A Monocular-Vision Based System for Detection of Drivable Road

Amitesh Patil  
Department of Electronics and Telecommunication  
Vishwakarma Institute of Technology

Pune, India, 411038  
amitesh.patil19@vit.edu

Kshitij Patil  
Department of Electronics and Telecommunication  
Vishwakarma Institute of Technology

Pune, India, 411038  
kshitij.patil19@vit.edu

Vaishnavi Pingat  
Department of Electronics and Telecommunication  
Vishwakarma Institute of Technology

Pune, India, 411038  
vaishnavi.pingat19@vit.edu

***Abstract*—Road detection is a core part of environmental perception for driver assistance systems. The paper presents a system for detection of drivable and non-drivable roads. The system assists the driver to reduce traffic casualties. Fusion of Scale Inverse Feature Transform (SIFT) and Binary Robust Invariant Scalable Key-points (BRISK) algorithms are used for feature extraction. These extracted features are provided to machine learning models for classification. Seven different classifiers are used for training models. Support Vector Machine (SVM) – Radial basis function (RBF) provided the highest accuracy of 70.9%. SVM provides the highest F1 score of 75%. Voting classifier provided accuracy of 70.2%.**

***Keywords— Computer Vision, Drivable Road Detection, Driver Assistant System, Machine Learning.***

1. INTRODUCTION

Drivable road detection is the priority task [1] for advanced driver assistance system (ADAS). It helps in route planning and deciding motion control strategies. More than 10 million people face traffic casualties every year [2]. Detection of drivable roads is challenging in situations such as complex traffic scenarios [3], changing lighting conditions, and changes in texture of the roads. It is difficult to adapt the strategies in diverse situations. It takes immense human work to create new annotations for training the model.

1. LITERATURE REVIEW

LiDAR is used to detect road boundaries in [4]. It has low resolution and limited perception. Feature-level fusion method of LiDAR data [5] is used for the detection of the drivable region. It provides 93% accuracy. Chip-Net is a CNN model using LiDAR data, implemented in [6] for drivable road detection. It provides better accuracy than other models on the

KITTI dataset. Disparity-proposal-based detection is implemented in [7]. It collects candidate frames of detecting objects using stereo camera under various perturbation. Region of Interest (ROI) is generated in [8] for drivable road detection. ROI is generated from dynamic threshold search method and drag process. Image-based classification, segmentation, and detection in [9] surpass earlier standard methods in complicated scenarios. More than one features are used in [10] for image segmentation based on shape, grayscale, colour, edge, and texture. Histogram of gradient (HOG) with support vector machine (SVM) model is used in [11] to classify road images. Illumination invariant feature space with likelihood classifier is used in [12]. It is ineffective in scenarios of complex traffic scenes. Road area detection using texture information [13] and road vanishing points solved this problem. A region-expanding method with seed point selection is proposed in [14]. It is used to extract drivable areas. It provided accurate results on shadowed or marked roads. Discriminating Feature Analysis model [15] is implemented for drivable road segmentation. It is used with ResNet-101 framework to boost performance. Fully evolved CNN and inverse perspective mapping are used in [16] to detect drivable roads and lane marks. It provides 86.9% pixel accuracy and 55.23% mean accuracy.

Drivable road is detected using V-disparity method in [17]. It does not count roadside vegetation as drivable area. It is addressed by detecting road and vegetation separately and then processing them together. It increases accuracy to 88.33%. A model based on deep neural network is used in [18] to detect, track and count vehicles. It provides better counting accuracy when compared with YOLO. The YOLOv3 model is used in [19]. The system divides the image into remote and proximal areas for getting better accuracy in vehicle detection. Detection of road borders is implemented in [20]. Canny edge detection followed by Hough Transform (HT) provides 75.5% accuracy. Boundary extraction followed by HT provides 86.5% accuracy. Edge detection is performed on vehicles in [21] and the morphological operations are used to remove noise. Colour based road segmentation is performed in [22] using CNN and SVM. Inclusion of image coordinates in the feature vector provides better accuracy. SVM outperformed neural networks slightly. Very steep roads are detected using a coarse environmental model [23] and digital elevation map.

