Comparison of handcrafted features extraction techniques for traffic sign recognition

Prestone Simiyu[1*]
Raphael Angulu[2]
Daniel Otanga[3]

[1*]sprestone@mmust.ac.ke
[2]rangulu@mmust.ac.ke
[3]otanga@mmust.ac.ke

[123]Masinde Muliro University of Science and Technology, Kenya

https://doi.org/10.51867/ajernet.6.4.2

ABSTRACT

Traffic Sign Recognition (TSR) is a critical computer vision task for the advancement of Autonomous Driver Assistance Systems (ADAS) and autonomous vehicles, directly impacting road safety. While deep learning dominates current research, handcrafted feature extraction techniques remain relevant due to their interpretability, lower computational demands, and suitability for embedded systems. This paper presents a systematic empirical evaluation and comparison of five predominant handcrafted feature extraction methods: Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Bag of Words (BoW), Local Binary Patterns (LBP), and Gabor Filters for traffic sign recognition. The techniques are assessed on the German Traffic Sign Recognition Benchmark (GTSRB) using three classifiers: Support Vector Machine (SVM), Random Forest, and Decision Tree. Experimental results demonstrate that the choice of feature extractor and classifier significantly impacts performance, with HOG combined with an SVM classifier achieving the highest accuracy (95.2%). The study provides a clear performance hierarchy, revealing that HOG and LBP offer the best balance of accuracy and computational efficiency for this domain. We conclude that the optimal selection of a handcrafted feature extraction strategy is problem-dependent and provide concrete recommendations for implementing effective and efficient TSR systems.

Keywords: Bag of Visual Words (BoW), Computer Vision, Feature Extraction, Gabor Filters, Handcrafted Features, Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), Machine Learning, Scale-Invariant Feature Transform (SIFT), Traffic Sign Recognition (TSR)

I. INTRODUCTION

Traffic sign recognition (TSR) research is attracting enormous attention and is considered an important element of Automatic Driver Assistant System (ADAS) and autonomous vehicles (Saadna & Behloul, 2017). Autonomous vehicle’s ability to promptly and accurately detect and recognize a traffic sign is considered as a matter of human life (Almaskati et al., 2023). Traffic signs convey crucial information that could be missed out by drivers due to many factors including distractions and tiredness. Drivers driving in unfamiliar places may pay less attention to traffic signs; making automatic detection and recognition of traffic signs an important aspect towards improving road safety. Traffic Sign Recognition is the process of automatically detecting, localizing and recognizing (classifying) a traffic sign. It involves accurately detection of a traffic sign amidst dynamic status of the road and correctly recognizing the traffic sign, within a reasonable time. This requires TSR systems to not only be accurate but also fast because it can be catastrophic for TSR system to correctly recognize a traffic sign after the vehicle has already significantly passed the sign. TSR can be done in two different ways; using machine learning and deep learning.

The importance of traffic sign recognition (TSR) in improving road safety and enabling autonomous driving technologies has made it one of the most studied topics in computer vision. Two concurrent trends have dominated the conversation globally: the continuous relevance of handmade feature extraction techniques and the emergence of deep learning algorithms, which today achieve state-of-the-art performance on benchmarks like GTSRB. According to Maletzky et al. (2023), deep learning-based techniques, specifically Convolutional Neural Networks (CNNs), have shown accuracies of over 98% on the GTSRB dataset. However, these techniques are very resource-dependent, requiring high energy consumption, strong GPUs, and huge datasets, which restricts their use on low-power or embedded devices (Omid-Zohoor et al., 2018). Because they are interpretable, computationally lighter, and more flexible in limited settings, handcrafted features like Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Bag of Words (BoW), Local Binary Patterns (LBP), and Gabor filters have seen a resurgence in popularity due to this conflict between accuracy and computational efficiency.

Compared to Europe, the US, and Asia, where there are several benchmarks and industrial deployments, TSR research is still comparatively underdeveloped at the regional level (Africa) (Ellahyani & El Ansari, 2021). Variable lighting and weather conditions, faded or damaged signs, non-standardized signage, and inadequate traffic infrastructure upkeep are some of the particular difficulties that African environments provide that have a direct impact on recognition accuracy. These circumstances highlight how crucial it is to have reliable feature extraction methods that perform effectively in less-than-ideal circumstances. Furthermore, there has been little development of localized datasets that capture the actual heterogeneity of African roads; instead, the majority of TSR studies in Africa have depended on imported datasets like GTSRB. Because of this disparity, it is challenging to determine if current feature extraction techniques can be directly used or require modification.

Road safety continues to be a significant economic and public health concern in Kenya. According to Kiran et al. (2009), human error, a lack of compliance with traffic regulations, and inadequate road infrastructure are the main causes of Kenya's high traffic accident rates, which rank among the highest in Sub-Saharan Africa. Thus, trustworthy TSR systems may help both human drivers and semi-autonomous systems, thereby lowering the number of accidents. However, there hasn't been much study on TSR in Kenya, and the majority of computer vision applications concentrate on medical imaging, facial recognition, and agriculture rather than intelligent transportation. The absence of empirical assessments of handcrafted feature extraction techniques in Kenyan road settings, where signs are frequently obscured by dust, vegetation, or commercial signage, is a significant gap. The paper discusses how machine learning techniques can be applied to images with the aim of extracting handcrafted features.

1.1 Statement of the Problem

The necessity for dependable Traffic Sign Recognition (TSR) systems has increased due to the quick development of autonomous cars and Advanced Driver-Assistance Systems (ADAS), as a malfunction could have catastrophic safety repercussions. Because of their interpretability, reduced processing requirements, and efficiency in resource constrained embedded systems, handcrafted feature extraction techniques continue to be very relevant even if deep learning approaches now dominate contemporary research (Almaskati et al., 2023). Nevertheless, there is a notable lack of useful advice for academics and developers on how to choose suitable handcrafted features. For the particular problem of traffic sign recognition, there is a lack of comparative documentation regarding the computational tradeoffs and performance of well-known methods such as HOG, SIFT, LBP, BoW, and Gabor filters. Without a clear, empirical analysis of these methods on a standardized benchmark, developers face challenges in selecting the optimal feature extraction strategy that balances accuracy, speed, and implementation complexity for their specific application requirements. This study addresses this gap by systematically evaluating and comparing these five handcrafted feature extraction techniques to provide a definitive performance hierarchy and practical recommendations.

1.2 Research Objective

To conduct a comprehensive empirical evaluation and comparison of five handcrafted feature extraction techniques for traffic sign recognition.

II. LITERATURE REVIEW

2.1 Theoretical Review

Domain knowledge is used to extracted features from images for handcrafted methods (Wali, 2015), (Fu & Huang, 2010), (Kiran et al., 2009). Handcrafted features that can be extracted from traffic sign images are shape, color, texture and spatial feature. This section discusses the four features in details.

