Calculation:** Using geodesic distance to measure spatial relationships, the system calculates the distance between consecutive road nodes and between vehicle GPS points and road nodes. This step is crucial for accurate mapping and distance-based matching.
- **Coordinate Filtering:** Outliers in GPS data are removed through a filtering process based on reasonable geographic bounds and speed thresholds, ensuring only realistic vehicle paths are retained.
- **Timestamp Synchronization:** Data synchronization is performed to align the vehicle path and geo datasets by timestamps, creating a coherent dataset where vehicle positions are matched to corresponding road network locations.

4.2 Advanced Feature Engineering:

Feature engineering in this project involves extracting and refining attributes to improve the map-matching accuracy and allow better clustering of vehicular movement patterns.

- **Road Segment Identification:** Using road nodes as graph nodes, edges between nodes are constructed based on geodesic distance to form segments. This method allows the system to build a continuous map where each segment represents a part of the road network.

- **Directional Analysis:** By calculating directional vectors from GPS data, the system determines the likely movement direction of each vehicle, helping to predict road type based on typical flow patterns associated with highways and service roads.
- **Speed-Based Filtering:** Speed data is used to further refine vehicle-road matches, as vehicles on highways generally travel at higher speeds compared to those on service roads.

4.3 Enhancement Techniques for Low-Quality Data:

In many real-world scenarios, GPS data may have inaccuracies due to signal interference or multipath errors. To handle these challenges, the project incorporates several enhancement techniques:

- **Spatial Smoothing:** A spatial smoothing technique is applied to reduce noise in the vehicle trajectory, making it easier to align GPS points with the nearest road segments without abrupt jumps in the data.
- **Temporal Smoothing:** Vehicle paths are smoothed temporally by averaging GPS points over short time windows, which improves accuracy in areas with inconsistent GPS readings.
- **Dynamic Thresholding:** Dynamic distance thresholds are used to adaptively match vehicle points to road nodes, particularly useful in dense urban settings where roads and service paths are closely located.

4.4 Feature Selection and Dimensionality Reduction:

To optimize computational efficiency, the project utilizes feature selection and dimensionality reduction techniques:

- **Principal Component Analysis (PCA):** PCA reduces the dimensionality of features in the vehicle dataset, retaining the most important aspects, such as speed and direction, while discarding irrelevant or redundant data. This step not only speeds up processing but also reduces memory load.
- **Distance-Based Feature Selection:** Features such as proximity to nearest node, speed variance, and direction consistency are selected as primary indicators for determining the most suitable road segment.
- **Noise Reduction**: PCA helps in removing noise from the dataset by focusing on the most variance-explained components, ensuring the model is less sensitive to irrelevant variations in the data. This leads to a cleaner dataset and better model performance.
- **Visual Interpretation**: The reduced dimensionality after PCA can be used for easier visualization of high-dimensional data, helping analysts gain insights into data trends and patterns, which can aid in interpreting the behavior of vehicle movements on different road types.

4.5 Custom Features for Clustering and Classification:

Specialized features are developed to enhance the model's ability to cluster vehicles by road type:

- **Spatial Proximity Clustering:** Spatial clustering is used to group vehicle points based on proximity to road nodes, allowing the system to classify vehicles into different clusters based on probable road usage patterns.
- **Road Type Indicator:** Each road node is tagged with a type indicator (highway or service road), which is used as a classification target in the clustering algorithm. This feature enables the system to recognize and separate highways from service roads, improving the accuracy of vehicular movement predictions.
- **Density-based Clustering**: Implementing density-based spatial clustering (DBSCAN) can help better separate vehicles based on their concentration in specific road areas, improving clustering accuracy for varying traffic densities.
- **Dynamic Road Adaptation**: The system can adapt to new road segments by continuously updating clustering results as new data is processed, ensuring that clustering remains accurate as road conditions and traffic patterns change over time.

5. Model Development and Training

The model architecture and training process are central to the success of the highway and service road distinction algorithm. Using a combination of clustering techniques and machine learning models, the system can accurately classify vehicles' positions on either highway or service roads, despite noise in GPS data and varied vehicle movement patterns.

5.1 Model Architecture:

The model used for this task is a two-part architecture that combines clustering for data segmentation with classification models to predict the type of road based on vehicle position and movement.

