Project Report

Project Title:

Detecting Driver Distraction Activities Using Machine Learning

Problem Statement:

Distracted driving is a major contributor to severe car accidents and an increasing public safety concern. Different types of distractions carry varying levels of risk, making it crucial to accurately recognize and categorize driver behaviors in real time. Traditional monitoring methods are limited in scalability and accuracy. The challenge is to develop a machine learning–based system that uses in-car images to detect and classify driver distraction activities. By preprocessing driver images and applying classifiers such as SVM, softmax regression, Naive Bayes, decision trees, and neural networks, the system aims to accurately distinguish between safe driving and various types of distraction, providing a foundation for improved road safety technologies.

High-Level Architecture:

The machine learning model follows this pipeline:

1. **Data Collection:**
   The State Farm Distracted Driver dataset was used, containing labeled images of drivers performing ten distinct activities.
2. **Preprocessing:**
   Images were resized, normalized, and converted to grayscale to reduce complexity ,noise and extracting feature faster and more effectively. We have also used PCA for reducing the number features and required features
3. **Feature Extraction:**
   The Histogram of Oriented Gradients (HOG) technique was applied to extract edge- and shape-based features representing driver posture and hand position.Well-suited for the ML models we have used (SVM, KNN, Decision Trees).
4. **Model Training:**

We have used ML algorithms like- Support Vector Machine (SVM), Decision Tree, and KNN were trained using the extracted features.

5. **Evaluation and Classification:**
   Each trained model we have evaluated using accuracy, confusion matrix, and classification reports to determine overall system performance.

Results:

- The SVM classifier achieved the highest accuracy of approximately 99.5%, outperforming Decision Tree(66%) and KNN(98.5%) models.
- The confusion matrix indicated that the system effectively distinguished safe driving from high-risk distractions such as texting and talking on the phone.
- The approach demonstrated that traditional ML methods combined with effective feature extraction can provide reliable performance while maintaining lower computational cost compared to deep learning models.

Conclusion:

We have developed a machine learning model that is capable of detecting driver distraction activities with accuracy.By combining HOG for feature extraction and PCA for dimensionality reduction, the system efficiently identifies key behavioral patterns. Models such as SVM, Decision Tree, and KNN were used to classify driver actions, with SVM delivering the highest accuracy. The results show that traditional ML techniques, when properly optimized, can reliably distinguish between safe and distracted driving. This work contributes toward developing intelligent driver monitoring systems that increase road safety and reduce accident risks.