ARABIC MANUSCRIPTS ANALYSIS AND RETRIEVAL

BY

Rashad Ahmed Abdullah Othman

A Dissertation Presented to the

DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

In

COMPUTER SCIENCE AND ENGINEERING

May 2015

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **RASHAD AHMED ABDULLAH OTHMAN** under the direction of his thesis advisor and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE AND ENGINEERING**.

Dissertation Committee

Prof. Sabri Mahmoud (Advisor)

Dr. Wasfi Alkhatib (Co-advisor)

Prof. Radwan Abdel Aal (Member)

Prof. Moustafa Elshafei (Member)

Dr. Mohammad Alshayeb (Member)

Dr. Adel F. Ahmed
Department Chairman

Dr. Salam A. Zummo
Dean of Graduate Studies

16/6/15
Date

Dedicated to

my country
Yemen

ACKNOWLEDGMENTS

First and foremost thanks to Allah who gave me strength, patience and ability to accomplish this research. Peace and blessing of Allah be upon his last messenger Mohammed, who guided us to the right path.

My deep appreciation and heartfelt gratitude goes to my thesis advisor **Prof. Sabri Mahmoud**, for his constant endeavour, guidance and the numerous moments of attention he devoted throughout the course of this research work. His continuous support and encouragement can never be forgotten. His valuable suggestions made this work interesting and knowledgeable for me.

I extend my deepest gratitude to my thesis co-advisor **Dr. Wasfi Alkhatib** for his support and valuable advice. Working with him in a friendly and motivating environment was really a joyful and learning experience.

Deep thanks to my thesis committee members **Prof. Radwan Abdel Aal, Prof. Moustafa Elshafei** and **Dr. Mohammad Alshayeb** for their suggestions and valuable comments. It was surely an honor to work with them.

I am also grateful to all faculty members for their encouragement and for imparting quality learning and knowledge with their valuable support and able guidance that has led my way through this point of undertaking my research work.

I also acknowledge the assistance of to complete this work especially for my brothers Dhiaa Musleh and Mohammed Al-Sayo'ony.

I owe thanks to my colleagues and my friends for their help, motivation and pivotal support. A few of them are Bashar Alhozaim, Talal Farhan, Hasan Alshehab, Ahmed Alshamiry, Fathi Alhammadi, Fahd Alhaidary and many others; all of whom I will not be able to name here. Special thanks to my brother Ahmed Al-Shaaby.

My heartfelt gratitude is given to my beloved father, mother, wife and my children, who always support me with their love, patience, encouragement and constant prayers. I would like to thank brothers, sisters, and all members of my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPREVIATIONS

AF Autocorrelation Feature

ANNs Artificial Neural Networks

AVD Area Voronoi Diagram

BCCs Basic Connected Components

BCN Barcelona Cathedral

BoF Bag of Features

BoWFs Bag of Word Fragments

BSM Blurred Shape Model

C-HMM Continuous HMM

CCHF Chain-Code Histogram Features

CS Correlation Similarity

DBH Distance-based Hashing

Docstrum Document Spectrum

DTW Dynamic Time Wrapping

ED Euclidean Distance

FRRH Fill Ratio and Run-length Histograms

GF Gabor Features

GHI Geometric Hashing Indexing

GHT Generalized Hough Transform

GSC Gradient, Structural and Concavity features

GW George Washington database

HKS Heat Kernel Signature

HMM Hidden Markov Model

HOG Histogram of Oriented Gradients

KNN K-Nearest Neighbor

LGH Local gradient histogram

LLC Locality Constrained Linear Coding

LP Lord Byron

LPI Locality Preserving Indexing

LSH Locality Sensitive Hashing

LSI Latent Semantic Indexing

NN Nearest Neighbor

OCR Optical Character Recognition

PAW Part of Arabic Word

QBE Query-By-Example

QBS Query-By-String

QE Query Expansion

RBF Radial Basis Function

RDA Regularized Discriminant Analysis

RLSA Run-Length Smearing Algorithm

RIFH Relational Invariant Feature Histograms

SC-HMM Semi-Continuous Hidden Markov Models

SCD Shape Context Descriptors

semi-CRFs Semi-Markov Conditional Random Fields

SF Spatial Features

SIFT Scale-Invariant Feature Transform

SPM Spatial Pyramid Matching

SS Shape Signature

SOM Self Organizing Map

SVM Support Vector Machines

TTFH Tamura Texture Features Histogram

VD Voronoi Diagram

VE Voronoi Edges

WVD Wigner-Ville distribution

THESIS ABSTRACT

NAME: Rashad Ahmed Abdullah othman

TITLE OF STUDY: Arabic Manuscripts Analysis and Retrieval

MAJOR FIELD: Computer Science and Engineering

DATE OF DEGREE: May 2015

Recent years have witnessed a dramatic growth of the amount of manuscripts that are preserved, processed and accessed in digital form. That creates the need for efficient analysis and retrieval techniques in order to extract the relevant information contained in these manuscript images. In practice, the problem of analyzing historical document images is closely tied to the problem of text region extraction. In this work, we address this problem by developing a new method for text region extraction. Similarly, the problem of retrieving document images is closely tied to the problem of image retrieval. Recent researches employ either global or local feature extraction approaches. Global features are affected by handwriting variability and variations. Similarly, local features focus on particular parts of the words and ignore others. In our effort to address this point we used a med-level feature representation model called Bag of Word Fragments (BoWFs),

that makes use of different parts of the word, such as contour and skeleton, and describes their local parts. The proposed model, evaluated on printed as well as historical documents, exhibited promising results. In particular, it achieved 99.20% in terms of precision when recall equals 1 for printed dataset and 89.60% in terms of precision when recall equals 0.5 for historical dataset written by one writer. We also evaluated the performance of the proposed model on a dataset written by two writers. We found that the proposed model is writer dependent.

ملخص الرسالة

الاسم الكامل: رشاد أحمد عبدالله عثمان

عنوان الرسالة: تحليل واسترجاع محتوى المخطوطات العربية

التخصص: علوم وهندسة الحاسوب

تاريخ الدرجة العلمية: ما يو 2105

شهدت السنوات الأخيرة زيادة هائلة في عدد المخطوطات التي يتم حفظها ومعالجتها إلكترونيا والذي بدوره أدى إلى ضرورة ابتكار أساليب كفوءة للتحليل والاسترجاع. تحتوي المخطوطت العربية بدورها على نصوص رئيسية وحواشي و أشكال رسومية وزخارف. يقصد بمشكلة تحليل المخطوطات استخلاص الأجزاء المختلفة من المخطوطة وتصنيفها بحسب النوع. كما يقصد بمشكلة استرجاع محتوى المخطوطة استرجاع المخطوطات التي تحتوي على نصوص معينة. تم في الدراسات السابقة استخدام العديد من المميزات العامة والخاصة لهذا الغرض. المميزات الخاصة تتاثر بالتغيير والاختلاف في الكتابة. والمميزات الخاصة تركز على أجزاء محددة من الكتابة.

تم في هذه الرسالة تطوير طريقة جديدة لاستخلاص النصوص الرئيسية والحواشي الموجودة في المخطوطات العربية. وتعتمد هذه الطريقة على اتجاه النص ، كثافة النص و وجود الفراغات بين أجزاء المخطوطة. كما تم تطوير طريقتين جديدتين لايجاد اتجاه النص. تعتمد الطريقتان على حساب تكرار اتجاهات مقاطع الكلمات.

كما تم تطوير سمة جديدة تسمى حقيبة أجزاء الكلمات ((Bag of Word Fragments (BoWFs). تعتمد هذه السمة على أجزاء مختلفة من الكلمات مثل الإطار الخارجي للكلمة (Contour) والإطار الداخلي للكلمة (Skeleton). يتم في هذه الطريقة تقسيم مقاطع الكلمة العربية إلى أجزاء صغيرة للكلمة (Word Fragments). ثم يتم توصيف هذه الأجزاء باستخدام سمات سياق الشكل (Shape Context Descriptors) و تكرار الاتجاهات (Histogram of Gradient).

تم اختبار طريقة استخلاص النصوص والحواشي باستخدام مجموعة من المخطوطات العربية وحققت الطريقة دقة تصل إلى حوالي 95% . كما تم اختبار السمة الجديدة باستخدام مجموعة من المخطوطات عربية بالإضافة الى الوثائق المطبوعة وحققت الطريقة نتائج متميزة للوثائق المطبوعة بمتوسط دقة يصل إلى 99.20% عندما يكون نسبة الاستراجاع 100% وجيدة للمخطوطات بمتوسط دقة يصل إلى 89.60% عندما يكون نسبة الاسترجاع 50%.

CHAPTER 1

INTRODUCTION

This thesis addresses historical documents analysis and retrieval, specifically Arabic historical documents. Intuitively, retrieval is the task of retrieving documents of a dataset which match a given user input keywords. In the literature, historical documents retrieval has been an active research area. Significant progress has been made. However, this research work has been mostly concerned with non Arabic historical documents. Historical documents retrieval has become a focus of interest for research due to the great challenges that are faced in the retrieval of historical documents. In the last two decades, word spotting has witnessed great interest as an emerging technology for document image retrieval applications. It is the task of locating specific keywords in a collection of document images. It has been introduced as an alternative to Optical Character Recognition (OCR) techniques for historical documents retrieval, since there is no efficient OCR designed for old historical documents. It is a challenging task due to infinite variations in handwriting styles. Word spotting locates keywords in a collection of document images by comparing the features that were extracted from the word images. It includes a matching process between a given keyword image and a collection of word images stored in the database using the extracted features. Most of the existing word spotting based systems locate keywords in documents as follows: firstly, they build indexes based on low level features that are extracted from word images. The corresponding features are also extracted from the keyword image. Second, they search through the whole database and measure the similarity between each stored word image and the keyword image. Finally the results are returned in a sorted order of the similarity matching level.

In general, the problem of retrieving document images is closely tied to the problem of image retrieval. In the past, several models have been proposed for image retrieval. Among these models is Bag of Features (BoF) model which is widely applied in various areas of computer vision. In this research work, we will make use of this technique to aid in the retrieval task.

In addition to historical document retrieval, we address Arabic historical documents layout analysis. Document analysis is one of the fundamental problems in document understanding and is now used in many practical applications such as OCR and word spotting. Layout analysis is the problem of extracting regions or objects of interest from document images. In the last decades, several works have been proposed for printed document analysis, especially for Latin documents. Although the proposed methods have achieved good results for machine printed document analysis, the state of the art of historical document analysis falls far behind, specifically for Arabic historical documents. Thus more work remains to be done in Arabic historical document analysis. Our aim in this thesis is to reduce this gap in research in Arabic historical document analysis.

1.1 Problem Statement

The problem to be addressed in this work can be stated as follows: the extraction of text regions of Arabic historical documents and the construction of a suitable representation for them. Then, given a keyword image, locate all its occurances in the collection of documents using this representation.

1.2 Motivation

Due to advances in information technology and communication, recent years have witnessed a dramatic growth in the amount of historical documents that are preserved, processed and accessed in digital form. They are valuable resources for scholars so their contents can be made available via the internet or other electronic media. The main problem is that such contents are only available in image formats, which makes them difficult to access. In this case, document image analysis and retrieval techniques can be used to extract the textual information from the digitized historical documents and make this information accessible to users. The state-of-the-art in Arabic historical documents analysis and retrieval falls far behind compared to Latin and other scripts. This thesis aims to fill this gap and to conduct research in the field of Arabic historical documents analysis and retrieval.

1.3 Thesis Objectives

The aim of this thesis work is to perform analysis and retrieval for Arabic historical documents. The first aim is to perform analysis to extract text regions such as the main body text and the side notes from these documents. The second objective is to develop a feature representation for Arabic historical documents retrieval. Using this representation, one can search for a given keyword image and locate its occurrences in a collection of historical documents.

1.4 Thesis Contributions

The main contributions of this thesis are

1. A literature survey of word spotting research for handwritten documents.

2. A text region analysis and extraction method. The proposed method is based on the fusion of background, foreground and window-based text analysis. The fusion of these techniques for text region extraction results in better segmentation results. The proposed method can be applied on a large variety of historical documents that have different layout structures. Furthermore, Introduction of the use of histograms for text orientation estimation. Two histogram based text orientation estimation methods are proposed. The first one is connected components based while the second one is pixel-based.

3. A novel feature representation appraoch called BoWFs, for Arabic historical documents word spotting. The proposed approach is inspired by the traditional BoF model. Moreover, Three methods for word fragment extractions (viz. contour-based, skeleton-based and batch-based methods) are proposed. The proposed methods make use of differen parts of the word such as the contour and the skeleton to build the BoWFs model.

4. The collected documents and their ground truth.

1.5 Thesis Outline

The remainder of this thesis is organized as follows. In the next chapter, we present background on historical document analysis. Next, we review the related work and summarize different features and methods for word spotting in Chapter 3. A survey of word spotting techniques of handwritten documents is also included. Chapter 4 addresses layout analysis of Arabic historical documents. A novel approach is introduced to extract text regions. In addition, this chapter presents two methods for text orientation estimation. In chapter 5, we address word spotting of Arabic historical documents. The BoWFs model is introduced and three methods for word fragment extraction are proposed. Finally, Chapter 6 concludes this thesis and suggests future work.

CHAPTER 2

BACKGROUND ON DOCUMENT LAYOUT ANALYSIS

This chapter presents general background on document image analysis. A framework for document image analysis is presented in Section 2.1. Section 2.2 presents challenges in Arabic historical document analysis. Then, we present the state of the art in document layout analysis. We begin with traditional page segmentation techniques proposed for printed documents in Section 2.3 , followed by a discussion of different zone classification approaches in Section 2.4. Finally, techniques proposed for historical document analysis are presented in Section 2.5.

2.1 Document Analysis Framework

A general framework for document analysis is shown Figure 2.1. It consists of two main tasks that can be subdivided into subtasks. The first task is the preprocessing. In this task, document images are binarized and noise is removed. The input for this task is a scanned document image and the output is a denoised binary

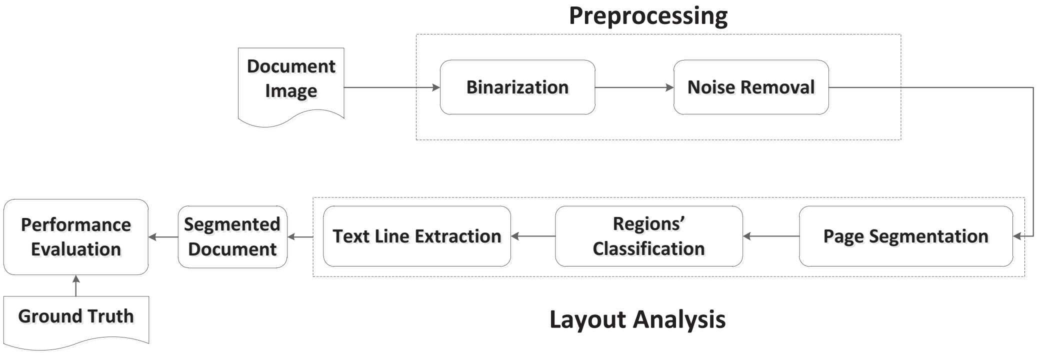


Figure 2.1: Document layout analysis framework

document image. The second main task is the layout analysis task. This task is subdivided into the following subtasks: page segmentation, region classification and text-line extraction. The subtasks are summarized as follows

- **Preprocessing**

  Generally, the preprocessing stage in document analysis includes the following tasks: binarization and noise removal.

  - **Binarization** is used to convert gray scale manuscript images into black and white. Most of developed techniques use connected components which can be extracted from binary images. Generally, Binary images contain noise.
  - **Noise Removal** is the process of removing noise from an image. Different analysis algorithms apply preprocessing tasks in different order

and some algorithms might skip one or more of these tasks. However, layout analysis algorithms use these preprocessing steps in some form.

- **Page Segmentation**

  Page segmentation is the process that identifies and localizes regions within document images. The importance of this step originates from the fact that it is an essential step used in the process of document indexing, retrieval and understanding. It is a difficult task; the difficulty quickly increases when the documents have complex structure or suffer from noise and other artifacts.

- **Region Classification:**

  Zone classification is the process that identifies the types of extracted regions such as main text, side notes and decorative elements, etc.

- **Text Line Extraction:**

  Text line extraction divides text regions into separated lines of text. It is an essential step in most document understanding tasks such as text recognition and word spotting.

2.2 Challenges in Arabic Historical Documents Analysis

A large number of Arabic historical document images are available in digital form. These images are scanned from collection of historical documents in libraries or archives. In general, the structure of Arabic historical document can be divided into two classes: simple structure and complex structures. Samples of historical documents having simple and complex structure are shown in Figures 2.2 and 2.3, respectively. Historical documents may have various types of degradations including noise, uneven background, aging effects, degraded text regions, uneven illumination and local changes or non-uniformity in foreground and background colors, etc. Historical documents present the following types of challenges:

- Contain several interleaved text regions, Figure 2.3a.
- Contain text regions having several text directions, Figure 2.3.
- Have degraded text regions, Figure 2.3b.
- Have background drawing, Figure 2.2.

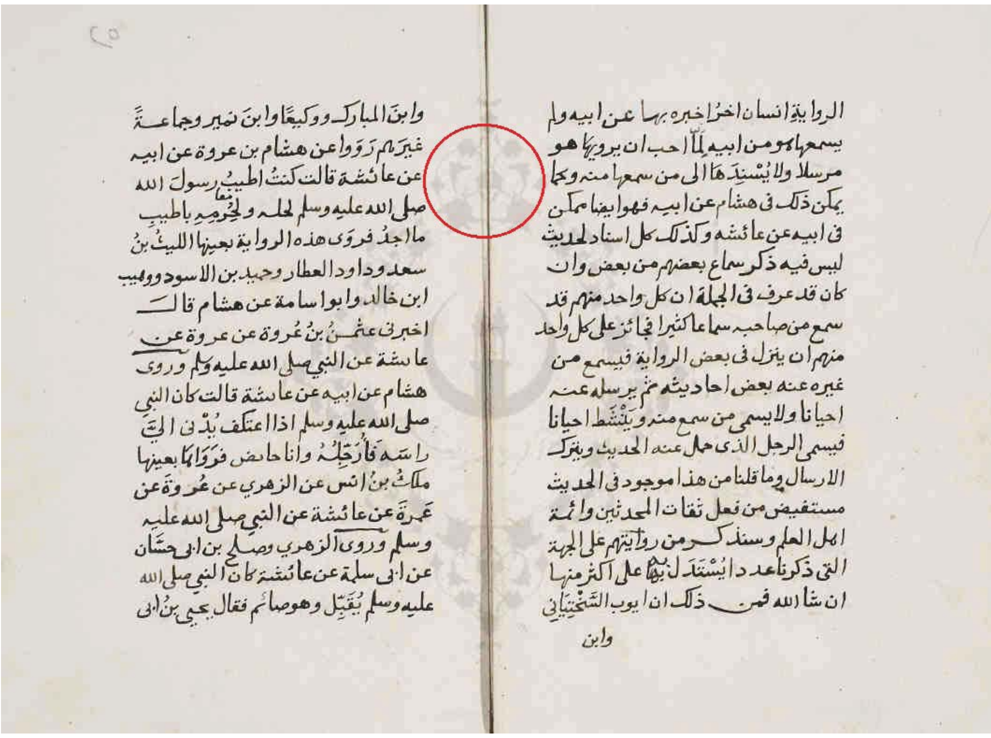


Figure 2.2: Arabic historical document with simple structure, containing background drawing (circled)

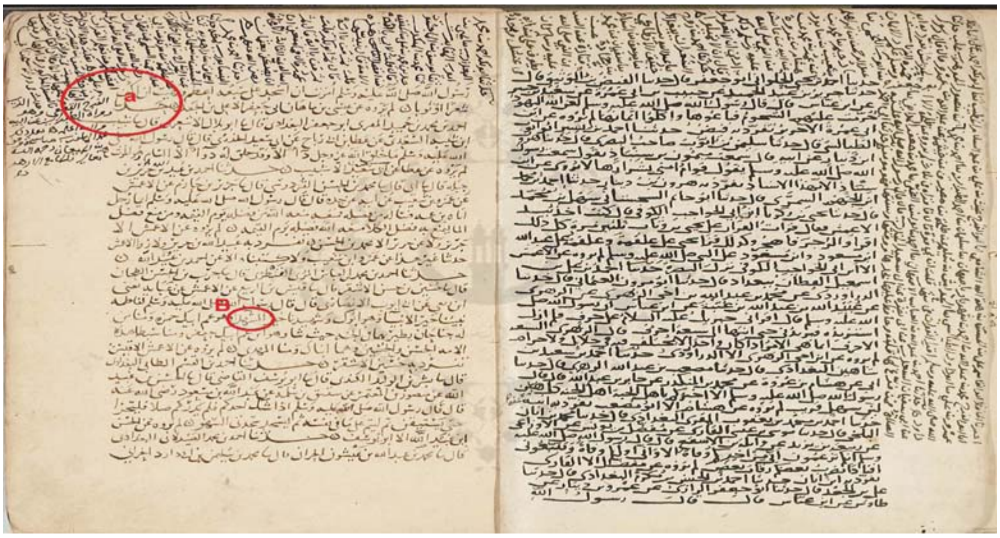


Figure 2.3: Arabic historical document with complex structure

2.3 Traditional Page Segmentation

Accurate page segmentation from document images is an essential step in document image understanding, and hence has attracted research attention. In the literature, there is a significant amount of work concerning printed document segmentations. However, due to the existence of variation and layout variability in handwritten documents, especially in historical documents, traditional segmentation algorithms proposed for printed documents may not work for handwritten documents' segmentations. Page segmentation strategies can be classified into top down, bottom up and hybrid approaches.

2.3.1 Top-Down Page Segmentation

This group of methods uses prior knowledge or a document model aiming at iteratively dividing the whole document image into smaller regions. These methods are fast. However, they are applicable for documents with well-defined layout. Projection based methods such X-Y Cut algorithm [1] and Antonacopoulos method [2] are examples of this group. X-Y Cut algorithm is a tree-based algorithm that divides the document image into regions by analyzing the projection histogram profile. The entire document is represented by the root of the tree. Each leaf node of the tree represents one segmented region. The intermediate nodes represent the regions obtained during the execution of the algorithm.

2.3.2 Bottom-Up Page Segmentation

This group of algorithms are based on low-level elements such as pixels or connected components of the image [3, 4, 5]. Majority of algorithms in this category work at the connected components level [4, 5]. Connected components are grouped into bigger elements such as lines and regions based on geometric relationship between neighboring connected components such as distance and overlap. The methods in this group are able to handle skew and arbitrary documents layout.

The pixel-based Run-Length Smearing Algorithm (RLSA) proposed in [3] links together neighbouring foreground pixels in a binary image, horizontally and vertically. Those foreground pixels are separated by less than a predefined number of background pixels. It yields two different bitmaps which in turn are combined in one bitmap image. Then, document regions are obtained by connected components analysis of the new bitmap image.

Another bottom-up algorithm is the Document Spectrum (Docstrum) proposed by O'Gorman [4]. Page segmentation is accomplished by nearest neighbor clustering of the extracted connected components obtained from the document image. The algorithm computes a histogram of distance and angles for each connected components using its k-nearest neighbors. The histogram is used to merge connected components into lines which in turn are grouped into blocks.

A graph based segmentation algorithm is proposed in [5] aiming at solving the problems that arise from the variability of the intra and inter-space of the regions. It starts by extracting points from connected components' boundaries which are used to construct a Voronoi Diagram (VD). Then, an Area Voronoi Diagram (AVD) is obtained by deleting some edges in the constructed VD. Finally, document regions are obtained by removing unnecessary Voronoi Edges (VE) of the AVD.

2.3.3 Hybrid and Other Methods

Top-down approaches and bottom-up approaches have been combined to deliver better segmentation performance [6, 7]. Wang and Srihari in [6] proposed to use a bottom-up RLSA algorithm to detect text and non-text regions followed by top-down X-Y Cut algorithm to merge the separated text lines into blocks. This method is applicable only to rectangular-layout documents.

An efficient split-merge based algorithm is proposed in [7]. The algorithm iteratively divides the document into subregions at the same time smaller regions are merged to form larger regions. This method is applicable to arbitrary layout documents.

2.4 Region Classification

Region classification is the process of identifying the classes of regions segmented from document images. It is important for applying proper domain-specific analysis approach on each region. For example, text regions may be prepared for OCR for further text recognition. Many algorithms for regions' classification are proposed in the literature [8 - 12]. These methods focus on identifying text and non-text regions in a document image via extracted features of these regions. Several features such as run length mean and variance, Autocorrelation Feature (AF) and Spatial Features (SF) are used in [8, 9]. Other features such as Tamura Texture Features Histogram (TTFH), Relational Invariant Feature Histograms (RIFH) and Fill Ratio and Run-length Histograms (FRRH) are used in [10].

Jiming Liu et al. [11] proposed to use Gabor Features (GF) and Chain-Code Histogram Features (CCHF). Different classifiers such as Artificial Neural Networks (ANNs) [12], Support Vector Machines (SVM) [11], Nearest Neighbor (NN) [10] classifiers are used for region classification.

2.5 Historical Documents Layout Analysis

In this section, related work concerning ancient historical documents is given. In [13], a system for meta-data retrieval from ancient manuscript image is proposed.

The system aims to assist researchers and historians to analyze digitized medieval manuscripts. The system starts by extracting some features of each connected component in the image. Then, connected components are grouped into larger elements and classified with K-Nearest Neighbor (KNN) classifier.

Text/non-text classification method for ancient printed documents is proposed in [14, 15]. Documents are segmented into zones and significant orientations are analyzed with a directional rose for each zone. The proposed method was qualitatively tested on 100 pages from five different books of the Renaissance and the reported accuracy was 95% and 88% for text and graphical parts respectively.

Ramel et al. [16] proposed a user-driven layout analysis algorithm for historical printed books. The algorithm is based on two maps: foreground and background maps. Foreground map is composed of all connected components extracted from the document image. Background map is created such that each pixel in the image is associated with the summation of the number of successive white pixels in the horizontal direction and the number of successive white pixels in the vertical direction. Connected components in the foreground map are merged into larger blocks if they are close to each other and the area between their centres of gravity is less than a specific threshold in the background map. For evaluation, the algorithm is applied to different datasets and they reported 93% accurate segmentation results.

The work of Garz et al. [17] is based on Scale-Invariant Feature Transform (SIFT) [18]. They considered layout elements such as text and decorative elements such as headlines and initials. Experimental results show that the proposed method has the ability to detect regular text area in ancient historical document images.

In Ouwayed et, al. [19] an approach for multi-oriented text line extraction of handwritten documents is presented. The local direction of the text is determined using an image meshing and the Wigner-Ville distribution (WVD) [20] on the projection histogram profile. Once these local orientations are determined, they are extended to neighbors having similar orientation. This operation is applied to all document meshes. Then, zones are constructed and text lines are extracted in each zone based on the follow-up of the lines orientation and the proximity of connected components. Finally, overlapping and touching connected components are separated. The proposed approach has been evaluated on 100 documents and it reported an accuracy of 98.6%.

Machine learning based approach for Arabic historical documents layout analysis has been proposed in [21]. In this work, layout analysis is formulated as a classification problem in which the historical document contents are classified into two distinct groups, main text and side notes. Evaluation is carried out using 38 documents. The results show that about 95% of main-text are classified correctly.

