Table of Contents

[Engineered Features 1](#_Toc215761922)

[Overview 1](#_Toc215761923)

[1. Edge-Based Features (5 features) 1](#_Toc215761924)

[2. Spatial/Region Features (11 features) 2](#_Toc215761925)

[3. Texture Features (7 features) 3](#_Toc215761926)

[Local Binary Pattern (LBP) 3](#_Toc215761927)

[Gray Level Co-occurrence Matrix (GLCM) 3](#_Toc215761928)

[4. Histogram/Distribution Features (11 features) 4](#_Toc215761929)

[5. Gradient/HOG Features (4 features) 5](#_Toc215761930)

[Most Discriminative Features for Steering 6](#_Toc215761931)

[Top Features 6](#_Toc215761932)

[Why These Features Work 6](#_Toc215761933)

[Logistic Regression 6](#_Toc215761934)

[Raw Data: 7](#_Toc215761935)

[Engineered Features: 8](#_Toc215761936)

[Summary: 8](#_Toc215761937)

[KNN 9](#_Toc215761938)

[Raw Data: 9](#_Toc215761939)

[Engineered Features: 10](#_Toc215761940)

[Naïve Bayes 11](#_Toc215761941)

[Raw Data 11](#_Toc215761942)

[Engineered Features 12](#_Toc215761943)

[CART Trees 13](#_Toc215761944)

[Raw Data 13](#_Toc215761945)

[Engineered Features 14](#_Toc215761946)

[Bagging 15](#_Toc215761947)

[Raw Data 15](#_Toc215761948)

[Engineered Features 16](#_Toc215761949)

[Random Forest 17](#_Toc215761950)

[Raw Data 17](#_Toc215761951)

[Engineered Features 18](#_Toc215761952)

[Boosting Methods 19](#_Toc215761953)

[Raw Data 19](#_Toc215761954)

[Engineered Features 20](#_Toc215761955)

[XG Boost 21](#_Toc215761956)

[Raw Data 21](#_Toc215761957)

[Engineered Features 22](#_Toc215761958)

[Support Vector Machines 23](#_Toc215761959)

[Raw Data 23](#_Toc215761960)

[Engineered Features 24](#_Toc215761961)

# Engineered Features

## Overview

The features are organized into **5 categories**:

| **Category** | **Count** | **Purpose** |
| --- | --- | --- |
| Edge-Based | 5 | Detect edges, lane markings, road boundaries |
| Spatial/Region | 11 | Capture where brightness is located (critical for steering) |
| Texture | 7 | Describe surface patterns using LBP and GLCM |
| Histogram/Distribution | 11 | Overall pixel intensity distribution |
| Gradient/HOG | 4 | Shape and edge direction information |

## 1. Edge-Based Features (5 features)

These features capture edge information using gradient operators, useful for detecting lane markings and road boundaries.

| **Feature** | **How It's Computed** | **What It Tells Us** |
| --- | --- | --- |
| edge\_mean | Mean of Sobel edge magnitude (√(Gx² + Gy²)) | Overall edge intensity in the image; higher = more edges/details |
| edge\_std | Standard deviation of edge magnitude | Variation in edge strength; higher = more heterogeneous edges |
| edge\_max | Maximum edge magnitude | Strongest edge in the image |
| edge\_hv\_ratio | Horizontal edge energy / Vertical edge energy | Ratio > 1 means more horizontal edges (e.g., road horizon); < 1 means more vertical (e.g., lane lines) |
| canny\_density | Fraction of pixels detected as edges by Canny | Overall "edginess" of the image; low = smooth scene |

## 2. Spatial/Region Features (11 features)

These capture **where** brightness is located in the image — critical for steering direction prediction.

| **Feature** | **How It's Computed** | **What It Tells Us** |
| --- | --- | --- |
| lr\_asymmetry | Mean(left half) - Mean(right half) | **Key feature!** Positive = left side brighter (suggests left turn) |
| lr\_ratio | Mean(left half) / Mean(right half) | Similar to above, but as a ratio |
| tb\_ratio | Mean(top half) / Mean(bottom half) | Sky vs road brightness; higher = brighter sky |
| tb\_diff | Mean(top half) - Mean(bottom half) | Absolute brightness difference top vs bottom |
| quad\_tl\_mean | Mean intensity of top-left quadrant | Brightness in upper-left region |
| quad\_tr\_mean | Mean intensity of top-right quadrant | Brightness in upper-right region |
| quad\_bl\_mean | Mean intensity of bottom-left quadrant | Brightness in lower-left region |
| quad\_br\_mean | Mean intensity of bottom-right quadrant | Brightness in lower-right region |
| diag\_asymmetry | (TL + BR) - (TR + BL) | Diagonal brightness pattern |
| brightness\_center\_x | Weighted centroid X (normalized to [-1, 1]) | **Key feature!** Negative = brightness center is left; positive = right |
| brightness\_center\_y | Weighted centroid Y (normalized to [-1, 1]) | Negative = brightness center is up; positive = down |

## 3. Texture Features (7 features)

These describe the surface texture patterns using Local Binary Patterns (LBP) and Gray Level Co-occurrence Matrix (GLCM).

