FACULTY OF ENGINEERING AND TECHNOLOGY
DEPARTMENT OF COMPUTER ENGINEERING

Computer Vision
ENCS5343

Assignment#2

Arabic Handwritten Text Identification Using Local Feature Extraction Techniques

Prepared by: Shereen Ibdah – Amr Halahla
IDs: 1200373 – 1200902

Instructor: Dr. Aziz Qaroush

December 23, 2024

Introduction

Identifying handwritten text is a crucial aspect of computer vision, with applications in digital archiving, automated document processing, and linguistic analysis. This report explores the use of the Scale-Invariant Feature Transform (SIFT) algorithm for recognizing Arabic handwritten text by extracting and matching local features from image data.

Feature extraction is a pivotal step in computer vision, enabling the analysis of images by detecting key points invariant to transformations such as scale, rotation, and illumination. In handwritten Arabic text, the variability in individual writing styles presents a unique challenge, emphasizing the importance of robust algorithms for accurate recognition. This study focuses exclusively on the SIFT algorithm, evaluating its performance on the AHAWP dataset, comprising diverse samples of handwritten Arabic words by multiple individuals.

The analysis assesses the algorithm's performance based on metrics such as accuracy, time efficiency, robustness to variations, and the number of detected key points. By focusing on SIFT, this report aims to provide detailed insights into its strengths and limitations in the context of handwritten Arabic text recognition.

Data Set

The AHAWP dataset is used for this study, focusing on word-level Arabic handwriting. It includes 10 unique words handwritten by 82 individuals, with each contributing 10 samples per word, totaling 8,144 images. The dataset’s diversity in handwriting styles makes it ideal for evaluating the SIFT algorithm's accuracy, efficiency, and robustness in handwritten Arabic text recognition.

Methodology

The methodology outlines the step-by-step approach taken in this study, including feature extraction, feature representation, and training the classifier for handwritten text recognition.

Feature Extraction

Two primary algorithms, SIFT (Scale-Invariant Feature Transform) and ORB (Oriented FAST and Rotated BRIEF) were employed to extract key points and descriptors from images:

- **SIFT:** The SIFT algorithm detects and computes image key points and descriptors that are invariant to transformations like scale, rotation, and illumination. This ensures robust feature matching across varying conditions. Using the OpenCV library.

- **ORB:** The ORB algorithm combines the FAST key point detector and the BRIEF descriptor. It offers similar feature extraction capabilities while being computationally less intensive, making it suitable for real-time applications. Similar to SIFT, ORB was implemented in OpenCV.

Feature Representation

The Bag of Words (BoW) model is used to transform feature descriptors (obtained from SIFT and ORB) into fixed-length histograms, enabling traditional machine learning techniques, such as Support Vector Machines (SVM), to be applied for image classification tasks.

Training and Classification

Support Vector Machines (SVM) were utilized for classification using Bag of Words (BoW) histograms, where the BoW model converted the image feature descriptors (from SIFT and ORB) into fixed-length histograms. After performing feature scaling to standardize the data, hyperparameter tuning was carried out to identify the optimal model parameters. This tuning process involved experimenting with different kernel functions, regularization parameters, and other SVM settings to enhance classification performance. The resulting model was then evaluated for accuracy using standard metrics, such as precision, recall, and F1-score.