Nikolaou et al. in [22] used adaptive run-length smoothing and skeleton segmentation. They tried to overcome two problems that exist in machine-printed historical documents, complex and dense layout and neighboring text columns or text lines. The evaluation is carried out using a dataset containing English, Roman Greek and French documents. The reported results show that the proposed approach achieved an F-measure of 84.5% for text line segmentation. Table 2.1 summarizes the proposed methods for historical document analysis.

Table 2.1: Analysis techniques for historical documents

Reference	**Technique**	**Purpose**	**Disadvantages**
Ramel et al. [16]	Background and foreground map	Regions extraction	Requires presence of space between regions
Garz et al. [17]	SIFT	Regions extraction	Only differentiate between text and non-text regions
Nikolaon et al. [22]	Adaptive run-length	Regions extraction	Limited to old printed documents
Ouwayed et al. [19]	Split-merge	Text line extraction	Fails when two neighboring text regions have the same orientation
Bukhari et al. [21]	Classification	Main text and side notes	Requires training process

CHAPTER 3

HANDWRITTEN DOCUMENT WORD SPOTTING: STATE OF THE ART

The explosive growth of the amount of handwritten documents that are scanned, processed and accessed in digital form, has lead to great interest in word spotting techniques attracting researchers from various research communities, such as pattern recognition, computer vision and information retrieval. Word spotting has been an active research area and significant progress has been made in the last few years. However, current word spotting techniques have not achieved acceptable performance on real-world handwritten documents that vary widely in writing style and quality. This chapter gives an overview of published research efforts in the area of word spotting and on the methods used in the field. We first start by describing a general model for document word spotting followed by discussing present challenges in handwritten document word spotting. Then the used databases for handwritten document word spotting and other handwritten text tasks are discussed. Next, research works on handwritten document word spotting are presented. Finally, several summary tables of published research work are provided for used handwritten documents databases and reported performance results on handwritten documents word spotting. These tables summarize the different aspects and reported accuracy for each technique.

3.1 Introduction

Retrieval is the task of retrieving documents of a dataset which fulfill a given user input keywords. There are two principal approaches to carry out word retrieval in document images. The first major approach is the traditional text search methods at the character level which requires efficient Optical Character Recognition (OCR) techniques. These methods are referred to as OCR-based techniques. Current OCR-based techniques have not shown great potential for extracting text from handwritten documents [23, 24, 25]. This is due to several challenging issues related to handwritten documents such as (i) poor quality documents, (ii) writing style variability, (iii) multiple writing styles, word writing variations, etc. The second category is to use word spotting techniques [26] to search in the image domain. Word spotting can be defined as the task of locating specific keywords in a collection of documents. In the last decades, word spotting has witnessed great interest as an emerging technology for document image retrieval applications and is becoming an eminent technique for this task. It has been introduced as an alternative to OCR-based techniques. The role of word spotting is especially emphasized in the case of historical document images because they are of poor quality and having large writing style variability. Word spotting finds specific keywords in document images by comparing features that are extracted from word images. It includes a matching process between word query image and a collection of word images stored in the database using the extracted features. Most of the proposed word-spotting based techniques find keywords as follows: firstly, they build indexes based on low level features that are extracted from word images. Same features are also extracted from the query image. Second, they search the entire database and measure how much the query image is similar to each word image in the database. Finally the results are sorted based on the similarity matching level. Document image word spotting presents problems worth solving. In this survey, we address the most frequently encountered problems when dealing with handwritten document image word spotting: (i) word image representation (features) (ii) word image matching and (iii) How to retrieve efficiently and accurately. The aim of this chapter is to provide a comprehensive survey of the current researches in handwritten document word spotting to date with emphasis on the period from 2001 to 2015. Based on this survey, it is possible for the researcher to find a good picture of the overall state of handwritten document image word spotting research, which facilitates: (1) finding of more information about different techniques, and (2) addressing weakness in current approaches to improve current techniques and possibly develop new ones.

The rest of this chapter is organized as follows: In Section 3.2, a general framework for document image word spotting is described. Challenges present in handwritten documents are discussed in Section 3.3. Databases used for word spotting are presented in Section 3.4. State-of-the-art techniques in word spotting are presented in Section 3.5. Other categorizations are presented in Section 3.6. Finally, we conclude the chapter in Section 3.7.

3.2 A General Framework for Document Image Word Spotting

A general framework for document image word spotting is shown in Figure 3.1. A document word spotting system can be conceptualized as a framework containing two parts: the index construction and keyword retrieval parts. Each part consists of a number of processes. These processes are pipelined, so that the output of one process forms the input to the next process. In the index construction stage, a set of features are extracted from word images and an index is built using the extracted features. In the keyword retrieval stage, a set of features are extracted from the input keyword and matched with the indexed features. A detailed description of the different stages of both parts is as follows.

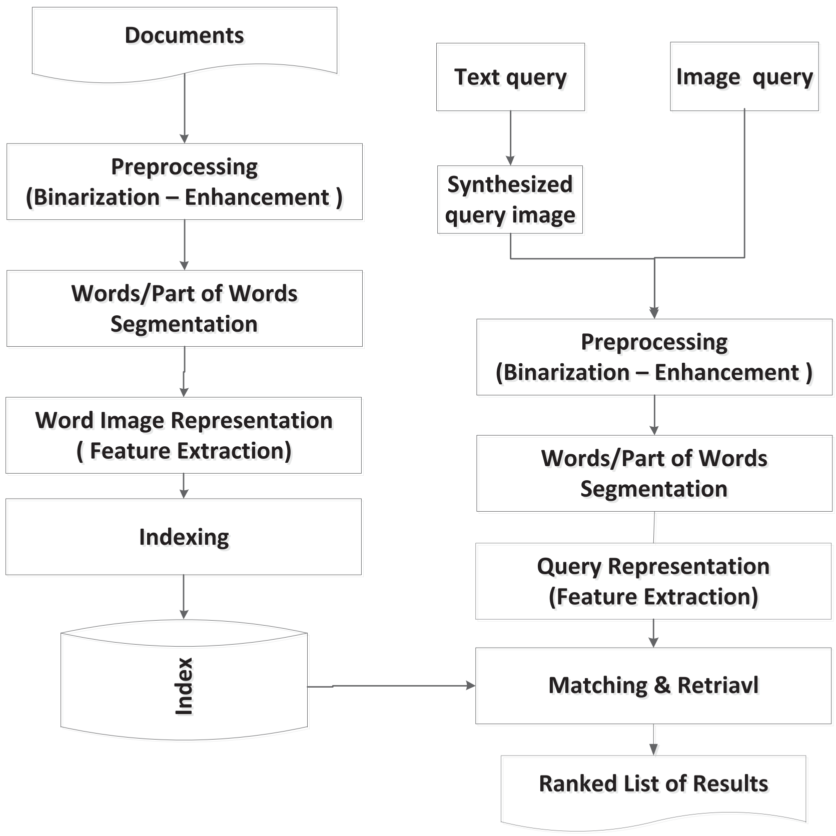


Figure 3.1: A general word spotting framework

3.2.1 Preprocessing

The preprocessing stage in document image word spotting consists, usually, of the following processes: binarization, noise removal, skew correction and line/word segmentation. Algorithms apply these tasks in different order and some algorithms might ignore one or more of these tasks if not needed. However, most of the document image word spotting systems use these preprocessing steps in some form or another [27, 28, 29].

3.2.2 Word Image Representation

When building a document image word spotting system, a key consideration is how to represent word images within document images. This is fundamental to providing acceptable performance results. One of the most important advantages of feature extraction is that it reduces the amount of required storage and hence the system becomes faster and effective. First, documents are segmented into words or sub-word. Then, a set of distinguishing features is extracted for each word/sub-word image. The extracted features are used to generate matching results in the subsequent matching process. In general, there are two main categories of features that can be applied to word images: local features and global features. Each of them has different capabilities of characterizing the shape of the word images and have discriminative power for handwritten word images word spotting.

Global features are computed over an entire word image while local features are computed on a specific parts of the images.

3.2.3 Feature Indexing

In this phase, word images are coded and indexed according to their extracted features. Indexing techniques have great impact on the searching time. Different techniques have been proposed in the literature including Latent Semantic Indexing (LSI) [30], Locality Preserving Indexing (LPI) [31], Self Organizing Map (SOM) [32], Geometric Hashing Indexing (GHI) [33], Locality Sensitive Hashing (LSH) [34] and Distance-based Hashing (DBH) [35].

3.2.4 Matching and Retrieval

The matching and retrieval module is the core of the word spotting framework. It performs the actual search, matching query image representations against document image representations and ranking the results. The aim of this process is identifying the word images of the documents which are similar to a given query word image by computing the similarity of the extracted feature vectors and the retrieval of similar words.

3.3 Handwritten Documents Word Spotting Challenges

Handwritten document word spotting presents a number of fundamental challenges. First, writing styles in handwritten documents exhibit large variability which significantly increase the difficulty of the task. Second, some handwritten documents are of poor quality. Furthermore, for handwritten historical documents, their age causes degradation problems such as ink bleed and similar effects.

3.4 Handwritten Document Image Databases

Development of robust document image word spotting systems requires databases of adequate sizes and diversity (many writers, multiple samples per writer, etc.) that contain an adequate amount of variations of several factors such as script, writing styles, font size and quality. In this section, we review several databases that have been used in the literature for various document image understanding tasks including word spotting task.

- **George Washington Database [36]:**

  In the literature, the George Washington database (GW) database has been widely used for various handwritten recognition and word spotting tasks. It consists of 20 pages of handwritten letters written by George Washington in the year 1755.

- **CENPARMI [37]:** The CENPARMI database includes 137 documents written by 13 writers. The dataset contains 2107 text lines. It is divided into two sets: testing set and validation set. The testing set contains 112 documents while the validation set contains 25 documents.

- **IAM Database [38]:** The IAM database includes 1539 pages of handwritten text written by 657 writers in the year 1978. It is divided into three sets: training, testing and validation. A good property of this data is that each set includes text lines written by several writers which makes it a good choice for word spotting with different writing styles.

- **Parzival Database [39]:** The Parvizal database includes 45 pages of manuscripts written using German language in the 13^{th} century. It is considered as a good choice for word spotting task as it is written by several writers.

Table 3.1: Summary of some of the databases used in word spotting and other handwritten text tasks

Databases	Descriptions	Writers
George Washington Database[36]	20 pages	2
CENPARMI[37]	137 documents	13
IAM Database[38]	1539 pages	657
Parzival Database[39]	45 manuscripts	-

3.5 State of the Art

In this section, we review various word spotting techniques proposed in the literature and categorize them based on the extracted features. Table 3.2 through Table 3.6 summarize these word spotting techniques.

3.5.1 Profile-based Features

Global shape features such as the lower and the upper profiles capture the outline of a word. A profile is represented by a one-dimensional vector corresponding to the column-wise distance from the top of the bounding box to the foreground pixel of a word. The distance between two profiles can be computed by any distance measure such as Euclidean Distance (ED) and Dynamic Time Wrapping (DTW) [36]. Generally, shorter profiles have less complexity than longer profiles and the chances of getting wrong matches are higher.

Manmatha and Croft [36], proposed to use word profile based features for historical document word spotting. In this approach, historical documents are segmented into words and matching is performed to retrieve relevant ones to a given query. In another work [40], they evaluated different techniques for handwritten words matching and proposed the use of DTW for matching word images. They also made an effort to select suitable features for the matching process [41]. The reported average precision rate was 72.65%.

Kolcz et al. [42] proposed the use of a line-oriented approach for handwritten documents word spotting based on matching profile-oriented features. They experimented with old Spanish manuscripts. They did not report the performance rate.

Sigappi et al. proposed a system for word spotting of Tamil Language [43]. They extracted projection profile, lower and upper word profile and Background-to-ink transitions features from each segmented word. Their approach make use of Hidden Markov Model (HMM) to characterize strokes' variations of handwritten characters. The reported accuracy was 80.75%. However, the test is done using small dataset containing only 400 words.

Shah and Suen [44] extracted two sets of shape features. The first set includes vertical, horizontal, left diagonal, right diagonal projection profiles and horizontal, vertical Background-to-Ink Transitions which are extracted from the entire word image. To extract the second set of features, the word image was segmented into regions and the features were extracted from each region. The extracted features are matched using the Correlation Similarity (CS). The evaluation was carried out using 168 queries. They reported about 94.57% and 60.25% precision and recall respectively, on 200 pages written by 200 writers.

Ali Abidi et al. [45] have used the height, width and convex area of each partial word as scalar features along with the vertical and horizontal word profiles. They have reported performance rate of 72% in terms of F-Measure on 115 queries and 90 Handwritten documents written by 90 writers.

Later Wie and Gao [46] introduced a system for locating words in the Mongolian Kanjur documents. They used four profile-based features for representing word images. Queries are synthesized using pre-extracted glyph. They evaluated their system using 200 Mongolian Kanjur document images where the average R-precision rate was 61.02%. The R-precision is the ratio of the number of relevant retrieved words in the top R results to the total number of relevant words.

Kesidis et al. [47] addressed the problem of accessing historical printed documents by employing word profiles' features and the character pixel density of each zone in the word image. Queries are synthetically created from the character templates. The reported performance rate was 87.5% on 12 handwritten pages.

Huaigu Cao et al. [48] introduced a probabilistic model for word spotting. The proposed model incorporates word segmentation and word recognition probabilities, obtained by modeling the conditional distribution of multivariate distance features of word gaps and the distances returned by the word recognizer, respectively [43]. They used scanned images of the New York State Pre-hospital Care Report (PCR) forms and achieved a mean average precision value of 4.7%.

Table 3.2: Various profile-based word spotting techniques

References	Matching Techniques	Language	Database size	#queries	Document Type	Accuracy
Rath and Manmatha [41]	DTW	English	GW: 15 pages	-	Historical	P:72.56%
Rath and Manmatha [49]	DTW	English	GW: 20 pages	-	Historical	P:62.41%
Huaigu Cao et al.[48]	DTW	English	PCR corpus	-	Handwritten	mAP:4.7%
Kolcz et al. [42]	Line-oriented	Spanish	13 pages	4	Handwritten	-
Sigappi et al. [43]	HMM	Indic	400 words	10	Handwritten	P:80.75%
Shah and Suin [44]	CS	Pashto	4200 words	168	Handwritten	P:94.57%
Ali Abidi et al. [45]	DTW	Urdu	90 pages	115	Handwritten	F:72%
Wei and Gao [46]	DCT	Mongolian	200 pages	-	Handwritten	P:61.02%
Kesidis et al. [47]	L_1 distance	Greek	12 pages	-	Handwritten	P:87.5%
P: Precision, R: Recall, mAP: mean Average Precision, F: F-measure and R-P: R-precision						

3.5.2 Gradient, Structural and Concavity (GSC) Features

These features are capable of measuring the characteristics of an image at global, intermediate and local ranges, respectively. Generally, Gradient, Structural and Concavity features (GSC) are suitable features for handwritten document word spotting since they are able to capture the shape of the written words.

In [50, 51], Srihari et al. make use of binary GSC features to spot words in handwritten documents. They segmented text lines into connected components. They used correlation distance for matching two GSC binary feature vectors. The method was tested on a dataset written by 8 writers and reported a precision rate of 70% at a recall of 50%. .

In other work [52, 53], GSC feature-based word spotting method was also presented. The method divides the word image into 32 regions and extracts 16 binary GSC features from each region. A total of 1024 bits of the GSC features are used to represent each word image. Similarity is computed using the bitwise matching of GSC features extracted from word images. The reported precision is 60 % obtained at a recall of 50% using English handwritten documents called CEDAR Letter. A precision of 90% was obtained at a recall of 50% for printed Sanskrit documents.

3.5.3 Shape Coding

Kefali and Chemmam [54] proposed an approach that is based on the coding of the words using topological features including diacritic, ascenders, descenders and loops. For each sub-word, a set of 4 features is extracted and each sub-word is represented by a sequence of codes. The query is also represented by a sequence of codes. During the matching process, the distance between the query code and the codes of each sub-word is computed using the DTW algorithm. Experimentation was done on an assembled database of 1100 images of Algerian postal envelopes written in Arabic, and yielded an accuracy of 75% while obtaining a recall of 87%.

Bai et al. [26] introduced a word shape coding based approach for keyword spotting in handwritten documents. A total of 7 features were extracted for each word namely, descenders, ascenders, eastward concavity, westward concavity, horizontal line intersection, i-dot connectors and holes. Then, based on the position and type of the features, each word was represented by a sequence of codes. Matching of codes is carried out using the DTW technique. The proposed method is evaluated on George Washington dataset achieving an F-measure of 93.37%.

In the context of historical documents, Liang et al. [55] introduced a novel character and word-based modelling approaches that allow the retrieval of vocabulary keywords which are not in the training set. Word images are segmented into small units, called graphemes [56], which represent the basis for synthesised character and word construction. They used the Self Organizing Map (SOM) to label the extracted graphemes. Then, the characters are modeled using the graphemes in all instances of the character, and subsequently words are synthesised using the models of all the characters in the word. Testing is done with three datasets including 3 pages from George Washington database (GW) database, 3 pages of Bargrave's Diary and 3 pages of a modern manuscripts. They reported a mean Average Precision of 67% on GW database.

Farrahi and Cheriet [57] introduced a language independent system that does not need line and word segmentation. In this system, the connected components are extracted from the images. During the matching process, each connected component of the query is compared to a list of Basic Connected Components (BCCs) that are generated from the connected components using Euclidean Distance and Dynamic Time Wrapping (DTW). Then, BCCs are clustered using the Kohonen network (Self organizing map). The system is evaluated on 160 pages from Juma Al Majid Center and the reported word spotting rate was 95%. The reported performance is promising. However, it is over-optimistic for the following reason. If the target BCC is found inside the cluster, the word is considered found. Meanwhile, each cluster contains a number of BCCs that may correspond to different words [55].

Another important contribution is the work described in [58] where keywords are considered as visual shapes. The authors make use of shape context [59] computed from skeleton points as features to describe word images. The shape context of skeleton points captures the spatial distribution of other points relative to it in polar coordinates. For each word, Shape Signature (SS) is formulated in terms of the shape context descriptors. The distance between two shape signatures is computed as the sum of the distances between the K most similar pairs of their shape contexts. Shape signatures are indexed using bit vectors which are simple and compact.

Alicia Fornés et al. [60] adapted Blurred Shape Model (BSM) for the task of keyword spotting. The BSM descriptors [61] encode the spatial probability of appearance of the shape pixels and their context information and extract a feature vector which describes the shape. The image is divided into a grid of $n \times n$ regions. For each region, a weighted sum of all pixels in the surrounding regions is computed. For keyword spotting, the BSM descriptors should also reflect the length of a word. This is done by creating a template blank image, where every word image of the database will be located in the center of this template image, according to its own centroid. They experimented with GW dataset. The proposed method was tested using 58 query images corresponding to 29 subjects. The reported performance was 53.14% of a mean average precision rate which is better than the mean average precision value obtained using the DTW algorithm.

The method in Lladós et al. [62] used the same set of features in Lladós and Sanchez [58]. The method was tested by the authors with 32 pages and a detection rate of 81% was reported.

3.5.4 Contour Matching

In the work of Adamek and O'Connor [63], a word spotting method for historical handwritten documents that is based on word contours matching is proposed. The contour matching technique proposed in [64] is utilized to compute the similarity between word contours. The proposed method is evaluated using a set of 20 pages taken from the GW dataset . The reported accuracy is 83%.

Can and Duygulu [65] proposed a line-based representation for historical document matching. Word images are represented with a set of contour segments which are approximated by lines using the Douglas Peucker algorithm [66]. Each line is described using the position, the length and the orientation. The distance between the two line descriptors is computed using the dissimilarity function as defined in [67]. They used two types of datasets. The first dataset consisted of arbitrarily selected 20 pages from GW dataset, while the second database consisted of 3 Ottoman pages. They reported a precision rate of 77.4% and 98.7% for the two datasets respectively.

Table 3.3: Various GSC-, shape coding and contour matching-based word spotting techniques

References	Techniques	Language	Database size	#queries	Document Type	Accuracy
GSC-based Features:						
Srihari et al. [50, 51]	Correlation Similarity (CS)	Arabic	100 pages	150	Handwritten	P:55%,R:50%
Srihari and G. Ball [52, 53]	CS	English	100 pages	100	Handwritten	P:60%,R:50%
Shape Coding:						
Kefali and Chemmam [54]	DTW	Arabic	1100 envelope	300	Handwritten	P: 75%, R: 87%
Farrahi and Cheriet [57]	Clustering of CCs	Arabic	20 pages	-	Historical	F:74%
Bai et al. [26]	DTW	English	30 pages	40	Handwritten	F:93.37%
Y. Liang [55]	Character modelling	English	GW: 15 pages	95	Historical	mAP:67%
Alicia Fornés et al. [60]	BSM	English	GW 20 pages	-	Historical	mAP:53.14%
Lladós et al [62]	BSM	-	32 Pages	7	Handwitten	A: 81%
Lladós and Sanchez [58]	SS	Spanish	-	-	Handwritten	-
Contour Matching:						
Adamek and O'Connor [63]	Multiscale representation	English	GW: 20 pages	-	Historical	A: 83%
F. Can and Duygulu [65]	Polygonal approximation	English	GW: 10 pages	-	Historical	P:77.4%, R:77%
		Ottoman	3	-	Handwritten	P:98.7%, R:100%
P: Precision, R: Recall, mAP: mean Average Precision, F: F-measure, A: Accuracy and R-P: Recall-Precision						

3.5.5 Bag of Features

The state-of-the-art bag of features model has been used for word spotting. Ataer and Duygulu [68] proposed an image retrieval method for historical document indexing based on Bag of Features (BoF) model. Visual words are referred as visterms. Matching is carried out with the use of bag-of-words powered by Scale-Invariant Feature Transform (SIFT) descriptors which are extracted from word images. Authors reported an average mAP value of 30% on small-printed and Rika datasets.

Recently, a segmentation-free approach for historical documents that is based on bag of features model is proposed in [69]. Instead of using clustering of the extracted descriptors using K-means clustering algorithm, latent semantic analysis is used to handle the redundancy and ambiguity of the individual words. The descriptors are compressed using the product quantization method. This approach was evaluated using three datasets (viz. GW, Barcelona Cathedral (BCN) and Lord Byron (LP)). The reported mAP values were 61.35%, 90.17% and 90.38% for the three datasets, respectively. Due to the use of compression and latent semantic analysis, this method is considered efficient in terms of time and memory.

Yalniz and Manmatha [70] proposed the use of SIFT descriptors for word spotting in noisy document images. They extracted SIFT descriptors from the corner points of each word image. Hierarchical K-Means algorithm is used to quantize the extracted features into visual terms (visterms). During the matching process, the common visterms between the query and the word images are identified and used to calculate the similarity score between the query and the word images. This technique achieved a mean average precision value of 93% using the book Adventures of Sherlock Holmes book and two books written in Telugu script.

The work reported in [71] is similar to the approach mentioned above in the sense that SIFT descriptors are used for word spotting. A notable difference is, however, the keypoints are represented using five-tuple entity, the x position, the y position, the scale, the orientation and the histogram of oriented gradients. The matching process employed compares the Histogram of Oriented Gradients (HOG) of the query word against those of the test images using the Euclidean Distance (ED). They reported a 46.9% and 71.5% of precision and recall on 20 pages extracted from LP dataset and 20 pages randomly chosen from the Journal of KIISE which is written using Korean language.

The work reported in Rothacker et al. [72] represents a significant advantage over other work belonging to this category. The work integrates the Bag of Features model with Hidden Markov Model (HMM) based statistical model that is estimated only from the query itself. When tested with 20 pages of GW dataset, the proposed approach has reported a mean average precision rate of 61.1% with vertical feature selection scheme that selects only those features whose local image descriptors do not extend beyond the lower or upper query boundary.

Rusiñol et al. [73] proposed a segmentation-free word spotting method based on a bag of features model that make use of SIFT. First, The document image is divided into a set of local patches. The local descriptors are extracted for each patch as well as for the query image. Then, the distance between the query image descriptors and patch descriptors is computed to find the similarity between the query image and each patch. Then, the document patches with a higher probability of containing instances of the query are retrieved. Spatial Pyramid Matching (SPM) method [74] is used to utilize the spatial information of the features. Feature descriptors are refined using the Latent Semantic indexing method. They used 3 datasets to evaluate the proposed method, 20 handwritten pages from GW database, 20 handwritten pages from the Lord Byron database and 20 typewritten pages of Persian dataset. The reported mean average precision values were 30.42% and 42.83% while the mean recall values were 71.1% and 85.86% for the GW and LP datasets respectively.

Czúni et al. [75] used local features based on different variations of SIFT descriptor to spot handwritten words. In this method, each feature point (q) of the query is only compared to the candidate points of the candidate words which lies inside a circle centered at the point q. Then, matching points set is constructed from all feature point pairs having the least Euclidean distance and the least difference in orientation. The similarity is computed with the use of the matching points for the query and candidate words. They reported a 85% accuracy on 22 pages.

Rodírguez and Perronnin [76] used rotation-invariant descriptors namely Local gradient histogram inspired by the SIFT keypoint descriptors. This approach makes use of a sliding window that moves in one direction from left to right over a word image. Then, a feature vector is computed using histogram of orientations at each position. These feature vectors are fed to HMMs. Evaluation is done using 630 images written in French. Experimentally, they showed a mean average precision value of 71.7%.

Another segmentation-free keyword spotting method, that is based on Heat Kernel Signature (HKS) which is proved to perform better than SIFT descriptor, is proposed in [77]. First, keypoints are located by the SIFT keypoint detector on the document pages and the query image. Then, for each keypoint, HKS descriptors are calculated from a local patch centered at the keypoint. Finally, a searching approach is employed to find a local patch containing enough matching keypoints corresponding to the query image . They used the GW dataset and reported an average mean precision value of 62.47% and the average recall value 92.38%. This method is not scalable for large dataset since it uses an expensive similarity matching technique.

The method proposed by Almazán et al. [78] avoids segmentation by using a sliding-window approach to locate the document regions that are most similar to the query. HOG descriptors and Exempler SVM framework are used to represent the documents and the query in supervised way. The proposed method is evaluated using 2 datasets, GW and LP. The reported results were 78.01% and 38.28% in terms of mAP for the two datasets.

Khayyat et al. [79] proposed word spotting method for Arabic handwritten documents. Their method is considered as a learning-based method that segments the documents into Part of Arabic Word (PAW) and integrates the language models with the partially segmented word to represent these words. During the testing phase, PAWs are passed to a hierarchical classifier which is composed of a set of classifiers trained on different PAWs. They reported a precision rate of 65% with a 53% recall rate on 43 documents of the CENPARMI Arabic handwritten documents database. They extended their work in [28] and used a hierarchical classifier which consists of Regularized Discriminant Analysis (RDA) and Support Vector Machines (SVM)s. When tested with the CENPARMI Arabic database, the reported performance was 84.56% precision rate at 50% recall. This appraoch circumvents the problem of having no clear boundaries between words.

Rodrigue and Perronnin [80, 81] used Local gradient histogram (LGH) features and Semi-Continuous Hidden Markov Models (SC-HMM) for handwritten word-spotting. Handwritten words are matched using synthesized typed text templates. The proposed method does not require prior training. Experimentation was done on an assembled database of 105 French letters. Results show that SC-HMM and LGH features outperforms the standard approaches using DTW.