Pedestrian detection system based on SVM classifier is used in [24]. It is trained using the HOG and Local Ternary Pattern (LTP) features. It provides better performance than existing multi-feature algorithms. The algorithm used in [25] is based on features extracted from the magnetic field distortion for vehicle detection. XG-Boost provides the highest accuracy of 80%.

1. METHODOLOGY

The proposed system detects drivable and non-drivable roads.It consists of a monocular camera and processor-based system. Camera captures information of surrounding environment and provides it as an input to a processor-based system. The system interprets real-time drivable roads using different classification algorithms. It translates this information in the form of audio format and conveys it to the driver through a Bluetooth speaker. A block diagram of the system is shown in Fig. 1.

Camera

Processor Based system

Classification and detection of drivable road

Non-Drivable Road

Drivable Road

Translation of detection road situation into audio feedback

Speaker

Fig. 1. Block diagram for drivable road detection.

*A. Dataset description*

A dataset containing 6000 images is curated for detection of drivable road. It consists of 3000 positive images (drivable road) and 3000 negative images (non-drivable road). Images of drivable road are curated from Berkeley Diverse Driving Dataset (BDD 100K) and Indian Driving Dataset**.** Negative images are acquired from BDD 100K. The details of distribution of images in dataset are presented in Table 1.

TABLE 1. Details of images in the dataset

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** |  | Image Class | No. of images |
| 1 | Training | Positive (Drivable Road) | 2000 |
| Negative (Non- Drivable Road) | 2000 |
|  |  | Positive (Drivable Road) | 1000 |
| 2 | Testing | Negative (Non- Drivable Road) | 1000 |
| Total Images | | | 6000 |

Sample images from the dataset are shown in Fig. 2.

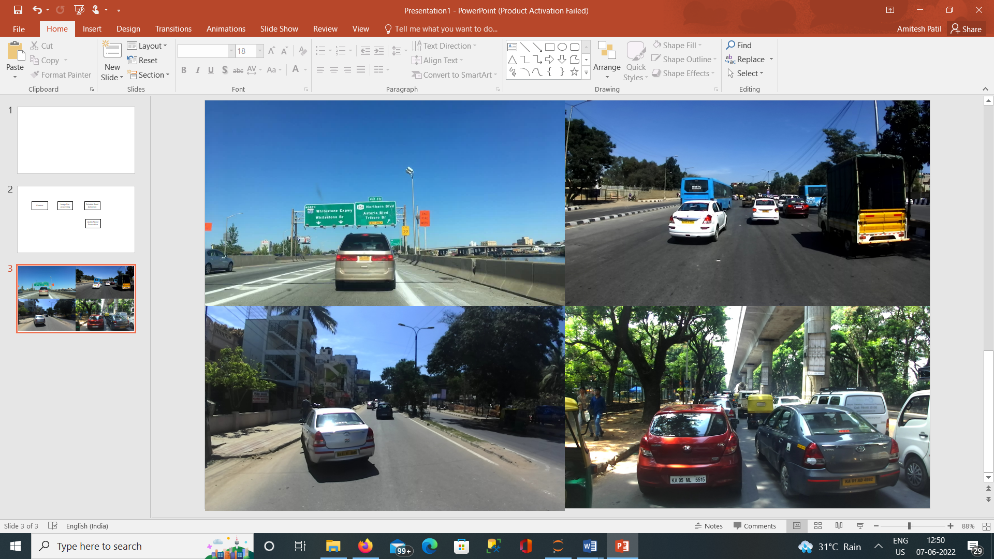


Fig. 2. Sample positive and negative image from the dataset

*B.* *Implementation of system*

All the images are resized to 280x430. The images are converted to grayscale. A 5x5 low pass filter is used to remove the noise from the image. A 3x3 high pass filter is used to sharpen the image and enhance the edges. The gamma transformation is applied to brighten the image. Histogram equalization is applied on the images to improve the contrast of the image. Image resizing ensures efficient training of the model.

Scale Invariant Feature Transform (SIFT) is used to extract features from 6000 images. SIFT generates a feature vector of dimension (X, 128). X is the number of key-points. SIFT generates a large number of key-points. All the key-points cover the image over a full range of scales and locations. SIFT is invariant to scale, illumination changes, rotations, and partially invariant to occlusions. Binary Robust Invariant Scalable Key-points (BRISK) is used for feature extraction. Invariance to scale and rotation is the reason for using BRISK. BRISK generates a feature vector of dimension (Z, 64). Z is the number of key-points. Process of feature extraction using SIFT is presented in Fig. 3.