Results and visualizations

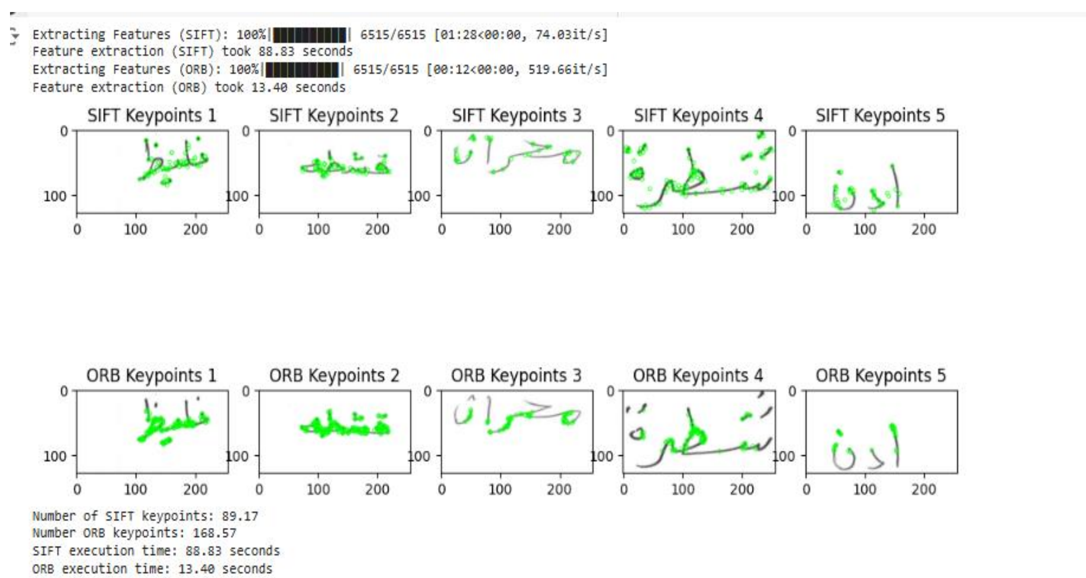
The dataset was successfully loaded and preprocessed for analysis. A total of **8,144 images** were loaded, each resized to uniform dimensions of **128x256 pixels**. These images represent handwritten text contributed by **82 unique users**, with user IDs serving as label

```
Number of images loaded: 8144
Image dimensions (after resizing): (128, 256)
Unique users (labels): 82
```

The dataset was divided into training and testing sets to ensure a robust evaluation of the classification model. The split allocated **80% of the data for training** and **20% for testing**:

- **Training Set Size**: 6,515 images and **Testing Set Size**: 1,629 images

The image shows the results of feature extraction:



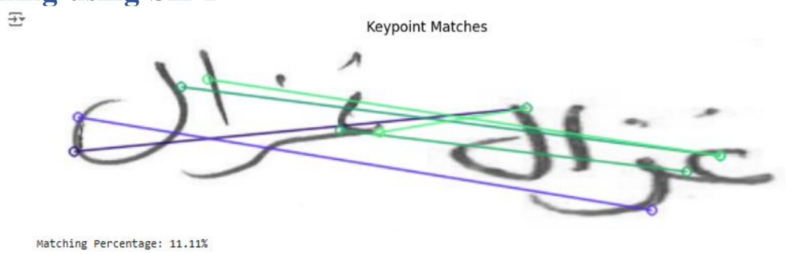
Summary of SIFT and ORB Variance:

- **SIFT**: Highly robust to **scale**, **rotation**, and **illumination**, making it suitable for detailed and complex handwriting recognition. It can handle variations in these conditions effectively, though it is more computationally intensive.

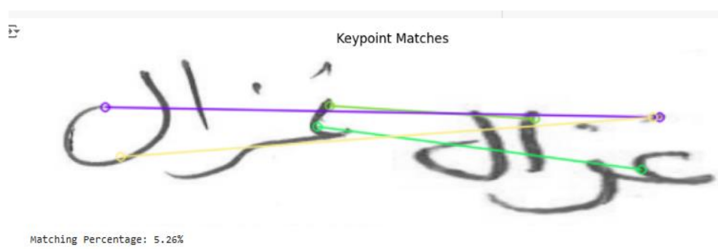
- **ORB**: Efficient and fast, ORB performs reasonably well with **scale** and **rotation**, but it is less robust to changes in **illumination** and **noise**, limiting its accuracy in more challenging conditions. It is ideal for simpler tasks or real-time applications.

Feature	SIFT	ORB
Key points Detected	**Fewer, but highly descriptive**	**More, with broader coverage**
Execution Time	**Slower, requiring more computational time**	**Faster, suitable for real-time tasks**
Precision	**High precision, robust to variations**	**Moderate precision, efficient**

Key point matching using SIFT



Key point matching using ORB



SIFT achieved a higher matching percentage compared to ORB, demonstrating greater precision and robustness to handwriting variations.