Table 3.4: Various local features based word spotting techniques

References	Descriptors	Language	Database size	#queries	Document Type	Accuracy
Khayyat et al. [79]	HOG	Arabic	CENPARMI	-	Handwritten	P:65%, R:53%
Khayyat et al.[28]	HOG	Arabic	CENPARMI	-	Handwritten	P:84.56%, R:50%
Rusiñol et al.[69]	SIFT	English	GW,BCN, LB	-	Historical	mAP:61.35%
Rothacker et al.[72]	SIFT	English	GW: 20 pages	-	Handwritten	mAP:61.1%
Rusiñol et al.[73]	SIFT	English	GW: 20 pages	-	Handwritten	P:30.42%, R:71.1%
L. Czúni et al. [75]	SIFT	English	22 pages	110	Handwritten	P:85%
Zhang and Tan [77]	HKS	English	GW	-	Handwritten	P:62.47%, R:92.38%
Almazán et al. [78]	HOG	English	GW: 20 pages	-	Historical	mAP:38.28%
Rodríguez et al. [80, 81]	HOG	French	105 pages	10	Handwritten	mAP:56.4%
Rodríguez et al. [76]	HOG	French	630 pages	10	Handwritten	mAP:71.7%

P: Precision, mAP: mean Average Precision, and F: F-measure,

3.5.6 Hybrid Features

Sagheer, et. al. [82] built an Urdu word recognition systems based on word spotting technique. They segmented the documents into connected components and generated words by combining neighboring components together using sliding window. Then, gradient features and profile based features are extracted for each generated word. They used SVM for recognition. The evaluation is performed on 40 pages obtained from the CENPARMI Urdu Database. They achieved a 50% precision rate at a recall rate of 70.1%.

Focusing on Arabic handwritten documents, Al Aghbari and Brook [27] proposed an algorithm for retrieving Arabic historical documents. They used several structural and statistical features. The structural features used are upper and lower profiles and projection profile, while the statistical features are the ratio between punctuation and main connected parts and punctuation count were extracted. They used neural networks to classify extracted features into word classes. The results showed that the proposed technique was robust to different styles and the reported accuracy was about 89.3% for 100 queries using 27 pages.

Fischer et al. [83] introduced a learning-based word spotting system based on HMM that utilized subword models. The proposed method can spot keywords even if they are not present in the training set. The proposed system does not need pre-segmenting of text lines into words. The authors proposed to represent text lines by a sequence of features extracted by a sliding-window approach. Based on a test set of 20 document images from the GW database, they reported a 31.5% average precision rate. One of the main drawbacks of their approach is that it needs large annotated data for HMM training. An extended version of this work is introduced in Fischer et al. [84]. They conducted several experiments using the IAM, the George Washington and the Parzival databases. By comparing the proposed system with a standard DTW-based matching technique, the authors claim that the proposed system outperformed the reference system on all data sets. The experiments yielded a reported recognition rate of 63.78%. In more recent work, Fischer et al.[85] improved the previous HMM based spotting technique [84] by integrating the character n-gram language models into the spotting system. They reported a 73.88% (+2.3) on the GW database consisting of 20 document images.

Wshah et al. [86, 87] proposed a language independent word spotting method that is based on HMM. The proposed method does not need word or character segmentation. Lines of the document are considered as a set of overlapped frames and gradient and intensity features are extracted for each frame. The proposed method has been tested on several languages including English, Arabic and Devanagari. The reported results show that it outperforms the previous line based algorithm.

Rodriguez and Perronnin [88] proposed a model-based approach for word spotting. Several features are extracted for all the images by moving a window from left to right. The following features are computed for each window, the Local gradient histogram (LGH) features [76]; the column features [89]; and the zoning features [90]. They conducted several experiments using three different datasets(viz. the IFN/ENIT dataset, the GW dataset and French letters dataset). The reported mAP values were 41.6%, 53.1% and 32.3% using the three datasets, respectively. This approach requires an error prone segmentation step to select the candidate windows.

In Şaykol et al. [91] a content-based retrieval system based on features in wavelet and spatial domain is proposed for historical Ottoman documents. The similarity between word images and queries is computed using the histogram intersection technique. Experiments are carried out using 102 documents. 16 query images are tested and the reported accuracy was 92%.

For Arabic historical documents, a computer aided retrieval framework is presented in [92]. Different combinations of five feature are used (viz. projection profiles, angular line features, concentric circle features and geometric features, Hu's moment). Four similarity measures have been tested (viz. Euclidean distance, angular separation, Manhattan distance, and histogram intersection). The developed system reported a matching rate of 72.5%.

Rodríguez and Perronnin [93] presented a statistical framework based on two types of HMMs, Semi-Continuous Hidden Markov Models (SC-HMM)s and Continuous HMM (C-HMM)s. They used query by "word-class" instead of querying by string or by example to overcome the downsides of both approaches. Column features [89], pixel count features [90] and Local gradient histogram (LGH) features [76] are employed in this work. Evaluation is done using 630 images written in French. Experimentally, they showed a mean average precision value of 87%. They also showed that the Semi-Continuous Hidden Markov Models (SC-HMM) is superior to the C-HMM and is consistently superior to DTW.

Huang et al. [94] described a novel approach for spotting Chinese documents. The proposed approach uses a contextual word model. Similarity between every candidate word and the query word is measured by fusion of the geometric context and character classifier as well as linguistic context. This fusion allows giving true positives and the negatives of query word high and low scores respectively. They reported a 85.87% accuracy on the CASIA-HWDB database [95].

Another hybrid feature extraction scheme is used in Kesidis et al. [96]. They used two types of features [41]. In the first one, word image is divided into a set of zones. Then, in each zone, the density of the character pixels is calculated. In the second type of features, the word is divided into vertical zones and word upper/lower profiles are computed for each zone. The word image is divided into two horizontal parts and upper/lower word profiles are computed for each part. Similarity between all segmented words and the query image is computed using the Euclidean distance metric. Natural language processing techniques have been used in order to improve the efficiency of the proposed technique. The experiments carried out in this work are based on a corpus of 110 images of historical Greek document. The experimental results show that using natural language processing techniques and user's feedback improve the retrieval performance.

The method in Konidaris et al. [29, 97] used the same set of features in Kesidis et al. [96] with synthetic data and user feedback. word images are synthesized from their ASCII representation. The method was tested with 100 document pages. Experimental results show that the combination of synthetic data and user feedback improves the performance and the hybrid feature scheme outperforms the single feature approach.

The method in Frinken et al. [98] used the same set of features in Fischer et al. [83] with Neural Networks model. The method was tested with IAM database and an average precision rate of 87.6% was reported.

Later, Frinken et al. [99] extended their work by introducing a modified version of the CTC Token Passing algorithm along with recurrent neural network. The method was evaluated by the authors with George Washington database (GW) database and an average precision rate of 71% was reported.

Table 3.5: Various hybrid features based word spotting techniques

References	Language	Database Size	#queries	Document Type	Accuracy
Al Aghbari and Brook.[27]	Arabic	2000 words	100	Historical	P:89.3%
Shahab et al.[92]	Arabic	-	-	Hndwritten	A:72.5%
Sagheer, et. al. [82]	Urdu	40 pages	56	Handwritten	P:50.75%, R:71.02%
Wshah et al.[86, 87]	AED	IAM: 180 pages	-	Handwritten	mAP:57.7%
Rodríguez and Perronnin[88]	English	GW	-	Historical	P:53.1%
Fischer et al.[83]	English	GW: 20 pages	-	Historical	P:31.5%
Frinken et al.[98]	English	IAM	-	Handwritten	P:87.6%
Frinken et al. [99]	English	GW IAM and Parzival	-	Historical Historical	P: 71% P: 76% P:92%
Fischer et al. [84]	English	GW IAM and Parzival	105	Historical Historical	mAP: 62.08% mAP: 47.75% mAP:85.53%
Kesidis et al.[96]	Greek	100 pages	32	Handwritten	-
Konidaris et al.[29, 97]	Greek	100 pages	25	Handwritten	-
Şaykol et al. [91]	Ottoman	102 pages	16	Historical	A:92%
Rodríguez andPerronnin [93]	French	630 pages	10	Handwritten	mAP:87%
Huang et al. [94]	Chinese	-	-	Handwritten	P:85.87%

P: Precision, R: Recall, mAP: mean Average Precision, F: F-measure, and A: Accuracy

3.5.7 Other Features

Moments features are used in Bhardwaj et al. [100] to spot keywords in Handwritten Document images . These features are capable of maintaining invariant representation of all word images. Cosine similarity is used for word matching and a relevance feedback technique is employed to refine the results. They used a database of 2163 word images: 707 from the IAM database, 763 from publicly available Million Book Project documents and 693 from 5 Sanskrit documents. The best performance is obtained for Sanskrit script (87.88% of precision).

Perronnin et al. [101] used a generic framework based on Fisher kernel for handwritten word spotting. This technique is widely known in the domain of face detection. It combines the benefits of generative and discriminative approaches to pattern classification. They reported a 89% average precision on the non-linear kernel tested against 630 images of French documents.

Zhang et al. [102] used local stroke direction histogram feature for Chinese word spotting. Document lines are segmented into primitive segments. Then, all consecutive segments are grouped to form candidate character patterns which are presented as a feature vector. During matching, the query is compared against each sequence of candidate patterns. The method yielded good performance in experiments on a database containing 550 pages by 110 writers. For one query, the precision, recall, and F-measure are 94.84%, 87.25% and 90.88%, respectively.

However, using one query does not give a robust performance evaluation.

In another work, Zhang [103] proposed a probabilistic graphical model based on the Semi-Markov Conditional Random Fields (semi-CRFs) [104] for Keyword Spotting in Online Chinese Handwritten documents. They reported a 95.62% F-Measure rate on the CASIA-OLHWDB database [95] consisting of 1020 images.

An eigenspace-based method is proposed in [105] for historical document image retrieval. This technique is widely used in the domain of face recognition. The proposed method is evaluated using one query on 22 images taken from The diary of Matsumae Kageyu. The reported recognition rate was less than 80%.

The work in [106], used Generalized Hough Transform (GHT) for Arabic handwritten historical document word spotting. They evaluated their method using 23 pages from historical documents of the Tunisian national Archive. Experiments show that the proposed technique has a quite satisfactory results as claimed by the authors.

In Sousa et al. [107], a fuzzy based approach is presented. They used oriented feature obtained from Gabor filter banks and a membership functions for determining the potential matches to the given word. They reported a 99.2% precision and 81.6% recall on 20 images provided by the Portuguese National Library.

Fernández [108] developed an approach based on characteristic Loci Features [109]. Characteristic Loci features encode the number of the intersections for a given key-point in the four directions. In this work, the authors added the two diagonal directions. Contour points, foreground pixels, background pixels or skeleton points can be used as key-points depending on the application. Once a Loci-based descriptor is extracted for word images, they are organized in a look up table. Matching is performed by a voting process using Euclidean and Cosine distance. Based on a test set of 30 images from the Cathedral of Barcelona archives, they reported a 82.5% precision rate with 67% recall rate.

Bilane et al. [110, 111] used directional roses features based on a selective sliding window technique to spot handwritten Syriac manuscripts. They used 14 pages for evaluation. They used only 6 queries which does not give a robust performance evaluation.

In [112], a word spotting system has been built which is based on biologically inspired features [113, 114]. The system consists of four layers. Each layer is composed of different unit types and receives its input from the previous layer. The neurons in the first layer and the third layer are modeled by Gabor functions and a function similar to Radial Basis Function (RBF) respectively [115]. Layer three performs local pooling while layer four performs global pooling. The experiments are performed on 1,146 pages of handwritten text taken from the Dutch National Archive [116]. The reported results is 89% of accuracy using Hamming distance.

Leydier et al. [117] proposed a word spotting method based on differential features. Differential features are favorable for handwritten document because they are a good representation of the complex shapes of the handwriting and robust to scale variation, lighting changes and noise. The method employs a cohesive elastic matching method to compare the extracted features. The proposed method was evaluated on medieval and more recent manuscripts. The authors reported a recall of 74.2% on George Washington dataset. This approach is not scalable for large datasets since it uses an expensive distance computation.

Later, Leydier et al. [118] extended their work by introducing the ability to search words using multiple grammatical variations and different graphic representations. The new algorithm has the ability to search for a word on multiple documents. This is done by transforming the ASCII query into a graph of graphemes using a predefined model that combines alphabet, the glyphbook and the grammar. They validated their appraoch on Latin, Arabic and Chinese manuscripts and reported a 62%, 77% and 70% for the three datasets, respectively. However, they use only 4 pages for all languages which does not give a robust performance evaluation.

Table 3.6: Various word spotting techniques using other features

References	Features	Language	Database size	#queries	Document Type	Accuracy
Aouadi and Kacem[106]	Hough Transform	Arabic	23 pages	-	Handwritten	-
Leydier et al. [117]	Visual cortex based features	English	GW	-	Handwritten	R: 74.2%
Perronnin et al.[101]	Fisher Kernel	French	630 pages	10	Handwritten	P:89%
Zhang [102]	Stroke direction histogram	Chinese	550 pages	-	Handwritten	P:94.84%, R:87.25%
Zhang [103]	semi-CRF	Chinese	CASIA-OLHWDB	-	Handwritten	F:95.62%
Terasawa et al.[105, 119]	Eigenspace	Japanese	22 pages	1	Handwritten	A: <80%
Sousa[107]	Gabor filter	Portuguese	20 pages	20	P-Historical	P:99.2%, R:81.6%
Fernández [108]	Loci features	Spanish	30 pages	10	Handwritten	P:82.5%,R:67%
Bilane et al.[110, 111]	Directional rose	-	14 pages	6	Handwritten	-
Zant et al. [112]	Fifferential features	Dutch	CASIA-OLHWDB	-	Handwritten	-

P: Precision, R: Recall, mAP: mean Average Precision, F: F-measure, and A: Accuracy

3.6 Other Categorizations

Word spotting can be classified also based on the format of the query, the matching and the layout analysis.

Word spotting techniques can be categorized into two groups based on the format of the query: Query-By-String (QBS) and Query-By-Example (QBE). In QBS, the user enters the query as a text. Based on trained model with a large number of characters/words, these approaches convert the text query into a word image which is in turn used in the matching process. QBS methods are user-friendly and able to spot words of arbitrary style [94]. On the other hand, in QBE, the user select an instance of the word in the document and uses it for locating other similar instances.

Similarly, word spotting techniques can be classified into template-based and learning-based. Template-based methods traditionally compare template-image with documents. Most of the early methods followed this paradigm. The main drawback of these methods is that they need a large number of templates for each keyword. Furthermore, they can not be generalized for new writing styles. Learning-based methods, on the other hand use supervised learning techniques for modeling words that users want to retrieve. They are able to handle writing style variations and allow the spotting of keywords that are not in the training data.

Word spotting methods can be also categorized based on the layout analysis. Two types of methods can be distinguished in the literature: the segmentation-based approaches and the segmentation-free approaches. The segmentation-based methods segments document images and query words into small parts called part of words Part of Arabic Word (PAW)s. This segmentation has a great effect on the final performance of the subsequent steps. Similarity is then measured between these PAWs and query word image. On the other hand, in segmentation-free methods [73], documents are represented as a list of overlapping patches. These methods avoid the prior segmentation problems.

3.7 Conclusions

This chapter presented an overview of the word spotting techniques used for handwritten documents. This chapter has started with a discussion of the different phases of a general framework for document word spotting. Then the most widely used features extraction techniques as well as the most used techniques for similarity matching are presented. Some of the reviewed research work have used global features such as word profiles, whereas the other work have used local features such as gradient. Some features are imported from other fields such as shape coding which are mainly used in shape matching. Several tables summarize the published results on document word spotting. Some of the proposed works focused on segmentation-based word spotting. While, other works mainly focus on segmentation-free based techniques for cursive text. Furthermore, most of the recent work has focused on bag of features model. Large number of works is devoted for Latin language. However, there are very few attempts on Arabic historical word spotting. This is clearly evident from the tables of Section 3.5. Therefore, more research work is needed for Arabic historical documents word spotting. It is noticed that the lack of benchmarking database for Arabic word spotting is a major problem. Language models are expected to significantly enhance the performance of word spotting systems for cursive text. Features for Arabic word spotting based on Arabic text characteristics are needed.

CHAPTER 4

HISTORICAL DOCUMENTS LAYOUT ANALYSIS USING WINDOW-BASED TEXT ANALYSIS APPROACH

One of the fundamental problems in document understanding field is document layout analysis, that is extracting regions or objects of interest from document images. It has been a core topic for the last decades, and is now involved in many applications such as optical character recognition, word spotting and other document understanding tasks. In Arabic historical documents, main and side notes text may coexist in the same document image. Therefore, it is necessary to separate these text regions for use in the subsequent stages of the document understanding tasks. In this chapter, we propose a new approach that addresses the problem of Arabic historical document layout analysis. The proposed approach is based on combination of foreground analysis, background analysis and window-based text analysis method that is able to separate the main text of the historical document from the side-notes. The combination of these techniques improves the segmentation results. Experimental results demonstrate that our approach has a strong capability in extracting text regions and achieving 90.41% correct segmentation rate.

4.1 Introduction

Layout analysis of document images is the first step in document understanding systems. The task consists of extracting text regions from the documents and identifying the proper class of these regions. Several layout analysis methods have been developed in the literature for printed documents [120, 4]. The existing layout analysis techniques can be generally classified into three categories: top-down methods, bottom-up methods and hybrid methods. The top-down methods such as the X-Y Cut algorithm [1] are based on the computation of the projection profile of the documents. These are applicable to printed documents which have well-known structures. Bottom-up methods such as [3, 4, 5] are connected components or pixel based. They make use of geometric relationship between neighbouring elements to form larger regions. These methods require fixed inter-space and intra-space variation which is in general not the case for the historical documents. Hybrid methods combine the top-down and bottom-up methods to improve the performance [6, 7, 22].

Research on Arabic historical layout analysis has focused on the extraction of text lines from simple structure documents [121-130] . To our knowledge, only two methods have been reported on the region extraction of complex Arabic historical documents in the literature [19, 21]. The first method [19] addressed the multi-oriented handwritten documents. The algorithm starts by dividing the image into rectangular regions. For each region, the local direction of the text is estimated using the Wigner-Ville distribution (WVD) on the signal representing the projection profile. Then, regions are merged based on their orientations to form larger text regions. This approach may not result in correct segmentation when the text in the window has multiple orientations. This is due to the fact that the separation point between different text orientations may not be detected as assumed by the authors. It also fails when the main text region has the same orientation as the side-notes and the two text regions have no clear space between them. The second method proposed in [21], formulates the layout analysis as a classification problem in which the historical document contents are classified into main text and side notes. However, it inherits the disadvantages of the classification techniques as it requires a training process.

Historical document analysis is a challenging problem. The difficulties come mainly from the complexity of the document structure (including irregular text regions), the diversity of the text (including text size and style variations, and curved baselines) and the complexity of the backgrounds (including noise), as shown in Figure 4.1. Different information can be utilized to segment historical documents into text regions like text orientation, white spaces(background), and pixel density, etc. However, this information is not present in all historical documents. For example, text orientation information can be used for segmenting the document shown in Figure 4.1(d). However, it is not useful for segmenting the document shown in Figure 4.1(c) as both main text and side notes have the same text orientation. Similarly, background information can be utilized to extract regions from the document shown in Figure 4.1(b). However, it is not useful for segmenting the document shown in Figure 4.1(a) as no clear background information can be computed from the document image. For the document shown in Figure 4.1(c) pixel density can be used instead of text orientation. Hence, it is necessary to combine available information in order to obtain good segmentation results of the different layout structures. In this chapter, we propose a new approach for Arabic historical document layout analysis that makes use of several attributes. It combines foreground analysis, background analysis, text orientation and pixel density. The proposed appraoch employs a window-based text analysis technique that takes into account text orientation, white space between regions and text density thus resulting in better segmentation performance. The novelty compared with previous approaches is two fold: (i) the fusion of background anal-

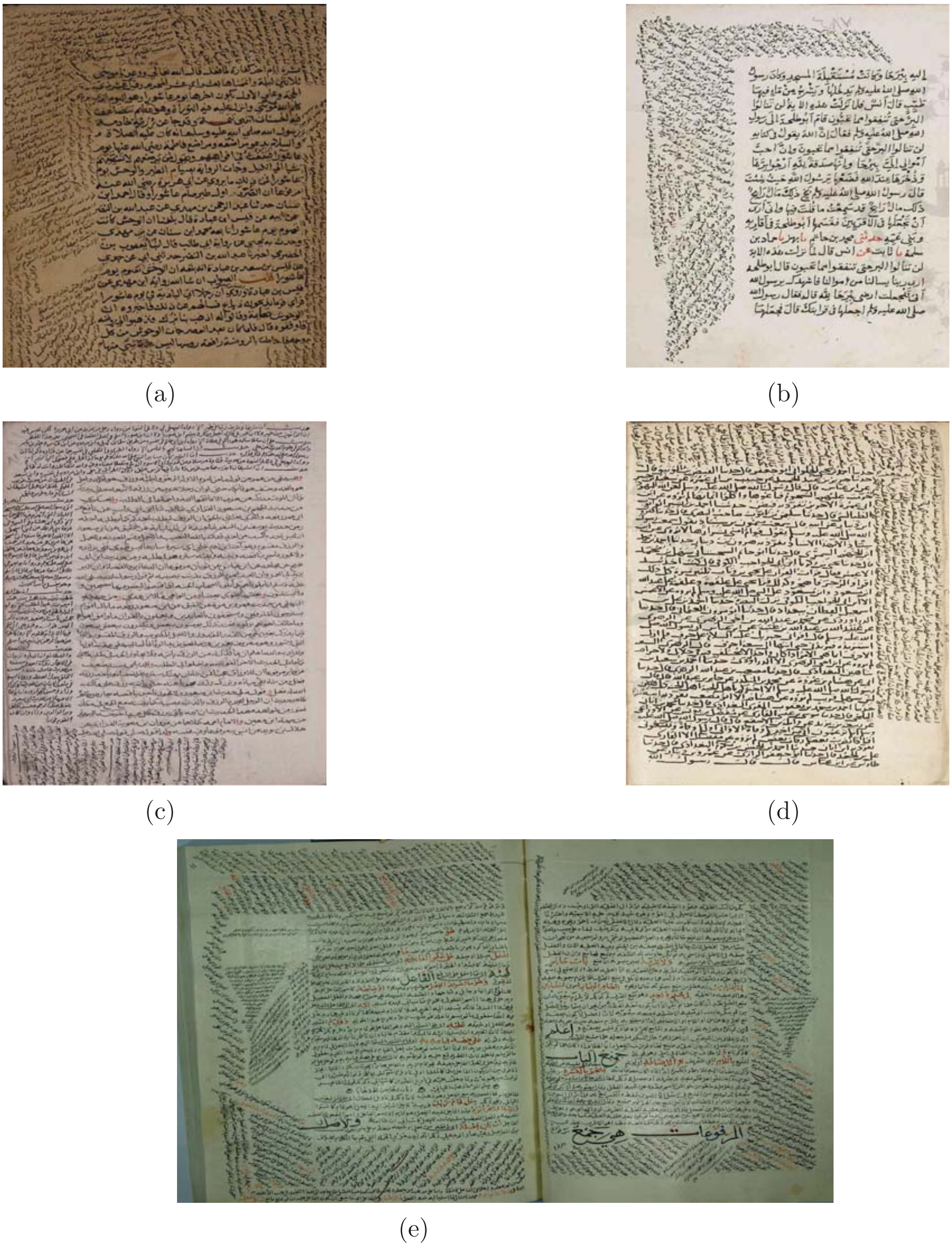


Figure 4.1: Arabic historical documents with different layout structures

ysis, foreground analysis, text orientation, white space between regions and text density. This fusion allows handling documents with different layout structures. (ii) a new method for text orientation estimation based on histogram of directions. The proposed approach can be integrated with Arabic historical text recognition, word spotting or historical document understanding. These applications require the extraction of text regions from historical documents before recognition or word spotting tasks.

The rest of the chapter is organized as follows. In Section 4.2, the proposed approach is discussed in detail. The experimental results are discussed in Section 4.3. Finally, Section 4.4 concludes the chapter.

4.2 Overview of the Proposed Approach

Foreground, background, and text analysis are combined to extract text regions. We first present a general overview of the proposed approach and then we give detailed description. The proposed approach consists of the following steps:

- **Pre-processing:** In this phase, images are binarized and noise is removed.
- **Initial analysis:** In this phase, analysis is used to identify a set of parameters that are employed in the next steps to separate main text region apart from side notes.
- **Text orientation estimation:** three methods are used to estimate text orientation. Based on the estimated orientations, main text regions are identified and separated from the side notes.
- **Sliding window-based analysis:** in this phase, main text regions are extracted and separated from the side notes using a moving-window-based method that makes use of space between regions, pixel density, and text orientation.

Details of each step of the proposed approach are presented in the following subsections.

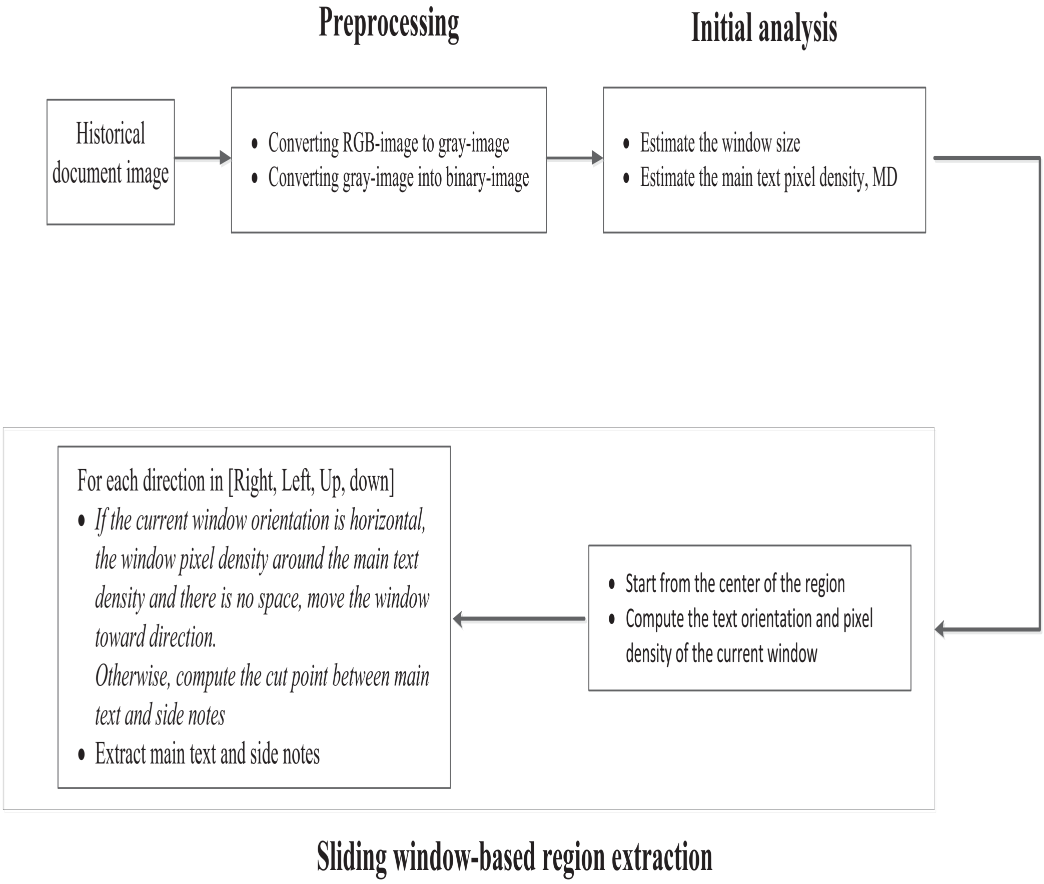


Figure 4.2: Block diagram of the proposed method

4.2.1 Pre-processing

Our scanned Arabic historical document images are saved in color format. They are first converted into gray-level images. Then, we use an adaptive binarization method [131] that finds a local threshold value for each region in the image based on the average gray value and the contrast difference in the region as follows:

$$T(x,y) = m(x,y).[1 + k.(\frac{s(x,y)}{R} - 1)] \tag{4.1}$$

where m is the average value of the window, s the standard deviation of the pixels in the current region, R is the dynamic range of the standard deviation, and k is a small positive value. Then, each pixel in the gray-level image is converted to binary by comparing each pixel with its computed threshold value. The window is set to 19×19 as suggested in [132]. Then, smoothing is applied to reduce the noise.