2.1.1 Shape Features

These features capture the shape of the traffic sign, such as the number of corners (Maletzky et al., 2023), the length of the sides, and the area of the sign. The number of corners in a traffic sign can be a useful feature for classification. For example, stop signs typically have four corners, while yield signs typically have three corners. The length of the sides of a traffic sign can also be a useful feature for classification. For example, speed limit signs typically have longer sides than warning signs. When it comes to area, the area of a traffic sign can also be a useful feature for classification. For example, larger signs are typically more important than smaller signs. The circularity of a traffic sign is the ratio of the area of the sign to the area of a circle with the same perimeter. Circular signs, such as stop signs, tend to have a higher circularity than non-circular signs. The compactness of a traffic sign is the ratio of the area of the sign to the square of the perimeter. Compact signs, such as yield signs, tend to have a higher compactness than non-compact signs. The convexity of a traffic sign is a measure of how smooth the edges of the sign are. Convex signs, such as stop signs, tend to have a higher convexity than non-convex signs. Finally, the symmetry of a traffic sign is a measure of how evenly the sign is divided into two halves. Symmetric signs, such as stop signs, tend to have a higher symmetry than non-symmetric signs (Maletzky et al., 2023). Therefore, shape features have an important role in the classification of traffic signs.

2.1.2 Color Features

Color features are based on the color of the traffic sign. They are one of the most common types of features used for traffic sign recognition. There are many different ways to extract color features from images (Maletzky et al., 2023). Some common features include Red Green Blue (RGB) features that involves extracting the red, green, and blue (RGB) values of each pixel in the image. Hue, Saturation Value (HSV) features is more robust to variations in illumination than RGB features. It involves extracting the Hue, Saturation, and Value (HSV) values of each pixel in the image. Another feature that can be extracted is the YCbCr features. These features are more robust to variations in illumination than HSV features. It involves extracting the luminance (Y), chrominance blue (Cb), and chrominance red (Cr) values of each pixel in the image. Histogram features can also be extracted from an image. These features involve extracting the histogram of the colors in the image. The histogram is a plot of the number of pixels in the image that have each color value. Lastly, color moments involve extracting the moments of the color distribution in the image. The moments are a set of statistical measures that can be used to describe the distribution of the colors in the image tune (Khalid et al., 2014) and (Azhar et al., 2015). Notably, color features can be effective for traffic sign recognition, but they can also be sensitive to variations in illumination and viewpoint. Additionally, they can be difficult to design and implement.

2.1.3 Texture Features

These features capture the texture of the traffic sign, such as the smoothness, roughness, and coarseness of the surface. Texture features are based on the texture of the traffic sign. (Maletzky et al., 2023) and (Khalid et al., 2014) notes that they are a type of features that is less sensitive to variations in illumination and viewpoint than color features. There are many different ways to extract texture features from images. Some common methods include Gabor filters are type of filter that is sensitive to specific frequencies and orientations of texture. They can be used to extract features that are invariant to changes in illumination and viewpoint. Haralick features are a set of statistical measures that can be used to describe the texture of an image. They are often used in combination with other features, such as color features. Finally, Local binary patterns (LBP) texture descriptor that measures the contrast between neighboring pixels. It is a simple yet effective texture descriptor that can be used with a variety of machine learning algorithms. Therefore, texture features can play a major role in TRS.

2.1.4 Spatial Features

These features capture the spatial relationships between different parts of the traffic sign, such as the distance between the edges of the sign (Khalid et al., 2014), or the position of the text in the sign. Spatial features are based on the spatial relationships between the pixels in the image. They are a type of feature that is less sensitive to variations in illumination and viewpoint than color and texture features. Some common methods include Edge features which are based on the edges in the image. Edges are discontinuities in the intensity or color of the image. They can be used to extract features that are invariant to changes in illumination and viewpoint. Corner features are based on the corners in the image. Corners are points where two edges meet. They can be used to extract features that are invariant to changes in illumination and viewpoint. Shape features are based on the shape of the object in the image. They can be used to extract features that are invariant to changes in illumination and viewpoint. Spatial features can be effective for traffic sign recognition, but they can also be difficult to design and tune.

2.2 Empirical Review

Texture, shape, color and spatial hand felt features of an image are the basis of Feature Extraction and Classification (FEC). (Azhar et al., 2015), (Omid-Zohoor et al., 2018), (Jaiswal & Banka, 2017) and (Sugimura et al., 2016) gives examples of different methods that are used for feature extraction in FEC including Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Bag-of-Words (BOW), Local Binary Patterns (LBP) and Gabor filters. According to (Omid-Zohoor et al., 2018) and (Jaiswal & Banka, 2017), the extracted features are arranged and prepared into a format that can be used as an input to a machine learning algorithm such as random forest, logistics regression, K- nearest neighbor, and support vector machine.

Histogram of Oriented Gradients (HOG) algorithm was first introduced by Navneet Dalal and Bill Triggs in 2005 in their paper titled "Histograms of Oriented Gradients for Human Detection" (Jaiswal & Banka, 2017). Since its introduction, the HOG algorithm has been extensively studied and applied in various fields such as pedestrian detection, facial expression recognition, vehicle detection, and texture classification (Jaiswal & Banka, 2017). Notably, in 2005, Dalal and Triggs demonstrated the effectiveness of the HOG algorithm in detecting pedestrians in real-world images. They compared the performance of the HOG algorithm with other feature descriptors such as SIFT and showed that the HOG algorithm outperforms other feature descriptors in terms of accuracy and speed. (Wang & Schmid, 2013) proposed a multi-scale HOG algorithm that can handle objects of different sizes. They achieved better performance in detecting small objects by using a finer grid for feature extraction in small scales. (Mikolajczyk & Schmid, 2005), proposed a HOG-based method for detecting salient objects in images. (Mikolajczyk & Schmid, 2005) used the HOG algorithm to extract features and combined them with saliency maps to improve the performance of salient object detection.

Scale-Invariant Feature Transform (SIFT) algorithm was introduced by David Lowe in 1999 in his paper titled "Object recognition from local scale-invariant features" (Omid-Zohoor et al., 2018). SIFT algorithm has been widely studied and applied in various fields such as object recognition, 3D reconstruction, and image retrieval (Sugimura et al., 2016). Mikolajczyk and Schmid (2005) proposed a method for detecting and recognizing objects in images using SIFT key points and an SVM classifier. They achieved good performance in detecting and recognizing objects in cluttered Images. Sadeghi et al (2018) proposed a modified SIFT algorithm that uses local contrast normalization to improve the robustness of SIFT features to illumination changes. They achieved better performance in object recognition under varying lighting conditions. Luus et al (2015) proposed a deep learning-based method for feature extraction that combines the strengths of SIFT and deep convolutional neural networks (CNNs). They achieved state-of-the-art performance in object recognition on several benchmark datasets. SIFT effectiveness and robustness to scale, rotation, and affine distortion have been demonstrated in various applications, and modifications have been proposed to improve its performance.

