Faculty of Engineering and Technology
Electrical and Computer Engineering Department

COMPUTER VISION
ENCS5343

Assignment 2 Report
Arabic Handwritten Text Identification Using Local Feature Extraction Techniques

---

Student Name: Raghad Murad Buzia
Student #: 1212214

Instructor: Aziz Qaroush

Section: 2

Date: 23/12/2024

Table of Contents

List of Figures

List of Tables

Introduction and Objective

Handwritten text recognition has always been a significant challenge in the field of computer vision and pattern recognition. The diversity in writing styles, variability in penmanship, and the influence of factors such as noise and distortion make the task complex, especially when dealing with Arabic handwritten text. Arabic script adds an extra layer of complexity due to its cursive nature and context-dependent character forms.

This project aims to identifying Arabic handwritten text using state-of-the-art local feature extraction techniques and classification methods. Specifically, the Scale-Invariant Feature Transform (SIFT) and Oriented FAST and Rotated BRIEF (ORB) are used for feature extraction, while the Bag of Visual Words (BoVW) model serves as the backbone for feature representation. To classify the extracted features, we employ three powerful classifiers: K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machines (SVM).

The objectives of this project are as follows:

- Design a reliable identification pipeline: Build a scalable model that leverages local feature extraction and BoVW for Arabic handwritten text recognition.
- Evaluate classification methods: Compare the performance of KNN, Random Forest, and SVM classifiers to determine the most effective approach for this task.
- Assess robustness: Analyze the model's ability to handle variations, such as noise and reduced training data, to ensure its practical applicability.
- Provide comprehensive evaluation and visualization: Use metrics, visual graphs, and detailed comparisons to present the results and insights gained from the experiments.

This research enhances the field of handwritten text recognition, especially for Arabic script, by examining the efficiency of local feature extraction methods.

Background

Handwritten text recognition is a fundamental problem in computer vision and pattern recognition that has been actively researched for decades. Its applications span a wide range of fields, including document digitization, automated mail sorting, and educational tools. Recognizing Arabic handwritten text is particularly challenging due to its unique characteristics:

- Cursive Nature: Arabic script is inherently cursive, with letters often connected within a word.
- Context-Sensitive Letter Forms: Arabic letters change shape depending on their position (initial, medial, final, or isolated) in a word.
- High Variability: Variability in handwriting styles adds complexity to feature extraction and classification.

Feature Extraction Techniques

Scale-Invariant Feature Transform (SIFT):

Introduced by David Lowe in 1999, SIFT (Scale-Invariant Feature Transform) is a powerful algorithm for detecting and describing local features in images. It is robust to changes in scale, rotation, and illumination, making it particularly suitable for handwritten text recognition. SIFT operates by identifying keypoints in an image and describing their surrounding regions using histograms of gradient orientations.

Oriented FAST and Rotated BRIEF (ORB):

ORB (Oriented FAST and Rotated BRIEF) is a fast and efficient alternative to SIFT and SURF, combining the FAST keypoint detector with the BRIEF descriptor. It is designed to be computationally less expensive while maintaining good accuracy. Due to its speed and effectiveness, ORB is widely used in real-time applications, making it an ideal choice for scenarios where computational resources are limited and quick processing is essential.

Bag of Visual Words (BoVW)

The Bag of Visual Words (BoVW) model, inspired by the Bag of Words model in natural language processing, is a widely used approach for image representation. It involves several steps: extracting local features from images using techniques like SIFT or ORB, clustering these features using an algorithm such as K-Means to create a vocabulary of visual words, and then representing each image as a histogram of visual word occurrences. This method allows for the compact and scalable processing of image data, making it particularly ideal for tasks like handwritten text recognition, where efficient and accurate representation of visual information is crucial.

Classification Methods

Three classification algorithms are evaluated in this study:

- K-Nearest Neighbors (KNN): A simple yet effective non-parametric classifier that assigns labels based on the majority vote of nearest neighbors.
- Random Forest: An ensemble learning method that constructs multiple decision trees and outputs the majority class for classification.
- Support Vector Machines (SVM): A powerful classifier that finds the optimal hyperplane to separate data points in high-dimensional spaces.

Challenges in Arabic Handwritten Text Recognition

Complexity of Arabic Script: The cursive nature and contextual letter forms increase the difficulty of segmentation and feature extraction.

Data Variability: Handwriting styles vary significantly across individuals.

Noise and Distortions: Handwritten text often contains noise due to scanning errors or writing artifacts.

Structure

The implemented solution for Arabic handwritten text recognition follows a structured pipeline, ensuring clarity and modularity. The main stages are:

Data Loading and Preprocessing:

- Images from the AHAWP dataset are loaded, resized, normalized, and analyzed.

Feature Extraction:

- Keypoints and descriptors are extracted using two local feature extraction techniques:
  - **SIFT (Scale-Invariant Feature Transform):** Robust to scale and rotation.
  - **ORB (Oriented FAST and Rotated BRIEF):** Efficient and computationally fast.

Bag of Visual Words (BoVW):

- Descriptors are clustered using K-Means to form visual words.
- Images are represented as histograms of visual word occurrences.