4.2.2 Initial Analysis

The aim of this stage is to identify a set of parameters used in the next step to separate the main text regions and side notes.

Choosing the window size

As suggested in [19], the window must approximately contain 3 lines to compute a projection profile histogram that is useful for text orientation estimation. This is necessary in order to have clear differences between the valleys and the peaks of the histogram. The window size is computed as follows. First, the height of the window is computed using equation 4.2.

$$window_h = 3 * CC_{avgh} + 2 * Gap \tag{4.2}$$

where CC_{avgh} is the average height of the connected components which are extraced using the built-in Matlab function 'bwconncomp'. The Gap is the estimated gap distance between three consecutive lines. Gap is estimated as follows. For each connected component CC_i, its closest two neighbours that are vertically located above and under CC_i are found (viz. CC_{above} and CC_{below}). Then the distance between the centers of CC_i and its closest neighbours, CC_{above} and CC_{below}, is computed, referred to as CC_i^{dist}. The average of all CC_i^{dist} is considered as the $Gap_{h.}$ Second, the width of the window is set to the estimated height. Figure 4.3 shows a graphical illustration for the computation of Gap value.

Algorithm 1 Window computation

1. Input a text region.
2. Extract connected components CCs from the input region.
3. Find the average height of CCs.
4. Find the average distance (Gap) between each CC_i and its closest two neighbours which are vertically located above and below CC_i
5. Compute the height of the window, $window_h$, using equation 4.2.
6. Set width value equal to window$_h$

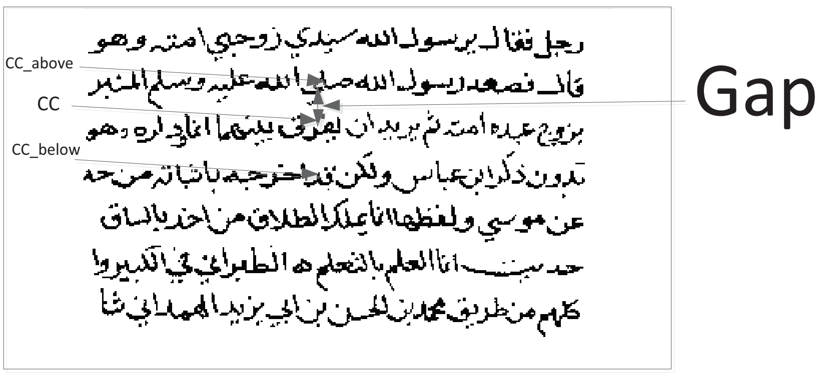


Figure 4.3: Estimation of the gaps between text lines

Main text pixel density estimation

In general, main text regions have less pixel density than side notes. So, pixel density can be utilized to distinguish between main text and side notes. Main text pixel density is estimated as follows. We select n windows in the middle of the page. Then, we compute the pixel density of each window as the total sum of the foreground pixel. Then, the maximum pixel density of all windows is computed which is considered as the estimated pixel density of the main text.

Algorithm 2 Main text pixel density estimation

---

1. Input a text page.
2. Select n windows at the middle of the page.
3. Compute the pixel density of the window as the total sum of the foreground pixel.
4. Compute the maximum pixel density of all windows and return it as the estimated pixel density of the main text.

4.2.3 Text Orientation Estimation

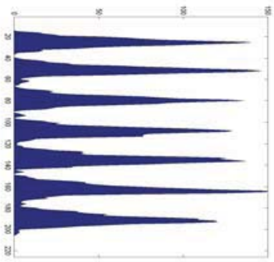
In the literature, projection profile is the most widely used method for text orientation estimation. Its principle relies on finding robust alternations between valleys and peaks. Figure 4.4 shows a multi line text region with a projection profile. The computed profile has seven sharp peaks which are corresponding to the text lines in the region. Generally, to determine the text orientation from the projection profile, differences between profile valleys and peaks are computed and averaged. Then, the profile that has the maximum difference reflects the text orientation [133]. This involves computing projection profile for 8 different angles for a given text window. Then, the profile with the maximum differences between valleys and peaks is chosen. Figure 4.5 shows projection profile computed for different angles for the text region given in Figure 4.4.

Algorithm 3 Projection-profile based orientation estimation

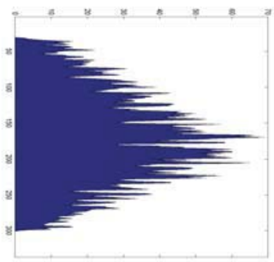
---

1. Input a text window called $txtRgn$.
2. For angle=-67.5 to 90, step 22.5
    newRgn=rotate(txtRgn,angle)
    Compute profile, PP_{angle}, for newRgn.
  end for.
3. Find the profile with the maximum differences between valleys and peaks, PP_{max} and return its corresponding angle as the estimated orientation of the input text window.

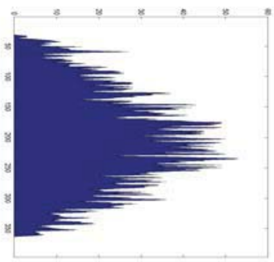
Figure 4.4: Projection profile computed for horizontal text regions



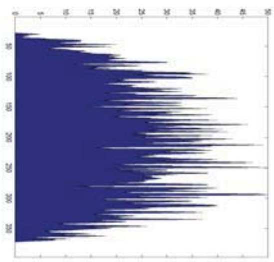
(a) Angle 0



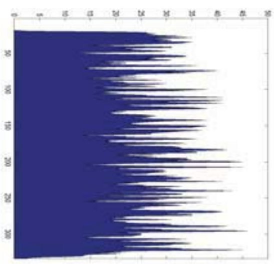
(b) Angle 22.5



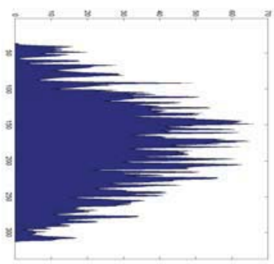
(c) Angle 45



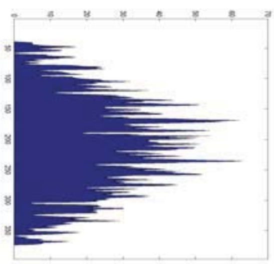
(d) Angle 67.5



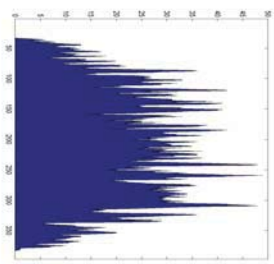
(e) Angle 90



(f) Angle -22.5



(g) Angle -45



(h) Angle -67.5

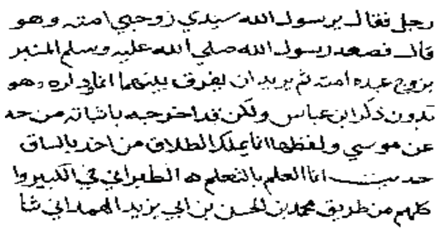
Figure 4.5: Projection profile computed for different angles

Proposed Methods for Orientation Estimation

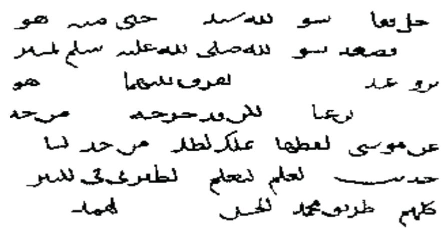
We propose two new methods for text orientation estimation. The first method is connected component based method while the second one is gradient based method. Both methods are histogram based. They estimate the orientation of text regions by constructing a histogram of connected components and gradient directions, respectively. The details of these methods are described in the following sections.

Connected Components based Orientation Estimation

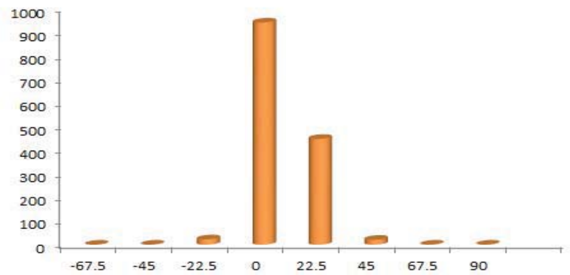
The first method is based on constructing a histogram of directions of the connected components extracted at an image. In Arabic script, words consist of at least one Part of Words (PAWs). Each PAW is composed of only one major connected component (CC) and some or no minor CCs. Minor and major CCs can be distinguished by their size. Most of the major CCs of horizontal text regions have larger width than height values. On the other hand, most of major CCs of vertical text regions have smaller width than height values. In general, text regions with orientation θ have major CCs whose length in the direction θ is larger than their length in the other directions. So, we start by extracting CCs of the input image and removing minor CCs as shown in Figure 4.6(b). Then, the major-axis length, minor-axis length and direction of each major CC are computed by using the Matlab built-in function 'regionprops'. Major CCs directions are quantized into 8 direction values, (-67.5, -45, -22.5, 0, 22.5, 45, 67.5 and 90), which are the common orientations of text. Due to writing style variations, quantizing the CCs direction into 8-bins tolerates some level of variations in writing directions. So, we employ soft quantization scheme as show in Figure 4.7. A histogram of the quantized directions is constructed. Then, the main direction of the text region is represented by the direction with the maximum histogram value as shown in Figure 4.6(c).



(a) Original text region

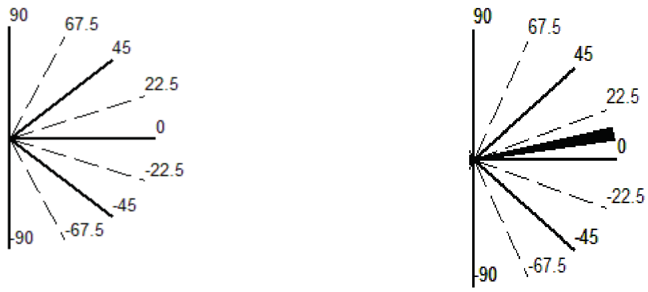


(b) Text region with major CCs only



(c) Histogram of major CCs major-axis

Figure 4.6: Orientation estimation using the histogram of CC directions



(a) Standard Angles

(b) The overlap between angle 0° and angle 22.5°

Figure 4.7: Angles quantization

Algorithm 4 Histogram of CCs directions based orientation estimation

---

1. Input a text region called $txtRgn$.
2. Find connected components in txtRgn.
3. Categorize CCs into major and minor CCs based on their size.
4. Remove minor CCs and build a new image that contains only the major CCs.
5. Compute major-axis length, minor-axis length and orientation CC_θ of each connected component in the new image.
6. Construct a histogram by adding major-axis length of all CCs into the corresponding bins of its orientation CC_θ using soft quantization.
7. Find the maximum value in the histogram and return its corresponding bin directions as the estimated orientation of the input text.

Gradient Based Orientation Estimation

In this section, we describe the second method for orientation estimation. It computes the histogram of gradient directions. The computed histogram is a 8-bin histogram constructed by quantizing the pixel gradient directions into 8 directions. Like our first method, the first step is to extract the connected components of a given text region. Then, we find the major CCs of the input region. The magnitude and the angles (orientations) of the gradient $f(x, y)$ are computed as in Equations 4.3 and 4.4. Then, a histogram of orientations is constructed from the magnitudes and the angles.

$$Direction : \theta(i, j) = tan^{-1}(\frac{v}{u}) \tag{4.3}$$

and

$$Magnitude : f(i, j) = \sqrt{u^2 + v^2} \tag{4.4}$$

where u and v are computed by convolving two 3×3 Sobel operators with the binary image, shown in Figure 4.8. After that, the gradient directions are quantized into 8 directions values (-67.5, -45, -22.5, 0, 22.5, 45, 67.5 and 90). A histogram of the quantized gradient directions is constructed. Finally, the main direction of the text region is represented by the direction with the maximum histogram value as shown in Figure 4.9.

$$
\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}
$$

Figure 4.8: X and Y Sobel operator masks

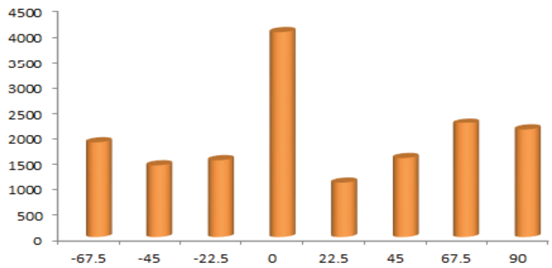


Figure 4.9: Histogram of gradient

4.2.4 Moving Window-based Region Extraction

The aim of this stage is to detect main text regions and separate them based on text orientation, pixel density and white space separation between regions. First, a window of size $h \times w$ is placed in the center of the region. The orientation of the window w is estimated, referred to as $w_{orientation}$. Simultaneously the pixel density of the current window is computed. Then, we continuously slide the window along the region in the horizontal direction from the center to right and from the center to left followed by sliding the window in the vertical direction as shown in Figure 4.10. At each position we estimate the current window orientation and pixel density. We compare the estimated orientation in the current window w_i to the estimated orientation of the previous window w_{i-1} and the pixel density of the current window w_i to the computed main text pixel density. We stop when the orientation of the window w_i is different than the orientation of the window w_{i-1}. This position will be the cut point where the region should be segmented. In addition to the window orientation, we also use the pixel density of the window as well as the white space to split the region into subregions. In other words, we continue moving the window until we encounter either a different orientation, a clear white space or large pixel density value. Figure 4.11 shows different criteria used to separate main text regions and side notes.

Figure 4.10: Sliding window text analysis, the window moves horizontally from left to right

(a) Different orientations

(b) Pixel density

(c) White space

Figure 4.11: Different situations for side notes detection

4.3 Experimental Results

In this section, we start by introducing the image representation, performance metrics and the used dataset in the evaluation process. Then, we discuss the experimental results based on the performance metrics. A popular method for representing images within the context of document analysis is the pixel-based representation approach [134]. In this representation, document images are represented as 24-bit RGB color images. Given an image of n regions $\{R_1, R_2, \ldots, R_n\}$, pixels which belong to a region R_i are assigned as their values the color of the region R_i. This representation is practical since it is independent of the region shape and can be saved using lossless color image format [135].

Figure 4.12: Pixel-based representation

To evaluate the proposed approach, we use Precision, Recall, and F-Measure as defined in [?]:

$$Precision = \frac{TP}{TP + FP} \tag{4.5}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.6}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4.7}$$

where TP, TF and FN with respect to the main text are defined as follows: TP is the main text correctly identified as main text. FP is the main text incorrectly identified as side notes. FN is side notes incorrectly identified as main text. Similarly, these metrics can also be identified with respect to the side notes.

We evaluated our method on 120 Arabic historical images collected from three different sources. The ground truth images are manually prepared using the pixel based representation. We evaluated the performance of our approach using different window sizes as shown in Tables 4.1-4.3. As shown in Table 4.2, the proposed methods, with projection profile, histogram of CC directions and histogram of gradients based orientation estimation, achieved an F-measure of 94.98%, 93.60% and 81.90% of correct segmented main text, respectively. Our method achieves an F-measure of 93.85% by combining the three methods for orientation estimation. By comparing the three text orientation estimation methods, it is clear that projection profile and histogram of direction based methods achieved higher accuracy than gradient based method. While the histogram of direction based method is comparable to the projection profile based method. Projection profile gives incorrect orientation estimation when there is no clear difference between valleys and peaks of the histogram. Gradient based method may give incorrect orientation estimation when large number of ascenders such as the letters "Alif" and "Lam" and descenders such as the letter "Mim" are present in the text. This makes the orientation estimation tends to be 90° opposite to the correct direction. Similarly, histogram of directions depends on the computed directions for connected components which is affected by the positions of the ascenders and descenders in the connected components. For example, words beginning with ascenders and/or ending with descenders. Connected components directions are also affected by handwriting variations. Two instances of the same word in the main text may have different directions due to handwriting variations. So, histogram of directions may give incorrect orientation estimation for text regions when most of its connected components have the foregoing characteristics.

Table 4.1: Overall performance results using a window of 3 lines

	Side notes			Main text		
Orientation estimation	P	R	F	P	R	F
Projection Profile (PP)	85.43%	90.81%	86.99%	94.81%	91.77%	**92.51**%
Histogram of CC (HCC)	84.16%	89.98%	84.91%	90.82%	89.42%	89.79%
Histogram of Gradient (HGD)	90.71%	69.24%	76.79%	68.71%	87.69%	73.87%
Combined (PP, HCC, HGD)	83.42%	95.22%	86.98%	96.56%	91.25%	**91.68%**

Table 4.2: Overall performance results using a window of 4 lines

	Side notes			Main text		
Orientation estimation	P	R	F	P	R	F
Projection Profile(PP)	84.42%	93.74%	88.41%	97.65%	92.84%	**94.98**%
Histogram of CC (HCC)	86.64%	95.67%	87.96%	96.34%	91.00%	93.60%
Histogram of Gradient (HGD)	90.06%	76.75%	78.95%	79.59%	93.68%	81.90%
Combined (PP, HCC, HGD)	90.29%	94.84%	88.33%	97.43%	91.53%	**93.85%**

Table 4.3: Overall performance results using a window of 5 lines

	Side notes			Main text		
Orientation estimation	P	R	F	P	R	F
Projection Profile (PP)	83.92%	92.98%	86.78%	94.76%	93.25%	**94.50**%
Histogram of CC (HCC)	85.07%	95.24%	87.66%	94.71%	93.11%	93.31%
Histogram of Gradient (HGD)	86.73%	77.82%	79.01%	83.22%	92.12%	84.81%
Combined (PP, HCC, HGD)	86.61%	91.65%	87.41%	97.02%	92.39%	**93.32**%

To study the effect of the size of the window, we experimented with three different window sizes, a window of 3 lines, 4 lines and 5 lines. It is clear that a window with size of 4 lines give better results using projection profile and histogram of connected components directions. While histogram of the gradient give the best result using a window of 5 lines. This may be contributed to fact that the larger the window, the less the percentage of the ascenders and descenders. Figure 4.13 shows the results of our appraoch on one document in our dataset.

Figure 4.13: Examples of extracted main text and side notes

4.3.1 Comparison with Previous Works

In order to show the viability of new techniques, one has to compare them with previous work that was done and tested on the same data. In this regard, the only research work that addressed analysis of Arabic manuscripts classifying their contents into main text and sid notes to the best of our knowledge was that of [21]. We did contact the authors to obtain the data they used in their experimentation, but without success. Therefore, it may not be fair to compare the results they achieved on their data to the results we achieved on our data. However, for the sake of completeness, we put forth the best results that we achieved next to their best results in Table 4.4. The performance of our algorithm depends on the text regions that are horizontally and vertically located at the middle of the image where the window moves. As a result, some side notes were identified as main text which lead to the low F-measure of the side notes compared to that of the main text. The reason for that is that side notes have the same orientation of the main text, see Figure 4.14. However, all sample pages

Table 4.4: Our results and results of [21] in terms of F-measure

Methods	Side notes	Main text
Our method	88.41%	94.98%
Bukhari et al. [21]	94.68%	95.02%

shown in [21] had side notes and main text with different orientations. One sample page is shown in Figure 4.15.

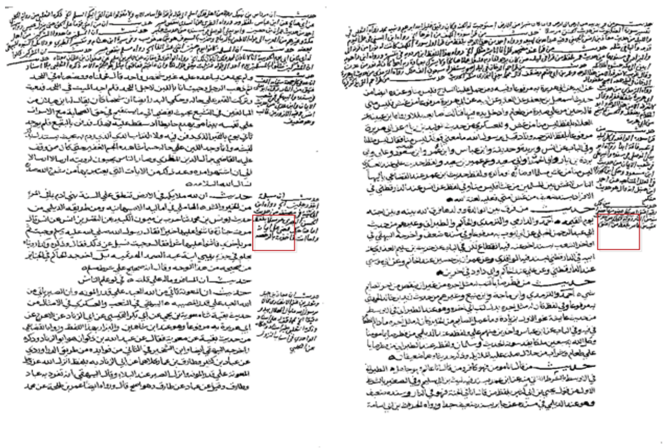


Figure 4.14: Sample pages from our data with side notes having same orientation as main text

Figure 4.15: Sample page of the data used in [21]

4.4 Conclusions

In this chapter, a new method for text region extraction is presented. The method is based on the fusion of foreground analysis, background analysis and text orientation estimation. A window based text analysis method is employed to separate the main text from side notes. Furthermore, two methods for text orientation estimation are proposed. They are based on constructing histogram of directions. The performance of the proposed methods are significantly affected by the characteristics of connected components. They are simple compared to the projection profile based method as finding the differences in the projection profile is not easy task [19]. Experimental results show that the proposed method has high performance results for main text and side notes extraction given the complex structure of Arabic manuscripts.

CHAPTER 5

ARABIC HISTORICAL DOCUMENTS WORD SPOTTING USING BAG OF WORD FRAGMENTS

Word spotting is one of the most challenging problems in pattern recognition. It has been a core topic in document image retrieval for the past two decades. Recent research on word spotting employ either global or local feature extraction approaches. In this chapter, we propose a novel Bag of Word Fragments (BoWFs) feature representation approach for Arabic handwritten word spotting, inspired by bag of features model and shape descriptors. In BoWFs, an Arabic word is decomposed into word fragments, each of which is described using Shape Context Descriptors (SCD) and Histogram of Oriented Gradients (HOG). These extracted word fragments are used to build our proposed model. Furthermore, we propose three word fragment extraction methods (viz. contour-based, skeleton-based and patch-based methods). Experimental results on a 50 Arabic historical documents has yielded promising results achieving a precision rate of 89.60% at 50% recall.

5.1 Introduction

Word spotting is the task of locating a specific keyword in a collection of document images where the query is given as a word image. Word spotting techniques have been introduced as an alternative to Optical Character Recognition (OCR) techniques for document image retrieval, since there is no efficient OCR technique for old historical documents. This is due to the fact that most historical documents are of poor quality and have vast variations in handwriting styles. In the literature, there are several works dealing with the problem of word spotting for handwritten document images and several features have been proposed. Nonetheless, the best method and set of features that deals with writing style variations have not yet been devised, as also witnessed by recent researches in this area [69, 28]. The main difficulties in this task arise from the style variation of the writing, including both writing variations and variability.

Historical document word spotting, as a specific case of document retrieval, can be viewed as an object matching problem in a feature space derived from the shape of the word. Therefore, it is well suited for local feature approaches. Local features are in general preferred over general features because the latter are not effective to handle writing style variations which may be available when the local features are extracted [136]. Many features have been explored, including profile-based features(such as upper and lower word profile and ink-transition[36], profile-oriented features [42]), structural and statistical features [27], Gradient, Structural and Concavity features [81, 80], Shape Coding [54, 26, 57], Contour Matching [63, 65], local features [68, 70, 72, 73] and hybrid features [82, 87]. The most common technique for feature matching is the Dynamic Time Wrapping (DTW) technique. Other matching techniques have received considerable attention as these methods are more effective and efficient than DTW [93]. Among the most recent results for Arabic handwritten word spotting, [28] implements a learning based model for historical document word spotting. It makes use of Regularized Discriminant Analysis (RDA) and Support Vector Machines (SVM). The authors claim that the proposed method accurately works when words in Arabic handwriting do not have clear boundaries between them. Recently, a huge amount of work and resources have been devoted to the introduction of different features and algorithms for the word spotting of handwritten images. The bag of features model has been extensively researched in and beyond the field of image retrieval, classification and categorization and there has been growing interest in using bag of features for word spotting. Even though it is still not that common in Arabic handwritten document word spotting now. In this chapter, based on bag of features model and shape context descriptors, a novel feature representation model, Bag of Word Fragments (BoWFs), is proposed. In this model, an Arabic word is decomposed into word fragments each of which is described using shape context descriptors based on the distribution of points on word skeleton, word contour and the word patches under polar coordinates.

The rest of this chapter is organized as follows: Section 5.2 gives an introduction about Bag of features model. Section 5.3 describes our BoWFs model. Section 5.4 describes our word spotting appraoch. The experiments and the performance results are reported in section 5.5. Finally, Section 5.6 concludes the chapter.

5.2 Bag of Features Model

Recently, bag of features model has been extensively researched in and beyond the field of image retrieval, classification and categorization and there has been growing interest in using bag of features model for word image retrieval. Even though it is still not very common in Arabic handwritten document word spotting now. Bag of features model, often powered with SIFT, is one of the most widely used techniques in the field of image classification, categorization and retrieval [137]. In this context, image retrieval aims at finding images similar to a respective query image. In contrast, image classification and categorization groups similar images into known categories. The bag of features model is borrowed from the text retrieval domain and is adapted for several applications in the computer vision domain such as image and video retrieval and classification[138].

To simplify the discussion, we will briefly introduce the bag of words model in text retrieval and then analogies used in the bag of features model will be given. In bag of word models, text retrieval systems consist of documents representing the retrieval database. Each document is represented by word vocabulary vector whose values corresponding to the count of words appearing in the document after excluding the most frequent and not discriminative words. Queries are also represented by word vocabulary vector. Document retrieval requires a way to measure similarities between word vocabulary vectors and query vector. Using the word vector representation described above, the distance can be computed by a suitable distance function such as the Euclidean distance of the word vocabulary vectors representation. The documents having a distance less than a predefined threshold will be returned in increasing order. More formally, the bag of words model can be defined as follows. Given a dataset D containing n text documents, a document in the database is represented by $W = \{w_1, w_2, w_3, ..., and\ w_m\}$ where w_i is a word, k documents can be retrieved based on word vocabulary vector W. Then we can summarize the data in a $n \times m$ matrix N where N_{ij} represents the frequency of the word w_i in document d_i.

In bag of features model, Visual words, called visual vocabulary, which are obtained by clustering of features(or descriptors) extracted from local image regions centered at points called key-points correspond to word vocabulary in the bag of word model [138]. For image representation purposes, the extracted features are quantized with respect to these visual vocabulary words obtained by clustering the whole features extracted from all images. The bag of features then reflects the frequency of each visual word. This representation does not have any spatial information of the visual words. The following figure presents a general flow diagram of image representation using the bag of features model.