The Bag-of-Words (BoW) model is a method for image and text classification in computer vision and natural language processing (Park et al., 2014). The BoW model has shown to be effective in various image classification tasks, such as Image recognition, object recognition, and face recognition (Ellahyani & El Ansari, 2021). In particular, the BoW model has been used in conjunction with classifiers such as support vector machines (SVMs) and random forests to achieve state-of-the-art performance in various benchmarks (Park et al., 2014), (Ellahyani & El Ansari, 2021) and (Zhao, 2013). The BoW model has also been applied in text classification tasks, where it represents each document as a bag of its constituent words. In this case, the BoW model creates a dictionary of words by collecting all unique words from the documents in the dataset (Park et al., 2014). Then, it represents each document as a histogram of the frequencies of the words in the document, where the words are the dictionary entries. The BoW model has been used in conjunction with classifiers such as Naive Bayes and SVMs to achieve state-of-the-art performance in various benchmarks. BoW model is a powerful and flexible method for feature extraction and classification in both image and text domains (Zhao, 2013).

Local Binary Patterns (LBP) was first introduced by Ojala et al. in 1996 as a simple yet effective texture descriptor, and since then, it has been extensively studied and applied in various computer vision applications (Kuo & Chiu, 2012). In texture classification, the LBP descriptor has been used in conjunction with classifiers such as SVMs and KNN to achieve good performance on standard texture datasets such as the Brodatz texture dataset (Nanni & Lumini, 2012). In object recognition, the LBP descriptor has been applied in conjunction with other feature descriptors such as SIFT and HOG on standard object recognition datasets such as the PASCAL VOC dataset (Peng et al., 2014). The Extended Local Binary Patterns (ELBP) descriptor (Zhang et al., 2013) extends the LBP operator by considering the circularly symmetric patterns, while the Rotation-Invariant Local Binary Patterns (RI-LBP) descriptor (Zhang et al., 2013) ensures the invariance of the LBP descriptor to rotation by encoding the local texture information using rotation-invariant patterns. LBP descriptor is a simple yet highly effective method for texture analysis and recognition in computer vision.

Gabor filters, named after physicist Dennis Gabor are based on the principle of decomposing an image into its constituent frequencies and orientations, similar to the human visual system (Ojala et al., 1996). Gabor filters have been extensively studied and applied in various computer vision applications one different datasets such as the Caltech 101 and Caltech 256 (Singh, 2016). Gabor filters have also been used in biometric applications such as face recognition and iris recognition, where they have been shown to be effective in capturing the local texture information of the face or iris. The Dual-Tree Complex Wavelet Transform (DT-CWT) (John et al., 2017) extends the Gabor filter by using a wavelet decomposition of the image, while the Log-Gabor filter (John et al., 2017) is a variant of the Gabor filter that is designed to be more efficient in terms of computational complexity. Gabor filter is a powerful tool for image and signal processing in computer vision

III. METHODOLOGY

3.1 Research Design and Approach

The performance of five popular handcrafted feature extraction techniques—Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Bag of Words (BoW), Local Binary Patterns (LBP), and Gabor Filters—in the context of traffic sign recognition (TSR) was systematically assessed and compared in this study using a quantitative experimental research design. The necessity to present actual data on the precision, effectiveness, and trade-offs of different techniques under controlled circumstances served as the driving force for the design.

The method coupled careful performance evaluation with various classifiers with experimentation on benchmark datasets. The German Traffic Sign Recognition Benchmark (GTSRB), an internationally known standard for TSR research, was chosen because of its complexity, diversity in illumination, occlusion, and picture resolution. It comprises 51,840 annotated traffic sign images in 43 categories.

3.2 The Dataset

The Germany Traffic Sign Detection Benchmark was used to carry out experiments on the five handcraft feature extractors. It has 51,840 traffic sign images that are grouped into 43 classes. These images are made up of different sizes, have varying occlusion and illumination making the dataset perfect for the comparison experiment. 80 percent of the images were used in training of the models and 20 percent for testing the performance of the model.

3.3 Experiment Procedure

The following section discusses how the experiment was carried out with the aim of comparing the five handcraft feature extractors. The experiment was done using python 3, windows 10, sklearn and CV libraries. Visual Studio IDE was used to conduct the experiment. All images were resized to 64 by 64 pixels before the process of feature extraction started. In order to normalize illumination and improve contrast, contrast limited adaptive histogram equalization was applied to the images. This process made features more robust to changes in light. Lastly, segmentation was done on the images with the aim of isolating the sign from the image. The classifier that will be used for training and testing the five handcraft features require a feature vector as a parameter. Therefore, each image was processed and the resulting feature vector was stored in as a set. Three classifiers were used to train and evaluate the feature sets that were extracted in section 3.2.1. These classifiers are Support Vector Machine, Random Forest and Decision Tree. All the three classifiers are part of the sklearn python library. Evaluation was done using accuracy, precision, recall and F1-Score. Validation was done using stratified k-fold with 4 folds and confusion matrix.

3.4 Justification of Design

This study's experimental research strategy is suitable since it allows for the controlled, repeatable testing of various feature extraction methods in consistent settings. Cross-validation and a combination of various classifiers guarantee that the study's conclusions are not skewed toward any one approach. Moreover, the GTSRB dataset offers a globally recognized standard by which the findings can be contrasted with earlier international research and used to guide local applications in Kenya.

IV. FINDINGS & DISCUSSION

4.1 Comparison of Handcrafted Features

Handcrafted features that can be extracted from an image are compared int the literature review and summarized in the table 1 outlining strength and weakness of each of these features.

Table 1

Summaries the Four Features stating the Advantages and Disadvantages

Feature	Strength	Weakness	Author
Color features	Easy to understand and interpret Can be used with a variety of machine learning algorithms Can be robust to variations in illumination to some extent	Can be sensitive to variations in viewpoint May not be able to capture all of the important information in the image	(Fu & Huang, 2010) and (Maletzky et al., 2023)
Texture features	Less sensitive to variations in illumination and viewpoint than color features Can be used to capture more information about the texture of the traffic sign	Can be difficult to design and tune May not be able to capture all of the important information in the image	(Fu & Huang, 2010) and (Kiran et al., 2009)
Spatial features	Less sensitive to variations in illumination and viewpoint than color and texture features Can be used to capture more information about the shape of the traffic sign	Can be difficult to design and tune May not be able to capture all of the important information in the image	(Kiran et al., 2009)
Shape features	Can be more robust to variations in illumination and viewpoint than color and texture features Can be used to capture more information about the shape of the traffic sign	Can be difficult to design and tune May not be able to capture all of the important information in the image	(Fu & Huang, 2010)

Although color features are easy to use and make use of the sign's color information, they are extremely sensitive to changes in lighting and perspective. Texture characteristics are more resilient to changes in perspective and illumination than color features. They also study surface patterns while spatial features are less impacted by environmental factors and capture interactions between pixels; however, their effectiveness depends on proper design. Shape aspects which emphasize geometric qualities like corners and contours that provide high resilience to visual changes, they also pose serious design difficulties. Even if each feature type has unique benefits for traffic sign identification overall, they are nonetheless limited by the same design complexity and information loss risk.

