PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
ECHAHID CHEIKH LARBI TEBESSI UNIVERSITY
FACULTY OF EXACT SCIENCES, NATURAL SCIENCES AND LIFE
DEPARTMENT OF MATHEMATICS AND COMPUTER
LABORATORY OF MATHEMATICS, INFORMATICS AND SYSTEMS (LAMIS)

DOCTORAL THESIS

OF THE THIRD CYCLE

Field: Computer science

Specialty: Artificial intelligence

By

Yousfi DOUAA

Title:

Keywords image retrieval in offline handwritten documents

This thesis was publicly defended on 2024 in front of the examination committee composed of:

MENASSEL Rafik	MCA, University of Tebessa, Algeria	President
HANNOUSSE Abdelhakim	MCA, University of Guelma, Algeria	Examiner
SARI Toufik	Professor, University of Annaba, Algeria	Examiner
METROUH Abdelmalek	MCA, National Higher School of Mathematics, Algeria	Examiner
GATTAL Abdeljalil	Professor, University of Tebessa, Algeria	Supervisor
DJEDDI Chawki	Professor, University of Tebessa, Algeria	Co-supervisor

Dedication

To myself, who has accepted challenges, strived with determination, and overpowered obstacles.

To my ever-supportive father.

To my great mother.

To my wonderful sisters, "Alaa and Sara".

To my dear brother and my supporter, "Houssem".

Abstract:

There are numerous clusters of historical and ancient documents in archives that are invaluable, as they are the most common way to share information. However, searching for this information is time-consuming due to its deteriorated condition and may be unusable. That is why, in recent years, digitization of these documents has become very popular, but numbering alone is not sufficient to make information accessible, particularly in historical manuscripts. Transcribing these documents is quite difficult due to poor preservation, different writing styles, etc.

An information retrieval technique called "keyword spotting" in document images has continued to get researchers' interest, which identifies word occurrences in document images. It represents an attractive alternative to transcription, which can be challenging, especially in the case of historical documents.

In this thesis, we study keyword spotting in handwritten historical documents using a Query-by-Example (QbE) approach type and a segmentation-based technique. The word images in the document are extracted and represented by a collection of textural features. These features are then used to match the image of the query word to the images in the reference base and then retrieve the relevant documents. Sundry textural metrics are used to capture the word shape, including oriented Basic Image Features (oBIFs) and its column scheme at different scales, Local Phase Quantization (LPQ), Local Binary Patterns (LBP), Local Directional Number Pattern (LDNP), Complete Local Binary Patterns (CLBP) and Completed Robust Local Binary Pattern (CRLBP). Likewise, multiple distance measurements are inspected for the matching phase. For the experiments, we used the ICFHR-2014 Word Spotting Competition database. The proposed technology evaluated in the database has yielded profitable results comparable to state-of-the-art technology.

Keywords: Keyword Spotting, Handwritten historical documents, Document images, Textural features, Segmentation-Based technique, Query-by-Example (QbE).

Résumé:

Il existe de nombreux groupes de documents historiques et anciens dans les archives qui sont inestimables, car ils constituent le moyen le plus courant de partager des informations. Cependant, la recherche de ces informations prend du temps en raison de leur état dégradé et peut s'avérer inutilisable. C’est pourquoi, ces dernières années, la numérisation de ces documents est devenue très populaire, mais la numérotation seule ne suffit pas à rendre l’information accessible, notamment dans les manuscrits historiques. La transcription de ces documents est assez difficile en raison d'une mauvaise conservation, de styles d'écriture différents, etc.

Une technique de recherche d'informations appelée « repérage de mots-clés » dans les images de documents a continué de susciter l'intérêt des chercheurs, qui identifie les occurrences de mots dans les images de documents. Elle représente une alternative intéressante à la transcription, qui peut s’avérer complexe, notamment dans le cas de documents historiques.

Dans cette thèse, nous étudions le repérage de mots-clés dans des documents historiques manuscrits en utilisant une approche de type Query-by-Example (QbE) et une technique basée sur la segmentation. Les images de mots dans le document sont extraites et représentées par une collection de caractéristiques texturales. Ces fonctionnalités sont ensuite utilisées pour faire correspondre l'image du mot requête aux images de la base de référence puis récupérer les documents pertinents. Diverses métriques texturales sont utilisées pour capturer la forme du mot, y compris les caractéristiques d'image de base orientées (oBIFs) et son schéma de colonnes à différentes échelles, la quantification de phase locale (LPQ), les modèles binaires locaux (LBP), les modèles de nombres directionnels locaux (LDNP), complets. Modèles binaires locaux (CLBP) et modèle binaire local robuste complété (CRLBP). De même, plusieurs mesures de distance sont inspectées pour la phase d'appariement. Pour les expériences, nous avons utilisé la base de données ICFHR-2014 Word Spotting Competition. La technologie proposée évaluée sur l'ensemble de données a donné des résultats rentables comparables à l’état de l'art.

Mots-clés: Repérage de mots clés, Documents historiques manuscrits, Images de documents, Caractéristiques texturales, Technique basé sur la segmentation, Requête par exemple (QbE).

الملخص:

هناك مجموعات عديدة من الوثائق التاريخية والقديمة في الأرشيفات التي لا تقدر بثمن، لأنها الطريقة الأكثر شيوعا لتبادل المعلومات .ومع ذلك، فإن البحث عن هذه المعلومات يستغرق وقتًا طويلاً نظرًا لتدهور حالتها وقد تكون غير صالحة للاستعمال .ولهذا السبب، أصبحت عملية رقمنة هذه الوثائق في السنوات الأخيرة شائعة جدًا، لكن الترقيم وحده لا يكفي لتسهيل الوصول إلى المعلومات، خاصة في المخطوطات التاريخية، فنسخ هذه الوثائق صعب للغاية بسبب: سوء الحفظ، واختلاف الكتابة الأنماط، الخ.

استمرت تقنية استرجاع المعلومات التي تسمى "اكتشاف الكلمات الرئيسية" في صور المستندات في جذب اهتمام الباحثين، والتي تحدد تكرارات الكلمات من صور المستندات. فهو يمثل بديلاً جذابًا للنسخ الذي قد يمثل تحديًا، خاصة فيما يتعلق بالوثائق التاريخية.

في هذه الأطروحة، قمنا بدراسة اكتشاف الكلمات الرئيسية في المستندات التاريخية المكتوبة بخط اليد باستخدام منهج الاستعلام بالمثال (QbE) والتقنية القائمة على التجزئة. يتم استخراج الصور الكلمات الموجودة في المستند وتمثيلها من خلال مجموعة من الميزات التركيبية. يتم بعد ذلك استخدام هذه الميزات لمطابقة صورة كلمة الاستعلام مع الصور الموجودة في القاعدة المرجعية ثم استرداد المستندات ذات الصلة. يتم استخدام مقاييس تركيبية متنوعة لالتقاط شكل الكلمة، بما في ذلك ميزات الصورة الأساسية الموجهة (oBIFs) ونظام الأعمدة الخاص بها بمقاييس مختلفة، وتكميم الطور المحلي (LPQ) ، والأنماط الثنائية المحلية (LBP) ، ونمط أرقام الاتجاه المحلي (LDNP) ، والأنماط الثنائية المحلية المكتمل (CLBP) والنموذج الثنائي المحلي القوي المكتمل (CRLBP). وبالمثل، يتم فحص قياسات المسافة المتعددة لمرحلة المطابقة. بالنسبة للتجارب، استخدمنا مجموعة بيانات مسابقة اكتشاف الكلمات ICFHR-2014 . لقد حققت التكنولوجيا المقترحة التي تم تقييمها في مجموعة البيانات نتائج مربحة يمكن مقارنتها بأحدث التقنيات.

الكلمات الرئيسية: اكتشاف الكلمات الرئيسية، المستندات التاريخية المكتوبة بخط اليد، صور المستندات، الميزات النصية، تقنية تعتمد على التجزئة، الاستعلام حسب المثال (QbE).

ACKNOWLEDGMENTS

First and foremost, we would like to express our gratitude to God Almighty for giving us the strength and patience to complete this research.

I want to thank my supervisor, Professor GATTAL Abdeljalil, for his continuous efforts and the guidance he devoted throughout this research work. I cannot forget his suggestions that made this work interesting. His immense knowledge, motivation, and patience gave me more strength to excel in finishing this research work.

I extend my sincere thanks to my Co-supervisor, Professor DJEDDI Chawki, for his support, comments, and encouragement.

I extend my sincere thanks to MENASSEL Rafik, HANNOUSSE Abdelhakim, SARI Toufik, and METROUH Abdelmalek for agreeing to review the thesis.

I would like to express my gratitude to the members of Lamis Laboratory, led by Professor BENDJENA Hakim, and to the entire family of the University of Tebessa, and my deepest thanks to all the professors of the Computer Science Department for their help and encouragement.

In the end, I am grateful to my parents, my sisters, and my brother, they provided me with moral support, encouragement, and sufficient motivation to achieve my goals. Thank you to all my relatives, friends, and acquaintances.

Content

LIST OF FIGURES

LIST OF TABLES

General Introduction

Handwritten documents are widely used in educational institutes, administrative offices, police stations, and courts because handwriting is the first method to document information. Therefore, these documents are kept and used at another time, resulting in the quality of these documents deteriorating over time.

Historical documents like old books and newspapers remain a necessary scientific and cultural reference for information retrieval. But when the collections of old documents are not transcribed into a digital format and due to the degradations that they present, the time of exploitation and the search for information in these documents is large, and they thus risk being unusable. For this and to maintain the heritage and make easy access to its collections, the archive services have been working in recent years to digitize the collections of its documents. Nevertheless, the presentation of the document in the form of a digital image remains insufficient to make the information accessible. As a result, several works have been done on digitization and information retrieval in the images of old documents, the main purpose of which is to develop techniques for handwriting recognition in image documents.

The first work is on traditional methods of text searching, which require effective Optical Character Recognition techniques (OCR). This technology converts handwritten document body text into machine-encoded form, enabling it to recognize words in images of printed documents. Currently, numerous OCR systems can work on high-quality printed and scanned documents. However, it is not the appropriate option for handwritten documents because it often has a variety of challenges, like the variability of the handwriting style, the overlap of letters, the poor-quality documents, the presence of noise, and image deformity.

As an OCR replacement resolution, a keyword spotting (KWS) system has been proposed, which is much quicker and more practical than an OCR solution and has obtained great attention as a technology for document image recovery from researchers. The objective of KWS is the determination and retrieval of all parts of document images in a database that contains an instance of a query word. More precisely, it matches the query word with the words found in document images and then classifies the target words according to how similar they are to the query word.

Techniques for keyword spotting have been used in different scripts, such as Arabic, Latin, Greek, etc. These handwritings vary in the direction of writing, the font, and the similarity between the characters. It can be either typed or handwritten.

In the literature, techniques for keyword spotting are classified according to three distinct aspects. First, depending on the type of query input, KWS approaches are organized into Query-by-Example (QbE) and Query-by-String (QbS). If the user enters the query term as an image, the approach is QbE. Conversely, if a query word is supplied as a string, the QbS approach is used. Secondly, one also classifies the KWS system as segmentation-free and segmentation-based techniques. Segmentation-free techniques pursue locating query instances across the whole document, while segmentation-based techniques depend on segmenting words from images and matching those with the input query, either through the use of word segmentation techniques or available ground truth. The segmentation-based technique is more effective than the segmentation-free technique because it focuses on particular sections of the document rather than the full document. But if the document is very noisy or very complex to apply the segmentation-based technique, then the segmentation-free methodology is the most appropriate technique. Third, the keyword spotting technique can also be categorized as learning-free [1], [2], [3], and learning-based [4], [5]. As the name indicates, the learning-based method trains the model on a large image database before searching for a specific set. In the learning-free methods, the similarity between the images of the database and the query word is calculated after the extraction features. Learning-based methods perform well when trained and tested on the same database, but learning-free methods are more practical and faster compared to learning-based methodologies.

Objectives of the research

The problem of keyword spotting in document images is widespread, but the results obtained are inadequate. As a result, we have chosen to focus on the problem of keyword spotting in handwritten documents using a segmentation-based and QbE framework. Keyword spotting systems involve several steps, including pre-processing, feature extraction, and matching. Our specific focus in this work is on the feature extraction step, as it is the most crucial aspect of our system. Our objective is to suggest a novel feature extraction method that is tailored to the challenges presented by handwritten documents. We aim to investigate the significance of textual features, such as oBIFs, oBIFs column histogram, LBP, LPQ, LDNP, CLBP, and CRLBP descriptors, in this problem. We will match the extracted features using simple matching methods, such as City-block distance, Correlation distance, Euclidean distance, and Cosine distance.

Outline of the thesis

This thesis is structured into two parts. The first part, containing Chapters 01 and 02, is dedicated to presenting the main concepts and work relating to the study. In the second part of the dissertation, represented by Chapter 03 and 04, we address in detail our conceptual choices, the implementation as well and the results obtained by the systems proposed for the keyword spotting system.

- **Part I. Fundamental Concepts and State-of-the-Art**
- **Chapter 01:** This chapter is devoted to the presentation of the generality of keyword spotting as well as how this system works, and we'll explain each of its steps.
- **Chapter 02:** This second chapter is committed to the state-of-the-art in the field of keyword spotting. In its first part, we concentrate on the presentation of the main research works in this field. Then, we compare the different works in the field, and finally, we end the chapter with an examination of the various competitions in the field of keyword spotting.

- **Part II. Contribution and Validation**

- **Chapter 03:** This chapter is dedicated to explaining the most important details of our contribution to the field of keyword spotting, especially the features used.
- **Chapter 04:** This chapter will be dedicated to the experimentation and analysis of the results obtained, followed by a comparative study of the proposed technique with state-of-the-art methods evaluated on the ICFHR 2014 Competition on Handwritten Keyword Spotting (H-KWS 2014) database.

- **General Conclusion and Research Perspective:**

At the end of this thesis, a general conclusion will be drawn about the research we have undertaken in the field of keyword spotting, as well as the perspectives and future prospects in this subject.

Part I

Fundamental Concepts and State-of-the-Art

Chapter 01: Overview of Keyword Spotting in Handwritten Documents

1.1 Introduction

In this chapter, we introduce the keyword spotting (KWS) system's basic idea. Preprocessing of images, segmenting images into smaller entities (lines and words), feature extraction, and matching are the steps that make up a KWS system. We will review some fundamental concepts and definitions related to keyword spotting, we also present the different methods of feature extraction and matching in the literature.

1.2 Keyword spotting

Spotting is the assignment of locating a particular query without recognizing the content. So, keyword spotting (KWS) is the assignment of recovering all occurrences of a specific query word in a handwritten or printed document image without needing a traditional OCR step. Keyword spotting facilitates the retrieval and indexing of information presented as a query in historical or modern documents when they are complex and degraded.

The term keyword spotting was initially the domain of audio processing, where it was employed to detect certain keywords in an audio stream. It was also proposed in speech recognition [6]. This task was applied later in different applications, such as querying textually handwritten [7] and printed [8], [9] documents, as well as information retrieval and indexation in handwritten document databases. Keyword Spotting systems have been developed for various scripts like Latin, Greek, Arabic scripts, etc. These scripts vary from each other in things like the alphabet, the number of characters, the direction of writing, the shape, and the cursiveness. They can be either handwritten or printed.