Image 1

Scale invariant feature transform of image 1

Scale invariant feature transform vector for all 4000 images

Image 2

Scale invariant feature transform of image 2

Image 3999

Scale invariant feature transform of image 3999

­­

Scale invariant feature transform of image 4000

Image 4000

Fig 3. Feature Extraction Using Scale Invariant Feature Transform

The size of the feature vector formed is very large. K-means clustering is used for reducing the number of rows in data. K-means clustering divides the data into k groups. The value of K is determined by using the elbow method.

Principal Component Analysis (PCA) is applied in order to reduce the number of columns. PCA constructs new features by taking the linear combination of the original components. These combinations are uncorrelated.

Algorithm 1: Dimension Reduction

Input:

1. Feature vector of SIFT (4603650 x 128)

2. Feature vector of BRISK (10494548 x 64)

Output: Optimized feature vector (4000 x 42)

1. Data1 = [], Data2 = []

2. For each image Feature vector of SIFT in F

3. X = Pre- Trained K-Means [k = 23]

4. N = Normalize(X)

5. Data1.append(N)

6. End For

7. For each image Feature vector of BRISK in F

8. Y = Pre- Trained K-Means [k = 25]

9. M = Normalize(Y)

10. Data2.append(M)

11**.** End For

12. Final data = Concatenate (Data1, Data2)

// Dimension of Final data is 4000x48

13. Apply standard scalar to Final data.

14. Reduced feature vector = PCA with [n = 42]

15. Return reduced feature vector

The feature vector for all images using SIFT has dimension 4603650 x 128. The feature vector for all images using BRISK has dimension 10494548 x 64. K-Means model is trained based on the SIFT feature vector and BRISK feature vector. Value of K for SIFT K-Means model is chosen as 23. Value of K for BRISK K-Means model is chosen as 25. K is determined by using elbow method. Each image is predicted using pre-trained K-Means model. The data is normalized by dividing with the number of key-points. The size of SIFT feature is reduced to 4000 x 23. The size of BRISK feature vector is reduced to 4000 x 25. The two feature vectors are concatenated to form a feature vector of size 4000 x 48. This data is standardized to bring all the features to a common scale. PCA is applied and it is observed that 99% of the information lies in the first 42 components. 42 features are considered in the final feature vector. The size of final feature vector is 4000 x 42.

The data is split into training and testing sets. The training data is 80% of training images. Testing data is 20% of all training images.

The main objective of this system is to classify images into drivable road or non - drivable road. It is a binary classification problem. Seven classification algorithms implemented in the system are i) Decision tree (DT) ii) Support vector machine (SVM) iii) K-Nearest Neighbour (KNN) iv) Random Forest (RF) v) Logistic Regression (LR) vi) XG Boost vii) Naïve Bayes.

SVM uses hyperplane for classification of data into different classes. Dimension of hyperplane depends on number of input features in dataset. The RBF kernel function for two points X₁ and X₂ computes the similarity or how close they are to each other. This kernel can be mathematically represented as follows:

where, ‘’ is the variance and our hyperparameter and is the Euclidean distance between two points X1

and X2.

The Voting Classifier used for classification is an ensemble approach . It uses multiple classifiers to make predictions for each data input. It carries out predictions based on the aggregated majority of votes of each class. Voting classifier is a powerful method as it considers multiple models at a time and is non – bias in nature. There are two types of voting classifiers - hard and soft. Soft voting classifier is used in this system. The output class in soft voting is the prediction based on the average probability assigned to that class. Classification and detection of drivable road is presented in Fig .4.

Input image

Image resizing

Extracting SIFT, BRISK features

K-Means clustering

Principle component analysis

Classifier

Drivable road

Voting

Audio feedback

Non-drivable road

Fig 4. Classification and detection of drivable road.

Process of detecting drivable road is presented in Algorithm 2.