4.2 Feature Extraction and Classification Techniques

The section compares the feature extraction techniques that can be applied on an image.

Table 2

Summary of feature Extraction and Classification

Technique	Strength	Weakness	Author
Histogram of Oriented Gradients	Effective for capturing shape and texture information in images. Robust to changes in illumination and background clutter.	May struggle with complex background textures and cluttered scenes. Sensitivity to variations in object appearance and viewpoint	(Jaiswal & Banka, 2017) (Wang & Schmid, 2013) and (Mikolajczyk & Schmid, 2005)
Scale-Invariant Feature Transform	Provides distinctive and invariant local features robust to scaling, rotation, and affine transformations. Well-suited for object recognition and image matching tasks.	Computational complexity increases with the number of key points and descriptors. Requires careful parameter tuning for optimal performance.	(Omid-Zohoor et al., 2018) and (Luus et al., 2015)
Bag-of-Words	Efficient representation of image features using visual word histograms. Suitable for classification and retrieval tasks in large-scale image datasets.	Loss of spatial information due to histogram aggregation. Limited ability to capture fine details and geometric information.	(Ellahyani & El Ansari, 2021) (Zhao, 2013) and (Kuo & Chiu, 2012)
Local Binary Patterns	Simple and computationally efficient texture descriptor. Effective for texture classification and segmentation tasks.	Limited discriminative power compared to more complex descriptors. May struggle with complex texture patterns and variations.	(Nanni & Lumini, 2012) (Peng et al., 2014) and (Ojala et al., 1996)
Gabor Filters	Sensitive to texture variations and edges in images. Useful for tasks such as texture analysis, segmentation, and edge detection.	Parameter selection and tuning can be challenging. Computationally intensive, especially for large filter banks and image sizes.	(Banik et al., 2022), (Singh, 2016) and (John et al., 2017)

HOG is highly effective for capturing shape and texture information while demonstrating robustness to illumination changes, though it can struggle with cluttered backgrounds. SIFT provides excellent invariance to scale and rotation transformations, making it ideal for recognition tasks, but at the cost of high computational complexity. Bag-of-Words offers efficient representation for large-scale datasets through histogram aggregation, but sacrifices spatial information and fine details. LBP is computationally simple and effective for texture analysis, though it has limited discriminative power for complex patterns. Gabor Filters are sensitive to texture and edge variations, useful for detailed analysis, but require significant parameter tuning and computational resources. Each method presents a trade-off between descriptive power, computational efficiency, and implementation complexity.

4.3 Experimental Results

The first experiment main aim was to compare the features of the five feature descriptors. A grayscale image was used as an unput to the functions that extracts features. Figure 1 and 2 shows the extracted features for each extractor.

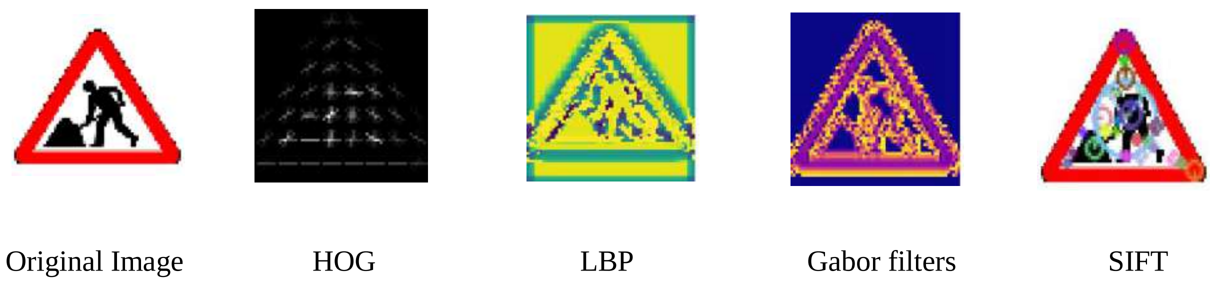

Figure 1
Visualization of Handcrafted Features Extracted from a sample Traffic Sign

From left to right: Preprocessed grayscale input image, HOG feature visualization, LBP texture map, and Gabor filter response magnitude.

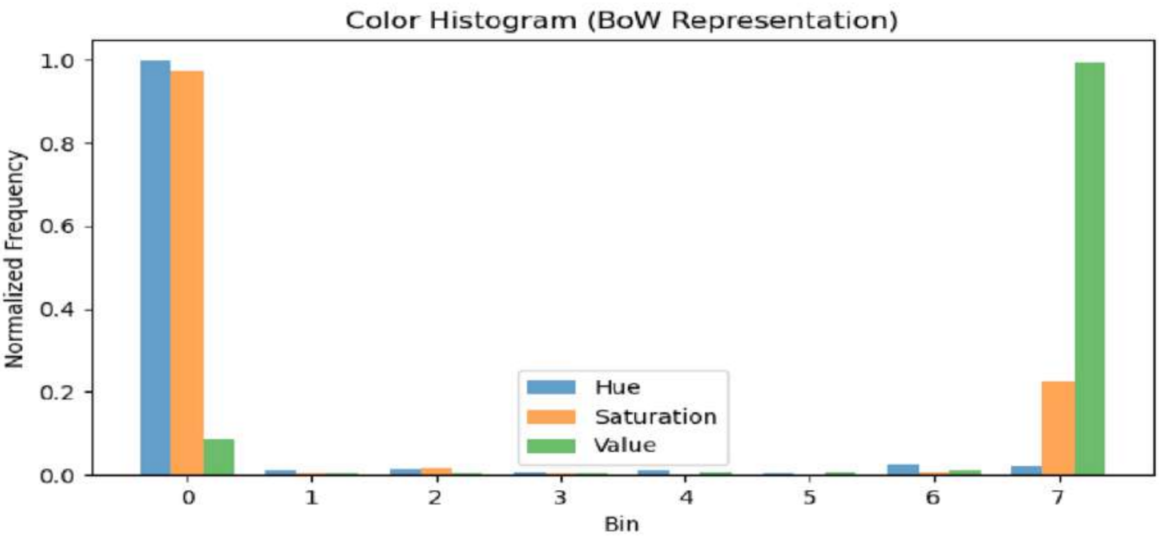

Figure 2
Visualization of Handcrafted Features Extracted from a Sample Traffic sign using Bag of Words

The second experiment was creation and evaluation of the model. Testing results on how the five feature extractors performed when subjected to three classifies are summarized by table 1, 2 and 3 below.