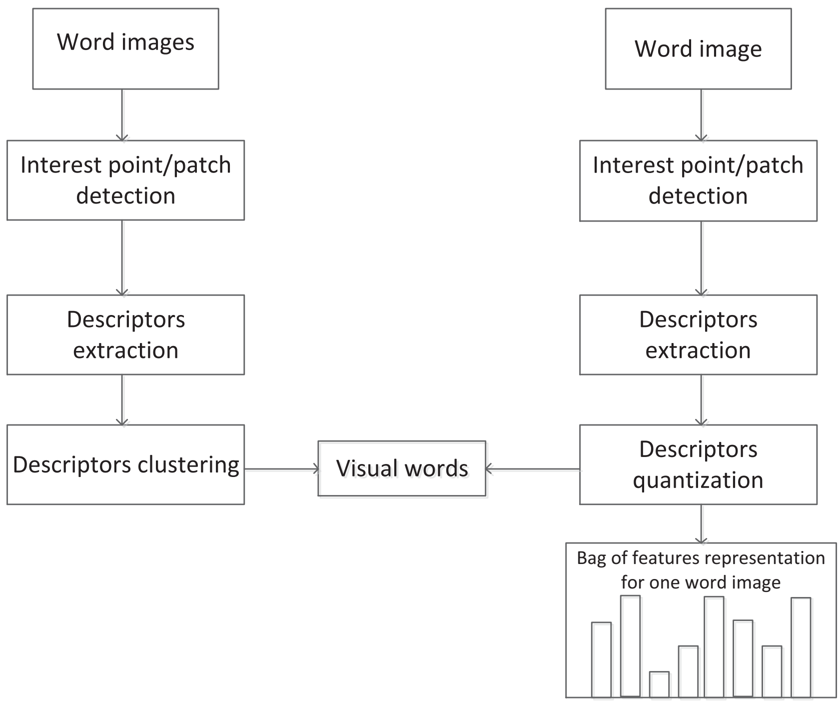


Figure 5.1: Bag of features model

5.2.1 Interest-point Detection and Description

Interest point detection and description are crucial steps in bag of features model. It is first applied to detect local interest points or local interest regions and to extract image descriptors of these points. These descriptors are used to represent images as a collection of features locally computed from a set of small regions, called patches, which in turn are considered as candidates for visual words [139]. In contrast to global feature extraction approaches in which features are extracted from the whole image to obtain a holistic representation of the image, local image features are extracted from specific image patches as different image parts are not equally important. This process is preferred as it allows to focus on the prominent parts of the image and it is robust to occlusion and clutter. After this detection process, feature descriptors are extracted for each patch. These descriptors should be discriminative for that patch and being robust to intensity changes caused by translation, scale, rotation or illumination changes.

Interest-point Detection

Interest point detection aims to find regions that are discriminative for a given input image. Several region detectors, that have been proposed in the literature, are described below[140, 141].

- *Laplacian-of-Gaussian (LoG):* in LoG, feature points are localized at local scale-space maxima of Laplacian-of Gaussian which is obtained by successive smoothing of an image with different-sized Gaussian based kernels.

- *Difference-of-Gaussians (DoG):* in DoG, scale-space representation is constructed by a combination of smoothing and subsampling operation of two successive smoothed images. Regions are detected at scale-space maxima of difference-of-Gaussian.

- *Harris-Laplace:* involves the localization of regions at scale of Harris function and scale-space maxima of Laplacian-of-Gaussian.

- *Hessian-Laplace:* involves the localization of regions at local maxima of Hessian determinant and in scale-space maxima of Laplacian-of-Gaussian

- *Salient Regions:* involves the localization of regions at local-space maxima of the entropy. For each image position and different-sized circular regions, histograms of the entropy of pixel intensity is computed.

- *Maximally Stable Extremal Regions (MSER):* are regions of connected pixels in binary image. Region boundaries that are stable over several threshold values are segmented using segmentation algorithm.

- *SIFT detector* is one of the most popular methods that have been widely used for point detectors [137]. In SIFT, points are detected in a difference-of-Gaussian scale representation of the image. The goal of using different scales is to achieve the scale invariance property. The image is convolved with Gaussian functions to construct scale space representation as defined in Equation 5.1 and shown in Figure 5.2.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{5.1}$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \tag{5.2}$$

where I is the image, $*$ is the convolution operation and the Gaussian G is given in equation 5.2. Then, edges are extracted using a difference-of-Gaussian kernel which results from subtracting two Gaussian functions that have two different σ as shown in equation 5.3.

$$D_G(x, y, \sigma) = G(x, y, k\sigma) - G(x, y, \sigma) \tag{5.3}$$

where D_G approximates the Laplacian-of-Gaussian function that is found to be suitable for interest point detection [142]. Then, the image is convolved with the kernel D_G to obtain the difference-of-Gaussian scale space representation in which interest points are detected by using a 3-D non-extremum-suppression. Finally, neighbourhood of each point in the same and adjacent scales is examined and some points are discarded based on edges contrast.

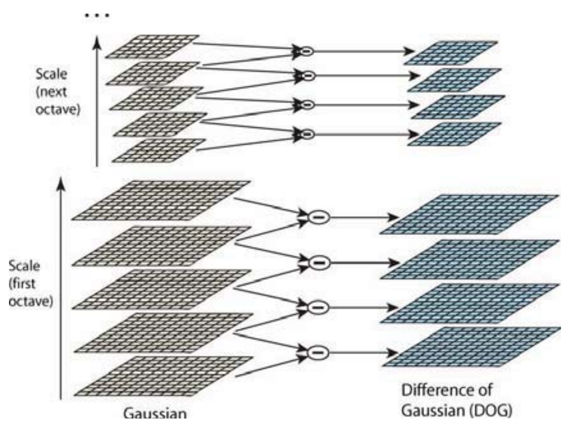


Figure 5.2: scale space representation

Interest-point Description

One of the key issues for bag of features model is to find an effective feature description for interest points, that is a local description of the neighbourhood of the point that is expected to have some invariance properties. A number of powerful descriptors which can be used in region description were proposed in the literature(SIFT[18], SURF[143], BRIEF[144], BRISK[145], HOG[146], LBP[147], HOG-LBP[148], SIFT-LBP[149]). Generally, two types of methods can be distinguished in the literature: intensity based and gradient based [137]. In the subse-

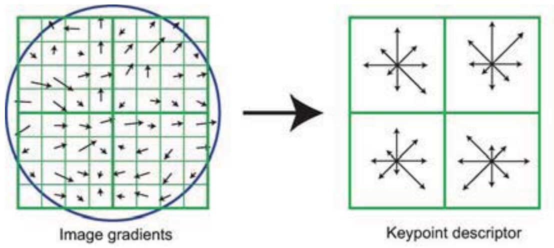


Figure 5.3: Image gradients and key point descriptors

quent section, SIFT descriptor will be discussed as an example of a well-known gradient based method.

5.2.2 SIFT Descriptors

SIFT is a well-know approach to compute a discriminative representation of patches within an image. It was originally proposed by Lowe [18] and shows the alignment and the distribution of the local gradients at different scales around a point [150]. SIFT descriptors have been found very effective at representing the image patches and they are rotation, scale and illumination invariant. The computation of SIFT descriptors is shown in Figure 5.3. Image gradient and orientation are first sampled with respect to key point location. Then, key points neighbourhood is divided into subregions that are arranged relative to the key point orientation. Gradient magnitudes are also rotated relative to the orientation and weights are assigned to the gradient magnitude using a Gaussian weighting function. Finally, Gradient magnitudes are accumulated in each sub-regions with respect to eight histogram orientations. As a result, a 128-dimensional descriptor vector is obtained using 4×4 subregions.

5.2.3 Clustering and Quantization

Clustering is defined as the process of grouping similar feature vectors into one group called cluster. Clustering plays an important role in bag of feature model as they determine the length of the final feature vector that describes the image patches or the key points. In general, an essential choice is the number of clusters. A large number of clusters is more accurate and needs more space while small number of clusters might be required for some applications. Assigning descriptors to constructed clusters is referred to as quantization. It is application dependent to determine the proper number of clusters as it is an important factor when performing SIFT quantization. The most widely used vector quantization algorithm is the k-means algorithm. A critical issue is the size of the vocabulary of visual words used in the model. Big-size vocabulary is not efficient due to the fact that they do not represent discriminative details. In contrast, a vocabulary which is too small ignores important details of the image[151].

5.3 Bag of Word Fragments

In our model, we treat an Arabic word as a shape represented by a set of parts (fragments) extracted from the interior and/or exterior part of the word. Our method is based on segmenting a given Arabic word into smaller parts called word fragments. Each word fragment is described by using a set of descriptors. Given an Arabic word w, we show how to build BoWFs(w) for w and use BoWFs(w) for word spotting.

5.3.1 Word Fragments Extraction

In this section, we describe how to segment a given word into smaller pieces called word fragments. We propose three methods to extract these fragments. The first method is based on extracting the contour of the word and the second method is based on the skeleton of the word. The third method is a patch based method that segments a word into small patches and extract each patch as a word fragment. The details of each method is given in the following sections.

Contour-based Method

In shape matching, contour fragments represent powerful shape features [152]. In this work, we use word contour fragments as a basic unit for building our word representation since they contain both global and local word information. The contour of a word represents the boundary and exterior parts of a word more than the interior part. For segmenting the whole word contour into a meaningful word fragments, we use the neighbourhood of each contour point as the basic units to build our model. We start by extracting the contour of the word using the Moore-Neighbor tracing algorithm [153]. Then, for each contour point p, the contour fragment whose length is n centered at p is considered as a word fragment. Figures 5.4 and 5.5 show examples of word fragment extraction using our contour-based method.

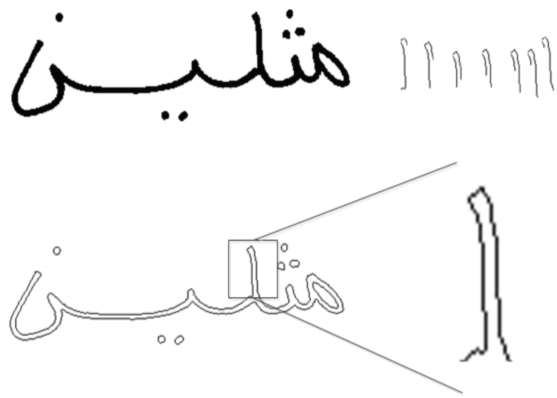


Figure 5.4: Fragments extraction using contour-based fragment extraction method

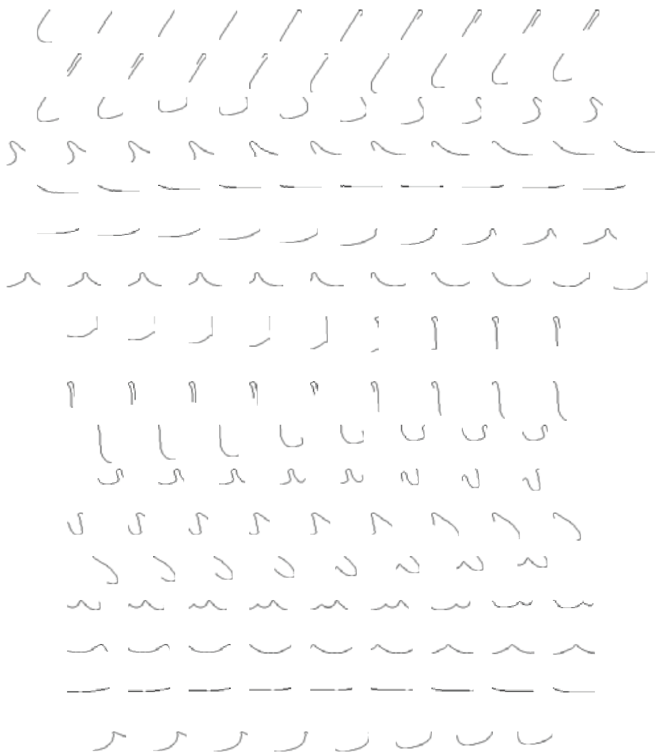


Figure 5.5: Example of contour based fragments extraction from the Arabic word shown in Figure 5.4 sampled at every fifth point

Skeleton-based method

Skeletons are often used for shape description and matching [154]. In the context of handwritten recognition, it can be used to represent words. A skeleton is defined as a one-pixel thick representation that shows the centerline of a word. Unlike the contour, the skeleton represents the central part of the word more than the boundary. It is useful for feature detection, word classification and segmentation. In this section, we describe how we use the skeleton of a word to obtain its word fragments. We propose two skeleton-based methods to extract word fragments.

Skeleton-grapheme-based method

To extract word fragments from the skeleton of the words, we make us of a method that is based on the segmentation method proposed in [155]. Given the skeleton image of a word as shown in Figure 5.6, the word fragments can be extracted as follows. Let $I(x, y)$ be the thinned image of the word image W, let P be a foreground pixel in $I(x, y)$ and $N_8(P)$ the 8-neighbourhood set of P. A set of pixels, called possible segmentation pixels ($PSPs$), is identified. ($PSPs$) contains all pixels with 3-4 pixels in their 8-neighbourhood.

$$PSPs = \{P|(N_8(P) = 4 \vee N_8(P) = 3)\} \tag{5.4}$$

All PSPs points are used to segment the word into fragments. Figure 5.7 shows a thinned word image with all possible ($PSPs$). The foreground pixels of ($PSPs$) in the thinned word image are converted to background pixels. This allows to separate the thinned word image into several connected components. Then, connected component analysis is performed to obtain a set of connected components in the thinned image. These connected components are the extracted word fragments of the word w. Figure 5.9 describes an example of word fragment extraction using the skeleton -based method.

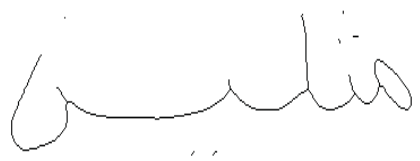


Figure 5.6: Skeleton image of an Arabic word

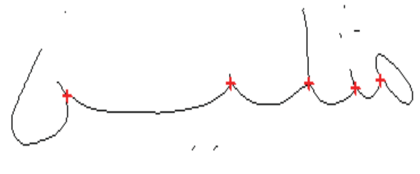


Figure 5.7: Skeleton image of an Arabic word with all possible ($PSPs$)

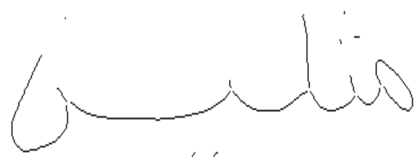


Figure 5.8: Skeleton image of an Arabic word with all possible (PSPs) converted to background pixels

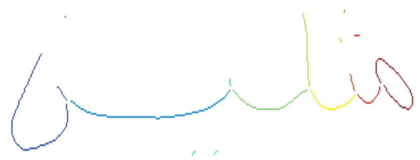


Figure 5.9: Example of graphemes based fragments extraction

Skeleton-points based method

For segmenting the whole word skeleton into meaningful word fragments, we use the skeleton points as the basic units to build our model. For each skeleton point p, the skeleton part whose length is n centered at p is considered as a word fragment. Figures 5.10 and 5.11 describes an example of word fragment extraction using our skeleton-based method.

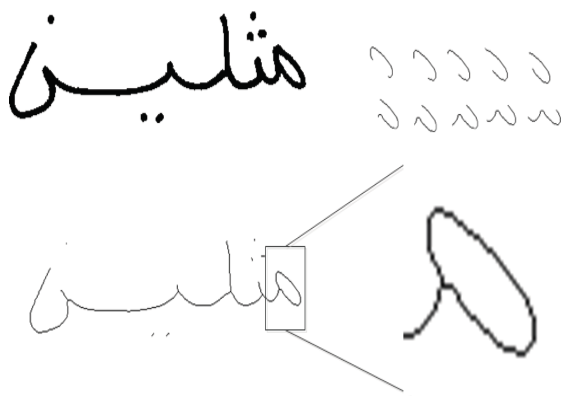


Figure 5.10: Fragments extraction using skeleton-points fragment extraction method

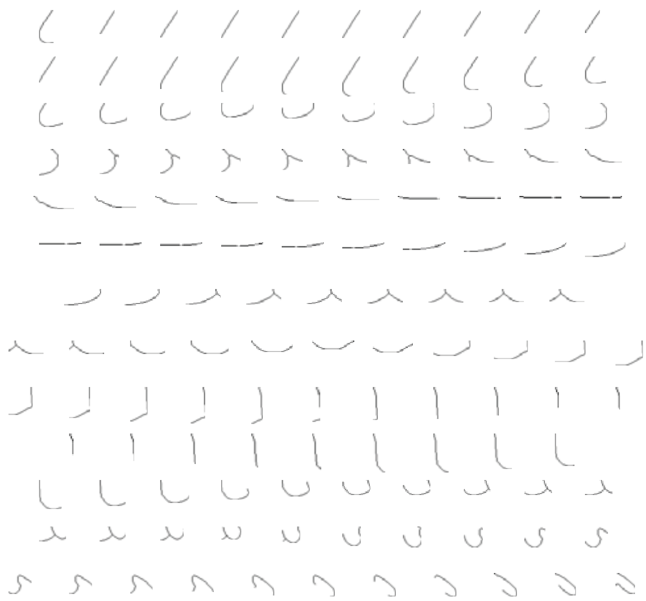


Figure 5.11: Example of skeleton-points based fragments extraction from the Arabic word shown in Figure 5.10, sampled at every fifth point

Patch-based Method

In contour-based methods, the boundary pixels of a word are considered to obtain the word representation. Similarly, in skeleton-based method only the interior pixels of a word are taken into account. However, in patch-based methods, all the pixels within a word are used to create the word representation. The image is partitioned into smaller regions of size $n \times n$. Each region represents one word fragment. In all fragment extraction methods, the number of extracted word fragment differs from one word to another depending on the characteristics of the word.

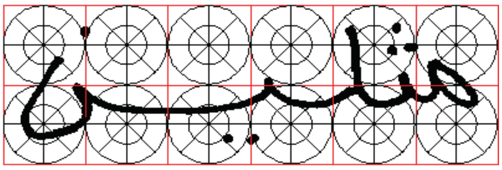


Figure 5.12: Word description using the patch-based fragment extraction method

5.3.2 Word Fragments Description

In this section we describe how to encode the obtained word fragments. After word fragments extraction, a set of features for each word fragment is extracted. Several features have been proposed in the literature. We employed two types of features for word fragment description. The first one is the Shape Context

Descriptors (SCD) while the second one is the Histogram of Gradient (HOG).

Shape context descriptors

Different shape code features can be employed to describe each word fragment. Interested readers are referred to [156] for detailed descriptions of such methods. In this study, we make use of the shape context descriptor proposed in [59]. This descriptor captures the distribution of points of the shape and describes global characteristics of the shape. Shape context descriptors are computed as follows. We divide the word image into several circles and we partition the circles into several bins using equal angle. Each bin is described using the triple $H_i = (r_i, \theta_i, n_i)$ where r_i is the radius of the circle, θ_i is the angle space and n_i is the number of points inside the bin. Figure 5.13 shows a graphical illustration of shape context description computation for an Arabic word. The resulting descriptors are shown in Figure 5.14. However, we compute the SCDs for each word fragment.

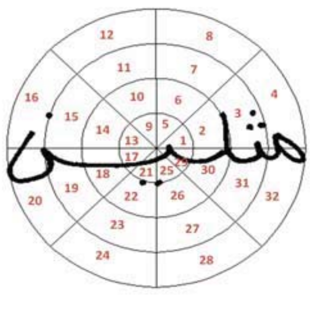


Figure 5.13: Shape context descriptors of an Arabic word

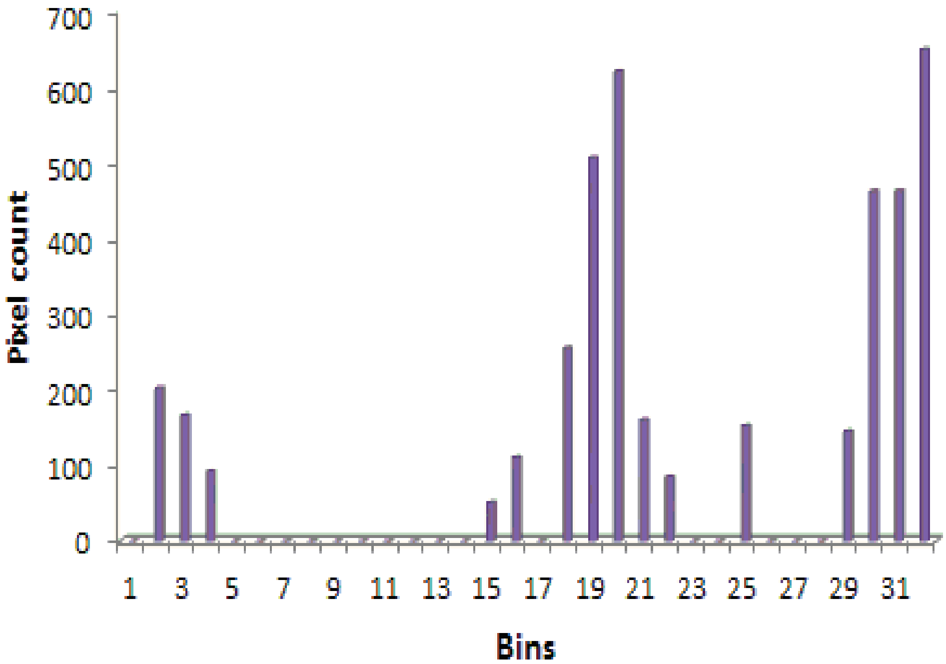


Figure 5.14: Histogram of the Shape context descriptors of the Arabic word shown in Figure 5.13

The histogram of oriented gradients(HOG)

HOG descriptors can be defined as a histogram of the occurrences of the gradient orientations in images. The angles (orientations) of the gradient $f(x,y)$ are computed as in Equations 5.5.

$$Direction : \theta(i,j) = tan^{-1}(\frac{v}{u}) \tag{5.5}$$

where u and v are computed by convolving two 3×3 Sobel operators with the binary image. Each word fragment is divided into 4 cells. Gradient directions in each cell are quantized into eight directions as shown in Figure 5.15. Finally, the Word fragment description using HOG is constructed by concatenating the four histograms as shown in Figure 5.16.

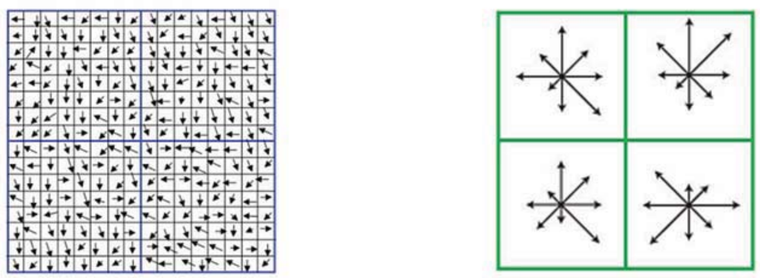


Figure 5.15: The histogram of oriented gradients

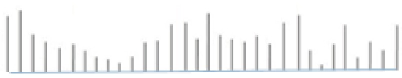


Figure 5.16: Word fragment description using HOG descriptors

5.3.3 Codebook Generation

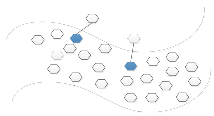
This stage is devoted to creating a codebook using the descriptors obtained in the previous stage. Researchers have proposed many codebook learning methods including supervised methods and unsupervised methods [157, 158, 159]. In this work, we choose k-means clustering algorithm for its simplicity and stability [160]. After feature extraction of each word fragment, k-mean clustering is performed. The obtained class centers are defined as the visual words which form the codebook. The learned codebook contains all visual words that correspond to word fragments obtained during word fragments extraction process. It is used for describing the whole word fragments space.

Algorithm 5 Codebook generation

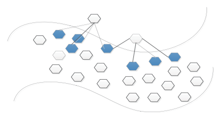
1. Input document images.
2. Segment the training data into PAWs and segment PAWs into small pieces called word fragments using one of the 3 mentioned methods.
3. Extract SCDs and HOG for each word fragment.
4. Cluster the extracted shape descriptors of all word fragments using K-Mean clustering algorithm into n clusters which are defined as the visual words.
5. Create the codebook using the obtained visual words.

5.3.4 Construction of BoWFs

The aim of this stage is to represent each word using the learned codebook. To represent words, the most widely used method is vector quantization [74]. In vector quantization, a word fragment feature vector is assigned to its nearest neighbour in the codebook. Let W be a word segmented into n word fragments, i.e $W = \{f_1, f_2, \cdots, f_n\}$. Each word fragment f_i is described by a feature vector of length m, i.e $f_i = \{d_1, d_2, \cdots, d_m\}$. Given a codebook with m entries, $CB = \{c_1, c_2, \cdots, c_m\}$, each word fragment f_j of a given word is represented by the most similar visual word c_k where the distance D_{jk} between f_j and c_k is the smallest distance among all distances between f_j and all visual words. It is re-



(a) Hard assignment



(b) Soft assignment

Figure 5.17: Hard vs. soft assignments

ported in the literature that assigning local descriptors of an interest point to its nearest neighbour is not an optimal choice[151]. This is due to the fact that clustering algorithm may put two similar points into different clusters. To tackle the drawbacks of the conventional assignment scheme, researchers proposed a variant of the conventional assignment scheme. Instead of assignment of a given local descriptors to one visual word, it is assigned to k-nearest neighbour visual words, soft assignment scheme as shown in Figure 5.17. In our work we use soft assignment scheme and we choose 5 as the value of k. To summarize, the Bag of Word Fragments histogram is employed to represent the distribution of word fragments in a give word. It represents the frequency of word fragment occurrences in the word and provides a concise description of the distribution of the word fragments of the word. By comparing BoWFs representation of two word images, one can determine the similarity between these two words. The complete algorithm of building the BoWF can be summarized as follows.

Algorithm 6 $BoWF$ Construction

1. Input a word image.
2. Segment the input image into word fragments using one of the 3 mentioned methods.
3. Extract SCDs and HOG for each word fragment.
4. Assign each word fragment descriptors into the k-nearest neighbours using the learned codebook constructed using algorithm 5.
5. Represent each word using a histogram of word fragment occurrences whose elements are the frequency of each word fragment in the word.

5.3.5 SCD-HOG feature integration

SCD descriptors describes word fragments by constructing histograms of the occurrences of the pixel while HOG describes word by constructing histograms of the occurrences of the pixel orientations. We believe that the shape of a word can be better captured by combining these two descriptors. We propose a histogram-based SCD-HOG integration method. We propose to compute SCD and HOG descriptors for all word fragments. Then, for each type of descriptors, we independently construct a BoWFS representation for SCD and HOG. We represent each word by concatenating the two representation. The complete algorithm of SCD-HOG feature integration can be summarized as follows.

Algorithm 7 Histogram-based SCD-HOG descriptor integration

---

1. Input a word image w.
2. Segment w into word fragments using one of the 3 mentioned methods.
3. Compute SCD for each word fragment.
4. Compute HOG for each word fragment.
5. Build BoWFs representation for w using SCD descriptors, call it $hist_{SCD}$.
6. Build BoWFs representation for w using HOG descriptors, call it $hist_{HOG}$.

7. $hist_{HOG}$ is directly concatenated to the end of $hist_{SCD}$, thus w can be described as SCD-HOG=[$hist_{SCD}$ $hist_{HOG}$].

5.3.6 Similarity of the BoWFs

Histogram based representation such as BoWFs can be compared using bin-wise similarity and dissimilarity techniques such as Histogram Intersection (HI) and Earth Movers Distance (EMD). Histogram intersection [161] is a suitable measure to compute the similarity between two BoWFs representation of two Arabic words. Histogram intersection between histogram H and histogram K finds the total overlap between histogram H and histogram K. The similarity can be expressed using the following Equation [91].

$$Dist\,(H,K) = \frac{\sum min\,(H,K)}{min\,(|H|,|K|)} \tag{5.6}$$

where $|H|$ and $|K|$ denotes the norm of the vectors H and K respectively. Another dissimilarity measure can be used to compute the dissimilarity between two BoWFs representation is the Earth Movers Distance (EMD) [162]. It captures the difference between two histograms and has been successfully used in many fields of image matching and retrieval. EMD between histogram H and histogram K is defined using equation 5.7.

$$Dist\,(H,K) = \sum \frac{(H_i - m_i)}{m_i} \tag{5.7}$$

where $m_i = \frac{H_i+K_i}{2}$

5.4 Proposed word spotting approach

This section presents a word spotting approach for Arabic handwritten documents. We fist present a general overview of the proposed approach and then we give a detailed description. The proposed word spotting system basd on the BoWFs feature representation is illustrated in Figure 5.18. It contains two phases: training and testing. In the training phase, we start by decomposing words into word fragments. Then SCD and HOG descriptors are used to encode the generated fragments. After that, the descriptors of all word fragments are clustered using k-means clustering algorithm and a codebook is generated. For each word in the testing data, it is decomposed into word fragments. Then SCD and HOG descriptors are used to encode the generated fragments. The BoWFs representation is constructed by quantizing their word fragments using the learned codebook. Similarly, a given query image is decomposed into word fragments and its BoWFs representation is constructed by quantizing their word fragments using the learned codebook. Details of each step are explained in the following subsections. Using the similarity measures, the similarity between the BoWFs representation of the query and the BoWFs of all words in the testing data is computed. Finally, the stored word images are retrieved and ranked according to the computed similarity. These steps are summarized as follows.

