جامعة بيرزيت

BIRZEIT UNIVERSITY

Computer Vision
ENCS5343

Assignment No. 2

Arabic Handwritten Text Identification Using Local Feature Extraction Techniques

Prepared by:

Name: Jouwana Daibes ID: 1210123 Section: 2

Name: Lana Mussafer ID: 1210455 Section: 1

Instructor:

Dr. Aziz Qaroash

Date:

December 22, 2024

1 Introduction

The objective of this assignment is to develop and evaluate a system for writer identification using images of handwritten words. The approach involves extracting features from the images using ORB and SIFT algorithms, representing them as Bag of Visual Words (BoVW) using KMeans clustering, and classifying the writers with machine learning models such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM). The performance of these methods is analyzed to identify the most effective combination for the task.

2 Dataset Description

The dataset used in this assignment consists of handwritten words from multiple writers. It includes 53199 alphabet images organized into folders corresponding to each writer. The data set was divided into 80% for training to train the model and 20% for testing to evaluate it. Preprocessing steps included resizing all images to a uniform size of 128x128 pixels, converting them to grayscale, and applying Gaussian blur for noise reduction. This consistent preprocessing ensured that the features extracted from the images were reliable and standardized for classification tasks.

3 Feature Extraction

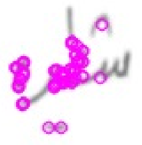
3.1 ORB Feature Extraction

The ORB algorithm was employed to detect and describe keypoints in images. ORB combines the FAST keypoint detector and BRIEF descriptor while adding rotation invariance and noise resistance. During experimentation, various parameters of ORB were adjusted, such as the number of features and scaling factor. However, the results showed minimal variation in accuracy, so the default parameters were ultimately used. The generated descriptors were stored as .npy files for efficient retrieval during the classification process.

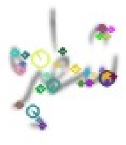
3.2 SIFT Feature Extraction

The SIFT algorithm was used to extract robust keypoints and descriptors invariant to scaling, rotation, and illumination changes. Several parameters, including the number of octaves and contrast threshold, were experimented with to observe their impact on accuracy. Similar to ORB, the results showed minimal differences with parameter adjustments, leading to the use of default values. The 128-dimensional descriptor vectors were saved as .npy files to facilitate efficient processing in subsequent steps.

Below is a demonstration of Number of key points detected using ORB and SIFT and Time Efficiency for each Algorithm:

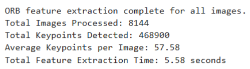


[ORB Keypoints]



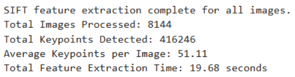
[SIFT Keypoints]

[ORB Algorithm]



```
ORB feature extraction complete for all images.
Total Images Processed: 8144
Total Keypoints Detected: 468900
Average Keypoints per Image: 57.58
Total Feature Extraction Time: 5.58 seconds
```

[SIFT Algorithm]



```
SIFT feature extraction complete for all images.
Total Images Processed: 8144
Total Keypoints Detected: 416246
Average Keypoints per Image: 51.11
Total Feature Extraction Time: 19.68 seconds
```

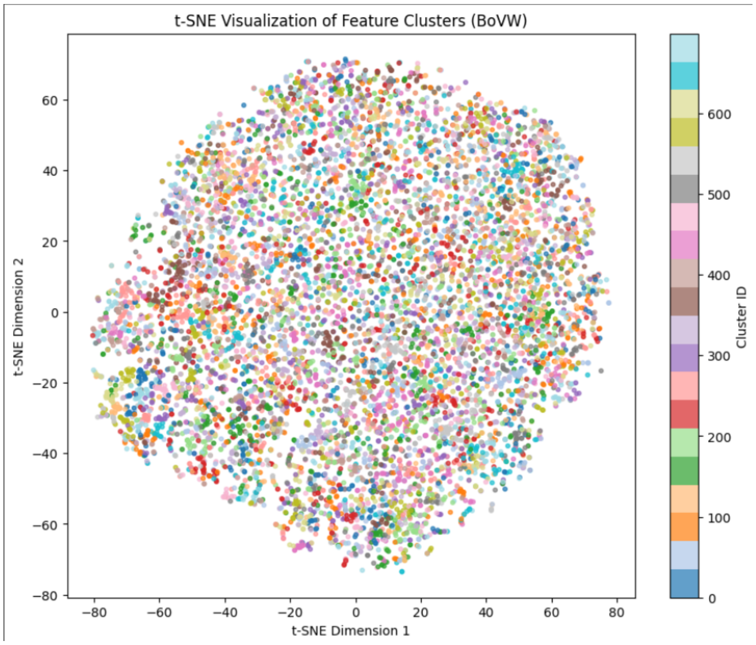
4 Bag of Visual Words (BoVW)

The Bag of Visual Words (BoVW) approach represents images as collections of visual words, which are formed by clustering local feature descriptors. KMeans clustering is employed to group similar descriptors into distinct clusters, where each cluster represents a unique visual word. This technique effectively reduces the complexity of feature data while maintaining discriminative information for classification tasks.