Keyword spotting has garnered a significant amount of attention from the community these past years. Especially in the fields of analysis of handwritten documents [11] and indexing historical documents [10].

The two important components can often be identified for the keyword spotting systems proposed: the first component is an assemblage of documents or databases, and the second component is an input element indicated as a query (see **Figure 1.1**). The result of a keyword spotting system should be localization in the collection of documents that are similar to the query.

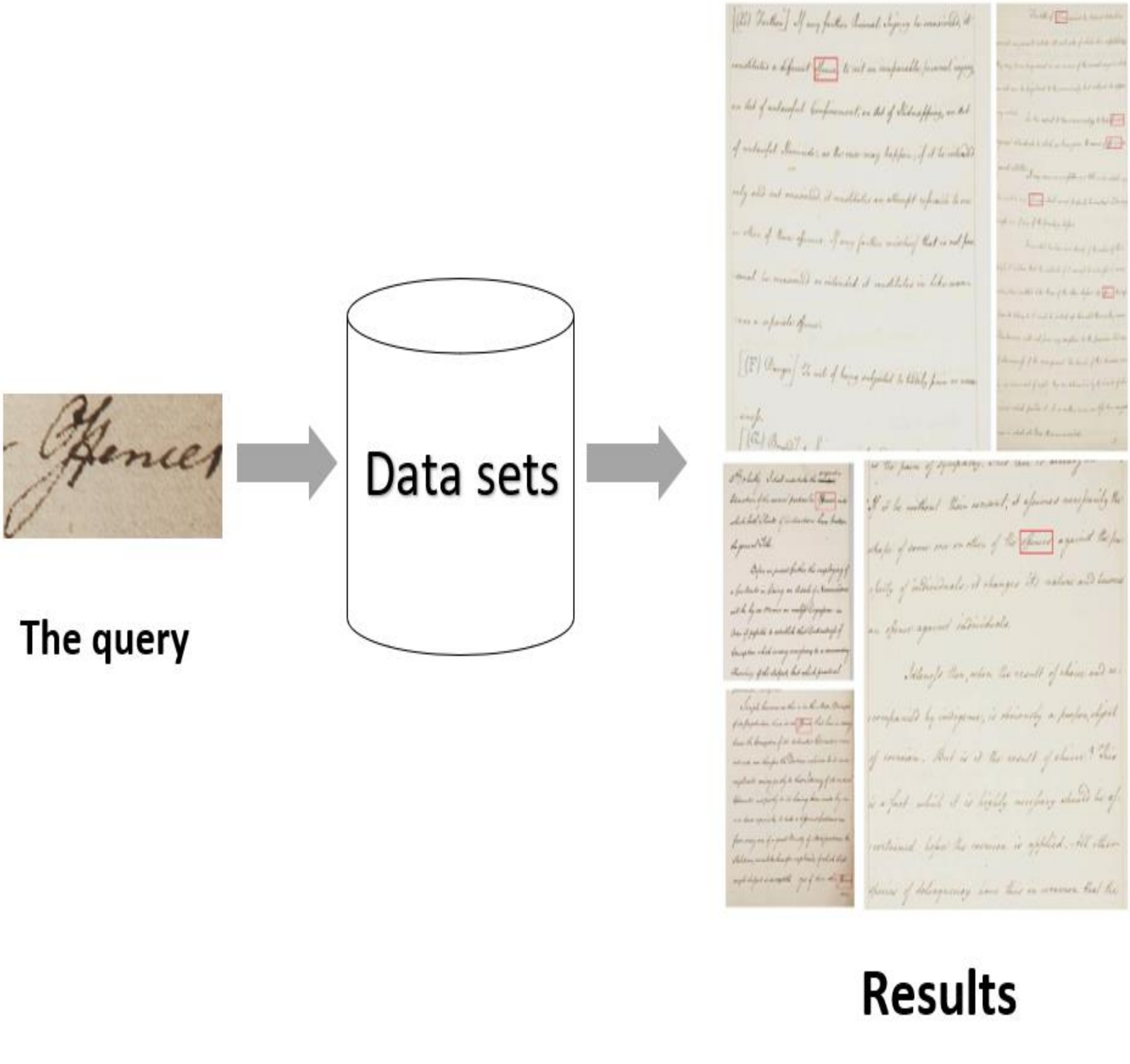


Figure 1.1 The components of keyword spotting system

1.3 Challenges in document image keyword spotting

Keyword spotting in document images raises numerous challenges connected to the quality of the original documents. First, severe degradation involved in historical documents, such as ink bleeds, slows the overall performance of a keyword spotting system. Second, handwritten documents exhibit large differences in style when written, which means the exact query word can be very different from its examples, which greatly increases the problem of the task.

1.4 Applications of keyword spotting

Keyword spotting has numerous practical applications, including indexing and retrieving documents. For example:

- It can be used to locate scenes or specific objects within graphical documents such as maps [12], and it can be useful in other fields such as surveillance, where identifying specific people or vehicles within a large collection of images is important.
- In historical document analysis, keyword spotting can help researchers identify and transcribe handwritten documents from different periods. It also helps transcribers recognize words in degraded documents, particularly those that appear for the first time.
- In the retrieval of documents with a given set of important keywords or phrases within a large oeuvre of documents in enterprise files.
- In performing an internet search within cultural heritage collections housed in libraries worldwide.
- In the medical domain, it can be used for retrieving keywords from reports on pre-hospital treatment (PCR forms) [13] and the analysis of medical records.
- In the automatic classification of handwritten mail correspondence with important words (like revocation, complaint, and urgent) [14].
- In the legal field, keyword spotting can be utilized to identify relevant documents and evidence during a case.

- Recognizing the figures and captions that correspond with them [15].
- In retrieving cuneiform structures constellations from the collections of digitized cuneiform tablets [16].

1.5 Different aspects of keyword spotting

In the literature, Keyword spotting methods are classified according to distinct aspects.

1.5.1 On-line and off-line modes

Handwritten documents can be broadly categorized into two types depending on the method of data collection used: Off-line and On-line handwritings. On-line handwriting is when data is obtained while writing, while Off-line handwriting is data acquisition using a scanner.

It is generally agreed that off-line mode is more intricate than on-line mode for several causes.

1.5.1.1 On-line mode

In the instance of on-line handwriting, the system receives data entry images in real-time (while writing), and symbols are recognized as they are entered manually. It is generally devoted to handwriting. The acquisition of writing requires using equipment such as a graphics tablet or smartphone equipped with an electronic pen.

Online mode has a significant advantage: the possibility of correcting and modifying writing interactively, given the continuous response of the system [17].

1.5.1.2 Off-line mode

Off-line handwriting refers to the process of inputting pre-existing text that has been obtained through a scanner or camera. This results in a grayscale or binary image. It is important to note that this method is similar to the traditional reading task performed by humans [18]. However, it is worth considering that in the off-line case, all temporal information regarding the sequence of plot points is lost. Additionally, the issue of variability in the thickness and form of cursive writing [17] must also be taken into account.

A brief comparison of the two approaches is provided in the table below:

Comparison measure	Off-line	On-line
Acquisition Tools	Scanner or Camera	Graphics tablet or Smartphone
The noise of the image	The existence of big noise	Feeble or none
Available information	There is no information available	The direction of movement, Position, Start points Stop points

Table 1.1 Comparison between types of mode

1.5.2 Keyword spotting techniques

Based on the type of target being searched, the KWS technique can be classified as either segmentation-free or segmentation-based.

1.5.2.1 Segmentation-based keyword spotting

The segmentation-based technique is predicated on the segmentation of a document image into smaller units such as words, or lines. Segmentation-based systems are less practical considering that the problem of segmentation is as challenging. In a keyword spotting system that is highly dependent on the segmentation step and regardless of segmentation faults, many researchers do not implement a segmentation method but use databases where the segments are given.

1.5.2.2 Segmentation-free keyword spotting

The segmentation-free technique can be applied directly to full document images; it takes a general view of the word. In the keyword spotting system, without segmenting the image in any way, the complete image is checked for similarities between the document image and the query image patches. Methods of this technique, on the one hand, skip the segmentation step, but on the other hand, they cannot avoid searching for words in parts of the image that may not include a query. Therefore, methods without segmentation avoid failures due to bad segmentation, but the execution time increases considerably, so segmentation-free systems are usually slow.

1.5.3 Keyword spotting approaches

Two principal approaches to keyword spotting depend on the query's representation. They can be based on Query-by-Example (QbE) or on Query-by-String (QbS) (see **Figure 1.2**).

1.5.3.1 Approach based on Query-by-Example (QbE)

In the QbE approach, the input is an image of the word to search, and the output is a set of the most representative images in the database, including a similar query [7], [2], [19].

Word spotting in this approach does not require prior knowledge of documents and does not require learning. In these approaches, features are extracted from database images and query images. The distances between the images in the database and the image of the query are calculated based on these features. The response to this query is a collection of word images in the document that are most similar to the image of the query.

1.5.3.2 Approach based on Query-by-String (QbS)

The QbS approach [20], [18], [21] enables the user to supply his query in the form of text to search the documents for the corresponding images of words. The user can type on the keyboard the keyword he wants to look for or choose from a list of predefined words.

The QbS has the flexibility to look for any keyword. However, it requires a substantial number of training data sets because the characters are pre-learned.

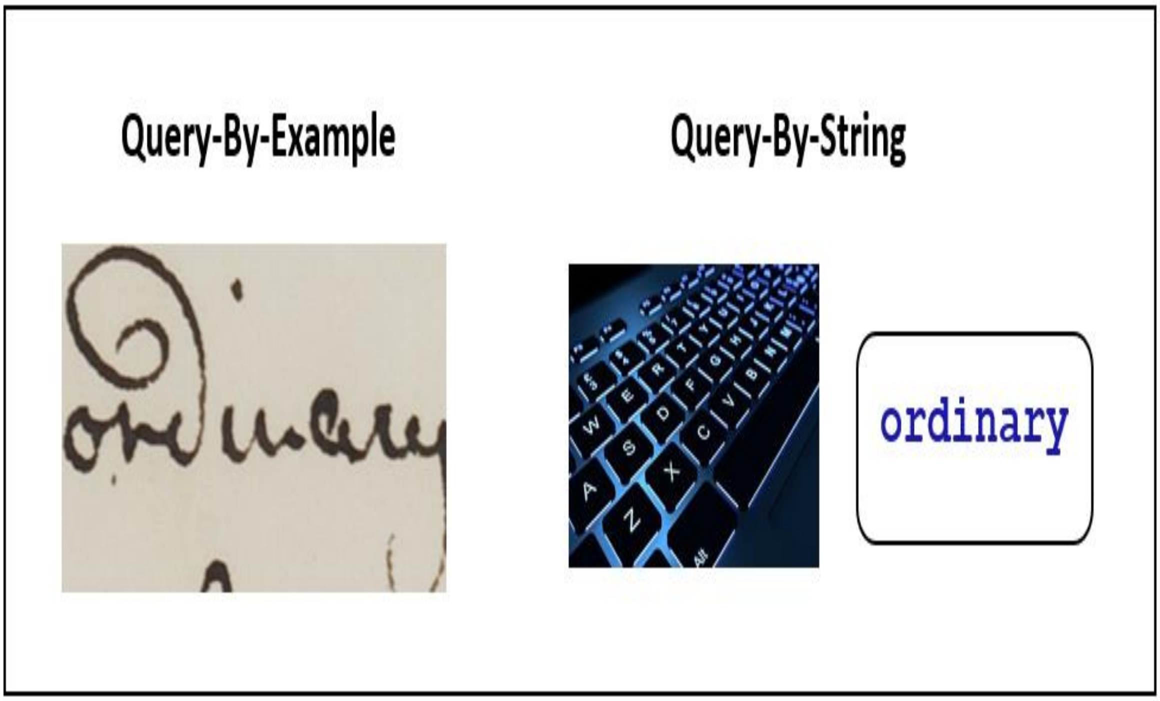


Figure 1.2 The keyword spotting approaches

1.6 General framework for keyword spotting in handwritten documents

The keyword spotting principle allows you to look for a word in a database by finding the best match between the features of the images and the features of the query in the database. It requires several preparation steps (see **Figure 1.3**), which can be summarized in four phases. The first phase is pre-processing the images. After the document images are pre-processed, the page images are segmented into lines or words, depending on the algorithm used. The images of the segmented pages and the request used are represented by digital vectors, which represent the features of these images. Finally, a distance is applied to locate the query in the document's images using their features.

In the sections that follow, we provide a study of the literature for each of the steps of a keyword spotting system.

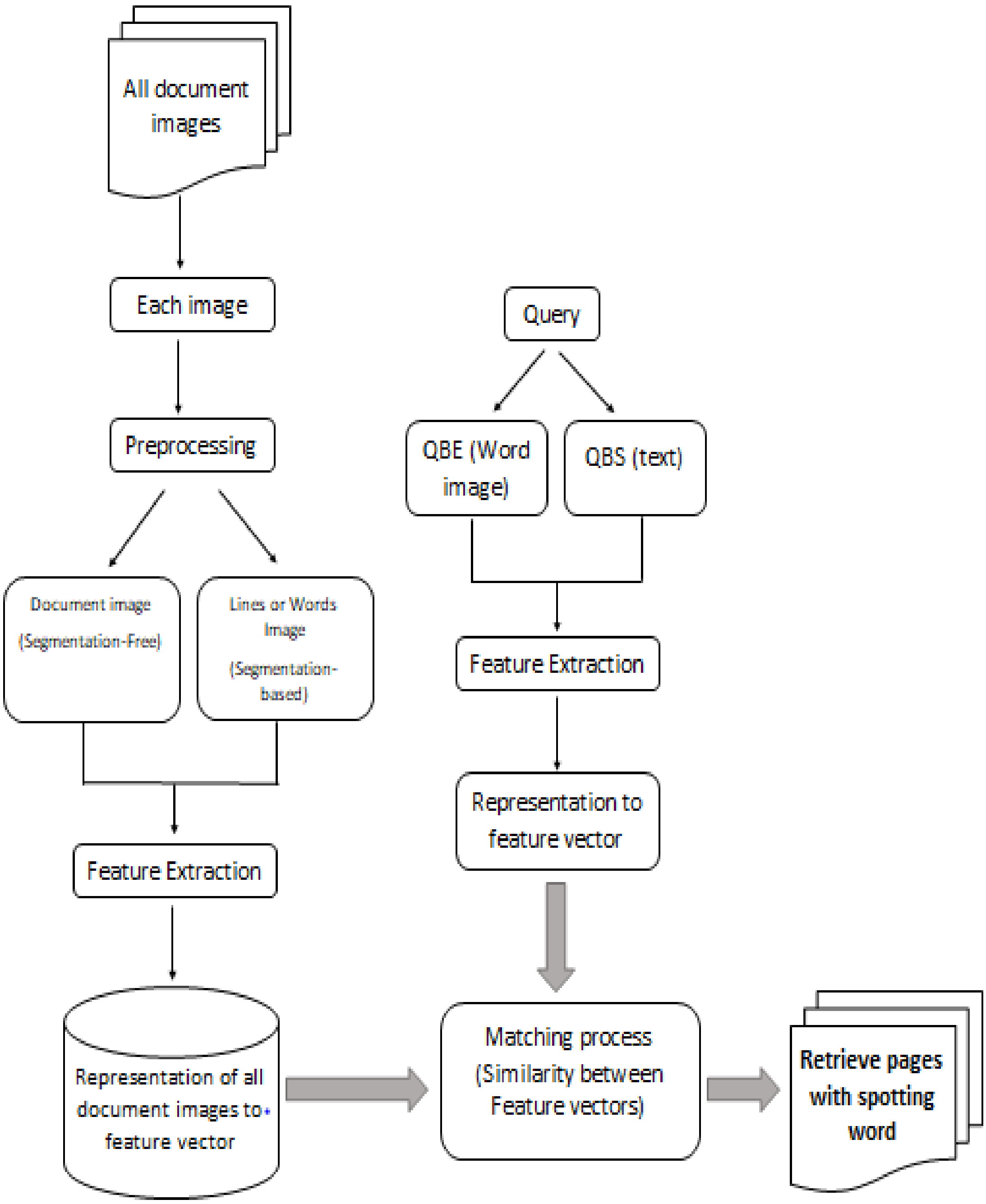


Figure 1.3 Principle of a keyword spotting system

1.6.1 Pre-processing

The pre-processing step is necessary in any keyword spotting system. So, it is the technique applied to format images to improve their quality or reduce the amount of information to process to keep only the most significant information. This step depends on the initial state in which the document is found, as well as on what treatment we wish to apply to it.

