Writer Identification on Historical Documents Using Oriented Basic Image Features

Abdeljalil Gattal[1], Chawki Djeddi[1], Imran Siddiqi[2], Somaya Al-Maadeed[3]
[1] Department of Mathematics and Computer Science, Larbi Tebessi University, Tebessa, Algeria.
[2] Bahria University Islamabad, Pakistan
[3]Department of Computer Science and Engineering, Qatar University, Qatar

Abstract*—This study addresses the problem of identifying the authorship of historical manuscripts, a challenging task that offers interesting applications for document examiners and paleographers. We exploit handwriting texture as the discriminative attribute characterizing the writer of a given document. The textural information in handwriting is captured using a combination of oriented Basic Image Features (oBIFs) at different scales. Classification is carried out using a number of distance metrics which are combined to arrive at a final decision. A comprehensive series of experiments is carried out using different configurations of the oBIFs and the realized classification rates are compared with the state-of-the-art techniques on this problem.*

1. Overview

- The objective of our study is to find a combination of features that can achieve high classification rates on historical manuscripts.
- Features considered in our study include oriented **Basic Image Feature (oBIF) columns** .
- Leave-one-out experimental protocol.
- Experiments: Database employed in the ICDAR 2017 Historical Document.

2. Feature Extraction

oriented Basic Image Features (oBIFs):

- An extension to **the Basic Image Features (BIFs)**
- Represents an effective **textural information** descriptor
- Every location in the image is categorized into one of the seven local symmetry classes
- Classification is based on the response of a bank of six Derivative-of-Gaussian (DoG) filters (up to second order) of size determined by the scale parameter
- No orientation is assigned: If the location is attributed to the dark rotations, light rotational or the flat class
- ***n*** possible orientations can be assigned: the dark line on light, light line on dark and saddle-like classes
- ***2n*** possible orientations can be assigned: for the slope class.
- Dimension of feature vector is ***5n+3***.
- Orientation quantization parameter ***n=4***.
- A total of 23 entries in the oBIFs dictionary.
- We combine oBIFs at two different scales to produce the oBIF column features by ignoring the symmetry type flat. The oBIFs are generated using different values of the parameter (σ∈{1,2,4,8,16}) while the parameter ε is selected from a set of three small values of ε∈{0.1, 0.01,0.001}.