A total of 700 clusters was selected for KMeans clustering. This number was determined experimentally, as it provided the best balance between accuracy and computational efficiency. Increasing or decreasing the number of clusters beyond 700 showed diminishing returns in performance.

Once the visual words are defined, each image is represented as a histogram of visual word occurrences. These histograms are generated by assigning each local feature descriptor of an image to its nearest visual word cluster. The resulting histogram captures the frequency of each visual word, creating a compact and informative representation for classification.

To better understand the clustering of feature descriptors into visual words, we used t-SNE to visualize the high-dimensional descriptor space in 2D after applying KMeans clustering. The figure below shows how feature descriptors from the training images are distributed among the 700 visual words (clusters). Each point represents a feature descriptor, and colors correspond to the cluster assignments. This visualization highlights the distinctiveness and separation of visual words, reinforcing their importance in creating robust image representations.



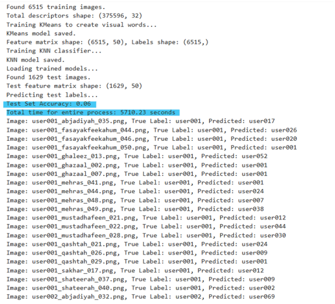
5 Experimental Setup and Classifiaction Models

First, we downloaded the AWAHP dataset and performed preprocessing steps such as resizing and applying Gaussian noise. The dataset was then split into training (80%) and testing (20%) sets, with the model being trained on the training set and tested on the test set. After that, feature extraction was performed separately using the ORB and SIFT algorithms. The keypoints from both algorithms were then applied to KNN and SVM classification models in order to compare the performance of each algorithm (ORB and SIFT) on these models.

5.1 K-Nearest Neighbors (KNN)

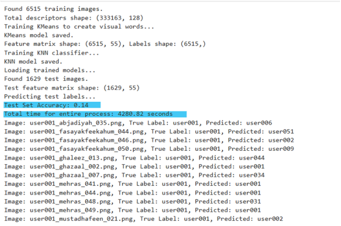
The K-Nearest Neighbors (KNN) classification model for image recognition is implemented in three main steps: first, KMeans clustering is used to extract descriptors from the tarining images and create a Bag of Visual Words (BOVW) by clustering the descriptors into a predefined number of visual words. The trained KMeans model is saved for future use. In the second step, histograms of visual words are generated from the descriptors, which are then used to train the KNN classifier. The trained KNN model is saved for making predictions. For parameter tuning, several tests were performed to determine the optimal number of clusters and neighbors for achieving the highest accuracy. After testing various configurations, it was found that the best results were achieved with 55 clusters and 20 neighbors. Finally, in the third step, the trained K Means and KNN models are loaded, and descriptors are extracted from the test images to generated histograms. These histograms are used to predict the labels (writer/user) of the test images, and the model's performance is evaluated based on accuracy and time efficiency. The results are displayed for each test image, showing the true and predicted labels.