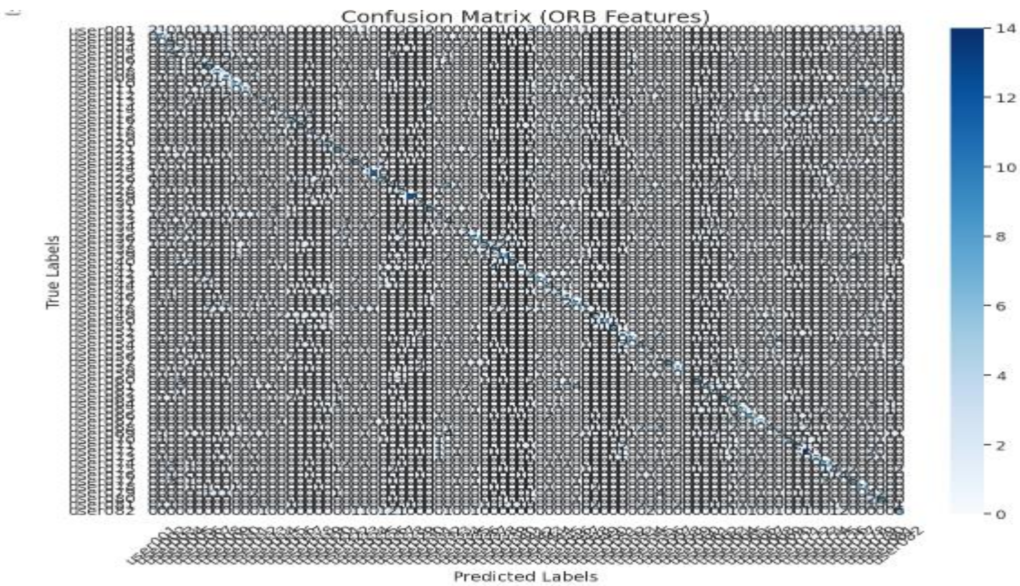
Training a Classifier using SIFT Features.

results of training the model by using the SIFT features:

```
Performing GridSearchCV for SIFT...
Fitting 3 folds for each of 12 candidates, totalling 36 fits
Best hyperparameters (SIFT): {'C': 10, 'gamma': 'auto', 'kernel': 'rbf'}
SIFT Test Accuracy: 38.92%
```

The tuning process optimized the SVM classifier with the best hyperparameters: C=10C = 10, Kernel = RBF, and Gamma = auto, achieving a test accuracy of 38.92%. The classification report shows an overall accuracy of 38%, with significant variability across 82 users. High-performing classes (e.g., user016, user067) demonstrate the model's ability to identify distinct handwriting patterns while low-performing ones (e.g., user006, user042) highlight challenges with ambiguous styles. The macro-average F1-score of 42% indicates balanced performance but difficulties with underrepresented classes.

Confusion matrix:



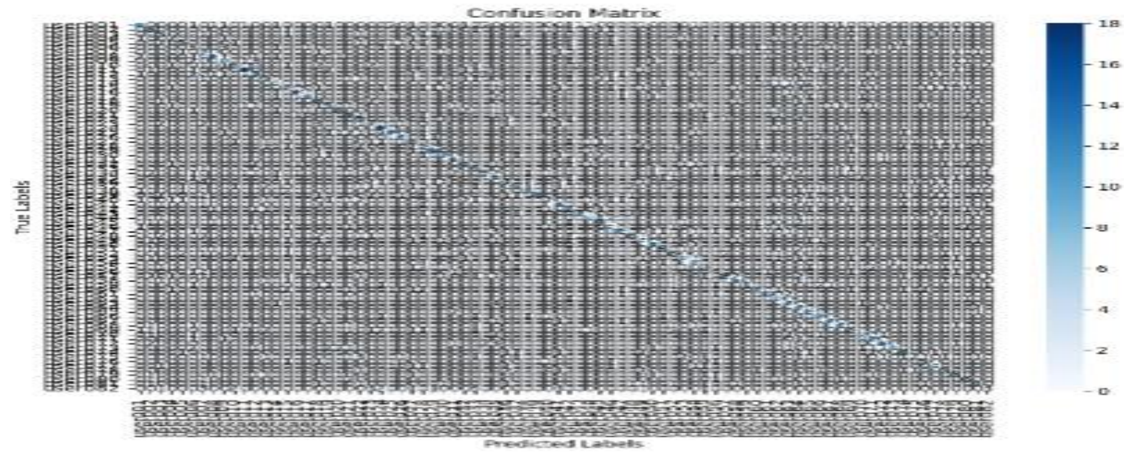
Training a Classifier using ORB Features.