Generally, the preprocessing phase includes the following tasks: noise reduction, binarization, and size normalization.

1.6.1.1 Noise reduction

The noise is a deterioration in the image signal due to external sources, especially historical images. The input quality of handwritten documents affects keyword spotting performance due to image noise such as document age, low-quality scanning machines, or archiving efforts, which causes the match rate to drop. In general, improving the image by denouncing the input image or document will always be an advantage of the matching system, whether the input image or document is of poor quality or not.

Numerous techniques for filtering have been presented [22]. The filter is chosen based on how significant the noise is. We find, for example, the Gaussian filter, which is utilized to completely remove low noise by low-pass filtering.

1.6.1.2 Binarization

In the preprocessing step, it is often essential to perform binarization in document analysis and keyword spotting, which converts a raw image into a grayscale image. Binarization aims to sharpen the object's foreground against its background. For the handwriting images, it converts a gray image coded on 256 values to a white binary image (1 or 0) and black (0 or 1) for the background and the handwritten text by a threshold. Although it significantly reduces the amount of data to be processed, significant information loss will occur.

In recent years, several binarization methods have been proposed. These methods can be categorized into two categories: global methods, which use a single threshold for the entire image, and local methods, which consider a different threshold for each pixel of the image.

- The Niblack method [23] involves moving a rectangle window across the whole image to calculate a local threshold at each pixel. The mean (m) and standard deviation (σ) of every pixel in the window (the pixel's immediate neighborhood) are used to determine the threshold. However, if the threshold is set too low, this approach may treat certain pixels as foreground, which could restrict its applicability.
- The Sauvola method [24] is a modification of the Niblack method to give more performance in documents with a background containing a light texture or too much variation and uneven illumination. Sauvola's method is more effective than Niblack's method in the case where the gray level of the text is close to 0 and that of the background is close to 255.
- There is a global Otsu method [25]. This approach presumes that a document image contains two distributions: one for the background and one for the foreground. To reduce the intra-class variation between these two distributions, the overall threshold is calculated. This approach is less effective for degraded images where there is a significant variation in foreground pixels, but it performs well in brighter images.

1.6.1.3 Size normalization

Size normalization is utilized to correct the dimension and position of the handwritten image. This step is applied to the images to reduce all types of variations and to obtain normalized data to facilitate feature extraction and improve their matching.

Many experimental studies have revealed that dealing with images of identical size can produce more homogeneous features and speed up document processing [26]. However, it can cause deformation or remove some useful information.

1.6.2 Segmentation

Segmentation-based keyword spotting methods include a segmentation preprocessing step to segment the document images to the line or word. This is an important step because it extracts the significant regions for the feature extraction, and thus we can extract the characteristics to be able to compare them with those of the query image.

Although segmentation can be deemed a simple task for modern documents, the segmentation of handwritten or historical documents is still a problem because of the significant challenges involved. These include strong proximity or big fragmentation because the problem will be to select the distances between lines or words. A document with different writing angles will significantly impact the difficulty of determining the appropriate angle to perform the segmentation. We also find punctuation marks, decorative letters, touching text parts, and overlapping.

1.6.2.1 Line segmentation

Line segmentation is the method in which from the image, we extract just lines. Horizontal projection is the method most employed to extract the lines from the document images. It will have separated valleys and peaks for the lines that are separated, which perform as the dividers of the text lines. These valleys are easily noticed and used to select the place of borders between the lines.

1.6.2.2 Word segmentation

Word segmentation is the method in which from the line segmentation, we extract solely words. As we know, there is a space between words, so the perpendicular projection helps separate the words by looking at minima in the perpendicular projection profile of a single line.

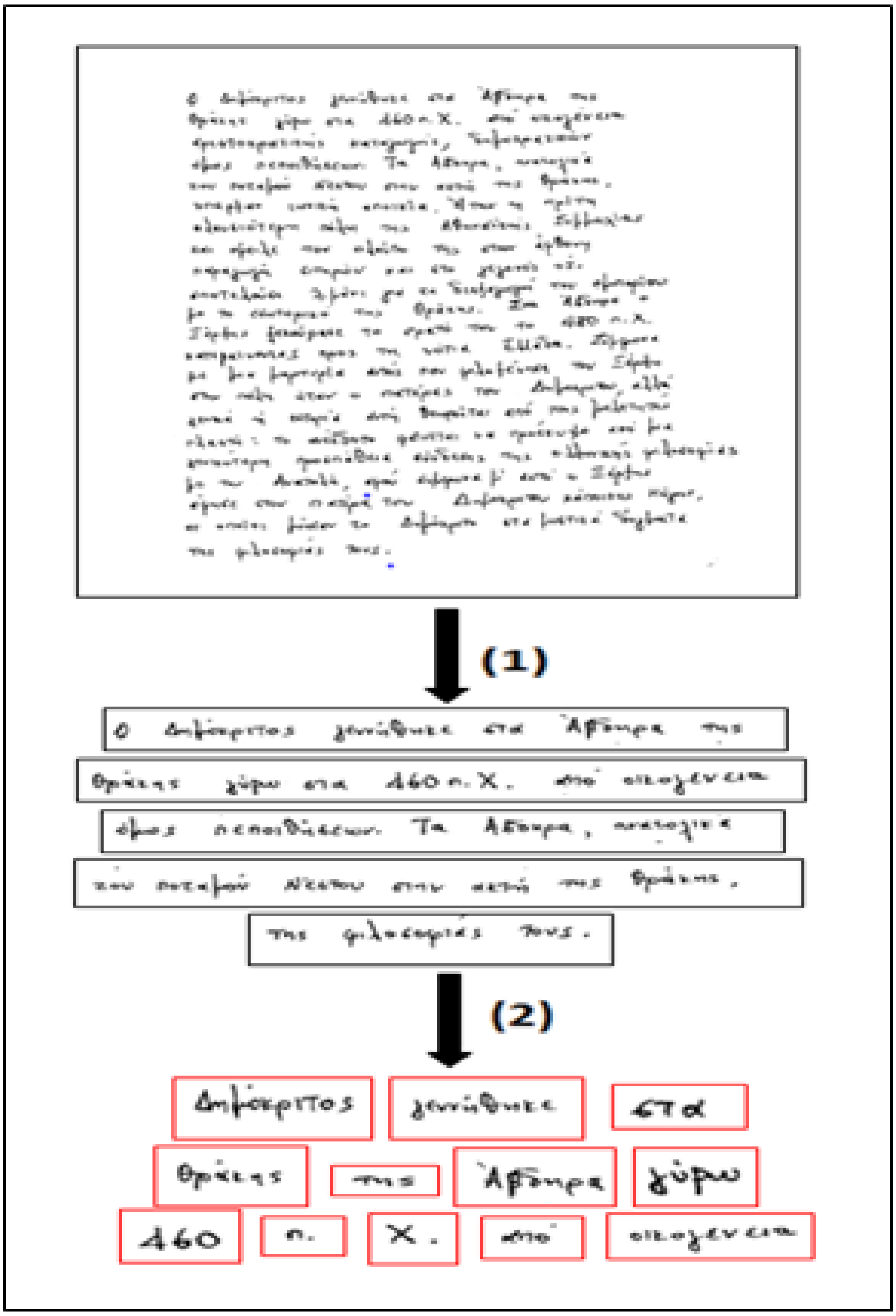


Figure 1.4 The different types of segmentation: (1) lines segmentation, (2) words segmentation

1.6.3 Feature Extraction

Feature extraction is a crucial step in keyword spotting systems because it reduces the dimensions of the authentic images while avoiding the risk of losing significant information, and therefore the system becomes more rapid and effective. Through the feature extraction step, a collection of determining features is extracted for each word image and then used to match results in the subsequent matching process.

There are many methods for feature extraction, and in this thesis, we will explain some of them.

1.6.3.1 Histograms of Oriented Gradients (HOG)

The HOG was first suggested by Dalal and Triggs [27] for human body detection, but now it is one of the considerable effectively used descriptors in keyword spotting, image processing, computer vision, and pattern recognition. The basic idea of this feature descriptor is to describe the local properties of the forms of objects, which are captured by the distribution of edge directions or intensity gradients. It is better than any descriptor for edge as it utilizes the angle of the gradient to calculate the features. It creates histograms for the regions of the image based on the gradient's size and orientation.

The HOG execution is comprised of a series of tasks. In the beginning, the image is separated into smallish “cells”, then for each pixel in the cells, edge orientation and a histogram of gradient directions are calculated and combined these histograms. Finally, blocks are formed by regrouping the cells and normalizing them to make the HOG values invariant to shadows and lighting.

1.6.3.2 Bag of Visual Words (BoVW)

The BoVW [28] is an extension of the Bag of Words [29] approach for digitized document images. This method is also referred to as a histogram of visible words.

The BoVW working method consists of three steps: the initial phase is to extract a specific numeral of image local interest points, and these key points are important image points. In the following step, these key points are employed to extract feature descriptions and combine these feature descriptors. Finally, each word image is represented by a vector that includes the occurrences of every visible word that appears in the image. A similarity measure is utilized to match the given image of the query and the collection of images in the database based on these feature vectors.

1.6.3.3 Scale Invariant Feature Transform (SIFT)

David Lowe developed the SIFT descriptor in 1999 [30]. It is a local, gradient-based image descriptor employed to extract information from an image area around a point. The SIFT descriptor is an extremely strong and robust local feature descriptor for many tasks in computer vision, such as video tracking, object recognition, and document image analysis. In the field of keyword spotting [21], [31], its use has led to good results.

The operation of a SIFT descriptor can be split into the following steps: The first phase is extracting interest points (key points) from labeled gray-level images by a Gaussian window, indicated by the overlaid circle (see **Figure 1.5**). In the second step and a unique aspect of SIFT, the feature vector is calculated by finding histograms of gradient directions in a local neighborhood about each key point. Next, lousy interest points such as borders and areas of low contrast are discarded, and orientation histograms and gradient modulus are assigned for the remaining interest points. Finally, with the rotation and scale invariance in place, a Hough transform is performed to select the clusters from a specific object. Next, the likelihood of a particular feature vector representing an object in the image is computed.

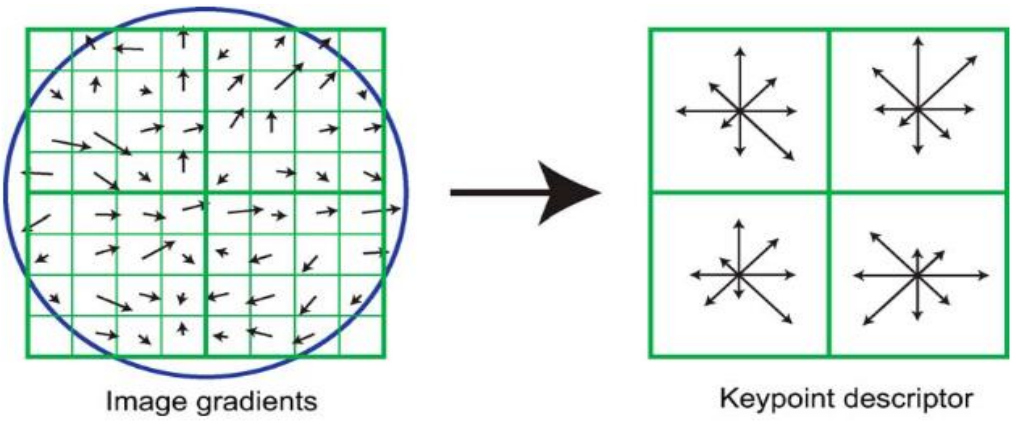


Figure 1.5 Image gradients and key point descriptors [32]

1.6.3.4 Speeded-up robust features (SURF)

The SURF proposed by Bay et al. [33] is a local feature descriptor and a detector. It has been used to locate, recognize objects, and extract points of interest. It is deemed an approximate version of SIFT. The difference is that SURF embraces integral images and computes Harr wavelet rather than orientation histogram, which makes SURF more computationally efficient than SIFT. SURF is typically utilized to address the issue of speed; it is several times faster than SIFT.

1.6.3.5 Pyramidal Histogram of Characters (PHOC)

The PHOC is a binary representation of words submitted by Almazán et al. [34]. It plays a part in representing word images and strings. It encodes if a certain character emerges in a certain spatial region of the string using a pyramidal disintegration, creating it discriminating (see **Figure 1.6**). The first level is just a basic histogram of the characters encoding the existence or absence of a certain character in the string. Then, new levels are counted, where at every level of the pyramid the word is further separated, and a new histogram of characters is counted for each new division to calculate for characters in different parts of the word.