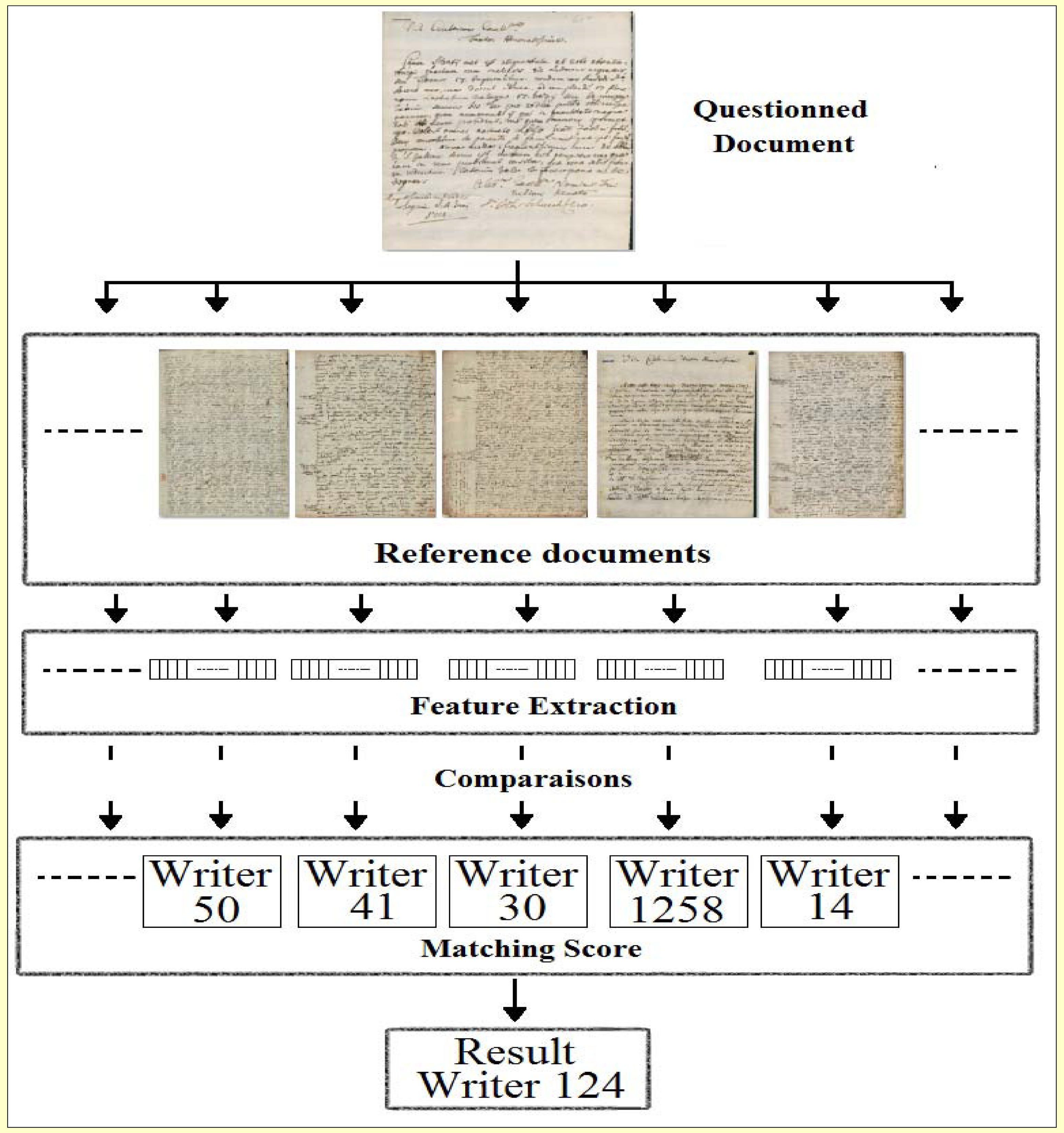

Fig 1: An overview of the matching scheme

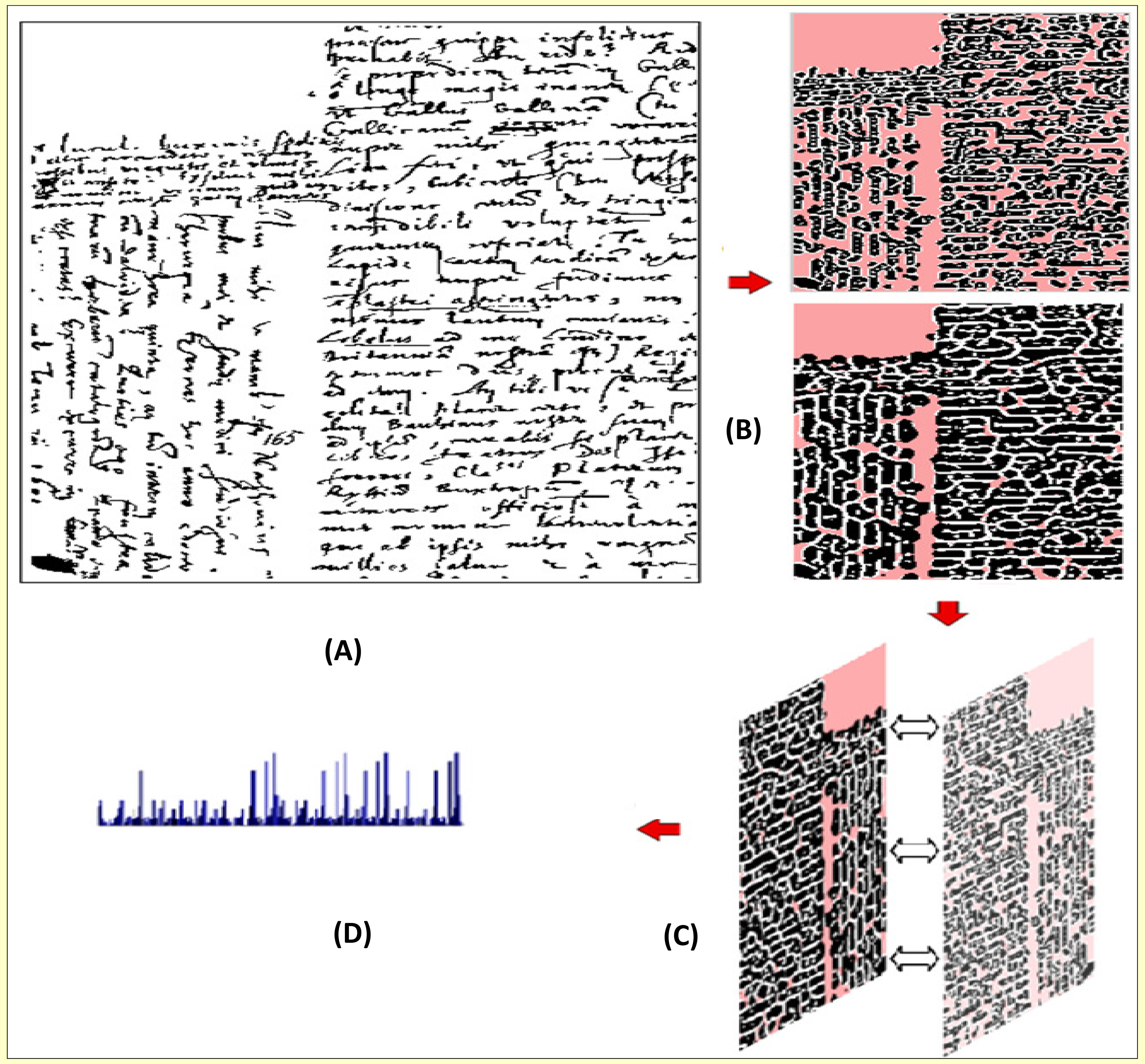

Fig 2: Different steps of the oBIF Column scheme (A) Original image (B) oBIFs computation for scale parameter σ=4 and σ=8 while ε =0.1 (C) The oBIFs at two scales are crossed to form columns at each location (D) Histogram is computed with non-flat columns.

3. Experimental Results

Dataset: ICDAR 2017 competition
Dataset Size: 3600 document samples contributed by720different writers (5 documents per writer)
Performance Measure: Top 1 Precision and the Mean Average Precision (mAP)

TABLE I: CLASSIFICATION PERFORMANCE WITH DIFFERENT CONFIGURATIONS OF OBIFS COLUMN FEATURES

Features	Parameters	Dim.	Classification Rate	
			Top1	mAP
f1	oBIFs column at **σ = 2** and **σ = 4 ; ε = 0.1**	484	73.42	52.19
f2	oBIFs column at **σ = 2** and **σ = 8 ; ε = 0.1**	484	72.36	51.11
f1+f2		968	75.53	54.14

TABLE II: CLASSIFICATION PERFORMANCE USING DIFFERENT DISTANCE METRICS

Features	Distance metrics		Classification Rate	
			Top1	mAP
f1	Eu	Euclidean Distance	73.42	52.19
	CB	City-Block Distance	74.28	52.36
	Corr	Correlation Distance	**74.75**	**53.42**
	Cos	Cosine Distance	73.94	53.25
	Sp	Spearman Distance	73.89	52.87
f2	Eu	Euclidean Distance	72.36	51.11
	CB	City-Block Distance	72.39	50.96
	Corr	Correlation Distance	73.11	52.71
	Cos	Cosine Distance	**75.03**	**53.80**
	Sp	Spearman Distance	72.19	51.68
f1+f2	Eu	Euclidean Distance	75.53	54.14
	CB	City-Block Distance	75.92	54.17