Algorithm 2: Detection of Drivable Road

Input: Images Dataset (D)

Output:Drivable Road or Non – Drivable Road

Foreach image I in D:

2. Image Pre-processing

3. Feature Extraction (F) from I

4. Dimension Reduction of feature

5. Extract Final Feature Vector (FV)

1. End for
2. Classifiers = [ Decision tree, SVM, KNN, Random Forest, Logistic Regression, XG Boost, and Naïve Bayes]
3. Foreach Classifier (C) in Classifiers:
4. Splitting the data
5. Model = C (data)
6. Model.predict()
7. Find accuracy and other evaluation metrics.
8. End For
9. Implementation of Soft Voting Classifier
10. if predicted value = 0

return drivable road.

else

return non-drivable road.

1. RESULTS

Seven different classifiers are implemented for classification of drivable roads. SVM-RBF model provided the highest test accuracy with 70.9 % followed by Random Forest with 70.4 %. The voting classifier provides accuracy of 70.2 %. The different performance metrics for various classification models implemented are presented in Table 2. and Table 3.

TABLE 2. Performance Metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classifiers** | DT | Random Forest | KNN | Logistic | SVM  (Linear) |
| **Train Accuracy** | 92.06 % | 96.7 % | 73.5 % | 67.7 % | 68 % |
| **Test Accuracy** | 61.6% | 70.4% | 67% | 68.8% | 67.9% |
| **Specificity** | 0.536 | 0.559 | 0.53 | 0.537 | 0.513 |
| **Precision Score** | 0.6 | 0.658 | 0.633 | 0.644 | 0.634 |
| **Recall Score** | 0.696 | 0.849 | 0.811 | 0.839 | 0.845 |
| **F1 Score** | 0.644 | 0.741 | 0.711 | 0.728 | 0.724 |

TABLE 3. Performance Metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classifiers** | SVM Poly | SVM Rbf | XG Boost | Naïve Bayes | Voting Classifier |
| **Train Accuracy** | 74.3 % | 74.1 % | 78.6% | 66.1 % | 83.8 % |
| **Test Accuracy** | 61.3% | 70.9% | 69% | 66% | 70.2% |
| **Specificity** | 0.502 | 0.542 | 0.576 | 0.499 | 0.547 |
| **Precision Score** | 0.634 | 0.656 | 0.654 | 0.621 | 0.654 |
| **Recall Score** | 0.865 | 0.876 | 0.804 | 0.821 | 0.857 |
| **F1 Score** | 0.732 | 0.75 | 0.721 | 0.707 | 0.741 |

SVM-RBF provides recall score of 87.6%. RF provided the highest precision of 65.8%. SVM-RBF has precision of 65.6%. Highest F1 score of 75 % is provided by SVM-RBF. Receiver Operating Characteristics (ROC) and Area Under Curve (AUC) for SVM- RBF classifier is shown in fig 5. The SVM-RBF model provides the AUC of 0.71.

Graphical user interface, application

Description automatically generated

Fig 5. ROC curve for SVM-RBF

1. CONCLUSION

Drivable road detection is crucial in planning routes and prevention of road accidents for autonomous vehicles. The system classifies images captured by camera into two possible scenarios. i) Drivable Road, ii) Non-drivable road. The proposed system focuses on detection of drivable roads in different traffic conditions. Fusion of two feature descriptors BRISK and SIFT are used in the system. BRISK and SIFT are rotation and scale invariant. SVM-RBF has provided highest accuracy score of 70.9 %. Computationally effective methods are implemented for classification and detection of drivable road. Accuracy of the system decreases in the scenes where lighting conditions are poor, and shadows are involved. Authors are working on larger dataset for multiple classes using deep learning approaches. Graphical user interface with sensors and hardware can improve the overall experience.

1. REFERENCES

[1] B. Wang, V. Frémont and S. A. Rodríguez, "Color-based road detection and its evaluation on the KITTI road benchmark," 2014 IEEE Intelligent Vehicles Symposium Proceedings, 2014, pp. 31-36.

[2] D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, “Survey of pedestrian detection for advanced driver assistance systems,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 7, pp. 1239–1258, Jul. 2010.

[3] W. Dong, Z. Yang, W. Ling, Z. Yonghui, L. Ting and Q. Xiaoliang, "Research on vehicle detection algorithm based on convolutional neural network and combining color and depth images," 2019 2nd International Conference on Information Systems and Computer Aided Education (ICISCAE), 2019, pp. 274-277.

[4] Liu, X., Deng, Z. “Segmentation of Drivable Road Using Deep Fully Convolutional Residual Network with Pyramid Pooling”. Cogn Comput 10, pp. 272–281, 2018.