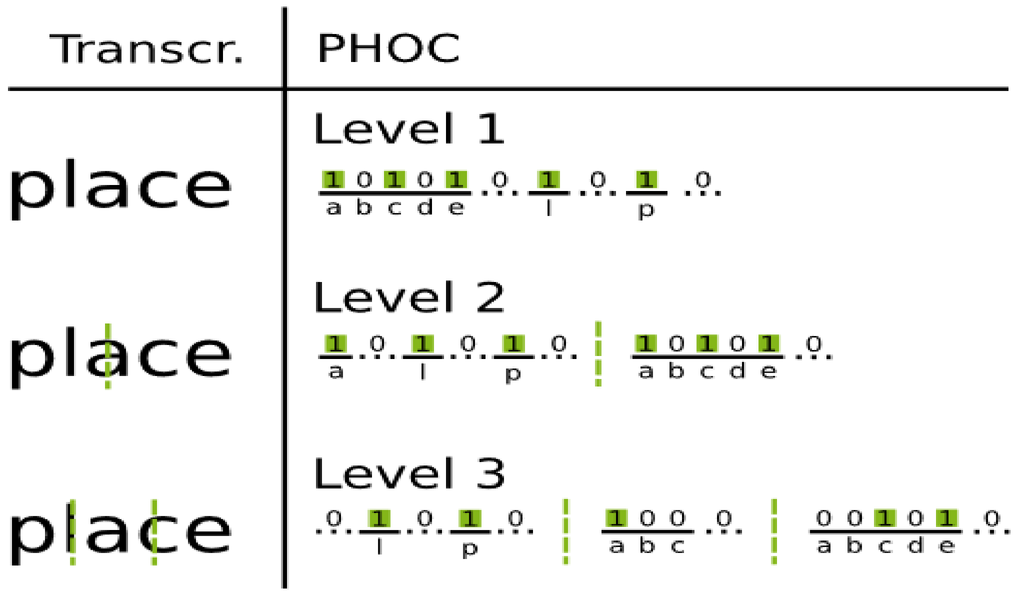


Figure 1.6 The extraction of a PHOC from a specific text string at levels 1, 2 and 3 [5]

1.6.4 Classification techniques

The classification step is the central component of the keyword spotting framework. Once a set of features has been extracted, this step involves searching and matching query image representations with document image representations, ultimately classifying the results. The main objective of this step is to identify word images within documents that are similar to a given query word image, achieved by calculating the similarity of the extracted feature vectors.

In the literature, two main techniques have been distinguished: Supervised and Unsupervised techniques.

1.6.4.1 Supervised techniques

Supervised techniques require a large number of training data to achieve more elevated retrieval performance. In this technique, features extracted from word images are represented by statistical models such as HMM, ANN, and others.

1.6.4.1.1 Hidden Markov Models (HMM)

The HMM is a statistical model that was first presented by Baum et al. in 1966 [36] and uses a Markov technique that includes hidden and unknown parameters. It is one of the most popular statistical models for sporting sequential and temporal data. They have simple and practical mathematical and theoretical foundations and have proven effective in solving problems like automatic speech recognition, handwriting recognition, and keyword spotting.

The principle of HMM is that the observed possibilities have no one-to-one correspondence with states but are linked to states via a probability distribution. It is a doubly stochastic process, which contains a Markov chain as the basic stochastic process and describes state transitions and stochastic processes that describe the statistical correspondence between the states and observed values. From the viewpoint of observers, only the observed value can be viewed, while the states cannot. A stochastic procedure is operated to identify the presence of states and their features. Thus, it is dubbed a “hidden” Markov model.

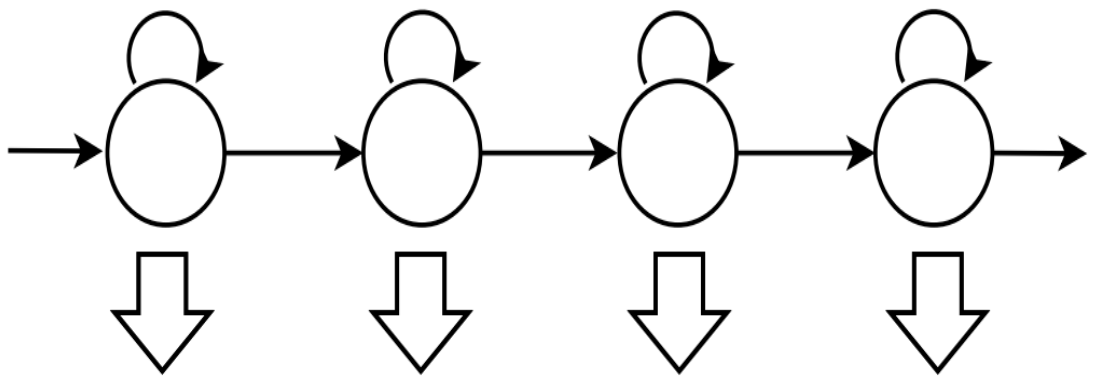


Figure 1.7 Building the models of the HMM

1.6.4.1.2 Support vector machine (SVM)

The SVMs are a set of learning-based techniques designed to solve classification (i.e., deciding which class a sample belongs to) and regression (i.e. predicting a variable's numerical value) problems from databases, including non-linear and linear, using a hyperplane in a high feature space.

The SVM classifier presents an example of a training set (x_i,y_i) where the xi are the data samples and the yi are the labels suggesting which class the sample belongs to. For the two-class pattern recognition problem, y_i = -1 or y_i = +1. Where, an example of a training (x_i,y_i) is called negative if y_i = -1 and positive otherwise.

For more complex problems, the characterization of a hyperplane can be very complicated and quite non-optimal. For instance, in a plane in which the (y_i = +1) points are clustered in a circle, with (y_i = -1) points scattered around, a two-dimensional hyperplane will not be able to accurately separate the two groups. This is known as a non-linearly separable problem. To address this issue, SVM uses kernel tricks to transform the data and apply linear classification to non-linear problems.

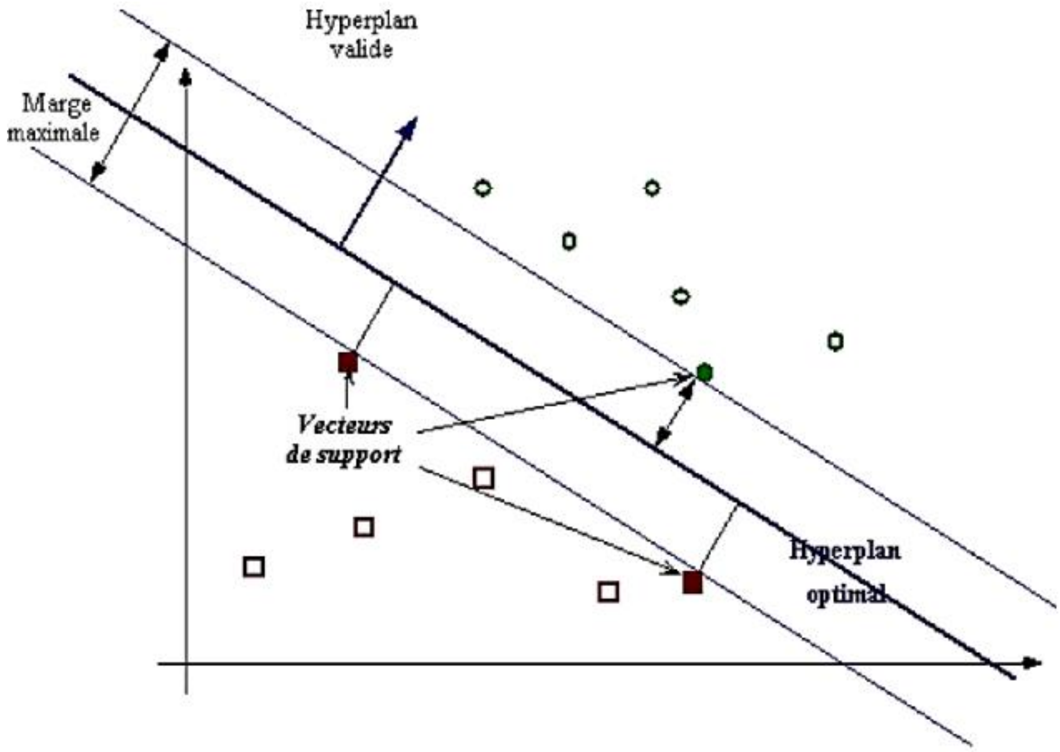


Figure 1.8 Example of SVM

1.6.4.2 Unsupervised techniques

The unsupervised techniques utilize matching to identify images of words that are similar to the query, making it a convenient option when training data is difficult to obtain, such as with ancient manuscripts. The similarity (distance) between two feature vectors can be calculated using various measures, including DTW distance, distance classifiers (such as Correlation, Cosine, Euclidean, City-block, etc.), and others.

1.6.4.2.1 Dynamic time warping (DTW)

The DTW algorithm [35] is an algorithm that permits measuring the degree of similarity between two sequences that can vary over time. It has extensively been applied such as graphics, video, bio-informatics, audio, speech processing, and in other domains. DTW is often used to match word images because it provides additional flexibility to compensate for handwriting differences.

In general, DTW is a method that seeks an optimal matching between two-time series, subject to certain constraints. The time series are deformed by the non-linear transformation of the temporal variability to determine a measure of their dissimilarity, apart from some non-linear transformation of time.

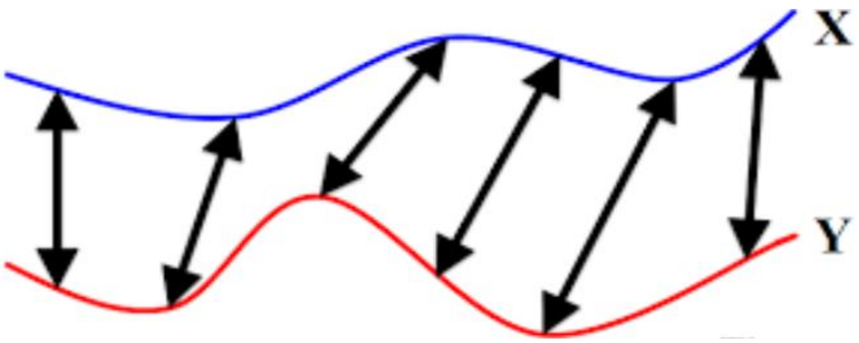


Figure 1.9 Building the models of the DTW

1.6.4.2.2 Distance Measures

Some keyword spotting systems make decisions based on a distance measure. Distances between feature space representations are utilized as the basis for the matching step.

Table 1.2 summarizes the equations for some distance measures.

The distance classifier	Equation
City-block	$D_{city-block}(x,y) = \sqrt{\sum_{i=1}^{l} \lvert x_i - y_i \rvert}$
Euclidean	$D_{eucl}(x,y) = \sqrt{\sum_{i=1}^{l} \lvert x_i - y_i \rvert^2}$
Correlation	$D_{corr}(x,y) = \frac{cov(x,y)}{\sqrt{Var(x).Var(y)}}$
Cosine	$D_{cos}(x,y) = \frac{x * y}{\lvert\lvert x \rvert\rvert \lvert\lvert y \rvert\rvert}$
Spearman	$D_{spe}(x,y) = 1 - \frac{6\sum_{i=1}^{n}[r(x_{i)} - r(y_i)]^2}{n^3 - n}$ Where n: number observations

Table 1.2 The equations for distance measures

Where:

x and y are feature vectors of two documents.

l: is the size of each feature vector.

1.7 Conclusion

In this chapter, firstly we presented an overview and the different phases of keyword spotting systems on which the work of this thesis is based, where we concentrate on extracting the features in the handwriting document and matching them. These features are needed to create an efficacious system of keyword spotting.

In the chapter that follows, we will provide a description and comparison of the most state-of-the-art contributed works in this field.

Chapter 02: Keyword Spotting in Handwritten Documents: A state-of-the-art

2.1 Introduction

The current state-of-the-art for keyword spotting in handwritten documents is summarized in this chapter. The first sections focus on presenting the main research studies in this field. Additionally, the chapter discusses various competitions related to keyword spotting, and we end the chapter with a comparison of the different works in the domain.

2.2 Related works

Keyword spotting remains a fascinating field of study for both handwritten and printed documents. However, current works can be separated into two classes: segmentation-free and segmentation-based techniques:

2.2.1 Segmentation-Based methods

A statistical technique is presented by Rothfeder et al. [45]. It is a method that depends on the matches segmented word images by recuperating similarities between interest points, the suggested system tracks a segmentation-based technique according to a query-by-example approach. Using the Harris corner detector, this system extracts the similarities between the two images' points of interest. It then leverages these similarities to build a similar measure by utilizing Euclidean distances.

They reported an average precision of 62.57% on a database of historical documents with 2372 images of good quality and 15.49% on 3262 images from documents of poor quality.

In an additional piece of work, Rodríguez et al. [46] presented a local descriptor for unconstrained handwritten keyword spotting. They depended on a Local Gradient Histogram (LGH) to extract features where, over a word image, a sliding window locomotive from left to right. At every position, the window is split into cells, and a histogram of orientations is compiled in every cell. This step is considered the most essential phase of their work. All tests will be taken out using two scoring mechanisms: HMM and DTW to find the degree of similarity between the query and the word image, and this is in the matching step.

The effectiveness of the suggested approach was verified via an experiment executed on the database that a company's customer service received, which had 630 scanned letters written in French. Where they used the QbE approach. The outcomes obtained using HMM were better than those obtained by DTW.

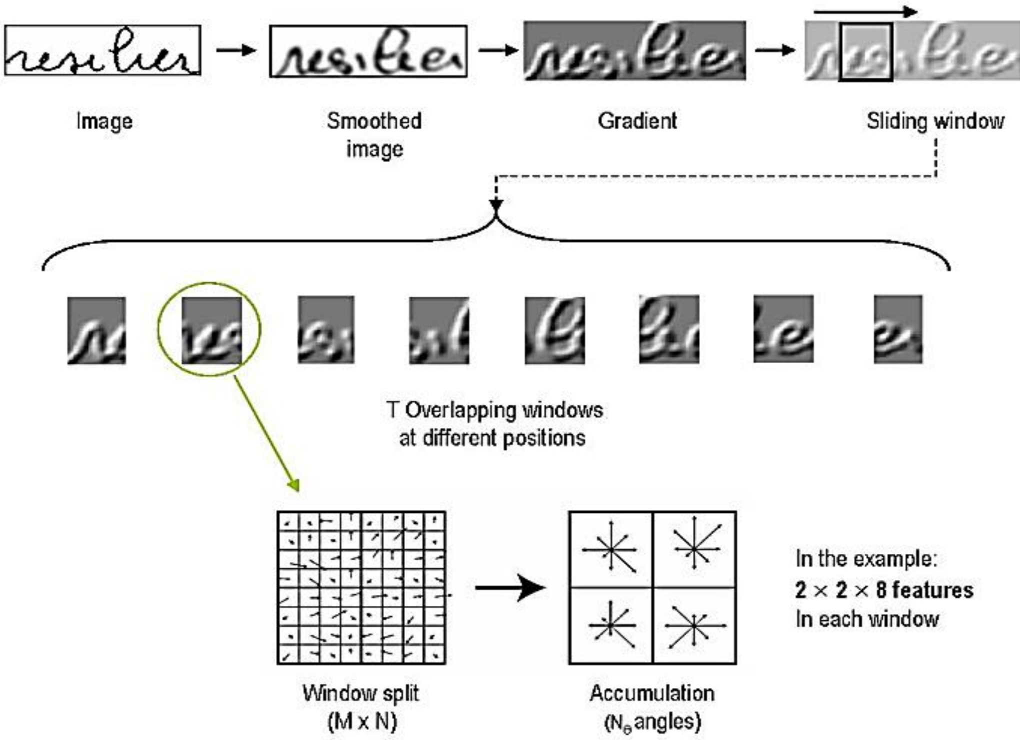


Figure 2.1 Example of the Extraction of LGH [46]

Another keyword spotting method based on online segmentation and the query-by-example approach is presented by Terasawa et al. [47]. They relied on a sliding window, DTW-based matching, and a slit-style HOG feature (SSHOG). A sliding window is used for every line of text to extract feature vectors; this window is moved in the direction of writing. For each sub-image determined by the window, a HOG feature vector is computed. Features are compared using the DTW method.