1. The input is 50 Arabic historical document images.

2. Preprocessing: Noise removal, binarization, text line and PAWs extraction.

3. The input data is split into 30 pages for training and 20 pages for testing.

4. Codebook generation: the training set is used to build the the codebook as described in algorithm 5.

5. $BoWF$ construction: the testing set is used to build the $BoWF$ model as described in algorithm 7.

6. Query representation: the $BoWF$ of the query is constructed as described in algorithm 7.

7. Matching: the similarity between the BoWFs representation of the query and the BoWFs of all words in the testing data is computed.

8 Results ranking: the stored word images are retrieved and ranked according to the computed similarity.

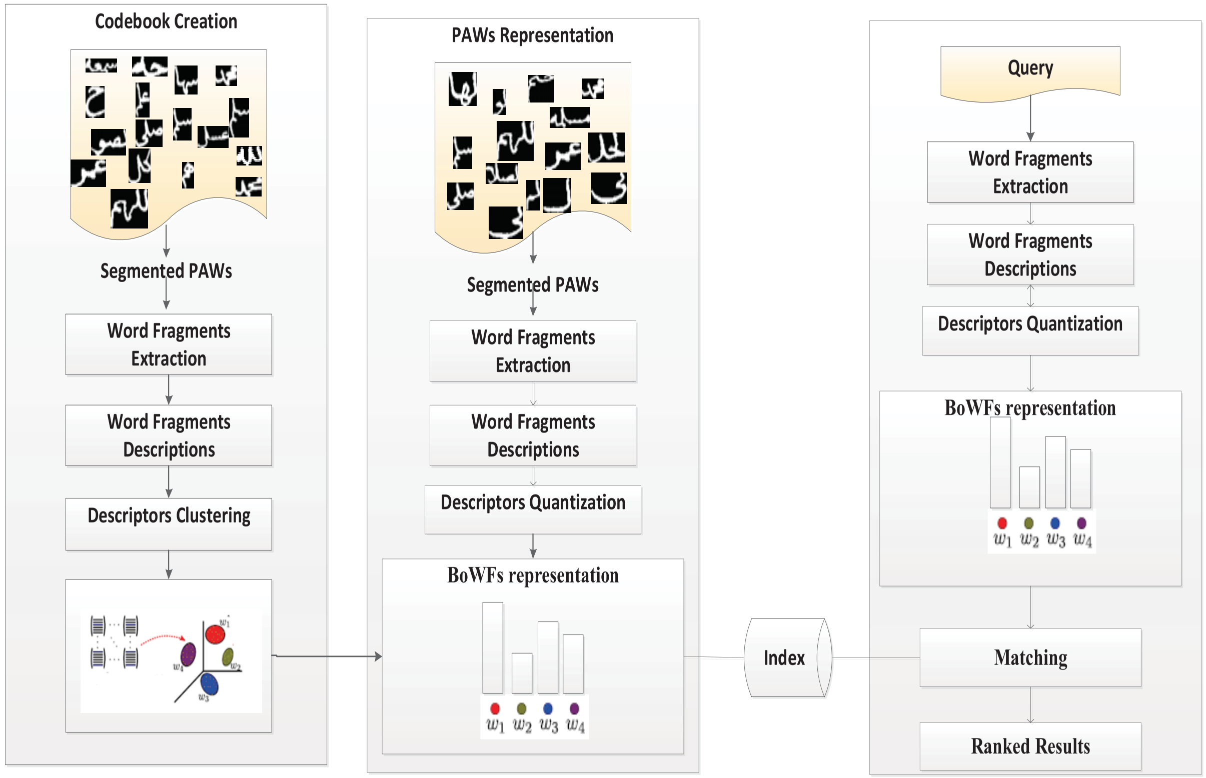


Figure 5.18: Word spotting using th prposed BoWFs representation

5.5 Experimental Results and Analysis

After presenting our model for word spotting, in this section we describe the experimental methodology, datasets and performance metrics used in the experimental analysis. The proposed method is first evaluated using printed dataset. Then, we apply the proposed method to historical documents corresponding to the printed dataset. Each dataset consists of 50 document images. The training and testing data contains 8238 and 5191 words respectively. We used 20 queries as shown in Table 5.1. To evaluate the performance of the proposed word spotting, three common measures of precision, recall and F-measure are used. Precision is defined as the ratio of the the number of the correctly spotted words over the total number of spotted words.

$$Precision = \frac{No\ \ of\, correctly\ \ spotted\ \ words}{No\ \ of\, spotted\, words} \tag{5.8}$$

Recall is defined as the ratio of the the number of the correctly retrieved words over the total number of correct words.

$$Recall = \frac{No\ \ of\, correctly\ \ spotted\ \ words}{No\ \ of\, correct\, words} \tag{5.9}$$

$$F - Measure = \frac{2 * precision * recall}{precision + recall} \tag{5.10}$$

Table 5.1: List of queries

ID	Query	Image
Q1	Haddathana	حدثنا
Q2	Salla	صلى
Q3	Alaihi	عليه
Q4	Sallam	سلم
Q5	Abd	عبد
Q6	Annabi	النبى
Q7	Mohammed	محمد
Q8	Saeed	سعيد
Q9	Shoaba	شعبة
Q10	Malik	مالك
Q11	Sameat	سمعت
Q12	Omer	عمر
Q13	Sufian	سفين
Q14	Hameed	حميد
Q15	Saad	سعد
Q16	Obaid	عبيد
Q17	Sulaiman	سلمان
Q18	Shoaib	شعيب
Q19	Yahia	يحي
Q20	Albait	البيت

5.5.1 Performance on Printed Dataset

This section evaluates the performance of the proposed method using printed dataset. The performance results, in terms of Precision at Recall=1, using histogram intersection and Earth Mover Distance are summarized in Tables 5.2 and 5.3 respectively. The results show that contour and skeleton based extraction methods perform significantly better than patch and grapheme based methods, achieving excellent precision of 99.20%. The precision value of skeleton-points method is competitive with the performance of the contour based method of 99.0%. Comparing the patch and grapheme methods, we can see that the patch based method performed better than the grapheme based method. The proposed approach with grapheme based extraction method has the lowest precision values. This may be attributed to the fact that the BoWFs representation using the grapheme based representation is sparse leading to the observed moderate precision value. Patch-based fragment extraction also has moderate precision values. The reasons for that is the presence of non-discriminative patches. To summarize, both contour and skeleton points are the most effective points that are used for word spotting. HOG descriptors has better precision values than the SCD descriptors. For all four fragment extraction methods, performances improve as we integrate SCD and HOG. The result of SCD-HOG is intuitive because words can be more distinguished by both pixels distribution and pixels orientations.

By comparing the EMD and HI, EMD measure is slightly better than HI and able to capture the similarity between keywords and word images better than histogram intersection for all fragment extraction methods except the grapheme based method.

Table 5.2: Precision at Recall=100% of the proposed method with Histogram Intersection on printed documents

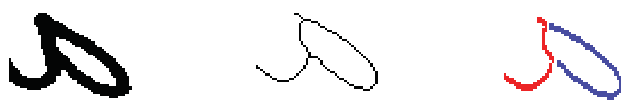
Methods	SCD	HOG	SCD-HOG
Patch-based	79.45%	81.13%	82.75%
Contour-based	**94.98%**	95.54%	**98.17%**
Grapheme	78.60%	80.87%	81.71%
Skeleton-Points	92.31%	**95.12%**	97.86%

Table 5.3: Precision at Recall=100% of the proposed method with Earth Mover Distance on printed documents

Methods	SCD	HOG	SCD-HOG
Patch-based	80.08%	81.43%	83.40%
Contour-based	**95.32%**	**97.12%**	**99.20%**
Grapheme	78.15%	78.08%	81.50%
Skeleton-Points	95.11%	96.10%	99%

5.5.2 Performance on Historical Dataset-1

In this section, we evaluate the performance of the proposed method using Arabic historical documents. In Table 5.4 we report the performance in terms of precision for the proposed methods for word fragment extraction. The best performance (precision = 89.60%) is obtained using the integrated SCD-HOG descriptors, with the contour based word fragment extraction method. The HOG descriptors presented more than 2% better performance than SCD considering the contour and skeleton based methods, while they presented about 1% better performance than SCD considering the batch and grapheme method. Both the batch and grapheme based methods generally gave the lowest precision values. The likely reason for this stems from a low number of word fragments per word in case of grapheme methods. This results in sparse BoWFs representation leading to the observed precision values. Furthermore, word fragment extraction depends on the skeletization algorithm. This is clearly shown in Figure 5.19 where two word fragments are extracted instead of one word fragment due to the handwriting style variation. This



(a) input image (b) skeleton (c) fragments

Figure 5.19: Effects of the skeletization algorithm: two word fragments are extracted instead of one word fragment

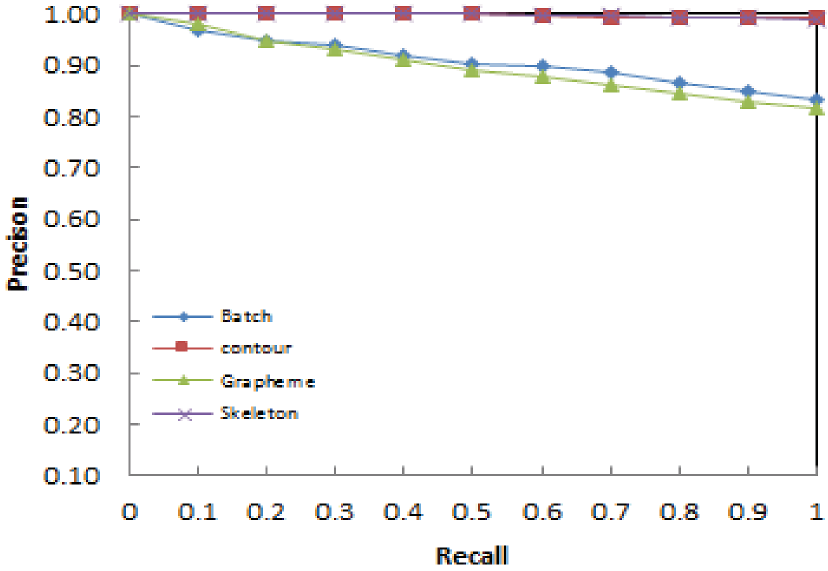
Table 5.4: Precision at Recall=50% of the proposed method with Earth Mover Distance on historical documents

Methods	SCD	HOG	SCD-HOG
Patch-based	46.10%	47.10%	50.70%
Contour-based	**83.18%**	**85.11%**	**89.60%**
Grapheme	43.15%	45.16%	48.19%
Skeleton-Points	81.10%	83.22%	87.40%

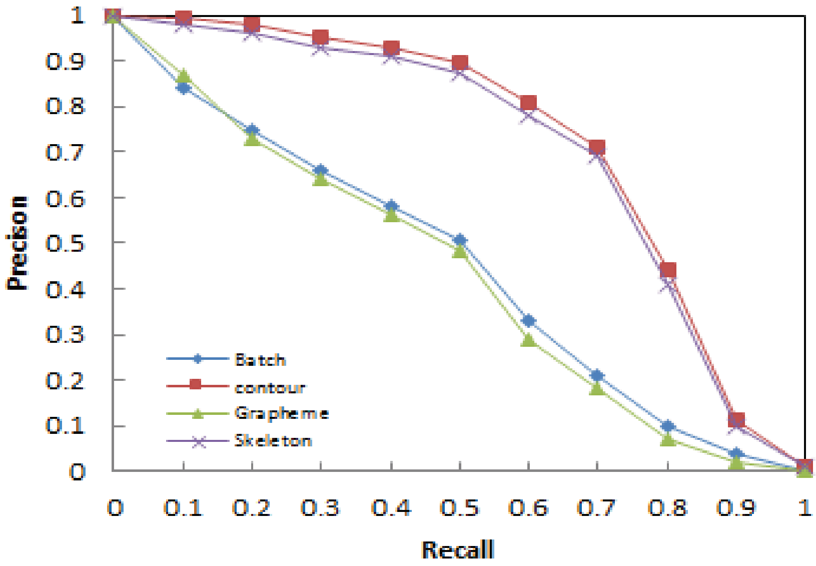
is an important result as it suggests that the performance of the BoWFs model imroves as the amount of word fragments extracted from the words increases. This is consistent with previous findings in the published literature regarding bag of visual words (BoVW) model' performance [163].

5.5.3 Analysis of the Results Considering the Recall and Precision

Figure 5.20(a) shows the precision-recall curve using printed dataset. The contour based fragment extraction method gave the best curve among the proposed fragment extraction methods. The skeleton-points based method gave the second best curve. It can be shown that the BoWFs using the grapheme based method performs better than using Patch based method with low recall values (less than 0.2) and vis-versa for higher recall values. The results show that both contour and



(a) Printed

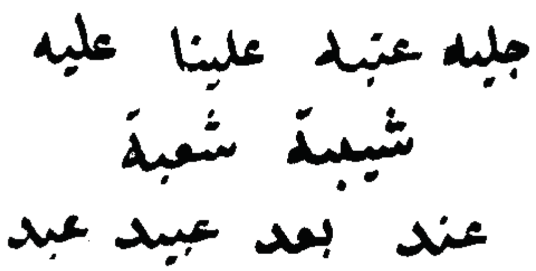


(b) Historical

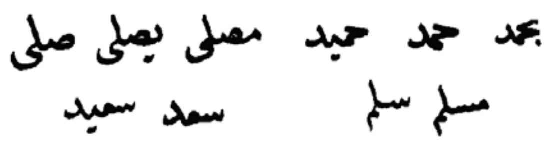
Figure 5.20: Precision-recall curves for the (a) printed dataset and (b) historical dataset

skeleton-points outperform the other methods and achieve almost equal precision and recall rate. This result shows the ability of BoWFs to spot words that perfectly match the keywords and to exclude words that do not match the keywords even though they are visually similar to the keywords .

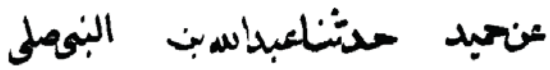
Figure 5.20(b) shows the precision-recall curve for the proposed methods using the historical dataset. Similar performance trends are obtained. The contour based fragment extraction method presented the best curve followed by the skeleton method. The Patch and grapheme fragment extraction methods have similar performance. The results show that both contour and skeleton-points methods achieve promising results with equal precision and recall rates of 70%. Most of the irrelevant spotted words are due to the confusion between words that differ in some character, the order of the characters and the position of the diacritic as shown in Figure 5.21(a). Furthermore, some of irrelevant spotted words have the same root of the keywords. This appears in the grammatical variations of the keywords as can be seen in Figure 5.21(b). Some missed relevant words are due to segmentation errors as can be seen in Figure 5.21(c).



(a) Similar Words with some differences.



(b) Grammatical variations



(c) Segmentation errors

Figure 5.21: Reasons for irrelevant words spotting and relevant words missing

5.5.4 Results Considering the F-Measure

Table 5.5 shows the performance in terms of F-measure for the two dataset. For printed dataset, the results show that contour based extraction method performs significantly better than patch and grapheme based methods, achieving excellent F-Measure (99.60%). On the other hand, the precision value of skeleton-points method is competitive with the performance of the contour based method (99.50%). Comparing the patch and grapheme methods, we can see that the patch based method performed slightly better than the grapheme based method. For the historical dataset, contour and skeleton based extraction methods achieved good F-Measure (70.52% and 69.55%), respectively. The patch and grapheme based methods achieved (50.35% and 49.08%), respectively. Tables 5.1 - 5.8 show used queries, top-10 retrieved images for some queries and the performance per query for printed and historical datasets.

Table 5.5: Overall performance results using F-measure

	Printed	Historical
Methods	P(R=1)	(P=R)
Patch	90.95%	50.35%
Contour	99.60%	70.52%
Grapheme	89.80%	49.08%
Skeleton-points	99.50%	69.55%

Table 5.6: Top -10 retrieved images for some queries in the historical dataset

Query	Results

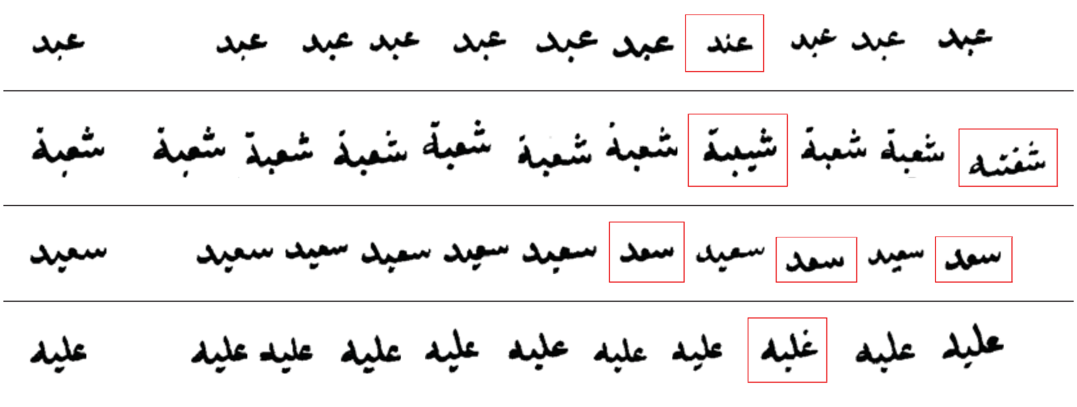


Table 5.7: Precision at Recall=50% per query of the historical dataset using the contour and patch fragment based extraction methods

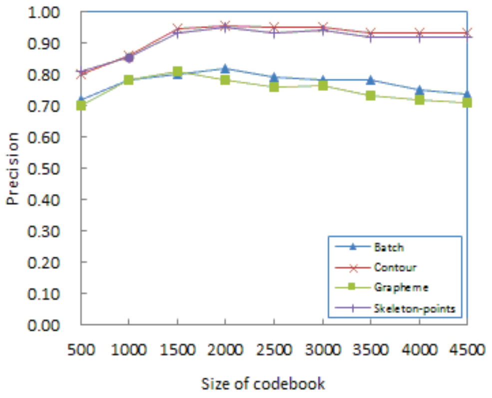
Keywords	Contour	Patch
Q1	85.48%	47.32%
Q2	93.48%	42.16%
Q3	89.36%	40.78%
Q4	89.13%	38.68%
Q5	83.33%	55.56%
Q6	90.91%	29.41%
Q7	88.89%	61.54%
Q8	77.78%	58.33%
Q9	87.50%	43.75%
Q10	80.00%	50%
Q11	100%	25%
Q12	100%	50%
Q13	100%	50%
Q14	85.71%	60%
Q15	85.71%	60%
Q16	100%	66.67%
Q17	71.43%	50%
Q18	100%	62.50%
Q19	83.33%	55.56%
Q20	100%	66.67%
Average	89.60%	50.70%

Table 5.8: Precision at (P=R$\simeq$ 70%) for each query of the historical dataset using the contour based extraction methods

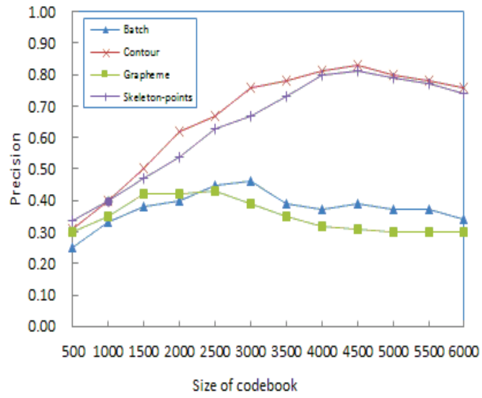
Keywords	Contour
Q1	71.15%
Q2	73.17%
Q3	74.68%
Q4	72.50%
Q5	71.43%
Q6	75.68%
Q7	64.71%
Q8	60%
Q9	69.23%
Q10	85.71%
Q11	83.33%
Q12	62.50%
Q13	66.67%
Q14	80%
Q15	66.67%
Q16	80%
Q17	58.33%
Q18	70%
Q19	63.54%
Q20	71.43%
Average	71.04%

5.5.5 Effect of Codebook Size

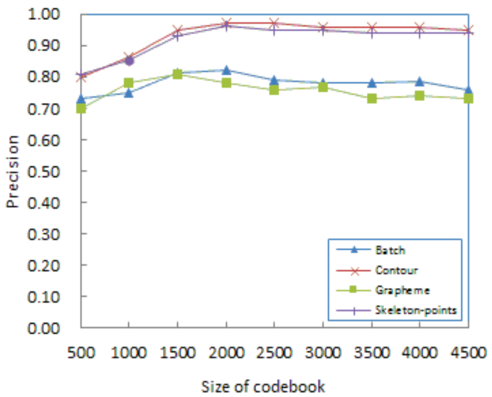
The bag of visual words model still reveals open problems and particulary the selection of the correct size of the codebook. To illustrate how the codebook size affects the performance, we do word spotting using word fragments codebooks of different sizes. We consider 9 and 12 different codebook sizes on the printed and historical datasets, respectively. The performance of BoWFs using codebooks of different sizes are reported in Figure 5.22. Generally, precision value improves as the size of codebook increases, but saturates when codebook size increases beyond 2000 and 4500 for the printed and historical datasets, respectively, with the contour and skeleton-points based methods. For Patch and grapheme based methods, the precision value improves until it saturated then precision decreases as we increase the codebook size. This decrease is more evident in case of the historical dataset. This indicates that the space of the word fragments of the handwritten documents is much larger than the space of word fragments of the printed documents. The reasons for that is the variations in handwriting style that is present in the historical documents even if it is written by the same author.



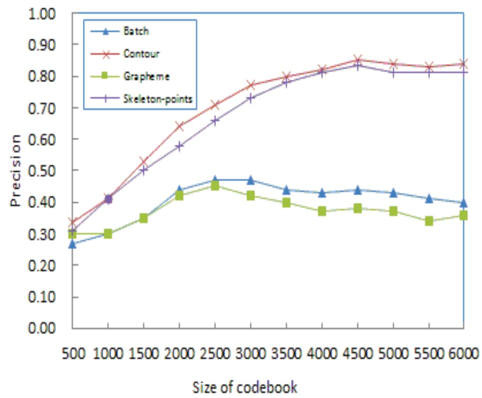
(a) SCD



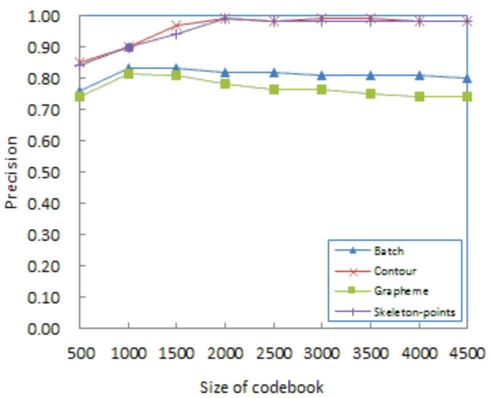
(b) SCD



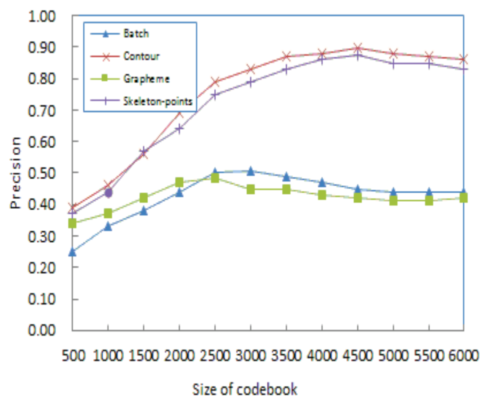
(c) HOG



(d) HOG



(e) SCD-HOG



(f) SCD-HOG

Figure 5.22: Mean average precision at different codebook sizes for the (a), (c) and (e) printed dataset, (b), (d) and (f) historical dataset

5.5.6 Performance on the Historical Dataset-2

In this section, we evaluate the performance of the proposed feature representation using the extracted main text in the analysis stage. We used 20 images for training and 10 images for testing. The result shows that the contour method using the integrated SCD-HOG descriptors achieve a precision rate of 76.63% at recall=50%. Table 5.9 shows the performance in terms of precision for each query. The decrease in the precision rates compared to previous dataset may be attributed to the conditions of the images. They are of low resolution and more complex writing style than the previous dataset.

Table 5.9: Precision at Recall=50% for each query using the contour based method and the integrated SCD-HOG descriptors on the extracted main text

Keywords		Precision
عليه	Alaihi	50%
سلم	Sallam	55.65%
محمد	Mohammed	100%
عبد	Abd	87.55%
السلطان	Assultan	100%
العلم	Alelm	100%
مسعود	Masoud	80%
الحسن	Alhasan	40%
Average		76.63%

5.5.7 Performance on Two Historical Datasets

In this section, we evaluate the performance of the proposed feature representation using 50 pages selected from the previous two datasets which are written by two different writers. We used 30 images for training and 20 images for testing. The training and testing images are selected equally from both datasets. For each query, we select its query image from the first dataset and the performance computed. Then, we select its query image from the second dataset and the performance is computed. So, each query is evaluated two times by selecting the query image from the two datasets. Finally, the average performance is computed for each query. Table 5.10 shows the performance in terms of precision for each query. The result shows that the contour method using the integrated SCD-HOG descriptors achieve a precision rate of 32.78% at Recall=Precision. We found that all retrieved instances of the given query are en by the same writer of the query. All instances of the query that written by the second writer are not retrieved. This indicate that the proposed model is writer dependent. This is due to the variations in the writing between the two writers as each writer has his own writing style.

Table 5.10: Precision at Recall=Precision for each query using the contour based method and the integrated SCD-HOG descriptors on historical datasets written by two writers

Keywords		Precision
صلى	Salla	19.40%
عليه	Alaihi	27.92%
سلم	Sallam	38.89%
عبد	Abd	37.23%
محمد	Mohammed	40.48%
Average		32.78%

5.6 Conclusions

In this chapter, we propose a novel feature representation for Arabic word spotting based on the bag of features model and shape descriptors. Arabic words are segmented into small pieces called word fragments using contour-based, skeleton-based and Patch-based methods. Word fragments are described using the shape context descriptors and histogram of gradients. We evaluated the performance of the proposed method using printed and historical documents. Based on the results, the contour-based extraction method achieves the best performance compared to other methods. The future work will focus on incorporating the spatial information into the model. The performance can be improved further by the fusion of the word fragments extracted by the contour and skeleton based methods. Word spotting using handwritten documents is not a trivial task due to the writing style variations and different words are visually similar. This variation is usually much higher in a multi-writer scenario. The BoWFs model should be studied under this scenario.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

This chapter summarizes the contributions of the thesis and discusses some issues related to the developed techniques which require further investigation. Limitations in the proposed methods and the used data are also discussed. Finally, future research directions relevant to analysis and retrieval of Arabic historical documents are discussed.