[5] Q. Li, L. Chen, M. Li, S. -L. Shaw and A. Nüchter, "A Sensor-Fusion Drivable-Region and Lane-Detection System for Autonomous Vehicle Navigation in Challenging Road Scenarios," in IEEE Transactions on Vehicular Technology, vol. 63, no. 2, pp. 540-555, Feb. 2014.

[6] Y. Lyu, L. Bai and X. Huang, "ChipNet: Real-Time LiDAR Processing for Drivable Region Segmentation on an FPGA," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 66, no. 5, pp. 1769-1779, May 2019.

[7] J. Chen, W. Xu, W. Peng, W. Bu, B. Xing and G. Liu, "Road Object Detection Using a Disparity-Based Fusion Model," in IEEE Access, vol. 6, pp. 19654-19663, 2018.

[8] A. M. Neto, A. C. Victorino, I. Fantoni and J. V. Ferreira, "Real-time estimation of drivable image area based on monocular vision," *2013 IEEE Intelligent Vehicles Symposium (IV)*, 2013, pp. 63-68.

[9] E. Shelhamer, J. Long and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640-651, April 2017.

[10] H. Kong, J.-Y. Audibert, and J. Ponce, “General road detection from a single image,” IEEE Trans. Image Process., vol. 19, no. 8, pp. 2211–2220, Aug. 2010.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 886-893 vol. 1.

[12] J. M. Á. Alvarez and A. M. Lopez, “Road detection based on Illuminant invariance,” IEEE Trans. Intell. Transp. Syst., vol. 12, no. 1, pp. 184–193, Mar. 2011.

[13] C. Rasmussen, "Grouping dominant orientations for ill-structured road following," Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2004, pp. I-I.

[14]Gao, Y., Song, Y., Yang, Z. “A Real-Time Drivable Road Detection Algorithm in Urban Traffic Environment,” in Computer Vision and Graphics. ICCVG 2012, vol. 7594.

[15] X. Liu, Z. Deng and G. Yang, "Drivable Road Detection Based on Dilated FPN with Feature Aggregation," 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1128-1134, 2017.

[16] Y. Chan, Y. Lin, and P. Chen, "Lane Mark and Drivable Area Detection Using a Novel Instance Segmentation Scheme," 2019 IEEE/SICE International Symposium on System Integration (SII), pp. 502-506, 2019.

[17] S. Lee and B. Moon, "Drivable Area Detection Method Capable of Distinguishing Vegetation Area on Country Road," 2018 International SoC Design Conference (ISOCC), 2018, pp. 80-81.

[18] J. Zhu et al., "Urban Traffic Density Estimation Based on Ultrahigh-Resolution UAV Video and Deep Neural Network," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 12, pp. 4968-4981, Dec. 2018.

[19] Song, H., Liang, H., Li, H. et al. “Vision-based vehicle detection and counting system using deep learning in highway scenes,” in Eur. Transp. Res. Rev. 11**,** pp51, 2019.

[20] Parameswaran, N.S., Revathi Achan, E., Subhashree, V., Manjusha, R.”Road Detection by Boundary Extraction Technique and Hough Transform.” Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB), 2018, vol 30.

[21] S. S. Kalyan, V. Pratyusha, N. Nishitha, and T. K. Ramesh, "Vehicle Detection Using Image Processing," 2020 IEEE International Conference for Innovation in Technology (NEOCON), 2020, pp. 1-5.

[22] P. Conrad and M. Foedisch, "Performance evaluation of colour based road detection using neural nets and support vector machines," 32nd Applied Imagery Pattern Recognition Workshop, 2003. Proceedings, pp. 157-160, 2003.

[23] A. M. Naguib, J. Kim and S. Lee, "3D environmental modeling and drivable road identification for a long-range rover," 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), 2017 pp. 697-700.

[24] Hongzhi Zhou, Gan Yu, “Research on pedestrian detection technology based on the SVM classifier trained by HOG and LTP features,” in Future Generation Computer Systems, Volume 125, pp 604-615, 2021.

[25] H. Dong, X. Wang, C. Zhang, R. He, L. Jia, and Y. Qin, "Improved Robust Vehicle Detection and Identification Based on Single Magnetic Sensor," in IEEE Access, vol. 6, pp. 5247-5255, 2018