Table 3
Comparison for LBP, LDP, HOG and LD-HOG on GTSDB dataset using SVM

Feature Extractor	Precision	Recall	F1- Score
HOG	96	95	93
SIFT	92	92	95
BoW	93	92	91
LBP	97	89	94
Gabor filter	82	85	83

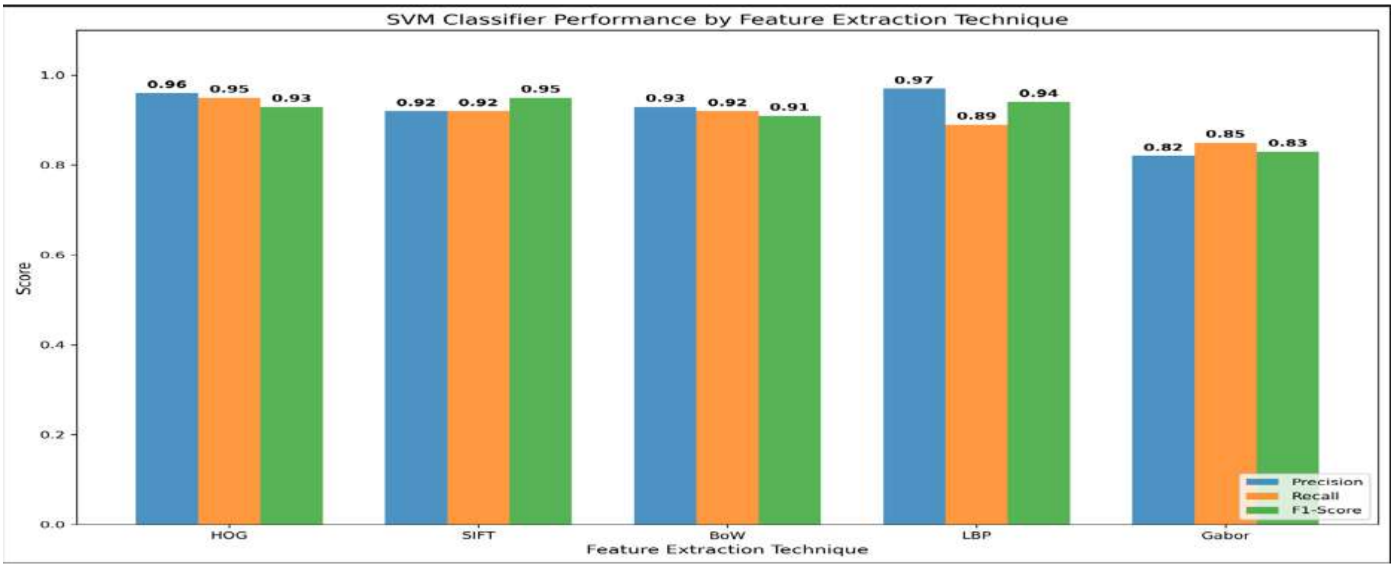

Figure 3
Visualization of Performance of Feature Extractors using SVM

HOG demonstrated strong and balanced performance with high precision and recall, confirming its robustness for capturing the defining shapes of traffic signs. SIFT achieved the highest overall F1-Score while LBP attained the highest precision, indicating minimal false positives, its lower recall suggests a weakness in detecting all true signs. The Bag-of-Words model delivered solid results, reflecting its general-purpose nature, whereas Gabor Filters performed significantly worse than all other methods.

Table 4
Comparison for LBP, LDP, HOG and LD-HOG on GTSDB dataset using Random Forest

Feature Extractor	Precision	Recall	F1- Score
HOG	89	89	89
SIFT	83	90	86
BoW	88	82	86
LBP	90	92	89
Gabor filter	78	80	81

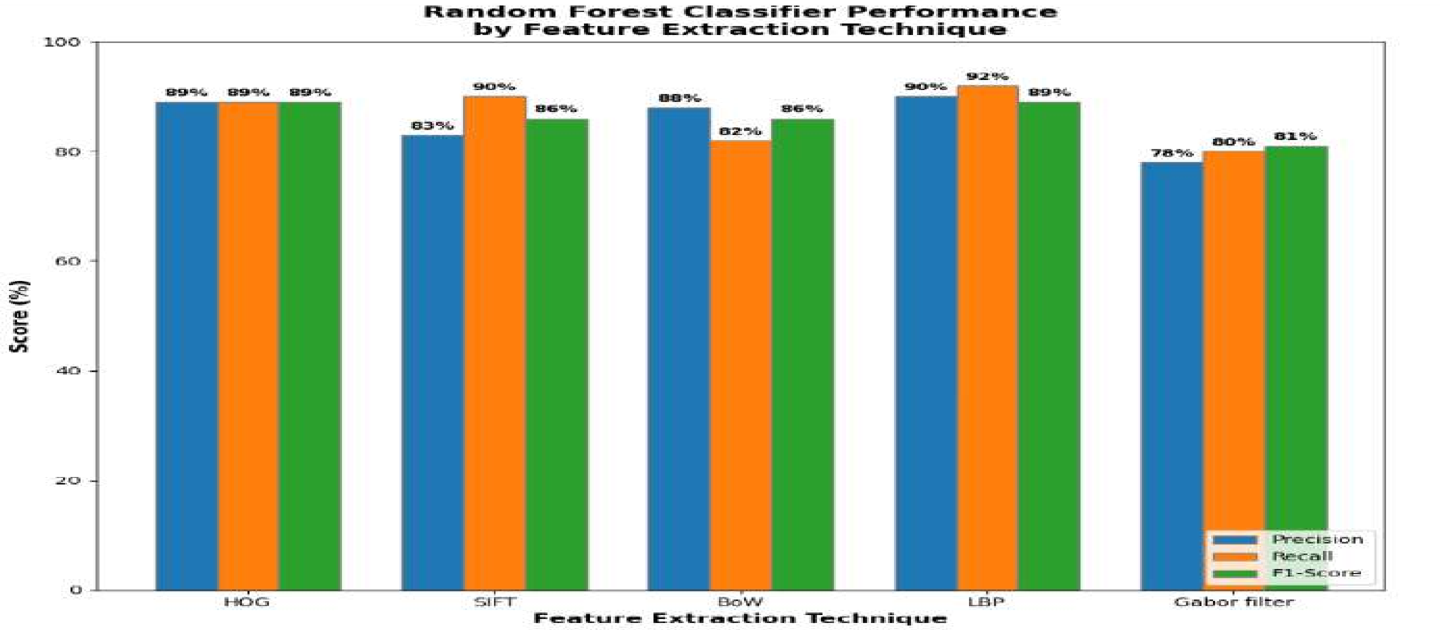

Figure 4
Visualization of Performance of Feature Extractors using Random Forest

Based on the results, LBP emerges as the top-performing feature extractor for the Random Forest classifier, achieving the highest combined scores with excellent precision (90%) and the best recall (92%), resulting in a strong F1-Score of 89%. HOG also performs robustly, demonstrating perfect balance across all three metrics (89% each), indicating consistent and reliable predictions. SIFT and BoW deliver comparable overall performance (F1: 86%), but with a key difference: SIFT is better at finding all relevant signs (high recall: 90%), while BoW is better at ensuring its predictions are correct (high precision: 88%). Finally, Gabor Filters are the least effective by a significant margin across all metrics (Precision: 78%, Recall: 80%, F1-Score: 81%), confirming they are not well-suited for this task compared to the other methods.