6.1 Summary of the Methods

So far, research on Arabic historical analysis has focused on the extraction of text lines from simple documents. Very few efforts have been reported on the region extraction of Arabic historical documents. Moreover, a large number of recent works on word spotting are based on global and/or local features word profiles, Gradient, structural features, etc. There are works on hybrid features based systems that are trying to integrate both different methods.

In this thesis, we have conducted research on analysis and retrieval of Arabic historical documents. The objective was to address the problem of Arabic historical documents analysis and retrieval. This objective was met by developing a new method for text region extraction and a new feature representation called Bag of Word Fragments (BoWFs) for word spotting. The general contribution of this thesis is the design and development of successful layout analysis and retrieval prototype for historical documents. The specific contributions of this thesis are as follows:

Historical Documents Analysis:

The first contribution is the proposal of a general text region extraction method presented in Chapter 4. The proposed method is based on the fusion of background, foreground and window-based text analysis. The fusion of these techniques for text region extraction gives better segmentation results. The performance of the proposed method depends on the text characteristics at the window positions. The proposed method can be applied on historical documents that have different layout structures. Furthermore, we introduced two methods for text orientation estimation in Chapter 4. The performance of the proposed methods is highly dependent on the characteristics of the connected components. Further investigation of the proposed methods with different writing styles is required.

Historical Documents Retrieval:

The second contribution is the design and implementation of a novel feature representation approach called Bag of Word Fragments (BoWFs), that is more adapted to the cursive nature of the Arabic language presented in Chapter 5. The proposed feature representation approach is inspired by the traditional bag of features model. furthermore, we introduced three methods for word fragment extractions (viz. contour-based, skeleton-based and batch-based methods). The proposed methods make use of the different parts of the word to build the BoWFs model. In BoWFs, an Arabic word is decomposed into small pieces called word fragments, each of which is described using SCD and HOG. A word spotting method for Arabic historical documents based on the proposed BoWFs feature representation is presented. Experimental results on a collection of 50 Arabic historical and printed documents has yielded promising results achieving a precision rate of 89.60% at 50% recall. The model is also tested using 50 documents written by two writers. It is found that the proposed model is writer dependent. In addition, a comprehensive literature survey for historical document retrieval is needed in the research community, and Chapter 3 has satisfied this need. These contributions and others is expected to help in advancing the field of historical documents retrieval in general and for Arabic historical documents in particular.

6.2 Limitations of the Work

Although the research has reached its aim, various limitations may exist in this work. First, the proposed method focused on the main text and side notes regions extraction. It can be extended to further segment the side notes into several regions based on the text orientation.

Second, for window size estimation, we assumed the gaps between two successive lines are almost equal and hence we used the average gap between two successive lines. However, they may be different. In this case, the standard deviation of the gaps can be employed to detect this case and improve the estimation.

Third, the performance of the proposed methods for orientation estimation is highly dependent on the characteristics of the connected components. Further investigation of the proposed methods with different writing styles and fonts is required.

Fourth, only two features are used for word fragment description. Different features such as Fourier and Gabor transforms can be used.

Fifth, only one word fragment size is considered. The best word fragment size can be determined experimentally. For example, the performance of the patch based method can be studied using different sizes.

Sixth, the query itself is counted in the results. This may make the results biased. This can be overcome by determining a set of images for query selection.

Seventh, the performance was evaluated only with query words known to exist in the used corpus. The performance can be studied using query words that do not exist in the used corpus.

6.3 Future Research Directions

There are a number of issues that researchers of Arabic historical documents analysis and retrieval need to address. The complex structure of documents, overlapping text regions, presence of noise, degraded quality of the documents, touching words, presence of diacritic, style writing variations and variability, cursive nature of the Arabic script and lack of benchmarking databases, etc. are among the reasons that make Arabic historical documents analysis and retrieval a challenging task.

Word spotting using handwritten documents is not a trivial task due to the writing style variations. This variation is usually much higher in a multi-writer scenario. The BoWFs model should be studied under this scenario.

The proposed method for word spotting can be enhanced by:

- Incorporating the spatial information into the model by integrating the model with the Spatial Pyramid Matching (SPM).
- Using the Query Expansion (QE) techniques.
- Using more sophisticated quantization techniques such as Locality Constrained Linear Coding (LLC).
- Using indexing techniques such Latent Semantic Indexing (LSI), Distance-based Hashing (DBH) to reduce the search time.

- Fusion of the word fragments extracted by the contour and skeleton based methods.
- Investigating different features and descriptors for word fragment descriptions in the BoWFs feature representation.

A promising research can be evolved in the direction of building word spotting systems for historical document images. The following also can be explored further

- Building a dataset of adequate sizes (many writers, multiple samples per writer, etc.) that contain proper amount of variations of several factors such as script, writing styles, font size and quality.
- Providing text query support using synthesis techniques.
- Providing spoken query support.

REFERENCES

[1] S. Bukhari, T. Breuel, A. Asi, and J. El-Sana, "Layout analysis for arabic historical document images using machine learning," in *2012 International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 639–644, Sept 2012.

[2] G. Nagy, S. Seth, and M. Viswanathan, "A prototype document image analysis system for technical journals," *Computer*, vol. 25, pp. 10–22, July 1992.

[3] A. Antonacopoulos and D. Karatzas, "Semantics-based content extraction in typewritten historical documents," in *Eighth International Conference on Document Analysis and Recognition Proceedings*, pp. 48–53 Vol. 1, Aug 2005.

[4] K. Wong, R. Casey, and F. Wahl, "Document analysis system," *IBM Journal of Research and Development*, vol. 26, pp. 647–656, Nov 1982.

[5] L. O'Gorman, "The document spectrum for page layout analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, pp. 1162–1173, Nov. 1993.

[6] K. Kise, A. Sato, and M. Iwata, "Segmentation of page images using the area voronoi diagram," *Computer Vision and Image Understanding*, vol. 70, no. 3, pp. 370 – 382, 1998.

[7] D. Wang and S. N. Srihari, "Classification of newspaper image blocks using texture analysis," *Computer Vision, Graphics, and Image Processing*, vol. 47, no. 3, pp. 327 – 352, 1989.

[8] J. Liu, Y. Y. Tang, and C. Y. Suen, "Chinese document layout analysis based on adaptive split-and-merge and qualitative spatial reasoning," *Pattern Recognition*, vol. 30, no. 8, pp. 1265 – 1278, 1997.

[9] Y. Wang, I. T. Phillips, and R. M. Haralick, "Document zone content classification and its performance evaluation," *Pattern Recogn.*, vol. 39, pp. 57–73, Jan. 2006.

[10] W. Abd-Almageed, M. Agrawal, W. Seo, and D. Doermann, "Document-zone classification using partial least squares and hybrid classifiers," in *ICPR 2008. 19th International Conference on Pattern Recognition.*, pp. 1–4, Dec 2008.

[11] D. Keysers, F. Shafait, and T. M. Breuel, "Document image zone classification - a simple high-performance approach," in *in 2nd Int. Conf. on Computer Vision Theory and Applications*, pp. 44–51, 2007.

[12] S. Chanda, K. Franke, and U. Pal, "Document-zone classification in torn documents," in *International Conference on Frontiers in Handwriting Recognition (ICFHR).*, pp. 25–30, Nov 2010.

[13] Z. Ibrahim, D. Isa, R. RajKumar, and G. Kendall, "Document zone content classification for technical document images using artificial neural networks and support vector machines," in *Second International Conference on the Applications of Digital Information and Web Technologies.*, pp. 345–350, Aug 2009.

[14] F. Le Bourgeois and H. Kaileh, "Automatic metadata retrieval from ancient manuscripts," in *Document Analysis Systems VI*, vol. 3163, pp. 75–89, Springer Berlin Heidelberg, 2004.

[15] N. Journet, R. Mullot, J.-Y. Ramel, and V. Eglin, "Ancient printed documents indexation: A new approach," in *Pattern Recognition and Data Mining*, vol. 3686, pp. 580–589, Springer Berlin Heidelberg, 2005.

[16] N. Journet, V. Eglin, J. Y. Ramel, and R. Mullot, "Text/graphic labelling of ancient printed documents," in *Proceedings. Eighth International Conference on Document Analysis and Recognition.*, vol. 2, pp. 1010–1014, Aug 2005.

[17] J. Ramel, S. Leriche, M. Demonet, and S. Busson, "User-driven page layout analysis of historical printed books," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 9, no. 2-4, pp. 243–261, 2007.

[18] A. Garz, M. Diem, and R. Sablatnig, "Local descriptors for document layout analysis," in *Proceedings of the 6th International Conference on Advances in Visual Computing - Volume Part III*, ISVC'10, (Berlin, Heidelberg), pp. 29–38, Springer-Verlag, 2010.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, pp. 91–110, Nov. 2004.

[20] N. Ouwayed and A. Belaïd, "A general approach for multi-oriented text line extraction of handwritten documents," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 15, no. 4, pp. 297–314, 2012.

[21] T. A. C. M. and W. F. G. Mecklenbrauker, "The wigner distribution: A tool for time-frequency signal analysis - part 2: Discrete-time signals," *Philips Journal of Research*, vol. 35, no. 4, pp. 276–350, 1980.

[22] N. Nikolaou, M. Makridis, B. Gatos, N. Stamatopoulos, and N. Papamarkos, "Segmentation of historical machine-printed documents using adaptive run length smoothing and skeleton segmentation paths," *Image and Vision Computing*, vol. 28, no. 4, pp. 590 – 604, 2010.

[23] D. Doermann, "The indexing and retrieval of document images: A survey," *Computer Vision and Image Understanding*, vol. 70, no. 3, pp. 287 – 298, 1998.

[24] T. Kameshiro, T. Hirano, Y. Okada, and F. Yoda, "A document image retrieval method tolerating recognition and segmentation errors of ocr using shape-feature and multiple candidates," in *Proceedings of the Fifth International Conference on Document Analysis and Recognition.*, pp. 681–684, Sep 1999.

[25] V. Lavrenko, T. M. Rath, and R. Manmatha, "Holistic word recognition for handwritten historical documents," in *Proceedings. First International Workshop on Document Image Analysis for Libraries.*, pp. 278–287, 2004.

[26] S. Bai, L. Li, and C. Tan, "Keyword spotting in document images through word shape coding," in *10th International Conference on Document Analysis and Recognition.*, pp. 331–335, July 2009.

[27] "HAH manuscripts: A holistic paradigm for classifying and retrieving historical arabic handwritten documents," *Expert Systems with Applications*, vol. 36, no. 8, pp. 10942 – 10951, 2009.

[28] M. Khayyat, L. Lam, and C. Y. Suen, "Learning-based word spotting system for arabic handwritten documents," *Pattern Recognition*, vol. 47, no. 3, pp. 1021 – 1030, 2014.

[29] T. Konidaris, B. Gatos, K. Ntzios, I. Pratikakis, S. Theodoridis, and S. Perantonis, "Keyword-guided word spotting in historical printed documents using synthetic data and user feedback," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 9, no. 2-4, pp. 167–177, 2007.

[30] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990.

[31] X. He, D. Cai, H. Liu, and W.-Y. Ma, "Locality preserving indexing for document representation," in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '04, (New York, NY, USA), pp. 96–103, ACM, 2004.

[32] S. Marinai, E. Marino, and G. Soda, "Font adaptive word indexing of modern printed documents," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, pp. 1187–1199, Aug 2006.

[33] T. Syeda-Mahmood, “Indexing of handwritten document images,” in *Workshop on Document Image Analysis.*, pp. 66–73, Jun 1997.

[34] A. Andoni and P. Indyk, “Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions,” in *47th Annual IEEE Symposium on Foundations of Computer Science.*, pp. 459–468, Oct 2006.

[35] V. Athitsos, M. Potamias, P. Papapetrou, and G. Kollios, “Nearest neighbor retrieval using distance-based hashing,” in *IEEE 24th International Conference on Data Engineering.*, pp. 327–336, April 2008.

[36] R. Manmatha and W. B. Croft, “Word spotting: Indexing handwritten archives,” 1997.

[37] H. Alamri, J. Sadri, C. Y. Suen, and N. Nobile, “A novel comprehensive database for arabic off-line handwriting recognition,” in *Proc. of the 11 th Int. Conference on Frontiers in Handwriting Recognition (ICFHR 2008*, pp. 664–669, 2008.

[38] U.-V. Marti and H. Bunke, “The iam-database: an english sentence database for offline handwriting recognition,” *International Journal on Document Analysis and Recognition*, vol. 5, no. 1, pp. 39–46, 2002.

[39] M. Wuthrich, M. Liwicki, A. Fischer, E. Indermuhle, H. Bunke, G. Viehhauser, and M. Stolz, “Language model integration for the recognition of handwritten medieval documents,” in *10th International Conference on Document Analysis and Recognition.*, pp. 211–215, July 2009.

[40] T. M. Rath and R. Manmatha, “Word image matching using dynamic time warping,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*, vol. 2, pp. 521–527, June 2003.

[41] T. M. Rath and R. Manmatha, “Features for word spotting in historical manuscripts,” in *Seventh International Conference on Document Analysis and Recognition.*, pp. 218–222 vol.1, Aug 2003.

[42] A. Kolcz, J. Alspector, M. Augusteijn, R. Carlson, and G. Viorel Popescu, "A line-oriented approach to word spotting in handwritten documents," *Pattern Analysis Applications*, vol. 3, no. 2, pp. 153–168, 2000.

[43] A. Sigappi, S. Palanivel, and V. Ramalingam, "Article:handwritten document retrieval system for tamil language," *International Journal of Computer Applications*, vol. 31, pp. 42–47, October 2011.

[44] M. Shah and C. Suen, "Word spotting in gray scale handwritten pashto documents," in *International Conference on Frontiers in Handwriting Recognition (ICFHR).*, pp. 136–141, Nov 2010.

[45] A. Abidi, A. Jamil, I. Siddiqi, and K. Khurshid, "Word spotting based retrieval of urdu handwritten documents," in *Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition*, ICFHR '12, (Washington, DC, USA), pp. 331–336, IEEE Computer Society, 2012.

[46] H. Wei and G. Gao, "A keyword retrieval system for historical mongolian document images," *International Journal of Document Analysis and Recognition (IJDAR)*, pp. 33–45, 2014.

[47] A. Kesidis, E. Galiotou, B. Gatos, A. Lampropoulos, I. Pratikakis, I. Manolessou, and A. Ralli, "Accessing the content of greek historical documents," in *Proceedings of The Third Workshop on Analytics for Noisy Unstructured Text Data*, AND '09, (New York, NY, USA), pp. 55–62, ACM, 2009.

[48] H. Cao, A. Bhardwaj, and V. Govindaraju, "A probabilistic method for keyword retrieval in handwritten document images," *Pattern Recognition*, vol. 42, no. 12, pp. 3374 – 3382, 2009.

[49] T. M. Rath and R. Manmatha, "Word spotting for historical documents," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 9, no. 2-4, pp. 139–152, 2007.

[50] S. Srihari, H. Srinivasan, P. Babu, and C. Bhole, "Spotting words in handwritten arabic documents," in *Document Recognition and Retrieval XIII: Proceedings SPIE*, 2006.

[51] S. Srihari, H. Srinivasan, P. Babu, and C. Bhole, "Handwritten arabic word spotting using the cedarabic document analysis system," in *Symposium on Document Image Understanding Technology*, 2005.

[52] S. Srihari and G. Ball, "Language independent word spotting in scanned documents," in *Digital Libraries: Universal and Ubiquitous Access to Information*, vol. 5362, pp. 134–143, 2008.

[53] B. Zhang, S. N. Srihari, and C. Huang, "Word image retrieval using binary features," 2003.

[54] A. Kefali and C. Chemmam, "A semi-automatic approach of old arabic documents indexing," in *CIIA'11*, pp. 1–1, 2011.

[55] Y. Liang, M. Fairhurst, and R. Guest, "A synthesised word approach to word retrieval in handwritten documents," *Pattern Recognition*, vol. 45, no. 12, pp. 4225 – 4236, 2012.

[56] R. G. Casey and E. Lecolinet, "A survey of methods and strategies in character segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, pp. 690–706, July 1996.

[57] R. Moghaddam and M. Cheriet, "Application of multi-level classifiers and clustering for automatic word spotting in historical document images," in *10th International Conference on Document Analysis and Recognition.*, pp. 511–515, July 2009.

[58] J. Llados and G. Sanchez, "Indexing historical documents by word shape signatures," in *Ninth International Conference on Document Analysis and Recognition.*, vol. 1, pp. 362–366, Sept 2007.

[59] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 24, pp. 509–522, Apr 2002.

[60] A. Fornés, V. Frinken, A. Fischer, J. Almazán, G. Jackson, and H. Bunke, “A keyword spotting approach using blurred shape model-based descriptors,” in *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, pp. 83–90, 2011.

[61] S. Escalera, A. Fornés, O. Pujol, P. Radeva, G. Sánchez, and J. Lladós, “Blurred shape model for binary and grey-level symbol recognition,” *Pattern Recognition Letters*, vol. 30, no. 15, pp. 1424 – 1433, 2009.

[62] J. Llads, P. Pratim-Roy, J. Rodrguez, and G. Snchez, “Word spotting in archive documents using shape contexts,” in *Pattern Recognition and Image Analysis*, vol. 4478 of *Lecture Notes in Computer Science*, pp. 290–297, 2007.

[63] T. Adamek, N. OConnor, and A. Smeaton, “Word matching using single closed contours for indexing handwritten historical documents,” *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 9, no. 2-4, pp. 153–165, 2007.

[64] T. Adamek and N. O’Connor, “A multiscale representation method for nonrigid shapes with a single closed contour,” *IEEE Transactions on Circuits and Systems for Video Technology.*, vol. 14, pp. 742–753, May 2004.

[65] E. F. Can and P. Duygulu, “A line-based representation for matching words in historical manuscripts,” *Pattern Recognition Letters*, vol. 32, no. 8, pp. 1126 – 1138, 2011.

[66] P. K. Agarwal and K. R. Varadarajan, “Efficient algorithms for approximating polygonal chains,” *Discrete Computational Geometry*, vol. 23, no. 2, pp. 273–291, 2000.

[67] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, “Groups of adjacent contour segments for object detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 30, pp. 36–51, Jan 2008.

[68] E. Ataer and P. Duygulu, "Matching ottoman words," in *IEEE 15th Signal Processing and Communications Applications.*, pp. 1–4, June 2007.

[69] M. Rusiñol, D. Aldavert, R. Toledo, and J. Llads, "Efficient segmentation-free keyword spotting in historical document collections," *Pattern Recognition*, vol. 48, no. 2, pp. 545 – 555, 2015.

[70] I. Z. Yalniz and R. Manmatha, "An efficient framework for searching text in noisy document images," in *10th IAPR International Workshop on Document Analysis Systems, DAS*, pp. 48–52, 2012.

[71] D.-R. Lee, W. Hong, and I.-S. Oh, "Segmentation-free word spotting using sift," in *IEEE Southwest Symposium on Image Analysis and Interpretation.*, pp. 65–68, April 2012.

[72] L. Rothacker, M. Rusiñol, and G. Fink, "Bag-of-features hmms for segmentation-free word spotting in handwritten documents," in *12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1305–1309, Aug 2013.

[73] M. Leydier, D. Aldavert, R. Toledo, and J. Llados, "Browsing heterogeneous document collections by a segmentation-free word spotting method," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 63–67, Sept 2011.

[74] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*, vol. 2, pp. 2169–2178, 2006.

[75] L. Czuni, P. Kiss, M. Gal, and A. Lipovits, "Local feature based word spotting in handwritten archive documents," in *11th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 179–184, June 2013.

[76] J. A. Rodrıguez and F. Perronnin, "Local gradient histogram features for word spotting in unconstrained handwritten documents," in *International Conference on Frontiers in Handwriting Recognition*, 2008.

[77] X. Zhang and C. Tan, "Segmentation-free keyword spotting for handwritten documents based on heat kernel signature," in *12th International Conference on Document Analysis and Recognition (ICDAR).*, pp. 827–831, Aug 2013.

[78] J. Almazán, A. Gordo, A. Fornés, and E. Valveny, "Segmentation-free word spotting with exemplar SVMs," *Pattern Recognition*, vol. 47, no. 12, pp. 3967 – 3978, 2014.

[79] M. Khayyat, L. Lam, and C. Y. Suen, "Arabic handwritten word spotting using language models," in *Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition.*, ICFHR '12, pp. 43–48, 2012.

[80] J. Rodriguez-Serrano and F. Perronnin, "Handwritten word image retrieval with synthesized typed queries," in *10th International Conference on Document Analysis and Recognition.*, pp. 351–355, July 2009.

[81] J. A. Rodriguez-Serrano and F. Perronnin, "Synthesizing queries for handwritten word image retrieval," *Pattern Recognition*, vol. 45, no. 9, pp. 3270 – 3276, 2012.

[82] M. Sagheer, N. Nobile, C. L. He, and C. Suen, "A novel handwritten urdu word spotting based on connected components analysis," in *20th International Conference on Pattern Recognition (ICPR)*, pp. 2013–2016, Aug 2010.

[83] A. Fischer, A. Keller, V. Frinken, and H. Bunke, "Hmm-based word spotting in handwritten documents using subword models," in *20th International Conference on Pattern Recognition (ICPR).*, pp. 3416–3419, Aug 2010.

[84] A. Fischer, A. Keller, V. Frinken, and H. Bunke, "Lexicon-free handwritten word spotting using character HMMs," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 934 – 942, 2012.

[85] A. Fischer, V. Frinken, H. Bunke, and C. Suen, "Improving hmm-based keyword spotting with character language models," in *12th International Conference on Document Analysis and Recognition (ICDAR).*, pp. 506–510, Aug 2013.

[86] S. Wshah, G. Kumar, and V. Govindaraju, “Script independent word spotting in offline handwritten documents based on hidden markov models,” in *International Conference on Frontiers in Handwriting Recognition (ICFHR).*, pp. 14–19, Sept 2012.

[87] S. Wshaha, G. Kumar, and V. Govindaraju, “Statistical script independent word spotting in offline handwritten documents,” *Pattern Recognition*, vol. 47, no. 3, pp. 1039 – 1050, 2014.

[88] J. Rodriguez-Serrano and F. Perronnin, “A model-based sequence similarity with application to handwritten word spotting,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, pp. 2108–2120, Nov. 2012.

[89] U.-V. Marti and H. Bunke, “Using a statistical language model to improve the performance of an hmm-based cursive handwriting recognition systems,” pp. 65–90, River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2002.

[90] H. Bunke, S. Bengio, and A. Vinciarelli, “Offline recognition of unconstrained handwritten texts using hmms and statistical language models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 709–720, 2004.

[91] E. Saykol, A. K. Sinop, U. Gudukbay, O. Ulusoy, and A. E. Çetin, “Content-based retrieval of historical ottoman documents stored as textual images,” *IEEE Transactions on Image Processing*, vol. 13, no. 3, pp. 314–325, 2004.

[92] S. Shahab, W. Al-Khatib, and S. Mahmoud, “Computer aided indexing of historical manuscripts,” in *International Conference on Computer Graphics, Imaging and Visualisation.*, pp. 287–295, July 2006.

[93] J. A. Rodrguez-Serrano and F. Perronnin, “Handwritten word-spotting using hidden markov models and universal vocabularies,” *Pattern Recognition*, vol. 42, no. 9, pp. 2106 – 2116, 2009.

[94] L. Huang, F. Yin, Q.-H. Chen, and C.-L. Liu, “Keyword spotting in unconstrained handwritten chinese documents using contextual word model,” *Image and Vision Computing*, vol. 31, no. 12, pp. 958 – 968, 2013.

[95] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "Casia online and offline chinese handwriting databases," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 37–41, Sept 2011.

[96] A. Kesidis, E. Galiotou, B. Gatos, and I. Pratikakis, "A word spotting framework for historical machine-printed documents," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 14, no. 2, pp. 131–144, 2011.

[97] B. Gatos, T. Konidaris, I. Pratikakis, and S. Perantonis, "A holistic methodology for keyword search in historical typewritten documents," in *Advances in Artificial Intelligence*, vol. 3955, pp. 490–493, 2006.

[98] V. Frinken, A. Fischer, and H. Bunke, "A novel word spotting algorithm using bidirectional long short-term memory neural networks," in *Artificial Neural Networks in Pattern Recognition*, vol. 5998, pp. 185–196, 2010.

[99] V. Frinken, A. Fischer, R. Manmatha, and H. Bunke, "A novel word spotting method based on recurrent neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 211–224, Feb 2012.

[100] A. Bhardwaj, S. Setlur, and V. Govindaraju, "Keyword spotting techniques for sanskrit documents," in *Sanskrit Computational Linguistics*, vol. 5402, pp. 403–416, 2009.

[101] F. Perronnin and J. A. Rodriguez-Serrano, "Fisher kernels for handwritten word-spotting," in *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*, ICDAR '09, pp. 106–110, 2009.

[102] H. Zhang, D.-H. Wang, and C.-L. Liu, "Keyword spotting from online chinese handwritten documents using one-vs-all trained character classifier," in *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 271–276, Nov 2010.

[103] H. Zhang, X.-D. Zhou, and C.-L. Liu, "Keyword spotting in online chinese handwritten documents with candidate scoring based on semi-crf model,"

in *12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 567–571, Aug 2013.

[104] X.-D. Zhou, D.-H. Wang, F. Tian, C.-L. Liu, and M. Nakagawa, "Handwritten chinese/japanese text recognition using semi-markov conditional random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 2413–2426, Oct 2013.

[105] K. Terasawa, T. Nagasaki, and T. Kawashima, "Eigenspace method for text retrieval in historical document images," in *Eighth International Conference on Document Analysis and Recognition, 2005. Proceedings.*, pp. 437–441 Vol. 1, Aug 2005.

[106] A. K. Nabil Aouadi, "Word spotting for arabic handwritten historical document retrieval using generalized hough transform," in *The Third International Conferences on Pervasive Patterns and Applications*, pp. 67 – 71, Sep 2011.

[107] J. Sousa, J. Gil, and J. Pinto, "Word indexing of ancient documents using fuzzy classification," *IEEE Transactions on Fuzzy Systems*, vol. 15, pp. 852–862, Oct 2007.

[108] D. Fernández, J. Lladós, and A. Fornés, "Handwritten word spotting in old manuscript images using a pseudo-structural descriptor organized in a hash structure," in *Pattern Recognition and Image Analysis*, vol. 6669, pp. 628–635, 2011.

[109] H. Glucksman, "Classification of mixed-font alphabets by characteristic loci," in *Proceeding of IEEE Computer Conference*, p. 138141, 1967.

[110] P. Bilane, S. Bres, K. Challita, and H. Emptoz, "Indexation of syriac manuscripts using directional features," in *16th IEEE International Conference on Image Processing (ICIP)*, pp. 1841–1844, Nov 2009.

[111] P. Bilane, S. Bres, and H. Emptoz, "Robust directional features for wordspotting in degraded syriac manuscripts," in *International Workshop on Content-Based Multimedia Indexing*, pp. 526–533, June 2008.

[112] T. van Der Zant, L. Schomaker, and K. Haak, "Handwritten-word spotting using biologically inspired features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1945–1957, Nov 2008.

[113] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 411–426, March 2007.

[114] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 994–1000 vol. 2, June 2005.

[115] M. J. D. Powell, "Algorithms for approximation," ch. Radial Basis Functions for Multivariable Interpolation: A Review, pp. 143–167, 1987.