The figures below show the Accuracy and Time Efficiency of SIFT and ORB Algorithms in KNN classification:



```
Found 6515 training images.
Total descriptors shape: (375596, 32)
Training KMeans to create visual words...
KMeans model saved.
Feature matrix shape: (6515, 50), Labels shape: (6515,)
Training KNN classifier...
KNN model saved.
Loading trained models...
Found 1629 test images.
Test feature matrix shape: (1629, 50)
Predicting test labels...
Test Set Accuracy: 0.06
Total time for entire process: 5710.23 seconds
Image: user001_abjadiyah_035.png, True Label: user001, Predicted: user017
Image: user001_fasayakfeekahum_044.png, True Label: user001, Predicted: user026
Image: user001_fasayakfeekahum_046.png, True Label: user001, Predicted: user020
Image: user001_fasayakfeekahum_050.png, True Label: user001, Predicted: user001
Image: user001_ghaleez_013.png, True Label: user001, Predicted: user052
Image: user001_ghazaal_002.png, True Label: user001, Predicted: user001
Image: user001_ghazaal_007.png, True Label: user001, Predicted: user001
Image: user001_mehras_041.png, True Label: user001, Predicted: user001
Image: user001_mehras_044.png, True Label: user001, Predicted: user024
Image: user001_mehras_048.png, True Label: user001, Predicted: user007
Image: user001_mehras_049.png, True Label: user001, Predicted: user038
Image: user001_mustadhafeen_021.png, True Label: user001, Predicted: user012
Image: user001_mustadhafeen_022.png, True Label: user001, Predicted: user044
Image: user001_mustadhafeen_028.png, True Label: user001, Predicted: user030
Image: user001_qashtah_021.png, True Label: user001, Predicted: user024
Image: user001_qashtah_026.png, True Label: user001, Predicted: user009
Image: user001_qashtah_029.png, True Label: user001, Predicted: user001
Image: user001_sakhar_017.png, True Label: user001, Predicted: user012
Image: user001_shateerah_037.png, True Label: user001, Predicted: user009
Image: user001_shateerah_040.png, True Label: user001, Predicted: user002
Image: user002_abjadiyah_032.png, True Label: user002, Predicted: user069
```

[ORB Algorithm]



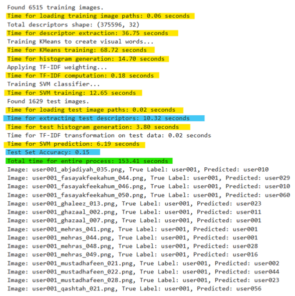
```
Found 6515 training images.
Total descriptors shape: (333163, 128)
Training KMeans to create visual words...
KMeans model saved.
Feature matrix shape: (6515, 55), Labels shape: (6515,)
Training KNN classifier...
KNN model saved.
Loading trained models...
Found 1629 test images.
Test feature matrix shape: (1629, 55)
Predicting test labels...
Test Set Accuracy: 0.14
Total time for entire process: 4280.82 seconds
Image: user001_abjadiyah_035.png, True Label: user001, Predicted: user006
Image: user001_fasayakfeekahum_044.png, True Label: user001, Predicted: user051
Image: user001_fasayakfeekahum_046.png, True Label: user001, Predicted: user002
Image: user001_fasayakfeekahum_050.png, True Label: user001, Predicted: user009
Image: user001_ghaleez_013.png, True Label: user001, Predicted: user044
Image: user001_ghazaal_002.png, True Label: user001, Predicted: user001
Image: user001_ghazaal_007.png, True Label: user001, Predicted: user034
Image: user001_mehras_041.png, True Label: user001, Predicted: user001
Image: user001_mehras_044.png, True Label: user001, Predicted: user001
Image: user001_mehras_048.png, True Label: user001, Predicted: user031
Image: user001_mehras_049.png, True Label: user001, Predicted: user001
Image: user001_mustadhafeen_021.png, True Label: user001, Predicted: user002
```

[SIFT Algorithm]

5.2 Support Vector Machines (SVM)

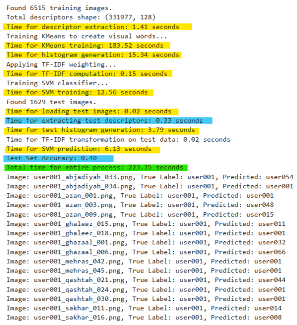
The Support Vector Machines (SVM) classification model for image recognition is implemented using the Bag of Visual Words (BoVW) approach with SIFT and ORB keypoints. The process starts by extracting keypoint descriptors from pre-saved files for training and test images. It then uses KMeans clustering to create a visual vocabulary (clusters of keypoints) with 700 clusters for both SIFT and ORB algorithms. Parameter tunning was conducted through a grid search to determine the optimal parameters in SVM for the highest accuracy and the best parameters found to be [kernel=rbf, c=10, gamma=1, class-weight=None], where several numbers of clusters were tested, and 700 clusters were found to be the best. Afterward, the descriptors are converted into histograms, which represent the frequency of visual words in each image. These histograms are then transformed using TF-IDF (Term Frequency-Inverse Document Frequency) weighting to enhance feature importance. The SVM (Support Vector Machine) classifier is trained using these TF-IDF weighted histograms, and the model is evaluated on the test set. The accuracy of the model is computed, and the classification time for both training and testing is logged for performance analysis.