- **Clustering with K-Means**: The clustering model processes the geographical data, grouping GPS coordinates into distinct clusters corresponding to highway and service road regions. By identifying natural patterns within location data, the clustering approach helps the model distinguish between the road types based on frequent paths and turns taken by vehicles.
- **Classification Model**: A classification model, such as a Decision Tree or Random Forest, further refines the prediction. Based on labeled training data, the model learns the characteristics associated with each road type, factoring in speed, direction, and vehicle density within clusters. This layered approach improves accuracy in ambiguous or dense areas, ensuring high differentiation.

5.2 Training Strategy:

The training strategy focuses on maximizing model accuracy and generalization by using structured data input and parameter tuning.

- **Batching Data**: Batches of GPS data sequences are used to optimize memory and processing, ensuring efficient handling of large datasets. By training in small chunks, the model reduces computational load and retains training stability.
- **Adaptive Learning Rate**: To balance initial model convergence and fine-tuning, an adaptive learning rate schedule starts high to speed up learning, then lowers gradually to refine predictions.
- **Data Augmentation**: Data augmentation is applied to simulate real-world GPS noise, including random shifts, scaling, and directional variations. This improves model robustness to location inaccuracies or drift.
- **Early Stopping and Checkpoints**: Early stopping prevents overfitting by halting training when the model's performance plateaus. Checkpoints allow saving the best-performing model during training.

5.3 Optimization Techniques:

Several optimization methods are employed to ensure the model's performance under various data conditions.

- **Hyperparameter Tuning**: Using techniques like grid search or randomized search, hyperparameters (e.g., number of clusters, tree depth) are tuned for optimal accuracy and generalization.
- **Regularization**: L2 regularization is added to the classification model, encouraging simpler models that generalize better to new data. Dropout is used for models prone to overfitting.
- **Gradient Clipping**: To avoid gradient explosions in GPS sequence processing, gradient clipping is applied, ensuring stable and consistent model training.

5.4 Evaluation Metrics:

Evaluating the model involves using multiple metrics to gauge its accuracy and reliability:

- **Accuracy**: Measures the percentage of correctly classified GPS points as highway or service road locations.
- **F1 Score**: Combines precision and recall, particularly important for datasets with class imbalances.
- **Clustering Cohesion**: Evaluates the tightness of clusters, ensuring the model forms well-defined geographic groups for each road type.
- **Computational Efficiency**: Analyzes memory and speed requirements, crucial for real-time applications.

5.5 Quality Assurance and Robustness Testing

- **Cross-validation**: The model undergoes cross-validation across multiple datasets, confirming its ability to generalize.
- **Edge Case Handling**: Scenarios like intersections and areas with ambiguous classifications are included in testing, ensuring model reliability.
- **Performance Monitoring**: After deployment, the model's accuracy and processing time are continuously monitored to maintain consistent performance.

6. User Interface and Integration

Frontend Development: The user interface, developed with React.js, is designed to be intuitive, providing map views of vehicle movements and classifications. Users can view individual paths, check for real-time classification updates, and navigate clusters to understand vehicle densities. Accessibility features ensure the platform is usable by individuals with varying technical skills, and progress indicators give feedback on processing tasks.

Backend API Integration: A RESTful API ensures smooth backend processing and communication. The backend is built to support scalability, securing data exchanges with HTTPS and JSON Web Tokens (JWT) for authentication.

- **Data Validation**: All incoming data undergoes rigorous validation, ensuring consistency with expected formats and preventing misclassifications.
- **Rate Limiting and Authentication**: The system implements rate limiting based on user tiers and uses JWT for secure access, ensuring fair use and secure data handling.
- **Model Serving and Scaling**: The model is deployed on a cloud platform, with automatic scaling that adjusts to real-time demands. This ensures the system remains responsive, even with high traffic, and supports easy updates and version control.
- **Logging and Monitoring**: The backend includes comprehensive logging, tracking API usage, model performance, and error rates. This enables proactive responses to issues, ensuring a reliable system.
- **User Customization and Reporting**: The platform supports customizable views and reporting options, allowing users to filter data based on specific vehicle types, routes, or timeframes. Users can generate detailed reports on traffic patterns, including peak hours, route preferences, and vehicle density. These reports can be exported for further analysis or shared with relevant stakeholders, enhancing the platform's utility for decision-making and urban planning.