Table 5
Comparison for LBP, LDP, HOG and LD-HOG on GTSDB dataset using Decision Tree

Feature Extractor	Precision	Recall	F1- Score
HOG	76	76	76
SIFT	72	73	76
BoW	70	69	67
LBP	74	72	75
Gabor filter	68	70	69

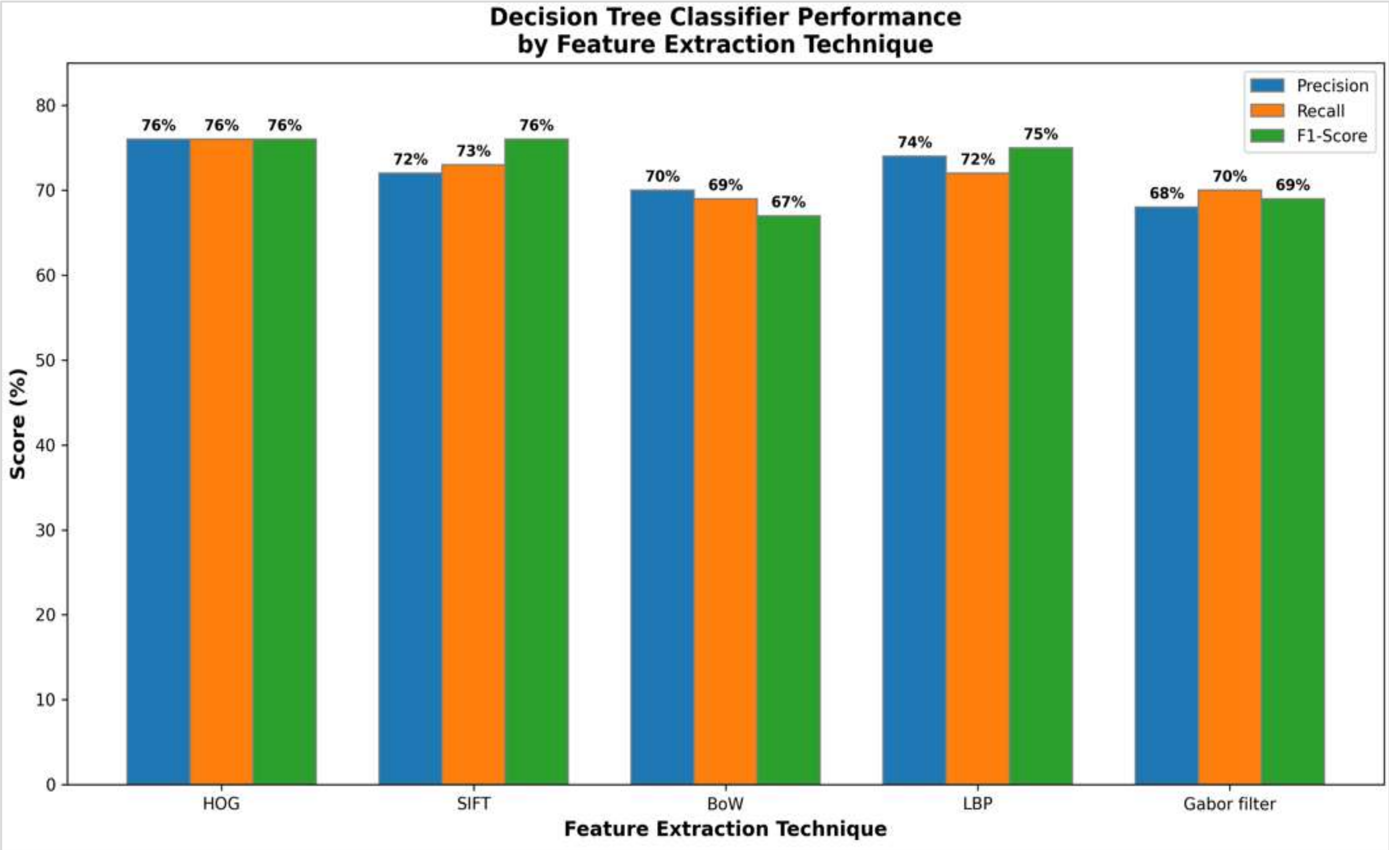

Figure 5
Visualization of Performance of Feature Extractors using Decision Tree

Validation of the evaluation results was done using stratified k fold with 4 folds on the HOG results for the three classifiers. Table 6 summaries these results.

Table 6

Validation of HOG Results on the Three Classifier.

classifier	Fold 1	Fold 2	Fold 3	Fold 4	Average
SVM	94.9	94.2	95.1	95.9	95.2
RF	88.3	88.3	90.1	89.3	89.0
DT	75.7	76.8	75.8	75.7	76.2

The results suggest that the evaluation criteria were right and accurate because the results were with the acceptable range of variation. This was generalized to the other feature extractors. Confusion matrix as shown by figure 6 further supports the validity of the evaluation results.