The figures below show the Accuracy and Time Efficiency of SIFT and ORB Algorithms in SVM classification:



```
Found 6515 training images.
Time for loading training image paths: 0.06 seconds
Total descriptors shape: (375596, 32)
Time for descriptor extraction: 36.75 seconds
Training KMeans to create visual words...
Time for KMeans training: 68.72 seconds
Time for histogram generation: 14.70 seconds
Applying TF-IDF weighting...
Time for TF-IDF computation: 0.18 seconds
Training SVM classifier...
Time for SVM training: 12.65 seconds
Found 1629 test images.
Time for loading test image paths: 0.02 seconds
Time for extracting test descriptors: 10.32 seconds
Time for test histogram generation: 3.80 seconds
Time for TF-IDF transformation on test data: 0.02 seconds
Time for SVM prediction: 6.19 seconds
Test Set Accuracy: 0.15
Total time for entire process: 153.41 seconds
Image: user001_abjadiyah_035.png, True Label: user001, Predicted: user010
Image: user001_fasayakfeekahum_044.png, True Label: user001, Predicted: user029
Image: user001_fasayakfeekahum_046.png, True Label: user001, Predicted: user010
Image: user001_fasayakfeekahum_050.png, True Label: user001, Predicted: user060
Image: user001_ghaleez_013.png, True Label: user001, Predicted: user023
Image: user001_ghazaal_002.png, True Label: user001, Predicted: user011
Image: user001_ghazaal_007.png, True Label: user001, Predicted: user001
Image: user001_mehras_041.png, True Label: user001, Predicted: user001
Image: user001_mehras_044.png, True Label: user001, Predicted: user001
Image: user001_mehras_048.png, True Label: user001, Predicted: user028
Image: user001_mehras_049.png, True Label: user001, Predicted: user016
Image: user001_mustadhafeen_021.png, True Label: user001, Predicted: user002
Image: user001_mustadhafeen_022.png, True Label: user001, Predicted: user044
Image: user001_mustadhafeen_028.png, True Label: user001, Predicted: user023
Image: user001_qashtah_021.png, True Label: user001, Predicted: user056
```

[ORB Algorithm]



```
Found 6515 training images.
Total descriptors shape: (331977, 128)
Time for descriptor extraction: 1.41 seconds
Training KMeans to create visual words...
Time for KMeans training: 183.52 seconds
Time for histogram generation: 15.34 seconds
Applying TF-IDF weighting...
Time for TF-IDF computation: 0.15 seconds
Training SVM classifier...
Time for SVM training: 12.56 seconds
Found 1629 test images.
Time for loading test images: 0.02 seconds
Time for extracting test descriptors: 0.33 seconds
Time for test histogram generation: 3.79 seconds
Time for TF-IDF transformation on test data: 0.02 seconds
Time for SVM prediction: 6.13 seconds
Test Set Accuracy: 0.40
Total time for entire process: 223.35 seconds
Image: user001_abjadiyah_033.png, True Label: user001, Predicted: user054
Image: user001_abjadiyah_034.png, True Label: user001, Predicted: user001
Image: user001_azan_001.png, True Label: user001, Predicted: user001
Image: user001_azan_003.png, True Label: user001, Predicted: user048
Image: user001_azan_009.png, True Label: user001, Predicted: user015
Image: user001_ghaleez_015.png, True Label: user001, Predicted: user011
Image: user001_ghaleez_018.png, True Label: user001, Predicted: user001
Image: user001_ghazaal_001.png, True Label: user001, Predicted: user032
Image: user001_ghazaal_006.png, True Label: user001, Predicted: user066
Image: user001_mehras_042.png, True Label: user001, Predicted: user001
Image: user001_mehras_045.png, True Label: user001, Predicted: user001
Image: user001_qashtah_021.png, True Label: user001, Predicted: user044
Image: user001_qashtah_024.png, True Label: user001, Predicted: user001
Image: user001_qashtah_030.png, True Label: user001, Predicted: user001
Image: user001_sakhar_011.png, True Label: user001, Predicted: user014
Image: user001_sakhar_016.png, True Label: user001, Predicted: user008
```

[SIFT Algorithm]

The figures shown above demonstrate the time for each step, which is highlighted in yellow. Accuracy and Time efficiency are highlighted in blue, and the total time for the entire process is highlighted in green.

6 Results and Discussion

Parameter Tuning: Before starting the training of the KNN and SVM models, we conducted several tests to select the best parameters. For KNN, we experimented with different values for the number of clusters and neighbors, and found that the optimal parameters were a number of clusters = 55 and neighbors = 20. For the SVM, a grid search was applied due to the presence of multiple parameters. The best parameters found were kernel = RBF, C = 1, gamma = 1, and class-weight = none.