The authors conducted the experimental test using an English manuscript from the GW database and images of Japanese manuscripts from "Akoku Raishiki's" scanned diary.

A novel system for handwritten word spotting of mail documents utilizing a statistical framework is presented by Rodríguez et al. [14]. They relied on HMMs to model the extracted feature vector sequences using the LGH feature of word images and the query from the database, and a GMM was utilized to normalize scores.

During the phase of matching the query with the word images in the database, the similarity between the word image and the query is acquired by the posterior probability produced by the HMM of the query. The efficacy of the offered method was verified through an experiment executed on the database that included 630 scanned letters that were written in French and delivered to a company's customer service department.

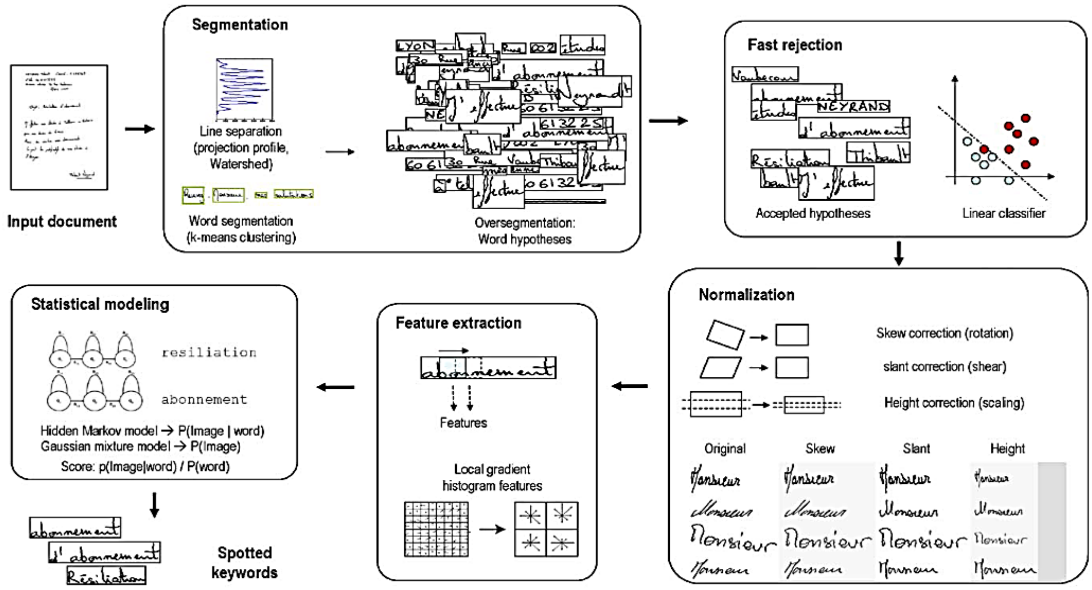


Figure 2.2 Summary of the proposed system by **Rodríguez et al.** [14**]**

In one of the techniques employing the query-by-string framework, Aldavert et al. [49] use two representations: textual and visual. The textual representation is based on n-grams of characters, while the visual representation depends on the Bag of Visual Words (BoVW) method. The two representations are merged and projected into a sub-vector space using the latent semantic analysis method. The GW database has been utilized to offer the suggested technique.

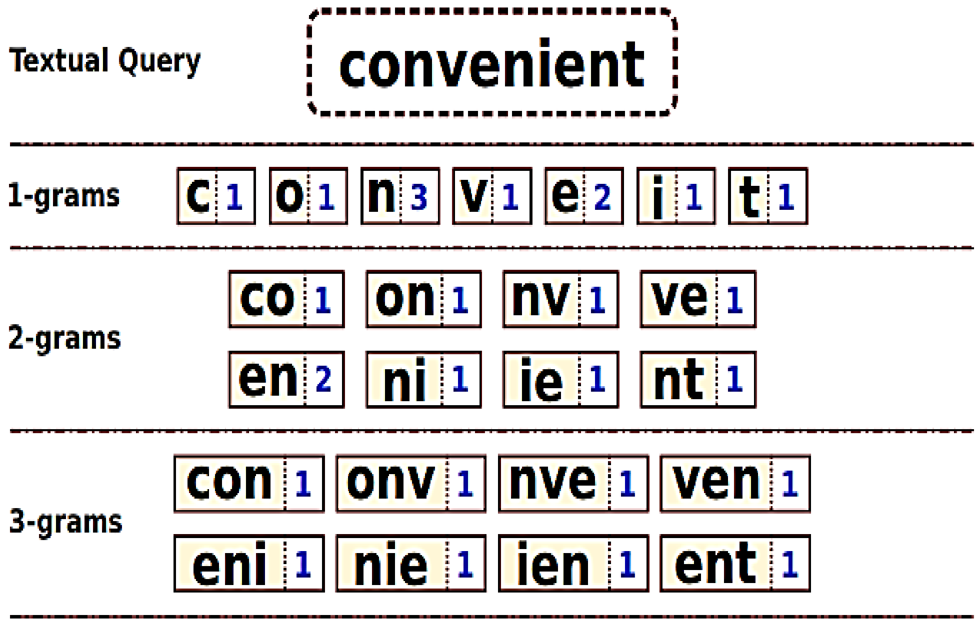


Figure 2.3 An illustration of an n-gram textual description [49]

A novel approach that allows efficient keyword spotting in handwritten documents based on typical keypoint detection is proposed by Zagoris et al. [50]. They are called Document Specific Local Features (DSLF). The suggested system tracks a segmentation-based technique according to a query-by-example approach. Their strategy comprises two independent stages: the Online and the Offline. In the Offline stage, document images are divided into word images, and DSLF extracts and indexes the features in the database. In the Online stage, the query word image's DSLF is extracted, and each indexed word image's feature set is compared against the query's features using a Local Proximity Nearest Neighbor (LPNN) search.

Experimental results are displayed on two historical handwritten databases for segmentation-based keyword spotting context (Bentham and GW databases).

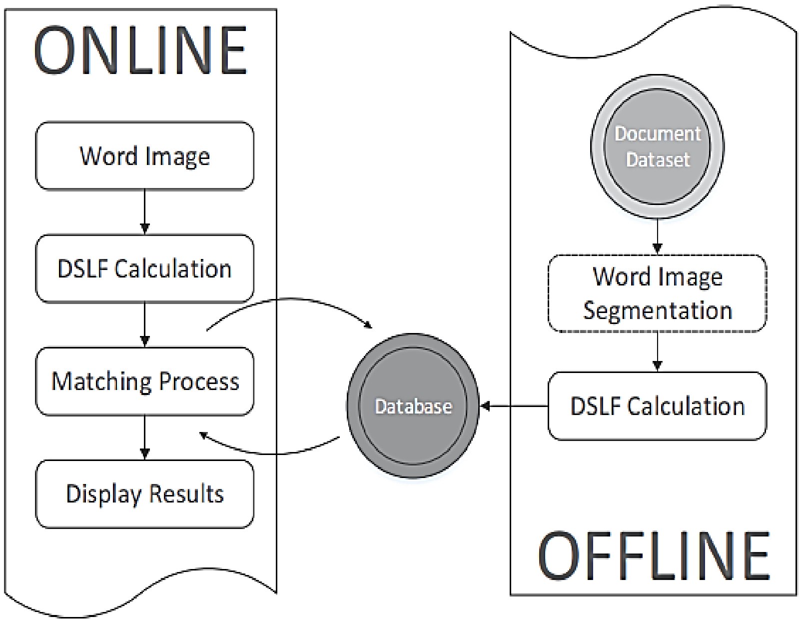


Figure 2.4 System diagram offered by **by Zagoris et al.** [50]

Retsinas et al. [1] suggested a method for segmentation-based and query-by-example approach keyword spotting on handwritten documents. Their method is broken down into three phases: preprocessing, feature extraction, and matching. In the feature extraction process, a sequence of descriptors is created by combining a modified Projections of Oriented Gradients descriptor with a zoning scheme. An uneven zoning scheme is established by creating only denser zoning for the query images for a more comprehensive representation, resulting in a substantial decrease in the document collection's storage requirements. In the matching step, the suggested MISM method effectively determines the distance between the word sequence and the query. An experiment conducted on various databases proved the effectiveness of the suggested method, and the outcomes were satisfactory.

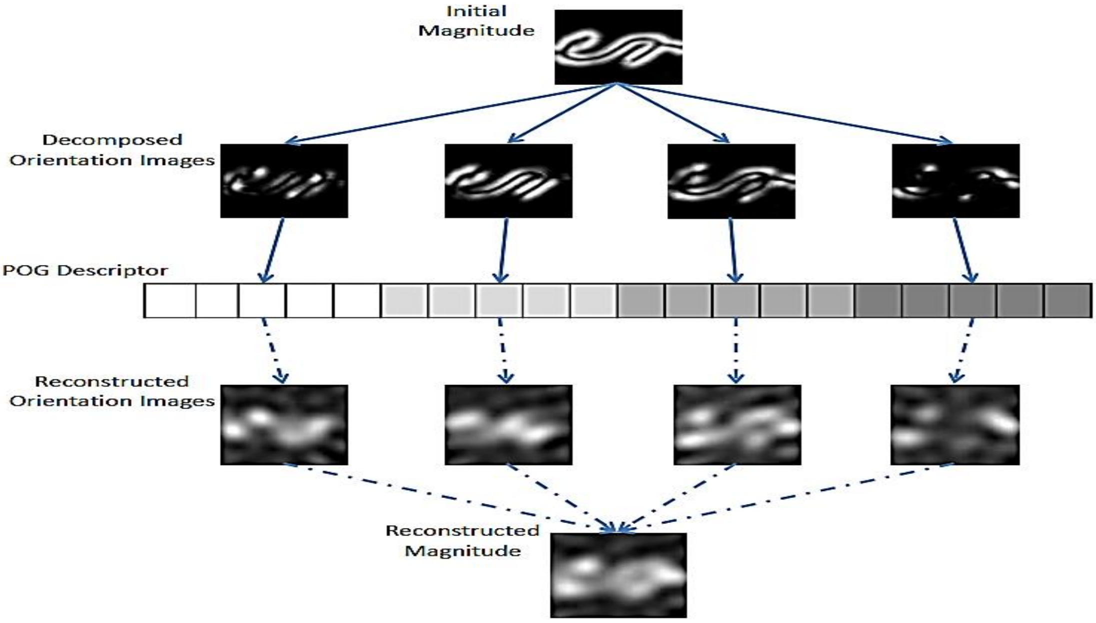


Figure 2.5 The extraction of mPOG descriptors from an image part and their corresponding reconstruction [1]

In a template-based keyword spotting technique, Stauffer et al. [56] captured the structural information of the segmented word images employing a graphical representation. Graph-matching techniques are then employed to compare reference images and the query. The presented technique was evaluated on four different databases: GW, Alvermann Konzilsprotokolle (AK), Parzival (PAR), and Botany (BOT).

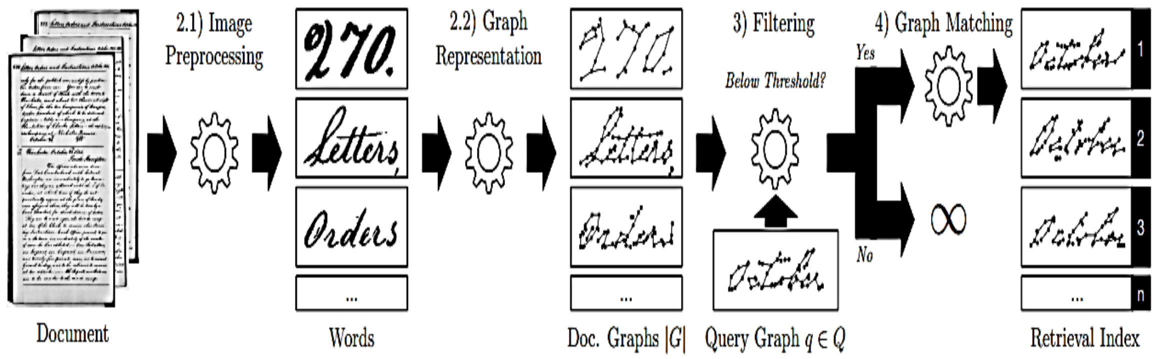


Figure 2.6 Method of graph-based keyword spotting of the word "October" [56]

Among the methods that relied on deep learning to crack the keyword spotting problem, we find Serdouk et al. [57] employ a Siamese neural network (trained on triplets) to retrieve words similar to the query image by Euclidean distance in the matching step. The suggested system tracks a segmentation-based technique according to a query-by-example approach. The results received for the GW database indicate the efficacy of the keyword spotting method they presented.

Figure 2.7 The SNN architecture with three inputs [57]

In one of our studies [58], we proposed another method for keyword spotting problems, which relies on a segmentation-based technique according to a query-by-example approach where the location of words in the document images is provided in the ground truth. In the first step, we used oriented basic image features (oBIFs) for the extraction of features from the word images. Using a variety of distance measures, features extracted from segmented handwritten words are compared with those of the query images in the matching process, which is the second step. The experiment of the system was conducted using some of the images from the modern database of the ICFHR 2014 keyword spotting competition, and the mean accuracy precision (mAP) scored 76.86 % when using City-block distance for matching.

They suggest Kundu et al. (2021) [3], a segmentation-based method utilizing the query-by-example approach for keyword spotting in handwritten papers. Pre-processing, vertical zone division, feature extraction, and feature matching are the four main steps that make up their methodology. In the feature extraction steps, they used the Hough transform-based angular feature to extract features from each of the segmented regions of both the target word images and query images. Using a DTW-based metric, they determine the feature matching step score by comparing the descriptors of features of a target word image with a query word image. To complete the needed experiments, they used three databases: ICDAR KWS 2015, Qatar University Writer Identification (QUWI), and IAM. We obtained mAP scores of 45.01%, 53.99%, and 86.40%, respectively.

In another deep learning-based method, Daraee et al. (2021) [59] employ a DNN with Monte-Carlo dropout to estimate the certitude of the extracted features, which they utilized in both QbS and QbE keyword spotting. In the QbE, the threshold is assigned based on the confidence of the words in each class through the training step. Using the cosine distance, the distance between the anticipated certainty of the query picture and those in the reference base is compared with each class's threshold during the matching step. In the QbS, the class of the query image is determined and contrasted with the retrieval set class, which is obtained by certainty prediction.

Experiments on four public databases confirm the notability of this method in comparison to the current methods.