	Corr	Correlation Distance	75.86	55.42
	Cos	Cosine Distance	**76.58**	**55.65**
	Sp	Spearman Distance	75.42	54.72

TABLE III: CLASSIFICATION RATES FOR COMBINATION OF DIFFERENT METRICS

Combination schemes	Classification Rate	
	Top1	mAP
Prod-Min(*f1*.Corr, *f2*.Cos)	**77.39**	**56.82**
Prod-Min(*f1*.Cos, *f2*.Cos)	76.78	55.91
Prod-Min(*f1*.Corr, *f2*.Corr)	76.17	55.70
Prod-Min(*f1*.CB, *f2*.Cos)	76.31	55.24
Prod-Min(*f1f2*.Cos, *f2*.Cos)	75.92	55.10
Prod-Min(*f1f2*.Cos, *f1*.Corr)	76.89	56.29
Prod-Min(*f1f2*.Cos, *f1f2*.Corr)	77.28	56.63
Prod-Min(*f1f2*.CB, *f1f2*.Corr)	77.19	56.16
Prod-Min(*f1f2*.CB, *f1f2*.Sp)	76.94	55.84
Prod-Min(*f1f2*.Eu, *f1f2*.Sp)	76.89	55.94
Prod-Min(*f1f2*.CB, *f1f2*.Cos)	76.56	55.44
Prod-Min(*f1f2*.Eu, *f1f2*.Cos)	76.53	55.37

TABLE IV: CLASSIFICATION RATES FOR COMBINATION OF DIFFERENT METRICS

Rank	Method	Classification Rate	
		Top-1	mAP
1	**Proposed Method**	**77.39**	**56.82**
2	Tébessa II	76.4	55.6
3	Groningen	76.1	54.2
4	Tébessa I	74.4	52.5
5	Hamburg	67.1	46.9
6	Barcelona	67.0	45.9
7	Fribourg	47.8	30.7

4. Conclusion

- In this study, we employed different configurations of oBIF columns to capture the textural information from handwriting for writer identification purposes.
- For future extensions of the work, we intend to investigate the applicability of other textural measures to characterize writer from handwriting.
- Furthermore, investigation of feature selection techniques to identify the most appropriate textural descriptors for this problem is also planned.