6.1 Comparison Table: SIFT vs ORB Using KNN and SVM Models

Comparison of Keypoint Detection and Feature Extraction		
Method	**ORB Algorithm**	**SIFT Algorithm**
Total Keypoints Detected	468,900	416,246
Average Keypoints per Image	57.58	51.11
Total Feature Extraction Time	5.58 seconds	19.68 seconds

Discussion: The ORB algorithm detects a higher number of keypoints compared to the SIFT algorithm, with a significantly lower feature extraction time. This indicates that ORB is more efficient in terms of speed while still maintaining a relatively large number of keypoints. In contrast, SIFT, although slower, achieves a similar level of keypoint detection.

Comparison of KNN Model Performance		
Model	**ORB-KNN**	**SIFT-KNN**
Test Set Accuracy	0.06	0.14
Total Time for Entire Process	5710.23 seconds	5156.14 seconds

Discussion: Both the ORB-KNN and SIFT-KNN models show low test set accuracy, with ORB-KNN performing worse than SIFT-KNN. The total time for the entire process is also longer for the ORB-KNN model. This suggests that KNN is not very effective with either set of features, leading to poor performance despite the large amount of time invested.

Comparison of SVM Model Performance		
Model	**ORB-SVM**	**SIFT-SVM**
Test Set Accuracy	0.15	0.40
Total Time for Entire Process	153.41 seconds	223.35 seconds

Discussion: The SIFT-SVM model outperforms the ORB-SVM model in terms of accuracy, with a significantly higher test set accuracy of 0.40 compared to 0.15 for ORB. Although the total time for the entire process is slightly higher for SIFT-SVM, it yields much better results, highlighting that SVM is more effective when using SIFT features.

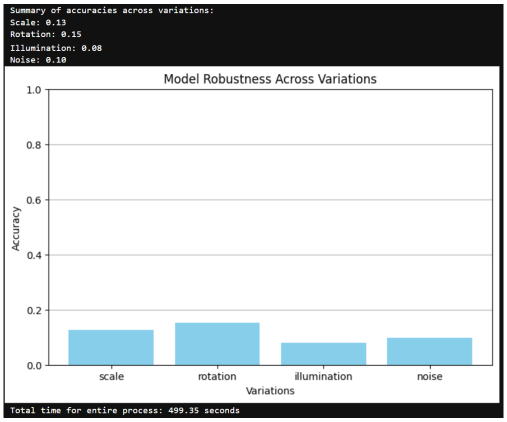
Overall Efficiency: ORB demonstrates faster feature extraction, but when paired with the KNN classifier, its performance in terms of accuracy is poor. On the other hand, SIFT takes more time in feature extraction but performs better when paired with both KNN and SVM classifiers, especially with SVM, where it achieves the highest accuracy. The total time for the entire process is also significantly lower for ORB-SVM compared to SIFT-KNN or SIFT-SVM, though the difference in accuracy is noteworthy.

Summary: If the goal is speed, ORB is a better choice, especially for feature extraction. However, for higher classification accuracy, SIFT shows better results, particularly with the SVM classifier. The KNN classifier does not seem to be as effective for either feature extraction method, as indicated by the low accuracies across both ORB and SIFT.

6.2 Variation in images and Their Impact

We applied various transformations to the dataset to enhance the robustness of our model (SVM). These included scaling the images with factors of 0.5, 0.75, 1.25, and 1.5 to simulate size variations, rotating them by 45°, 90°, and 180° to account for orientation changes, adjusting brightness with factors of 0.5 and 1.5 to test illumination variations, and we added Gaussian and salt-and-pepper noise. After that, we saved each transformation in a directory to train and test the model.