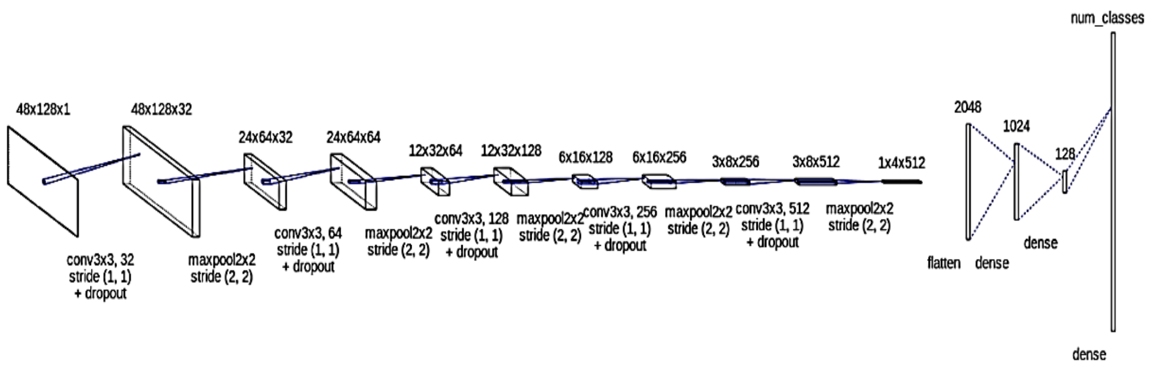


Figure 2.8 The Monte-Carlo dropout Architecture [59]

A novel method for spotting keywords in historical Vietnamese manuscripts has been presented by Scius-Bertrand [61]. They proposed an annotation-free KWS technique that learns from a printed font instead of labeled handwritten samples. Their method consists of two components. Initially, printed characters from synthetic pages are used to train a deep convolutional character detection system based on YOLO. Thereafter, the character images are described employing keypoint graphs and compared with query graphs utilizing the HED (Hausdorff edit distance) [62] to recover the most similar samples. They assessed their method for spotting Chu Nom logographic characters on the recently launched Kieu database, which is a historical Vietnamese manuscript with 719 digitized pages from the well-known Tale of Kieu, and reported a mAP of 77% for QbE and 63% for QbS.

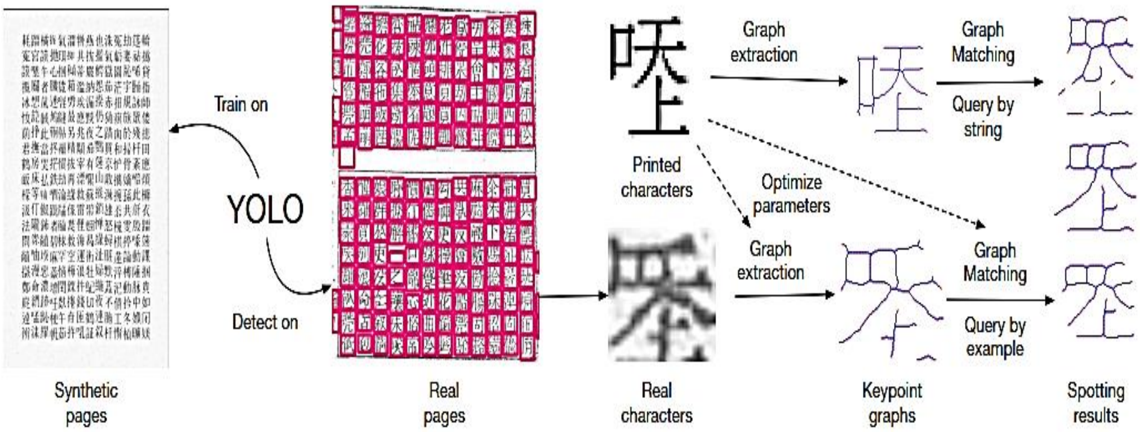


Figure 2.9 Overview of the system proposed by **Scius-Bertrand et al.** [61]

Another of our studies [92], uses query-by-example (QBE) and a segmentation-based technique for keyword spotting in historical documents. In the first step, we extracted features utilizing textural features to verify their effectiveness in historical documents, specifically LDNP, CLBP, and CRLBP. The second step is to match the extracted features from both images using the Euclidean distance. To improve the matching rate, we used the product (Prod) of different distance measures with matching features.

2.2.2 Segmentation-Free methods

In another relevant study [48], the authors employ SIFT descriptors with a BoVW model and get the similarity list utilizing the cosine distance, at the retrieval stage. They evaluate the technique on Latin in addition to Persian documents, both handwritten and printed, for the segmentation-free keyword spotting method.

In another study, Kovalchuk et al. [51] extracted fixed-size descriptors for each image using a HOG and a LBP, and they performed the retrieval step using a simple distance measure (Euclidean distance). Because of their simplicity, this is important for large collections of documents. The efficacy of the method proposed was demonstrated through an experiment executed on different databases: the GW database and the LB database, and the results were satisfactory.

Yao et al. [52] developed a technique for handwritten keyword spotting. The HOG based two-directional DTW is the foundation of this technique. They connected the HOG descriptors in the same column and row and got a sequence of feature vectors after they extracted the HOG descriptors from every cell in the normalized images. The two-directional DTW approach was utilized to define the distance between the query image and the existing normalized vectors from the HOG descriptors. This method produced more reliable results than the traditional DTW method. The results are displayed on the GW database and the CASIA-HWDB 2.1 database in a segmentation-free context and a query-by-example approach, reporting a mean Average Precision (mAP) of 57.2% and 70.5%, respectively.

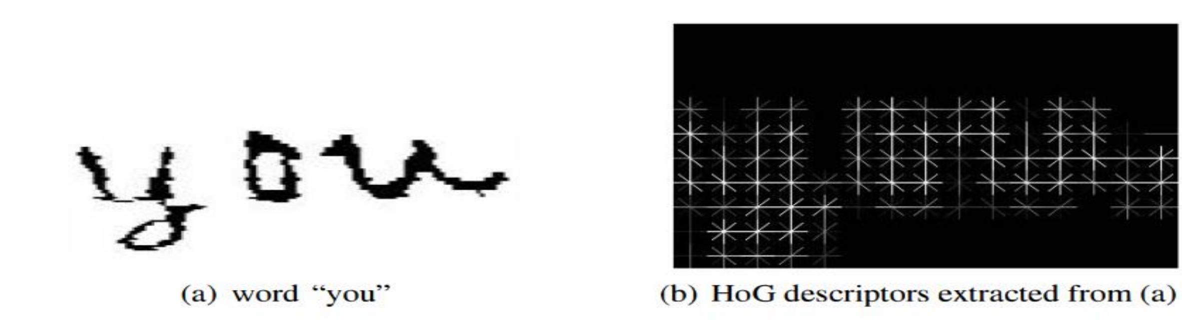


Figure 2.10 The words and their corresponding HoG descriptors. (a) word “you” in GW database. (b) HoG descriptors extracted from (a) [52]

Rabaev et al. (2016) [53] focus on a pyramid-based approach for keyword spotting in historical document images, and they chose a segmentation-free retrieval scheme and a query-by-example approach where locating the query word in a document is done in a scale-space pyramid. In this work, the HOG descriptors are extracted at each level of the pyramid. A hierarchical search is conducted beginning at the most elevated level of the pyramid for query matching. Four distinct historical document collections (GW, LB, CG, and AH) were used to test the proposed methodology.

In light of recent advancements in keyword discovery using deep learning, Ritsinas et al. [63] developed a segmentation-free system and the QbS paradigm, where no previous information about the location of the word is known on the document page. The basic idea of this work is “efficient bounding box estimation by counting character occurrences”. They first built a CNN-based system that turns a document's input image into a thumbnail map of possible characters. Then, by considering KWS as a character counting problem, they aim to find the bounding boxes that contain the required characters. To find the most similar regions, they calculated the count for each letter within the bounding box and compared it to the query count. The comparison is made by cosine similarity. Count-based retrieval cannot differentiate between various permutations of query characters. Therefore, they addressed this problem, offering two various techniques that can be combined into one method, as different steps, and have a common feature: the already appropriately trained network is utilized to expect more precise results effectively. A typical, non-maximum suppression step follows both steps. These steps are illustrated in full in **Figure 2.11**.

Experimental verification on two widely utilized databases: IAM and GW, shows that their method achieves promising results.

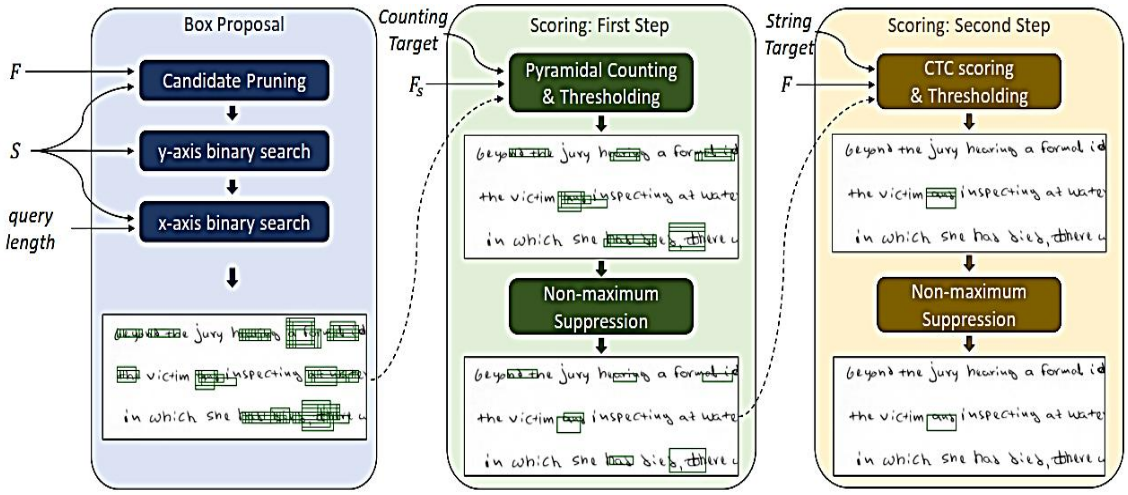


Figure 2.11 Overview of the suggested spotting pipeline [63]

2.2.3 Segmentation-Free and Segmentation-Based methods

Zagoris et al. (2017) [54] decided to improve their work based on the suggestions of Zagoris et al. (2014) [50]. Where they proposed a method for keyword spotting that utilizes DoLF (document-oriented local features). This method involves two steps: first, the extraction of document-oriented key points and the computation of instrumental information around these points; and second, the use of Nearest Neighbor Search (NNS) [55] for matching. The experimental outcomes of this study are displayed for two various techniques (segmentation-based and segmentation-free) according to a query-by-example approach using the Bentham, GW, and Barcelona Historical Handwritten Marriages databases (BH2M).

Due to the deteriorating and noisy historical documents, Mohammed et al. [60] suggested a new powerful technique for multilingual keyword spotting utilizing two various feature extraction methods: HOG and SURF features. They first extracted regions of interest (ROIs) and then re-ranked the extracted ROIs utilizing a combination of various feature extraction and matching methods using the Brute-Force algorithm, which is based on the L2-norm. This technology deals with two types of techniques: segmentation-free and Segmentation-based utilizing the query-by-example approach.

The proposed method showed enhanced performance, which conducted two GW and

HADARA databases utilizing the standard evaluation method.

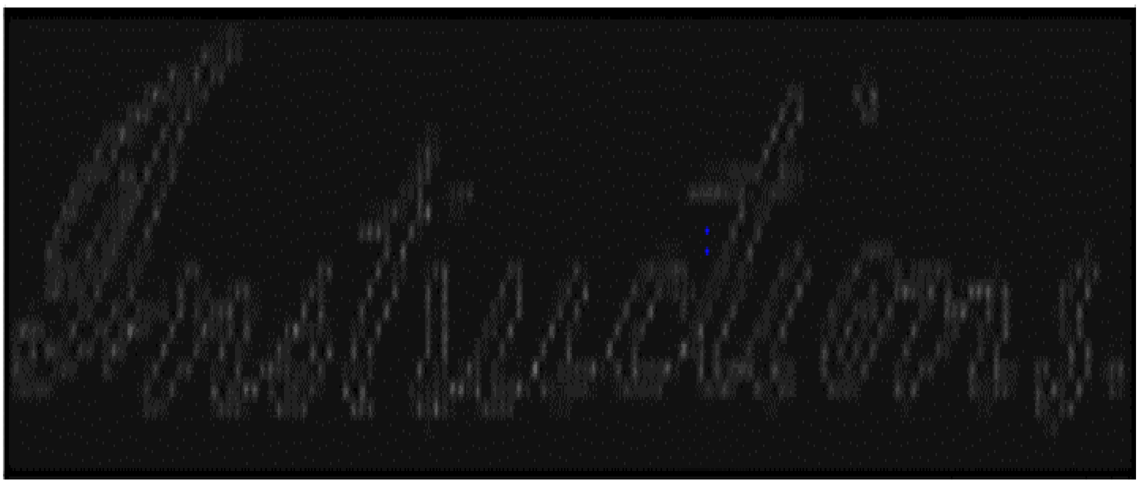


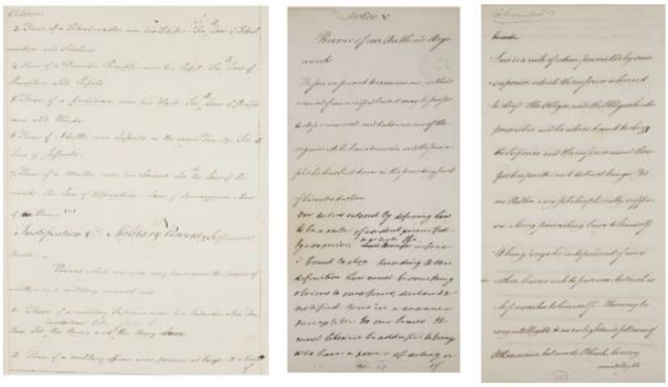
Figure 2.12 HOG descriptors of the query image with cell size (4, 4) [60]

2.3 Keyword spotting competition

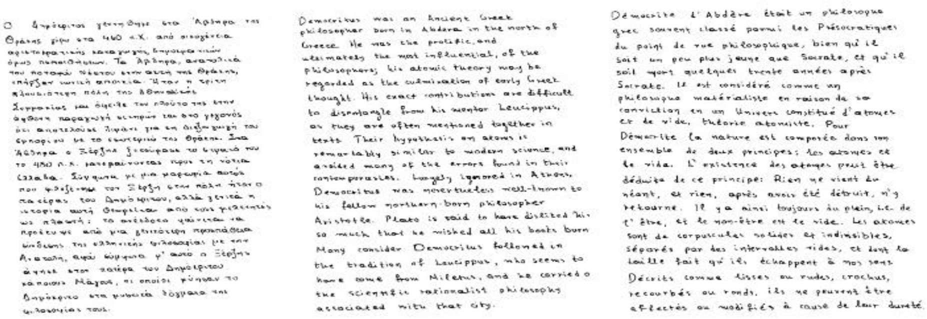
The interest of researchers developed in keyword spotting and in methods for evaluating these matching systems has grown dramatically in the last few years. Due to the significance of the domain, researchers have organized several competitions. The purpose of these competitions is to give a platform for the comparative evaluation of methods developed by researchers. It is also critical to note that all of these competitions took place on different-sized databases with different content and words, and the evaluation protocols are very different.