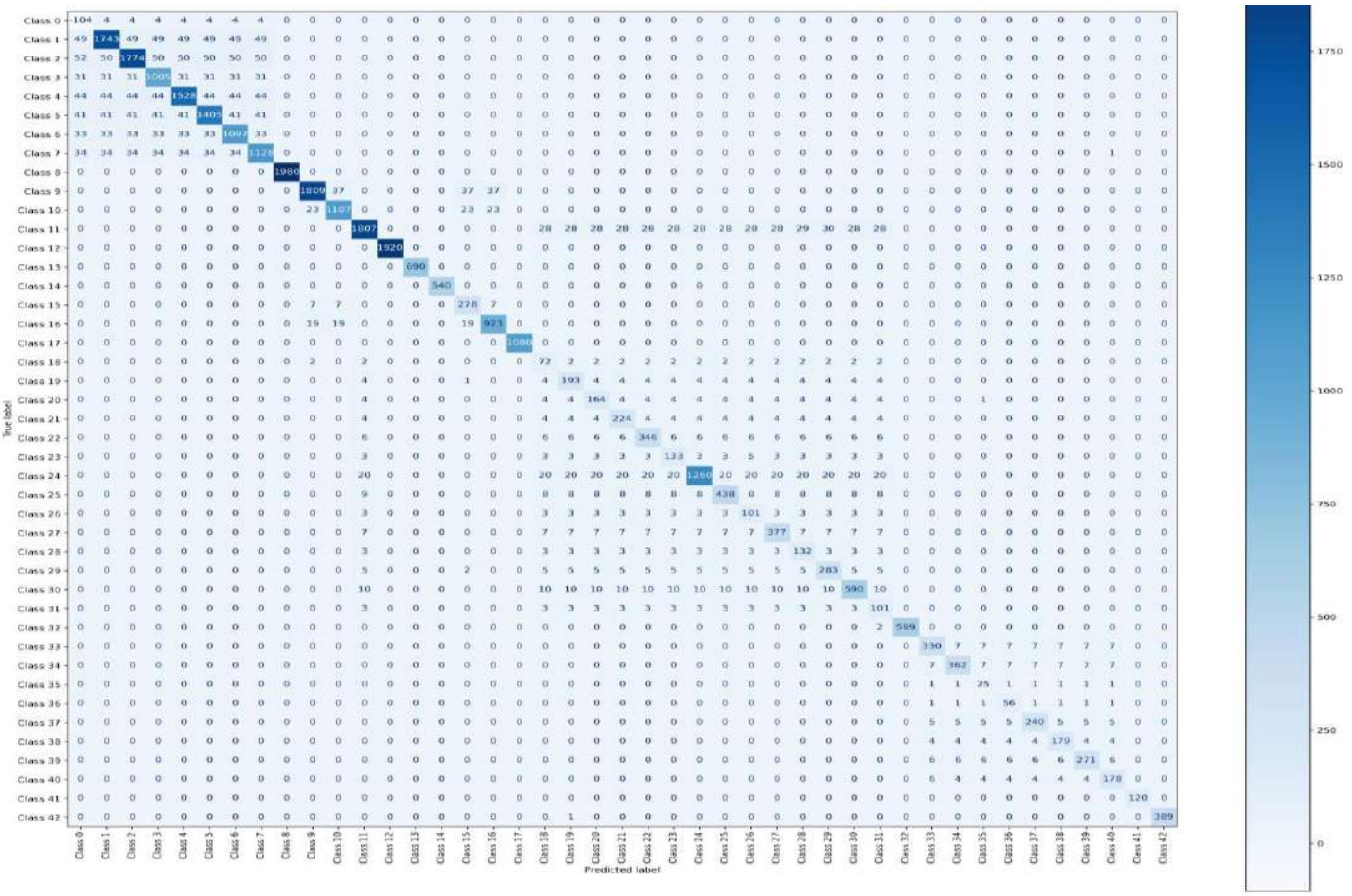

Figure 6

HOG and SVM Confusion Matrix for the GTSDB Dataset

With equal Precision, Recall, and F1-scores of 76%, the Decision Tree experimental results demonstrated that HOG features produced the most consistent and balanced performance. This supports the findings from the literature review, which showed that HOG's great capacity to capture shape and gradient information made it one of the best handmade feature descriptors for TSR. The empirical review also noted that several studies applying HOG to the GTSRB dataset achieved superior recognition rates, supporting the current findings. The resilience of HOG as a baseline feature extraction technique is further supported by the fact that it remained consistently strong across classifiers, even though the Decision Tree classifier generated lower overall scores than SVM (95.2%) and Random Forest (89.0%).

Both SIFT and LBP produced competitive results; SIFT matched HOG's F1-score while having poorer individual metrics, while LBP showed good precision and recall balance. While SIFT was praised for its invariance to scale and rotation but critiqued for its greater processing demands, this is consistent with other studies that highlighted LBP's effectiveness and applicability for real-time applications because to its low computing cost. Thus, the findings of this work are consistent with real data demonstrating that both LBP and SIFT continue to be formidable competitors in TSR tasks, particularly when resource limitations and transformation resilience are taken into account.

In contrast, BoW and Gabor filters were the least effective feature extraction techniques, with BoW producing the lowest F1-score (67%) and Gabor filters performing poorly across all metrics. This is consistent with the empirical review, where BoW was noted to lose important spatial information when representing traffic sign features, leading to weaker classification results. Similarly, Gabor filters, although theoretically effective in texture and edge detection, were reported to be computationally intensive and less adaptable in real-world TSR applications. The current study's findings therefore confirm these limitations in practice.

The stratified k-fold validation with four folds provided further confirmation of the reliability of these results. HOG in particular showed consistent performance across the three classifiers (SVM, Random Forest, Decision Tree), which strengthens confidence in its suitability for TSR tasks. This echoes the methodological practices highlighted in the reviewed empirical studies, where cross-validation was emphasized as critical for ensuring generalizable and unbiased performance evaluations.

Taken together, the results of this study reinforce the literature and empirical evidence: HOG remains the strongest handcrafted feature descriptor for TSR, particularly in terms of balanced performance across classifiers; LBP provides an efficient, lightweight alternative for resource-limited applications; SIFT offers transformation robustness but at higher computational cost; while BoW and Gabor filters are less suited for reliable TSR tasks, especially in contexts where accuracy and efficiency are both required.

V. CONCLUSION & RECOMMENDATIONS

5.1 Conclusion

Several important inferences can be drawn from the experimental data. With a performance difference of up to 29 percentage points in F1-score between the best and poorest approaches, the feature extraction strategy selection has a significant effect on classification performance. Second, out of all the classifiers, HOG features performed the best or very close to the best accuracy with SVM (95.2%), Random Forest (89.0%), and Decision Tree (76.2%). This illustrates how well it can capture the unique gradient and shape information found on traffic signs. Third, SIFT's computational complexity is a major disadvantage for real-time applications, even if it demonstrated exceptional invariance to transformation and obtained the greatest F1-score with SVM. Fourth, LBP was appropriate for systems with limited resources because it provided a good tradeoff between excellent performance and noticeably faster computation times. Ultimately, the findings demonstrate the importance of classifier selection, with SVM consistently outperforming Random Forest and Decision Tree as a competent but less accurate baseline across all feature types. With competitive accuracy of up to 96%, this study confirms that intricate handcrafted feature pipelines are still very successful at recognizing traffic signs.

5.2 Recommendations

This paper proposes several recommendations for improving traffic sign recognition systems. For applications where recognition accuracy is crucial, such as safety-critical autonomous driving systems, the use of HOG feature extraction combined with an SVM classifier is recommended for optimal accuracy. To ensure effective resource utilization, combining LBP features with a Random Forest classifier is suggested. Researchers are also encouraged to explore hybrid feature techniques that integrate the strengths of multiple extractors, particularly the combination of LBP's texture features with HOG's shape information. Moreover, special attention should be given to class imbalance in traffic sign databases, as performance on rare sign classes tends to lag behind more common ones, as indicated by confusion matrix analysis. Employing class-balanced sampling techniques or targeted data augmentation could significantly enhance system robustness. Finally, in scenarios where computing resources are limited or where feature interpretability is preferred over the black-box nature of deep learning, the handcrafted features evaluated in this study provide a reliable foundation for traffic sign recognition systems.