[116] NL-HaNA, "2.02.14, archief van het kabinet der koningin," 1903.

[117] Y. Leydier, F. Lebourgeois, and H. Emptoz, "Text search for medieval manuscript images," *Pattern Recognition*, vol. 40, no. 12, pp. 3552 – 3567, 2007.

[118] Y. Leydier, A. Ouji, F. LeBourgeois, and H. Emptoz, "Towards an omnilingual word retrieval system for ancient manuscripts," *Pattern Recognition*, vol. 42, no. 9, pp. 2089 – 2105, 2009.

[119] K. Terasawa1, T. Nagasaki, and T. Kawashima, "Automatic keyword extraction from historical document images," in *Document Analysis Systems VII*, vol. 3872 of *Lecture Notes in Computer Science*, pp. 413–424, Springer Berlin Heidelberg, 2006.

[120] T. M. Breuel, "Two geometric algorithms for layout analysis," in *Proceedings of the 5th International Workshop on Document Analysis Systems V*, (London, UK, UK), pp. 188–199, Springer-Verlag, 2002.

[121] A. Zahour, B. Taconet, P. Mercy, and S. Ramdane, "Arabic hand-written text-line extraction," in *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on*, pp. 281–285, 2001.

[122] A. Zahour, B. Taconet, L. Likforman-Sulem, and W. Boussellaa, "Overlapping and multi-touching text-line segmentation by block covering analysis," *Pattern Analysis and Applications*, vol. 12, no. 4, pp. 335–351, 2009.

[123] W. Boussellaa, A. Zahour, H. Elabed, A. Benabdelhafid, and A. Alimi, "Unsupervised block covering analysis for text-line segmentation of arabic ancient handwritten document images," in *20th International Conference on Pattern Recognition (ICPR)*, pp. 1929–1932, Aug 2010.

[124] R. Saabni and J. El-Sana, "Language-independent text lines extraction using seam carving," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 563–568, Sept 2011.

[125] A. Asi, R. Saabni, and J. El-Sana, "Text line segmentation for gray scale historical document images," in *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, HIP '11, (New York, NY, USA), pp. 120–126, ACM, 2011.

[126] S. Bukhari, F. Shafait, and T. Breuel, "Performance evaluation of curled textline segmentation algorithms on cbdar 2007 dewarping contest dataset," in *17th IEEE International Conference on Image Processing (ICIP)*, pp. 2161–2164, Sept 2010.

[127] R. Saabni, A. Asi, and J. El-Sana, "Text line extraction for historical document images," *Pattern Recognition Letters*, vol. 35, no. 0, pp. 23 – 33, 2014.

[128] M. Khayyat, L. Lam, C. Suen, F. Yin, and C.-L. Liu, "Arabic handwritten text line extraction by applying an adaptive mask to morphological dilation," in *10th IAPR International Workshop on Document Analysis Systems (DAS)*, pp. 100–104, March 2012.

[129] I. Messaoud, H. Amiri, H. Abed, and V. Margner, "A multilevel text-line segmentation framework for handwritten historical documents," in *2012 International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 515–520, Sept 2012.

[130] S. Bukhari, F. Shafait, and T. Breuel, "Ridges based curled textline region detection from grayscale camera-captured document images," in *Computer Analysis of Images and Patterns*, vol. 5702 of *Lecture Notes in Computer Science*, pp. 173–180, Springer Berlin Heidelberg, 2009.

[131] J. Sauvola and M. Pietikinen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225 – 236, 2000.

[132] K. Khurshid, I. Siddiqi, C. Faure, and N. Vincent, "Comparison of niblack inspired binarization methods for ancient documents," 2009.

[133] T. Pavlidis and J. Zhou, "Page segmentation and classification," *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 6, pp. 484 – 496, 1992.

[134] T. M. Breuel, "Representations and metrics for off-line handwriting segmentation," in *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition*, IWFHR '02, (Washington, DC, USA), pp. 428–, IEEE Computer Society, 2002.

[135] F. Shafait, D. Keysers, and T. Breuel, "Pixel-accurate representation and evaluation of page segmentation in document images," in *18th International Conference on Pattern Recognition*, vol. 1, pp. 872–875, 2006.

[136] B. G. Konstantinos Zagoris, Ioannis Pratikakis, "Segmentation-based historical handwritten word spotting using document-specific local features," in *14th International Conference on Frontiers in Handwriting Recognition*, pp. 9–14, 2014.

[137] S. O'Hara and B. A. Draper, "Introduction to the bag of features paradigm for image classification and retrieval," *CoRR*, vol. abs/1101.3354, 2011.

[138] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proceedings the Ninth IEEE International Conference on Computer Vision*, pp. 1470–1477, IEEE, 2003.

[139] C.-F. Tsai, "Bag-of-words representation in image annotation: A review," *ISRN Artificial Intelligence*, 2012.

[140] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Found. Trends. Comput. Graph. Vis.*, vol. 3, pp. 177–280, July 2008.

[141] K. Mikolajczyk, B. Leibe, and B. Schiele, "Local features for object class recognition," in *Tenth IEEE International Conference on Computer Vision*, vol. 2, pp. 1792–1799 Vol. 2, Oct 2005.

[142] K. Mikolajczyk and C. Schmid, "Scale affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[143] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008.

[144] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, "Brief: Computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1281–1298, July 2012.

[145] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proceedings of the 2011 International Conference on Computer Vision*, ICCV '11, (Washington, DC, USA), pp. 2548–2555, IEEE Computer Society, 2011.

[146] Y. Pang, Y. Yuan, X. Li, and J. Pan, "Efficient hog human detection," *Signal Processing*, vol. 91, no. 4, pp. 773 – 781, 2011.

[147] T. Ojala, M. Pietikinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51 – 59, 1996.

[148] X. Wang, T. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," in *IEEE 12th International Conference on Computer Vision*, pp. 32–39, Sept 2009.

[149] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, Jul 2002.

[150] L. Nanni and A. Lumini, "Heterogeneous bag-of-features for object/scene recognition," *Applied Soft Computing*, vol. 13, no. 4, pp. 2171 – 2178, 2013.

[151] Y.-G. Jiang, C.-W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, CIVR '07, (New York, NY, USA), pp. 494–501, ACM, 2007.

[152] X. Wang, B. Feng, X. Bai, W. Liu, and L. J. Latecki, "Bag of contour fragments for robust shape classification," *Pattern Recognition*, vol. 47, no. 6, pp. 2116 – 2125, 2014.

[153] R. E. W. Gonzalez, R. C. and S. L. Eddins, *Digital Image Processing Using MATLAB, New Jersey.* 2004.

[154] W.-B. Goh, "Strategies for shape matching using skeletons," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 326 – 345, 2008.

[155] M. Elzobi, A. Al-Hamadi, Z. Al Aghbari, and L. Dings, "IESK-ArDB: a database for handwritten arabic and an optimized topological segmentation approach," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 16, no. 3, pp. 295–308, 2013.

[156] M. Yang, K. Kpalma, and J. Ronsin, "A survey of shape feature extraction techniques," *Pattern Recognition, IN-TECH*, pp. 43–90, 2008.

[157] F. Moosmann, E. Nowak, and F. Jurie, "Randomized clustering forests for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, vol. 30, pp. 1632–1646, Sept 2008.

[158] J. Yang, K. Yu, and T. Huang, "Supervised translation-invariant sparse coding," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*, pp. 3517–3524, June 2010.

[159] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*, pp. 3360–3367, June 2010.

[160] P. E. H. Richard O. Duda and D. G. Strok, *Pattern classification and scene analysis.* 1995.

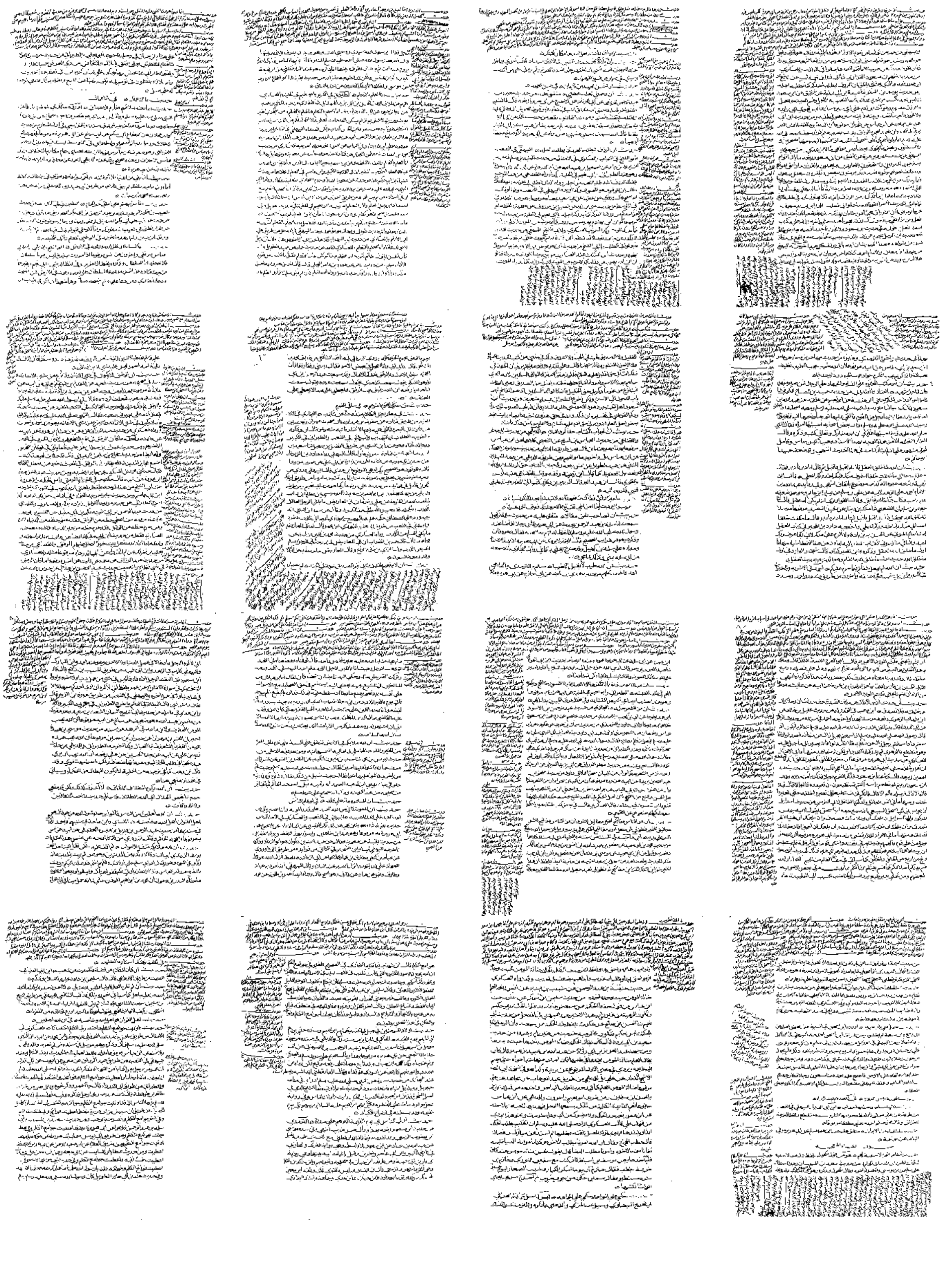
[161] M. Swain and D. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

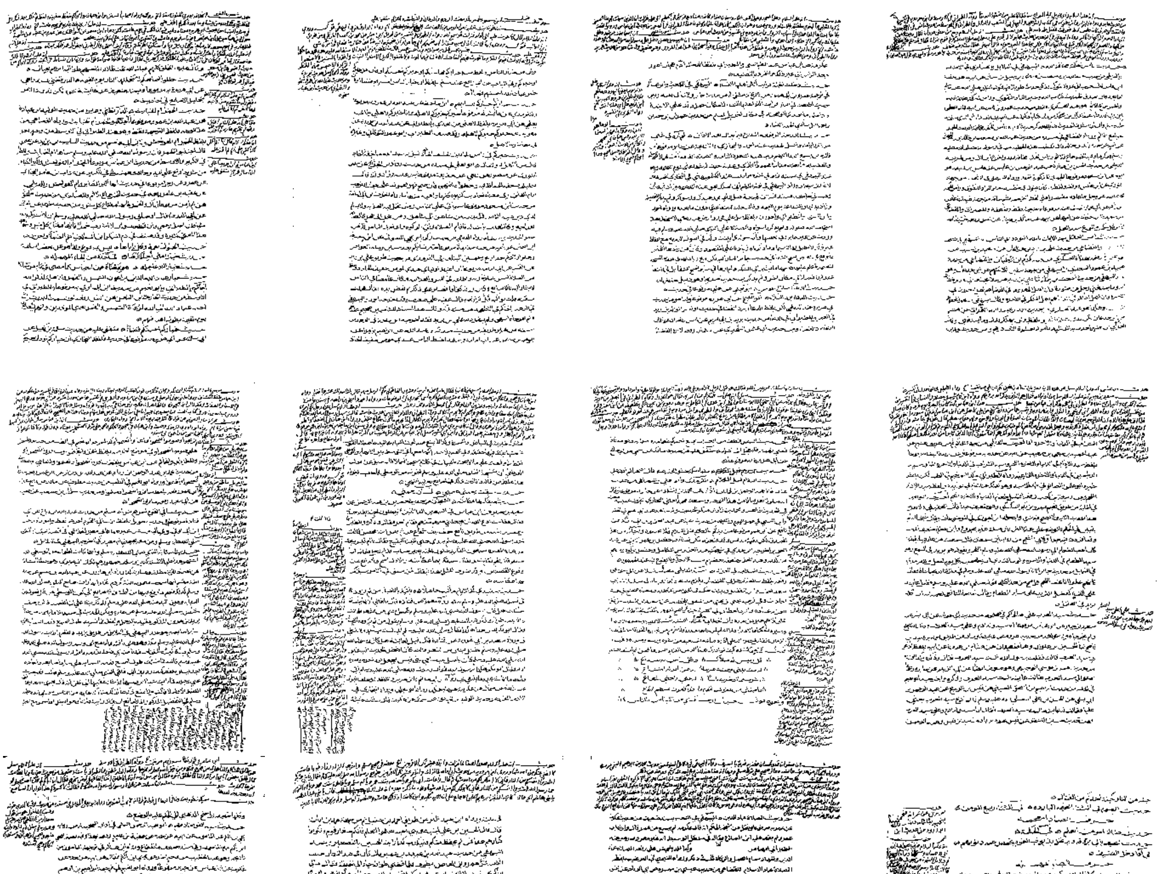
[162] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99–121, 2000.

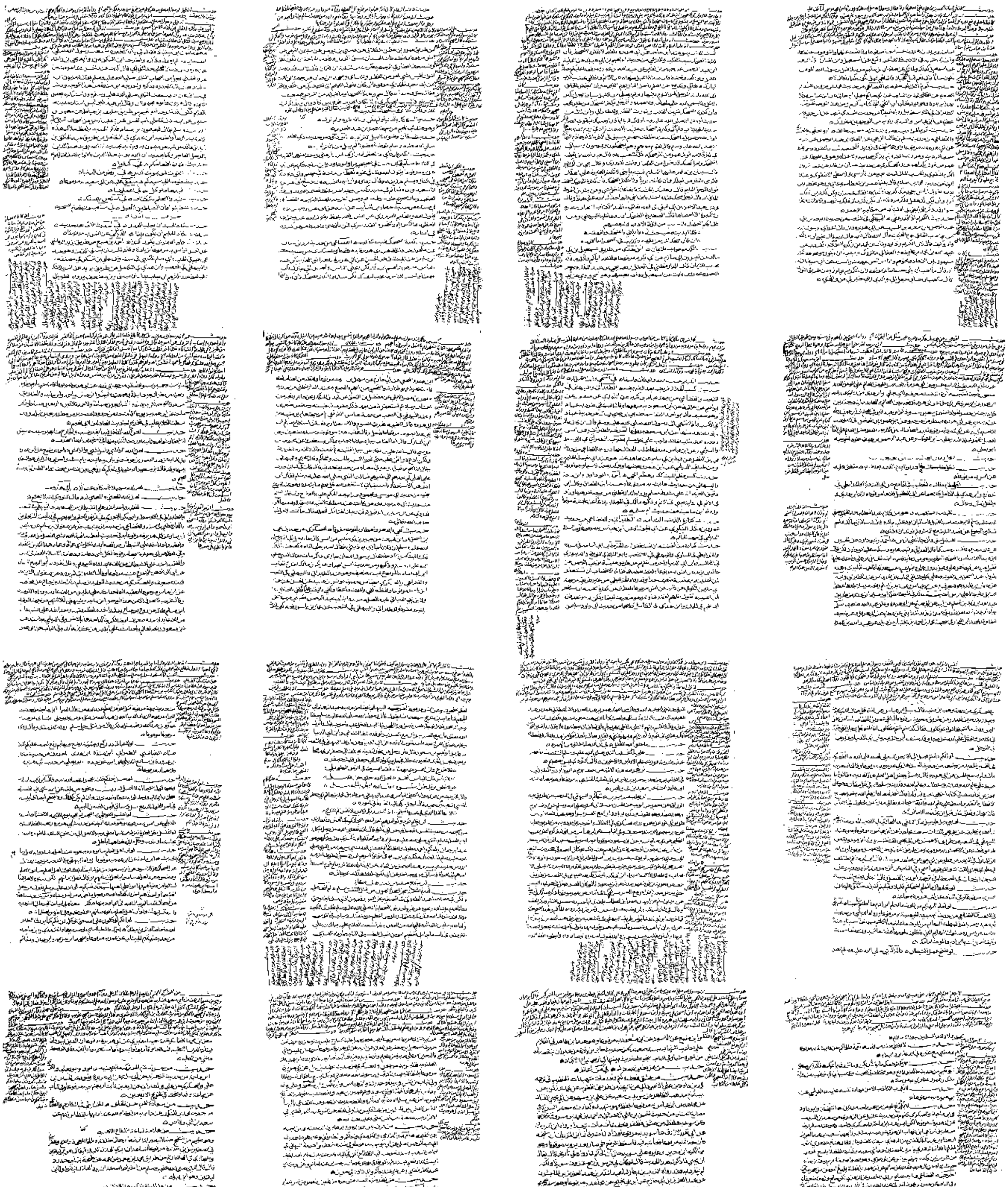
[163] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Computer Vision ECCV 2006*, vol. 3954, pp. 490–503, 2006.

Appendices

Appendix A: Historical Documents











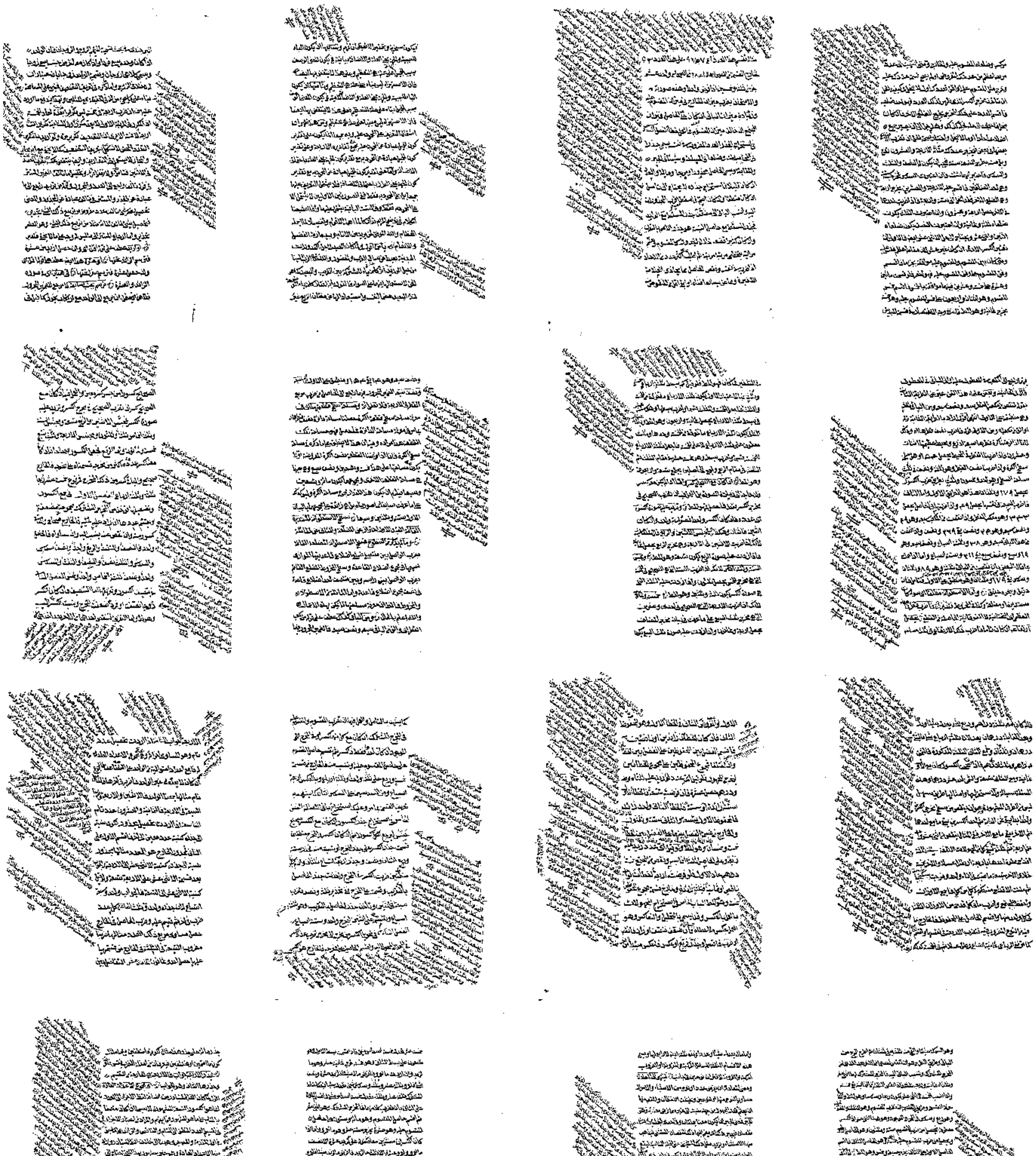


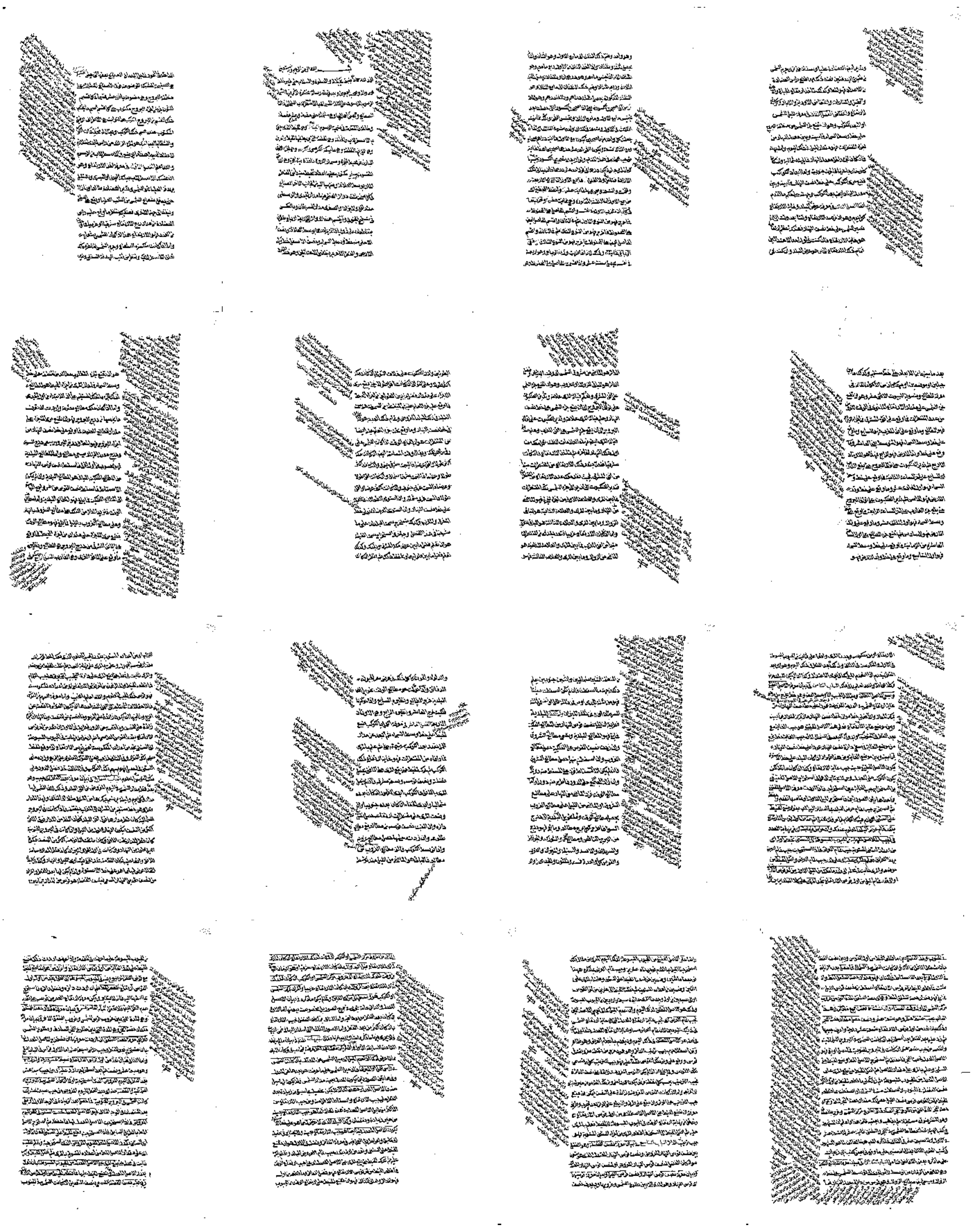










Vitae

- Rashad Ahmed Abdullah Othman.
- Born in Taiz, Yemen on October 4, 1978.
- Graduated from high school in 1997 with grade of 90.25.
- Received Bachelor of Science (BSc) degree in Computer Science (batch 2001-2002) from Al-Mustansiriya University, Baghdad, Iraq in 2002.
  - Secured overall 1^{st} position among 52 students.
- Appointed as a graduate assistant at Taiz University, college of Science in 2003, and I am still working there as a faculty member.
- Received Master of Science (MSc) degree in Computer Science from KFUPM, Dhahran, Saudi Arabia, in 2011.
- Submitted this dissertation to fulfil the requirements of his PhD degree in Computer Science and Engineering from King Fahd University of Petroleum & Minerals.
- **Publications**
  - Basem Almadani, Anas Al-Roubaiey, Rashad Ahmed, Manufacturing Systems Integration Using Real Time QoS-Aware Middleware, Advanced Materials Research, Vol. 711 (2013) pp 629-635.
  - Salah K, Al-Khiaty M-A-R, Ahmed R, Mahdi A. ,Performance Evaluation of Snort under Windows 7 and Windows Server 2008, Journal of Universal Computer Science 2011;17:160522.
- **Present Address:** Department of Information and Computer Science, King Fahd University of Petroleum and Minerals, P.O. Box 8585, Dhahran 31261, Saudi Arabia. E-mail : g200704590@kfupm.edu.sa
- **Permanent Address:** Department of Computer Science, Taiz University, Taiz, Yemen.E-mail rashadyousofi@yahoo.com
  rashadyousofi@gmail.com