2.3.1 ICFHR 2014 Competition on Handwritten Keyword Spotting (H-KWS 2014)

H-KWS 2014 is the Handwritten Keyword Spotting Competition [64]. It is one of the most prominent competitions that have been composed in intersection with the International Conference on the Frontiers of Handwriting Recognition (ICFHR 2014). In this competition, an assessment framework is created for comparing handwritten keyword spotting methods that answer the query by example problem. Where the competition featured two distinct tracks, namely, the segmentation-based track, in which the location of the word images in the document images of the database is given, and the segmentation-free track. The competition included five (5) distinct research groups, three (3) solutions for the segmentation-based track, and four (4) solutions for the segmentation-free track.



(a)



(b)

Figure 2.13 Sample document images from the (a) Bentham database, (b) Modern database of the H-KWS 2014

2.3.2 ICDAR 2015 Competition on Keyword Spotting for Handwritten Documents

The KWS-2015 Bentham database [65] was prepared by the tranScriptorium project [66]. It is a difficult database composed of 70 handwritten document pages, 15419 segmented word images from the Bentham collection, and 1421 query images, in which there are 243 various keyword strings of various lengths (6-15 characters), each of these strings is represented by 6 or fewer various query images. The papers have been written by various authors in differing styles, and font sizes, with crossed-out words.

It contains historical manuscripts on law and ethical philosophy handwritten by Jeremy Bentham (1748-1832) and some handwritten documents from his secretarial team. This competition was divided into two different tracks, namely, a training-free and a training-based track, and each track necessitated two elective assignments. Six participants proposed solutions to one or both assignments, relying on the limitations and/or capabilities of their systems.

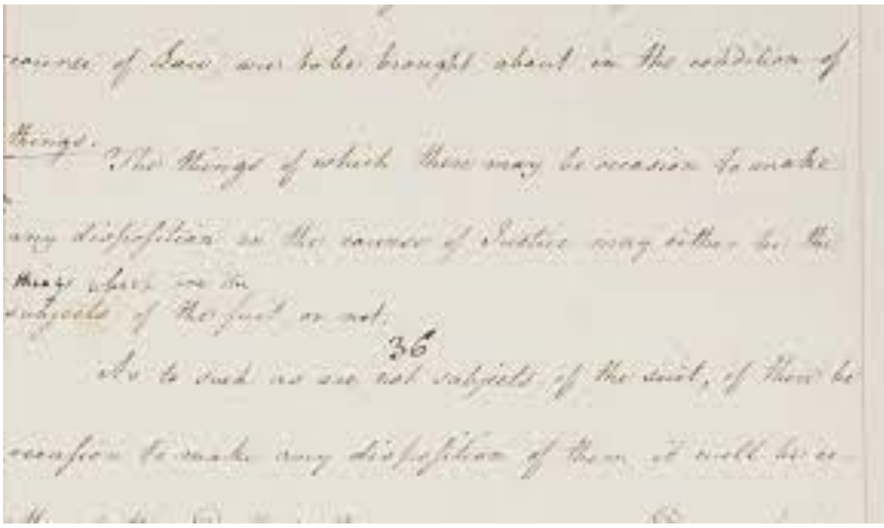


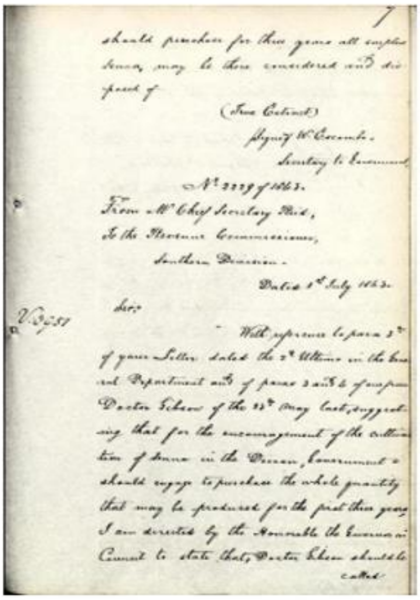
Figure 2.14 Sample images of document from the KWS-2015 Bentham database

2.3.3 ICFHR 2016 Handwritten Keyword Spotting Competition (H-KWS 2016)

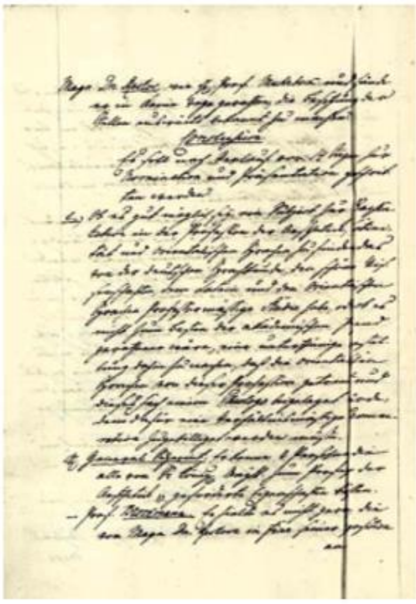
The H-KWS 2016 [67] was marshaled in the context of the ICFHR 2016 conference. The presented databases involve a series of documents from two various groups created in the European project READ: the Botany and the Alvermann Konzilsprotokolle in British India collections.

This competition was divided into two different tracks, namely, the Query-by-String and Query-by-Example tracks, and each track necessitated two elective challenges, i.e., Segmentation-based and Segmentation-free.

Four players offered solutions to one of the challenges, according to the limitations and/or capabilities within their systems. The data used in the competition consisted of historical English and documents with their intricacies and characteristics.



(a)



(b)

Figure 2.15 Sample document images from the H-KWS 2016 (a) Botany database, (b) Konzilsprotokolle database

2.4 Comparison of well-known keyword spotting systems

A synopsis of the method of keyword spotting and the performances of the techniques noted in the literature is shown in **Table 2.1**. In general, mAP results can be improved and offer a great possibility for research to find solutions to this difficult problem.

Segmentation type	Methods	Features (Descriptors)	Similarity	Database	KWS approaches		mAP(%)
					QbE	QbS	
Segmentation-Based	Rothfeder et al. (2003)	Harris corner detector	Euclidean distances	2372 images of good quality 3262 images of poor quality	×		62.57 15.49
	Rodríguez et al. (2008)	LGH	HMM DTW	630 scanned letters written in French	×		71.70 25.4
	Terasawa et al. (2009)	SSHoG	DTW	GW	×		79.14
	Rodríguez et al. (2009)	LGH	HMM	630 scanned letters written in French	×		87.00
	Aldavert et al. (2013)	Textual descriptor and BoVW	Cosine distance	All words as queries (GW) In vocabulary queries (GW)		×	56.54 76.20
	Zagoris et al. (2014)	DSLF	LPNN	Bentham GW	×		68.00 63.70

	Retsinas et al. (2018)	mPOG	MISM	GW Bentham Modern BOT AK	✕		87.70 71.10 49.10 58.30 76.20
	Stauffer et al. (2018)	Graph representations	Graph matching	GW PAR AK BOT	✕		70.56 79.38 84.77 68.88
	Serdouk et al. (2019)	SNN	Euclidean Distance	GW (hard) GW (soft)	✕		91,63 95,41
	Douaa et al. (2021)	oBIFS descriptor	City-block Distance	Six words as queries in Modern Database	✕		78.89
	Kundu et al. (2021)	Hough transform-based angular	DTW-based	IAM QUWI ICDAR KWS 2015	✕		86.40 53.99 45.01
	Daraee et al. (2021)	DNN with Monte-Carlo dropout	Cosine Distance	IAM GW BOT AK	✕		96.21 99.22 99.76 99.89
				IAM GW BOT AK		✕	99.89 100 100 100

	Scius-Bertrand et al. (2022)	YOLO-based deep convolutional character	HED	Kieu (QbE) Kieu (QbS)	✕		77 63
	Douaa et al. (2024)	Textural features	Euclidean Distance	twelve words as queries in Modern Database twelve words as queries in Bentham Database	✕		46.23 53.84
Segmentation-Free	Rusinol et al. (2011)	SIFT descriptors and BoVW	Cosine distance	GW LB	✕		30.42 42.83
	Kovalchuk et al. (2014)	LBP and HoG	Euclidean distance	GW LB	✕		50,10 90,70
	Yao et al. (2015)	HoG	Two-DTW	GW CASIA-HWDB 2.1	✕		57.20 70.50
	Rabaev et al. (2016)	HoG	The highest level of the pyramid	GW LB CG AH	✕		56.03 89.36 80.19 51.48
	Retsinas et al. (2023)	CNN	Cosine Distance	IAM GW		✕	59.2 66.3

Segmentation-Based And Segmentation-Free	Zagoris et al. (2017)	DoLF	NNS	Bentham (Segmentation-Free) BH2M	✗		51.70 53.00
				Bentham (Segmentation-Based) GW BH2M			58.40 69.20 60.70
	Mohammed et al. (2021)	HOG and SURF	L2-norm	GW (Segmentation-Free) HADARA	✗		80.00 52.00
				GW (Segmentation-Based)			81.30

Table 2.1 Performance comparison of well-known keyword spotting systems reported in the literature

2.5 Conclusion

In this chapter, we have delivered an overview of the state-of-the-art in the domain of keyword spotting. In the first part, we focus on presenting the main research works in this domain. Second, we discussed the different competitions. Then, we end with a comparison of the different works in the domain. The structures of the suggested system keyword spotting are covered in detail in the following chapter.

Part II

Contribution and Validation

Chapter 03: Keyword Spotting in Handwritten Documents using Textural Feature

3.1 Introduction

The task of keyword spotting in document images remains an active area of research due to its potential applications in information retrieval. Our contributions aim to explore a segmentation-based technique for keyword spotting in historical documents, with a focus on extracting and representing textual features.

This chapter presents the fundamental idea of our system, providing a detailed explanation of the features and the databases utilized to implement the proposed system.

3.2 Proposed system for keyword spotting

The method submitted in our study relies on a Query-by-Example (QbE) approach type and a segmentation-based technique, where the ground truth provides the word locations in the document images.

There are two important steps in our system. The first step is the extraction of features from the word images to represent them by the different methods based on feature vector sequence, where we choose the textual features to use in our proposed keyword spotting methodology.

The second step is Image Matching to count the similarity of the feature vector sequence.

An illustrated synopsis of the method is delivered in **Figure 3.1**.

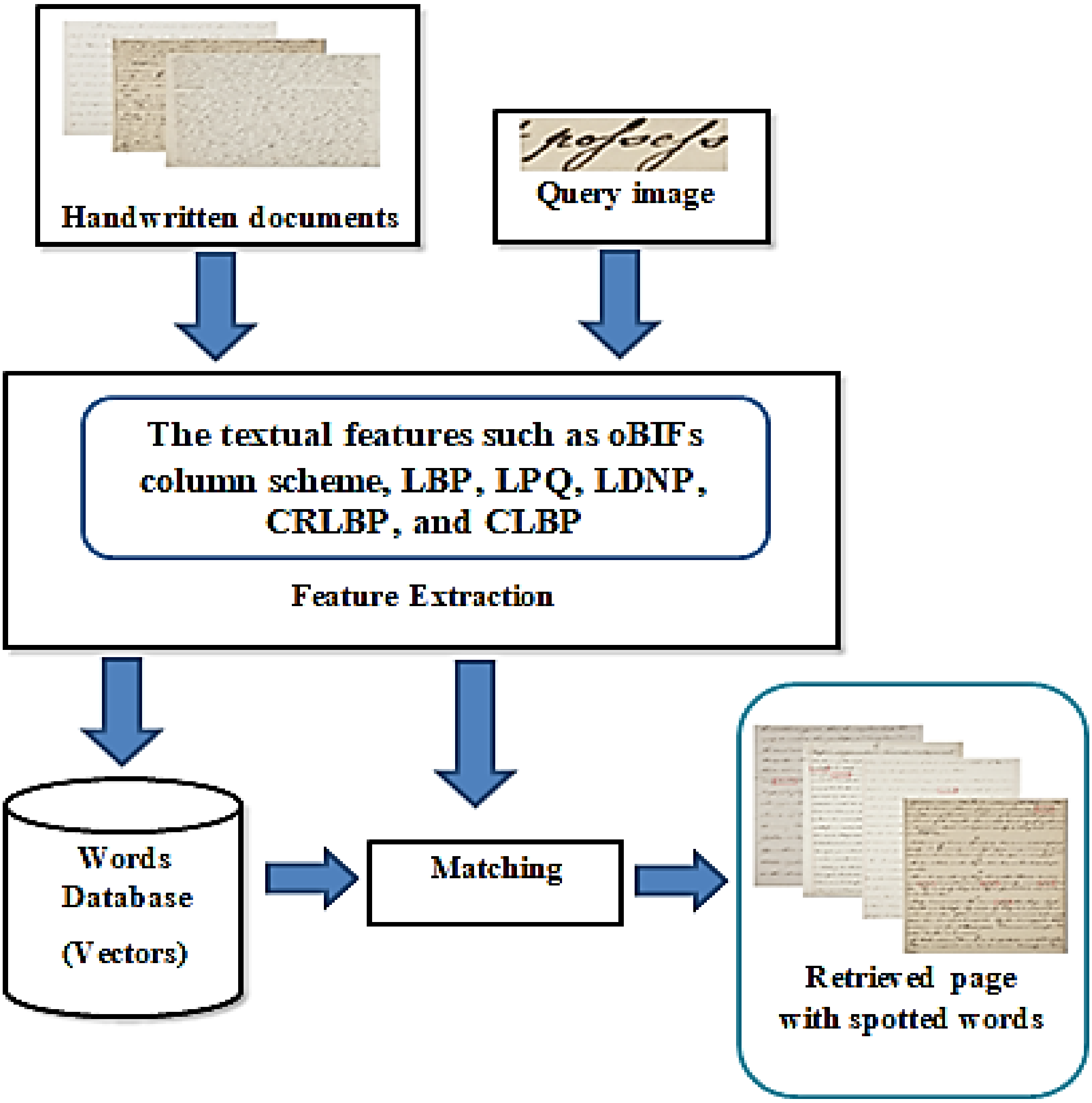


Figure 3.1 Synopsis of the submitted system

3.2.1 Feature Extraction

Effective textual feature representation is an important stage in every problem of pattern matching. In our keyword spotting from historical handwritten manuscripts, we choose the use of specific text feature descriptors, namely, the oBIFs, the oBIFs column, the LBP, the LPQ, the LDNP, the CLBP, and the CRLBP features. These feature descriptors enable the discriminatory visual representation of handwritten documents by capturing their textual and morphological information.

Many issues of pattern classification have been effectively solved with the use of these descriptors, like texture classification [73], digit recognition [75], [91] writer identification [74], [76], [77], [89], gender classification [78] [90], and handwriting-based personality identification [79]. We chose it for our experiments to determine how effective it is at keyword spotting.

In the subsections that follow, these descriptors are covered.

3.2.1.1 orientated Basic Image Features (oBIFs)

The oBIFs represent a texture-based descriptor, which is an extension of the Basic Image Features (BIFs) [73], [74], where every place in the image is designated to one of the seven local symmetry classes as part of the computation. The symmetry types contain slope, flat, dark on the light line, light line on the dark, dark rotational, light rotational, or saddle-like.

The algorithm has two adjustable parameters. A scale parameter σ, and a supplementary parameter ε. The scale parameter σ determines the response of a bank of six derivative-of-Gaussian (DoG) filters (one 0th order, two 1st orders, and three 2nd orders) of size. The parameter ε defines if a location is to be classified as flat.