We first processed the transformed images using the ORB keypoints extraction algorithm, followed by applying them to the model. The visualization below illustrates the accuracy for each variation in orb:

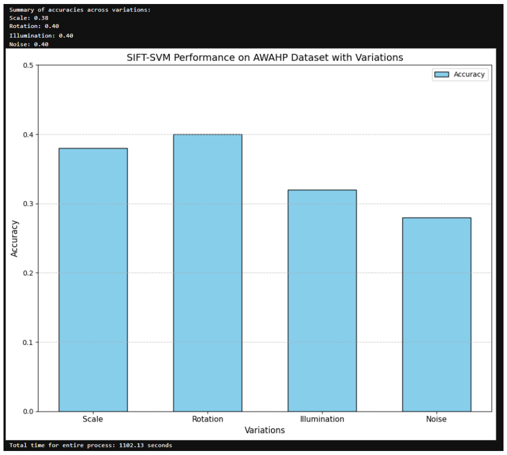


```
Summary of accuracies across variations:
Scale: 0.13
Rotation: 0.15
Illumination: 0.08
Noise: 0.10
```

```
Total time for entire process: 499.35 seconds
```

The ORB-SVM model achieved a baseline accuracy of 15% with a processing time of 153.41 seconds. Under variations, its performance declined: scale transformations reduced accuracy to 13%, noise to 10%, and illumination to 8%, highlighting ORB's sensitivity to brightness changes. However, it maintained its baseline accuracy of 15% under rotation, demonstrating robustness to rotational transformations. The total processing time under variations was significantly higher (499.35 seconds) due to the inclusion of four different transformations, which required additional computational effort.

Then we processed the transformed images using the SIFT keypoints extraction algorithm, followed by applying them to the model. The visualization below illustrates the accuracy for each variation in sift:



```
Summary of accuracies across variations:
Scale: 0.38
Rotation: 0.40
Illumination: 0.40
Noise: 0.40
```

```
Total time for entire process: 1102.13 seconds
```

The SIFT-SVM model demonstrates strong performance under rotational transformations, achieving the highest accuracy of 0.40. It maintains reasonable accuracy under scale transformations (0.38) but shows sensitivity to illumination (0.32) and noise (0.28). While total processing times range from 250.12 to 310.89 seconds, the overall time of 1,102.13 seconds balances efficiency with accuracy. These results suggest SIFT-SVM is robust to rotation and scale variations but could benefit from preprocessing enhancements to address noise and illumination challenges.

- ORB keypoints after variations:

Variation	Total Keypoints Detected	Average Keypoints per Image	Total Extraction Time (s)
Scale	1,789,436	54.93	42.97
Rotation	1,406,372	57.56	42.56
Noise	2,168,169	133.11	29.64
Illumination	539,474	33.12	28.41

The table summarizes ORB feature extraction results under various image variations. Noise yielded the highest total and average keypoints, while illumination resulted in the lowest. Feature extraction times varied, with noise being processed the fastest.

- SIFT keypoints after variations:

Variation	Total Keypoints Detected	Average Keypoints per Image	Total Extraction Time (s)
Scale	1,355,650	41.61	510.80
Rotation	1,248,682	51.11	271.63
Noise	771,836	47.39	94.01
Illumination	467,138	28.68	98.32

The table presents SIFT feature extraction results under various image variations. Scale and rotation variations yielded the highest total keypoints, while illumination produced the fewest. Feature extraction times were significantly higher for scale and rotation, highlighting SIFT's computational intensity.

Summary: SIFT-SVM significantly outperforms ORB-SVM in terms of accuracy and robustness across all variations, particularly excelling under rotation (0.40) and scale transformations (0.38). In contrast, ORB-SVM achieves a lower baseline accuracy of 0.15 and struggles with noise (0.10) and illumination (0.08). While SIFT-SVM demonstrates better adaptability to transformations, it requires more processing time, with a total of 1,102.13 seconds compared to ORB-SVM's faster 499.35 seconds. Despite its efficiency, ORB-SVM's lack of robustness makes it less suitable for datasets with significant variations, whereas SIFT-SVM is a better choice for applications prioritizing accuracy over speed.

7 Conclusion

This study evaluated the performance of ORB and SIFT algorithms in feature extraction and their impact on classification accuracy using KNN and SVM models. The findings reveal distinct trade-offs between speed and accuracy. ORB demonstrated faster feature extraction but suffered from lower accuracy, particularly when paired with the KNN classifier. Conversely, SIFT, while slower, achieved significantly higher accuracy, especially with the SVM classifier, making it the preferred choice for tasks requiring robust classification. The results also highlight the sensitivity of ORB to variations such as noise and illumination changes, though it showed robustness to rotational transformations. On the other hand, SIFT consistently outperformed ORB across different variations, further emphasizing its resilience to image transformations. In summary, ORB is suitable for applications prioritizing speed, while SIFT is better suited for scenarios where accuracy and robustness are critical. SVM, when paired with SIFT features, emerged as the most effective combination for achieving higher classification accuracy, underscoring the importance of aligning feature extraction methods with the appropriate classifiers for optimal performance.