### Local Binary Pattern (LBP)

LBP compares each pixel to its neighbors, encoding local texture as a binary pattern.

| **Feature** | **How It's Computed** | **What It Tells Us** |
| --- | --- | --- |
| lbp\_mean | Mean of LBP image | Average local texture pattern code |
| lbp\_std | Std deviation of LBP values | Texture pattern variation |
| lbp\_uniformity | Sum of squared LBP histogram bins | Higher = more uniform texture (fewer distinct patterns) |

### Gray Level Co-occurrence Matrix (GLCM)

GLCM analyzes how often pairs of pixels with specific values occur at specific spatial relationships.

| **Feature** | **How It's Computed** | **What It Tells Us** |
| --- | --- | --- |
| glcm\_contrast | GLCM contrast (averaged over 4 angles) | Local intensity variation; higher = rougher texture |
| glcm\_correlation | GLCM correlation | How correlated a pixel is with its neighbors; higher = more structured patterns |
| glcm\_energy | GLCM energy (angular second moment) | Texture uniformity; higher = more regular/repetitive |
| glcm\_homogeneity | GLCM homogeneity | Closeness of distribution to diagonal; higher = smoother transitions |

## 4. Histogram/Distribution Features (11 features)

These capture the overall pixel intensity distribution.

| **Feature** | **How It's Computed** | **What It Tells Us** |
| --- | --- | --- |
| intensity\_mean | Mean pixel value | Overall image brightness |
| intensity\_std | Std deviation of pixel values | Image contrast; higher = more variation |
| percentile\_10 | 10th percentile of pixel values | Darkest regions (shadows) |
| percentile\_25 | 25th percentile (Q1) | Lower brightness bound |
| percentile\_50 | Median pixel value | Central brightness |
| percentile\_75 | 75th percentile (Q3) | Upper brightness bound |
| percentile\_90 | 90th percentile | Brightest regions (highlights) |
| iqr | Q3 - Q1 (interquartile range) | Spread of middle 50% of intensities |
| skewness | Skewness of pixel distribution | Negative = tail toward dark; Positive = tail toward bright |
| kurtosis | Kurtosis of pixel distribution | Peakedness; high = concentrated around mean; low = flat distribution |
| intensity\_range | Max - Min pixel value | Dynamic range of the image |

## 5. Gradient/HOG Features (4 features)

Summary statistics from Histogram of Oriented Gradients (HOG), which captures shape and edge direction information.

HOG divides the image into cells, computes gradient orientations within each cell, and creates histograms of these orientations. The full HOG descriptor has 1764 values, so we summarize it with statistics.

| **Feature** | **How It's Computed** | **What It Tells Us** |
| --- | --- | --- |
| hog\_mean | Mean of full HOG descriptor (1764 values) | Average gradient orientation strength |
| hog\_std | Std deviation of HOG values | Variation in gradient orientations |
| hog\_max | Maximum HOG bin value | Strongest gradient direction |
| hog\_energy | Sum of squared HOG values | Total gradient energy; higher = more edges/structure |

## Most Discriminative Features for Steering

Based on ANOVA F-scores and Mutual Information analysis, the **most important features for predicting steering direction** are:

### Top Features

1. **lr\_asymmetry** — Left-right brightness difference is directly related to steering direction
2. **brightness\_center\_x** — Where the "center of mass" of brightness is located horizontally
3. **quad\_bl\_mean** / **quad\_br\_mean** — Bottom quadrant brightness patterns
4. **lr\_ratio** — Ratio version of left-right asymmetry
5. **edge\_hv\_ratio** — Balance of horizontal vs vertical edges

### Why These Features Work

These spatial features make intuitive sense for steering prediction:

* **Left turns**: The left side of the road typically shows different features (obstacles, curves, lane markings) than the right side, causing asymmetry in brightness.
* **Right turns**: The opposite pattern occurs.
* **Forward/Straight**: Both sides tend to be more balanced.

The brightness\_center\_x feature essentially asks: "Where is the visual 'weight' of the image?" If it's shifted left, the vehicle may need to turn left to center the road.

# Logistic Regression

## Raw Data:

A screenshot of a computer

AI-generated content may be incorrect.

## Engineered Features:

A screenshot of a computer

AI-generated content may be incorrect.

## Summary:

A close-up of a document

AI-generated content may be incorrect.

A graph of different sizes and colors

AI-generated content may be incorrect.

# KNN

## Raw Data:

A screenshot of a computer screen

AI-generated content may be incorrect.

## Engineered Features:

A screenshot of a computer

AI-generated content may be incorrect.

# Naïve Bayes

## Raw Data

A screenshot of a computer screen

AI-generated content may be incorrect.

## Engineered Features

A screenshot of a computer

AI-generated content may be incorrect.

# CART Trees

# Raw Data

A screenshot of a computer

AI-generated content may be incorrect.

## Engineered Features

A screenshot of a computer

AI-generated content may be incorrect.

# Bagging

## Raw Data

A screenshot of a computer

AI-generated content may be incorrect.

## Engineered Features

A screenshot of a computer

AI-generated content may be incorrect.

# Random Forest

## Raw Data

A screenshot of a computer

AI-generated content may be incorrect.

## Engineered Features

A screenshot of a computer

AI-generated content may be incorrect.

# Boosting Methods

## Raw Data

A screenshot of a computer screen

AI-generated content may be incorrect.

## Engineered Features

A screenshot of a computer screen

AI-generated content may be incorrect.

# XG Boost

## Raw Data

A screenshot of a computer

AI-generated content may be incorrect.

## Engineered Features

A screenshot of a computer

AI-generated content may be incorrect.

# Support Vector Machines

## Raw Data

A screenshot of a computer

AI-generated content may be incorrect.

## Engineered Features

A screenshot of a computer

AI-generated content may be incorrect.

# Model Summaries

A screenshot of a computer screen

AI-generated content may be incorrect.

A graph of different colored bars

AI-generated content may be incorrect.

A graph of blue and orange bars

AI-generated content may be incorrect.