Classification:

- BoVW histograms are classified using:
  - **KNN (K-Nearest Neighbors)**
  - **Random Forest**
  - **SVM (Support Vector Machines)**

Evaluation:

- Performance is assessed using metrics like accuracy, time efficiency, and robustness.
- Visualizations such as bar charts and confusion matrices summarize the results.

User Interaction:

- A user can input an image for classification, and the system predicts the label using pre-trained models.

Results and Discussion

Dataset Summary

The AHAWP dataset was loaded and analyzed. Below are the key statistics:

- **Total Images:** 8,144.
- **Total Users:** 82.
- **Images per User:** Most users had 100 images, except for user048 (94 images) and user059 (50 images).
- **Mean Word Count per User:** 1.00.
- **Standard Deviation:** 0.00 (consistent distribution across users).

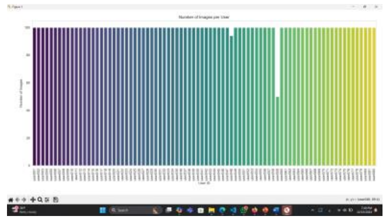

Figure 1: Dataset Visualization - Number of Images Per Users

The bar chart shows in `Figure 1` the distribution of images per user. It highlights the uniformity in data, with a few exceptions where certain users had fewer samples.

Feature Extraction and Matching Time

The time efficiency of feature extraction and matching for SIFT and ORB is summarized below:

Table 1: Feature Extraction and Matching Time Comparison

Algorithm	Feature Extraction Time (s)	Matching Time (s)	Total Time (s)
SIFT	51.44	13.70	65.14
ORB	9.96	12.04	22.00

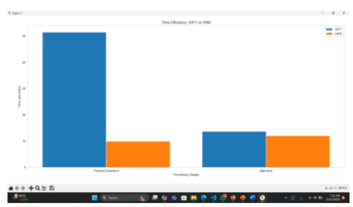

Figure 2: Feature Extraction and Matching Time - Time Efficiency

The ORB algorithm demonstrated superior time efficiency compared to SIFT, making it more suitable for applications requiring real-time performance.

Accuracy of Classifiers

Three classifiers were evaluated for both SIFT and ORB features using cross-validation:

- KNN (K-Nearest Neighbors)
- Random Forest
- SVM (Support Vector Machines)

Table 2: Accuracy of Classifiers Comparison

Model	Accuracy (SIFT)	Accuracy (ORB)

KNN	10.62%	7.44%
Random Forest	18.74%	9.36%
SVM	31.74%	20.22%

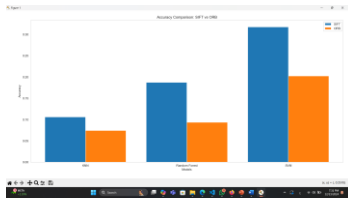

Figure 3: Accuracy of Classifiers - Accuracy Comparison

The SVM classifier achieved the highest accuracy for both SIFT and ORB features, highlighting its effectiveness in separating complex feature distributions.

User Input Classification

This step allows real-time testing of the trained models by classifying user-provided images. The system prompts the user to input an image path, processes the image (resizing and normalizing), and extracts features using either **SIFT** or **ORB**. The extracted descriptors are mapped to the pre-trained Bag of Visual Words (BoVW) model to generate a histogram, which is then classified using **KNN**, **Random Forest**, and **SVM**.

The classification results for user-provided images demonstrated varying performance across the tested classifiers and feature extraction methods. **SVM** consistently provided the most accurate predictions for both **SIFT** and **ORB**, successfully identifying the correct user ID in most cases. **Random Forest** showed reliable performance but was slightly less accurate than SVM, while **KNN** struggled with lower accuracy, especially for ORB features. These results highlight SVM's ability to handle the complex feature distributions of handwritten text. Additionally, ORB, while faster in feature extraction, exhibited more misclassifications compared to SIFT, reinforcing the trade-off between speed and accuracy.

Figure 4: User Input Classification Samples

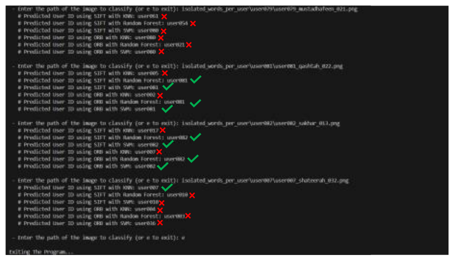

Figure 5: User Input Classification Samples

Conclusion

This project explored Arabic handwritten text identification using two feature extraction techniques, SIFT and ORB, coupled with the Bag of Visual Words (BoVW) model for feature representation. Three classifiers—KNN, Random Forest, and SVM—were evaluated to determine the most effective combination for recognizing user-specific handwriting patterns.

The results demonstrated that SIFT provided higher accuracy due to its robust feature detection, while ORB excelled in computational efficiency, making it suitable for real-time applications. Among the classifiers, SVM consistently outperformed others, showcasing its capability to handle the complex distributions of handwritten text features. However, Random Forest offered a balance between accuracy and speed, whereas KNN lagged in accuracy but maintained simplicity.

Overall, the project highlights the strengths and trade-offs between different techniques.