The results of training the model by using the ORB features:

```
Performing GridSearchCV...
Fitting 3 folds for each of 12 candidates, totalling 36 fits
Best hyperparameters: {'C': 10, 'gamma': 'scale', 'kernel': 'rbf'}
Test Accuracy: 26.46%
```

The ORB-based classification achieved a test accuracy of **26.46%**, with highly variable performance across users. While some users (e.g., **user025**, **user029**) performed well, many classes showed low precision and recall, reflecting ORB's limitations in capturing fine-grained handwriting details. The macro and weighted F1-scores of **26%** indicate overall weak performance. ORB's efficiency is offset by its inability to handle complex handwriting styles effectively.

Confusion matrix:



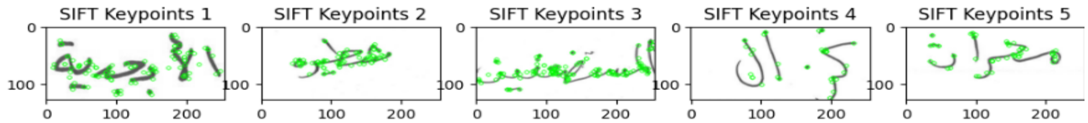
The SIFT-based model significantly outperformed the ORB-based model in terms of accuracy, F1 scores, and class-specific consistency. SIFT's ability to extract detailed and robust features made it better suited for Arabic handwriting classification. While ORB is faster, its lower accuracy and weaker performance suggest it is less suitable for tasks requiring fine-grained feature recognition.

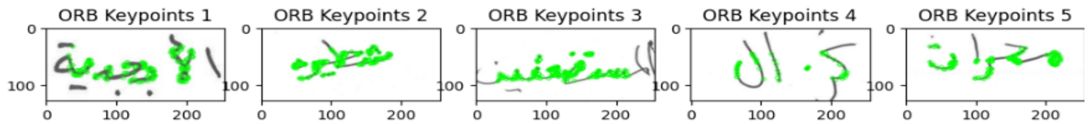
Data augmentation was applied to the training dataset to enhance its diversity and improve the model's generalization capabilities. The following transformations were used for each image:

1. **Rotation**: Each image was rotated by 15 degrees to account for variability in handwriting orientation.
2. **Noise and Illumination Adjustment**: Gaussian noise was added, and brightness/contrast was adjusted to simulate real-world variations such as poor lighting or noisy input.
3. **Scaling**: Images were scaled up by 20%, and then resized back to the original dimensions to introduce size variations.

The data augmentation process successfully expanded the original training dataset from **6515 images** to **26060 images**, quadrupling its size.

Key points after the Augmentation:






```
Average number of SIFT keypoints: 167.99
Average number of ORB keypoints: 201.19
SIFT execution time (Augmented Training): 54.22 seconds
ORB execution time (Augmented Training): 13.63 seconds
SIFT execution time (Test Data): 26.05 seconds
ORB execution time (Test Data): 0.78 seconds
```

SIFT offers better precision but is slower, whereas ORB is faster and computationally efficient.

- Model accuracy after augmentation for SIFT:

```
Performing GridSearchCV...
Fitting 3 folds for each of 12 candidates, totalling 36 fits
Best hyperparameters: {'C': 10, 'gamma': 'auto', 'kernel': 'rbf'}
Test Accuracy: 42.85%
```

- Model accuracy after augmentation for ORB:

```
Performing GridSearchCV for ORB...
Fitting 3 folds for each of 12 candidates, totalling 36 fits
Best hyperparameters (ORB): {'C': 10, 'gamma': 'auto', 'kernel': 'rbf'}
ORB Test Accuracy: 28.61%
```

After data augmentation, accuracy increased from 38% to 42% for SIFT and from 26% to 28% for ORB. Augmentation added variability through rotation, noise, and scaling, enhancing generalization. SIFT showed greater improvement due to its detailed feature capture, while ORB's gain was smaller, reflecting its efficiency but lower precision.

Conclusion

This study examined handwriting recognition using SIFT and ORB features with Support Vector Machines (SVM). SIFT showed greater accuracy and robustness, while ORB provided faster but less precise results. Data augmentation, through transformations like rotation and noise, improved the model's generalization and performance for both methods.