REFERENCES

Almaskati, D., Kermanshachi, S., & Pamidimukkala, A. (2023). Autonomous vehicles and traffic accidents. *International Scientific Conference "The Science and Development of Transport," 17*(3), 1–8. ELSEVIER. https://www.researchgate.net/publication/374947350_Autonomous_vehicles_and_traffic_accidents

Azhar, R., Tuwohingide, D., Kamudi, D., Suciati, N., & Sarjana, S. (2015). Batik image classification using SIFT feature extraction, bag of features and support vector machine. *Procedia Computer Science, 72*, 24–30. https://www.researchgate.net/publication/289735014_Batik_Image_Classification_Using_SIFT_Feature_Extraction_Bag_of_Features_and_Support_Vector_Machine

Ellahyani, A., & El Ansari, M. (2021). Traffic sign detection for intelligent transportation systems: A survey. *E3S Web of Conferences, 229*, 01040. https://doi.org/10.1051/e3sconf/202122901040

Fu, M.-Y., & Huang, Y.-S. (2010, December). A survey of traffic sign recognition. In *2010 International Conference on Wavelet Analysis and Pattern Recognition* (pp. 119–124). IEEE.

Jaiswal, A. K., & Banka, H. (2017). Local pattern transformation-based feature extraction techniques for classification of epileptic EEG signals. *Biomedical Signal Processing and Control, 34*, 81–92. https://www.sciencedirect.com/science/article/abs/pii/S174680941730006X?via%3Dihub

John, V., Boyali, A., & Mita, S. (2017). Gabor filter and Gershgorin disk-based convolutional filter constraining for image classification. *International Journal of Machine Learning and Computing, 7*(4), 55–60. https://doi.org/10.18178/ijmlc.2017.7.4.620

Khalid, S., Khalil, T., & Nasreen, S. (2014, November). A survey of feature selection and feature extraction techniques in machine learning. In *2014 Science and Information Conference* (pp. 372–378). IEEE. https://doi.org/10.1109/SAI.2014.6918213

Kiran, C. G., Prabhu, L., Varikkottil, A., & Kumaraswamy, R. (2009). Traffic sign detection and pattern recognition using support vector machine. In *Proceedings* (pp. 87–90). https://www.researchgate.net/publication/220781211_Traffic_Sign_Detection_and_Pattern_Recognition_Using_Support_Vector_Machine

Kuo, W. J., & Chiu, C. C. (2012). Content-based image retrieval with bag of visual words. *Journal of Information Science and Engineering*, *28*(3), 655–668.

Luus, F. P. S., Salmon, B. P., van den Bergh, F., & Maharaj, B. T. J. (2015). Multiview deep learning for land-use classification. *IEEE Geoscience and Remote Sensing Letters, 12*(12), 2448–2452. https://doi.org/10.1109/LGRS.2015.2483680

Maletzky, A., Hofer, N., Thumfart, S., Bruckmüller, K., & Kasper, J. (2023). Traffic sign detection and classification on the Austrian highway traffic sign data set. *Data, 8*(1), 16. https://doi.org/10.3390/data8010016

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(10), 1615–1630. https://doi.org/10.1109/TPAMI.2005.188

Nanni, L., & Lumini, A. (2012). Local binary patterns variants as texture descriptors for medical image analysis. *Artificial Intelligence in Medicine*, *56*(1), 25–35. https://doi.org/10.1016/j.artmed.2012.07.001

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, *29*(1), 51–59. https://doi.org/10.1016/0031-3203(95)00067-4

Omid-Zohoor, A., Almeida, B., Taori, R., Murmann, B., & Chandraiah, P. (2018). Toward always-on mobile object detection: Energy versus performance tradeoffs for embedded HOG feature extraction. *IEEE Transactions on Circuits and Systems for Video Technology*, *28*(5), 1102–1115. https://www.researchgate.net/publication/312476432_Toward_Always-On_Mobile_Object_Detection_Energy_Versus_Performance_Tradeoffs_for_Embedded_HOG_Feature_Extraction

Park, K. H., Wang, S.-W., & Wang, W.-H. (2014). Frequency-cepstral features for bag of words based acoustic context awareness. *The Journal of the Acoustical Society of Korea*, *33*(4), 248–254. https://doi.org/10.7776/ASK.2014.33.4.248

Peng, F., Qian, L., Yang, J., Zhang, C., & Zuo, W. (2014). Feature coding in image classification: A comprehensive study. *IEEE Transactions on Multimedia*, *16*(11), 2828–2840. https://www.researchgate.net/publication/239943155_Feature_Coding_in_Image_Classification_A_Comprehensive_Study

Rangayyan, R. M., Banik, S., & Desautels, J. E. (2010). Computer-aided detection of architectural distortion in prior mammograms of interval cancer. *Journal of Digital Imaging*, *23*(5), 564–577. https://link.springer.com/article/10.1007/s10278-009-9257-x

Saadna, Y., & Behloul, A. (2017). An overview of traffic sign detection and classification methods. *International Journal of Multimedia and Information Retrieval*, *6*(3), 193–206. https://doi.org/10.1007/s13735-017-0129-8

Sadeghi, B., Jafari, A. M., Vahdani, K., & Shahri, A. M. (2018). 2DIGH: A polar invariant local image descriptor based on joint histogram. *The Visual Computer*, *34*, 1579–1595. https://www.researchgate.net/publication/319661560_2DIGH_a_polar_invariant_local_image_descriptor_based_on_joint_histogram

Singh, B. (2016). Iris recognition using curvelet transformation based on Gabor filter & SVM. *International Journal of Engineering and Computer Science, 5*(8), 17731–17735. https://doi.org/10.18535/ijecs/v5i8.32

Sugimura, D., Fujimura, T., & Hakura, T. (2016). Enhanced cascading classifier using multi-scale HOG for pedestrian detection from aerial images. *International Journal of Pattern Recognition and Artificial Intelligence, 30*(3), 1655009. https://www.researchgate.net/publication/283617136_Enhanced_Cascading_Classifier_Using_Multi-Scale_HOG_for_Pedestrian_Detection_from_Aerial_Images

Wali, S. B. (2015). Comparative survey on traffic sign detection and recognition: A review. *Przeglad Elektrotechniczny*, *91*(10), 40–44.

https://www.researchgate.net/publication/283617136_Enhanced_Cascading_Classifier_Using_Multi-Scale_HOG_for_Pedestrian_Detection_from_Aerial_Images

Wang, H., & Schmid, C. (2013, December). Action recognition with improved trajectories. In *2013 IEEE International Conference on Computer Vision* (pp. 3551–3558). IEEE. https://doi.org/10.1109/ICCV.2013.441

Zhang, Z., Gao, S., & Zhao, B. (2013). Multi-scale local binary pattern for texture classification. *Journal of Computational Information Systems, 9*(7), 2561–2568.

Zhao, S. (2013). Image bag generator based on bag of visual words. *Journal of Information and Computational Science, 10*(5), 1453–1462. https://doi.org/10.12733/jics20101532