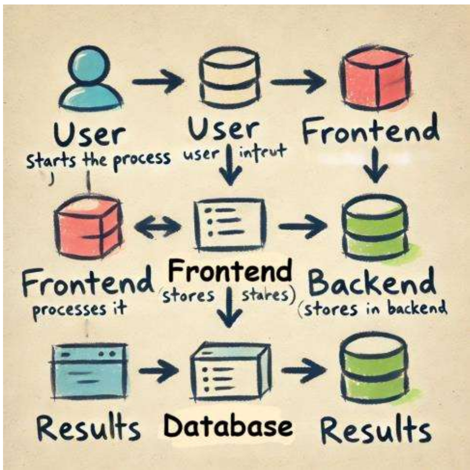
7. Use Cases and Applications

The road classification project has potential applications in various sectors:

- **Traffic Management**: Assists traffic controllers in monitoring vehicle distribution on highways versus service roads, optimizing traffic flow.
- **Environmental Monitoring**: The system can be used to analyze traffic patterns and their impact on air quality and noise pollution. By tracking vehicle distribution across highways and service roads, cities can identify areas of high traffic density, which can then be targeted for environmental interventions, such as green zones or traffic reduction measures.
- **Public Safety and Emergency Response**: The road classification system can be leveraged to enhance public safety by identifying congestion hotspots or accident-prone areas. This information can be used to optimize emergency response routes, allowing for quicker access during critical situations, improving response times and potentially saving lives.
- **Urban Planning**: Aids urban planners in understanding road usage patterns for future infrastructure development.

- **Logistics and Transportation**: Improves route optimization for logistics companies by distinguishing between highway and service road segments, leading to better fuel efficiency.

The flexibility and scalability of the system make it well-suited for high-demand applications, ensuring reliable performance across diverse urban areas and vehicle types.



8. Challenges and Future Work

While the highway and service road classification system is effective, it faces several challenges and potential areas for enhancement:

Challenges

- **Real-Time Data Processing**: Achieving real-time processing of large vehicle datasets can be challenging, especially when processing continuous GPS updates and adapting to fluctuations in traffic density.
- **GPS Noise and Accuracy**: GPS data is often noisy, with position drift or inaccuracies, which can impact classification accuracy. Managing noise while maintaining robust predictions is essential for reliable results.
- **Model Generalization**: While the current model is optimized for specific geographic regions, expanding its application across varied terrain and road networks may require additional model tuning and training with diverse datasets.
- **Data Privacy and Security**: As the system processes sensitive vehicle data, ensuring robust data privacy and security measures is critical. Implementing encryption, secure storage, and user consent protocols will be necessary to protect sensitive information and comply with data protection regulations.

Future Enhancements

- **Dynamic Route Prediction:** Integrating predictive analytics to anticipate vehicle routes based on past movement patterns, enhancing accuracy in ambiguous or busy areas.
- **Edge Computing for Real-Time Processing**: Implementing edge computing to enable real-time classification closer to the data source, reducing latency and improving response times.
- **Expanded Dataset Inclusion**: Adding more geographical data and traffic conditions to improve model adaptability, allowing it to handle a wider range of road types and environmental conditions.

- **Multimodal Data Integration**: Integrating additional data sources, such as camera feeds or weather information, could further improve the system's ability to classify roads accurately. This would help the model adjust to dynamic road conditions, like wet or icy roads, and provide more comprehensive insights into traffic behavior.

By addressing these areas, the classification system aims to improve its performance, robustness, and adaptability for diverse traffic conditions and regions.

9. Conclusion

The highway and service road classification project demonstrates significant potential for enhancing traffic management and urban planning by accurately distinguishing between highway and service road usage. With its layered architecture of clustering and classification techniques, the system can process and analyze large vehicle datasets effectively, providing actionable insights for sectors like logistics, transportation, and urban development. Future developments, including dynamic route prediction and edge computing, aim to expand its capabilities, making the system a valuable tool for traffic intelligence and infrastructure optimization in rapidly urbanizing environments.