To generate a collection of features known as oBIFs and identify the possible orientations, local symmetry is combined with local orientation.

The local orientation that can be assigned depends on the local symmetry type, as follows:

- n possible orientations are assigned to the dark line on the light, the light line on the dark, and saddle-like classes.
- Can assign 2n orientations if the location is of the slope class.
- No orientation is assigned if the location is designated to the dark rotations, light rotational, or flat class.

As a result, 5n+3 features determine the dimension of the oBIFs feature vector.

Figure 3.2 Sample of oBIFs calculation for a handwritten word image for σ=2 and σ=4 while ε=0.001

3.2.1.2 orientated Basic Image Features column (oBIFs column)

Has been improved the oBIFs descriptor's performance, by combining the oBIFs at two various scales and skipping the symmetry type flat (which is not likely to be discriminative). The oBIFs column [74], [80] are produced with a dimension of $(5n+2)^2$.

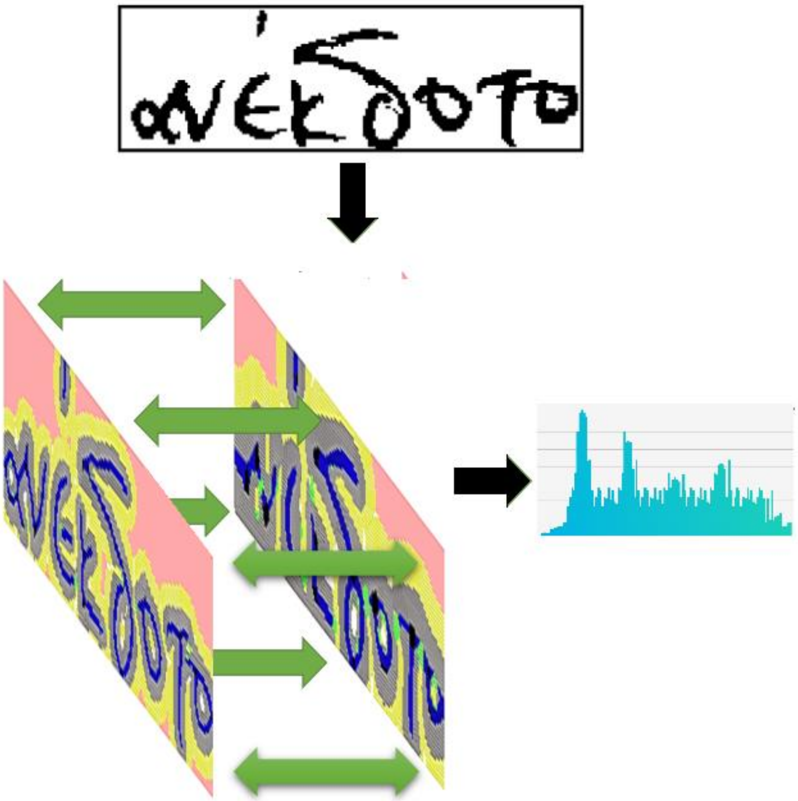


Figure 3.3 Model for oBIFs column scheme calculation for a handwritten word image: Two oBIFs images ($\boldsymbol{\sigma} = \mathbf{2}$ and $\boldsymbol{\sigma} = \mathbf{4}$ while $\boldsymbol{\varepsilon} = \mathbf{0.001}$) are scratched to form an oBIFs column at every location

3.2.1.3 Local Binary Pattern (LBP)

Ojala et al. [81] proposed Local Binary Patterns in 1996, which are to characterize the textures in gray-scale images. The idea of this texture operator is to give each pixel a code depending on the gray levels of its neighborhood. According to the following formula (1), the central pixel's gray level (i_C) is compared to that of its neighbors (i_P):

$$\mathbf{LBP}_{P,R} = \sum_{P=0}^{P-1} 2^{P}\, \mathbf{S}(\boldsymbol{i_P} - \boldsymbol{i_C}) \qquad (\mathbf{1})$$

$$\textbf{where} \begin{cases} \mathbf{S}(\mathbf{x}) = \mathbf{1}, \mathbf{x} \geq \mathbf{0} \\ \mathbf{S}(\mathbf{x}) = \mathbf{0}, \mathbf{x} < \mathbf{0} \end{cases}$$

By thresholding a neighborhood with the gray level of the central pixel, one can compute a binary code that represents the local texture of a region. All the neighbors will then take a value of 1 if their value is larger than or equal to the current pixel and 0 otherwise.

The LBP code of the current pixel is then produced by concatenating these 8 neighboring values to form a binary code. **Figure 3.4** gives an example of processing the LBP operator. This sequence is converted into a decimal value that represents its LBP code, so we obtain a matrix of LBP values containing intensity values from 0 to 255.

The LBP technique was later extended using neighborhoods of various sizes. The local neighborhood can be defined using the circular neighborhood, where a circle of radius R is drawn around the central pixel, and the values of the P sampled points on the circle's edge are obtained and compared to the value of the central pixel. To obtain the values of the P sampled points in the neighborhood for any radius R. Interpolation is necessary. We adopt the notation (P, R) to define the neighborhood of P points of the radius R of a pixel.

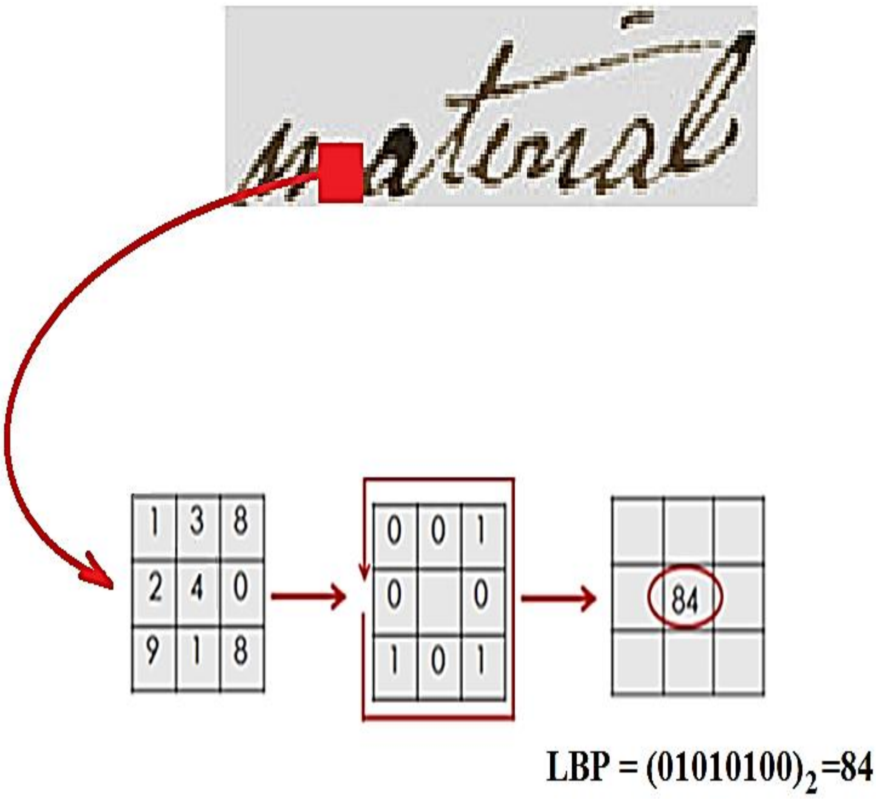


Figure 3.4 The computation of LBP code

The Local Binary Patterns (LBP) method for feature extraction has shown excellent performance in several comparative studies, both in terms of speed and in terms of discriminating between different textures.

3.2.1.4 Local Phase Quantization (LPQ)

Heikkila and Ojansivu [82] introduced the new descriptor known as the Local Phase Quantization operator, which is intended for use in texture classification for blurred images. It improves the classification of textures to be robust to artifacts generated by the blur present in an image. The LPQ descriptor is constructed in such a way as to retain in an image only the local information invariant to a certain type of blur.

The Fourier transform of the phase is the foundation of LPQ, from which selective frequency filters can be utilized to derive local frequency characteristics. More precisely, the local phase information extraction is utilized by the application of the short-term Fourier transform (STFT) calculated on a rectangle N_x of size M∗M neighbors for every pixel position x in the texture image f(x) defined by the equation:

$$\boldsymbol{F}(\boldsymbol{u},\boldsymbol{x}) = \sum_{\boldsymbol{y}\in \boldsymbol{N_x}} \boldsymbol{f}(\boldsymbol{x}-\boldsymbol{y})\boldsymbol{e}^{-\boldsymbol{j2\pi u^T y}} = \boldsymbol{w_u^T f_x} \qquad (\boldsymbol{2})$$

Where $N_x$is the neighborhood, f(x−y) is the function's value in the neighborhood, w_u is the basis vector of the 2D discrete Fourier transform (DFT) at frequency u, and f_x is a vector containing all the image specimens from N_x.

In LPQ, just four complex coefficients are considered: $u_1 = [a, 0]^T$, $u_2 = [0, a]^T$, $u_3 = [a, a]^T$ and $u_4 = [a, -a]^T$.

Where a = 1/M (M is window size). This yields the following vector for each pixel position:

$$\boldsymbol{F}(\boldsymbol{x}) = [\boldsymbol{F}(\boldsymbol{u_1},\boldsymbol{x}), \boldsymbol{F}(\boldsymbol{u_2},\boldsymbol{x}), \boldsymbol{F}(\boldsymbol{u_3},\boldsymbol{x}), \boldsymbol{F}(\boldsymbol{u_4},\boldsymbol{x})] \qquad (\boldsymbol{3})$$

Then, G_x (the discrete Fourier transformations, or DFT, of the blurry image) are calculated for every pixel, and what results from its vectors is quantized using a simple scalar quantizer.

$$q_j(x) = \begin{cases} 1, & g_j(x) \geq 0 \\ 0, & otherwise \end{cases} \qquad (4)$$

Where g_j is the j component of the vector (x) = [Re {(x)}, Im {(x)}], and Re{.} and Im{.} return the imaginary and real parts of a complex number, respectively.

Using binary coding, the resulting eight q_j binary coefficients are represented by values ranging from 0 to 255 as:

$$f_{LPQ}(x) = \sum_{i=1}^{8} q_j\,(x)2^{j-1} \qquad (5)$$

The diagram below depicts the steps required to construct the LPQ descriptor.

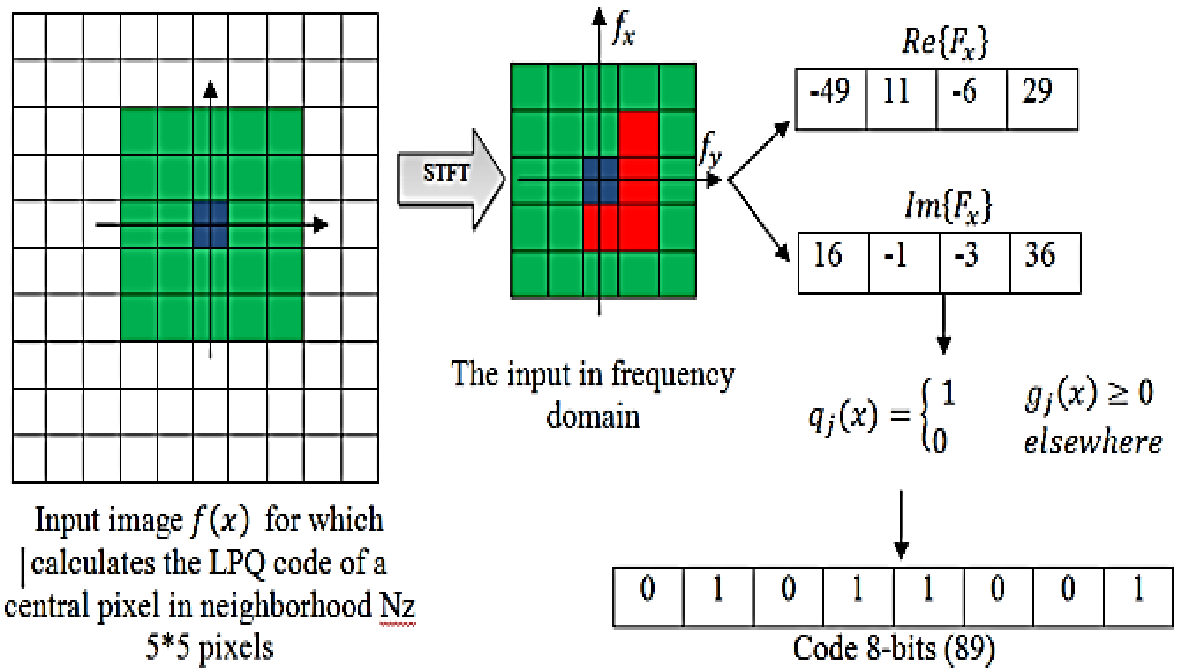


Figure 3.5 The illustration of the computing LPQ [83]

3.2.1.5 Local Directional Number Pattern (LDNP)

Rivera et al. [84] developed the Local Directional Number Pattern (LDNP) code Local. It is a six-bit binary code assigned to each pixel of an input image that represents the texture's format and its intensity changes. By examining the response of the edge of each mask, M0,..., M7, which mirrors the advantage importance in each direction, and adding the dominating directional values. The existence of a heightened negative or positive value suggests the existence of a prominent dark or bright area. Therefore, they implicitly utilize the sign information to encode these prominent regions, where the top positive directive number is assigned a fixed position, the three most important bits in the code, the three most important bits are the top negative directional number, and accordingly, the code is:

$$\boldsymbol{LDNP(x,y) = 8i_{x,y} + j_{x,y}} \qquad (6)$$

Where (i, y) is the neighborhood's central pixel to be encoded, $i_{x,y}$ is the directive number of the maximum positive response, and $j_{x,y}$ is the directive number of the minimum negative response defined by:

$$\boldsymbol{i_{x,y} = arg\,\max_{i}\{I^i(x,y)\big|0 \le i \le 7\}} \qquad (7)$$

$$\boldsymbol{j_{x,y} = arg\,\max_{j}\{I^i(x,y)\big|0 \le i \le 7\}} \qquad (8)$$

Where $\mathbb{I}^i$is the original image's convolution, I, and the i^{th} mask, M^i, described by:

$$\boldsymbol{\mathbb{I}^i = I * M^i} \qquad (9)$$

3.2.1.6 Complete Local Binary Patterns (CLBP)

Guo et al. [85] presented an upgraded version of LBP called Complete Local Binary Patterns (CLBP) to modify the discriminative capabilities of the local structure. Because LBP examines only the variation in two gray values, it frequently generates incorrect codes. Where the values of the middle gray level were blended with the local variations in each pattern's magnitude (m_p) and sign information (s_p). Two bits are utilized to represent both the sign and the magnitude of the change. Equation 10 is a summary of the computation.

$$\boldsymbol{s_P = s(i_P - i_c) \quad , \quad m_P = |i_P - i_c|} \qquad \textbf{(10)}$$

Where i_P is the neighboring pixel's intensity level, and i_c represents the central pixel's intensity level.

Three operators are CLBP-Center (CLBP-C), CLBP-Sign (CLBP-S), and CLBP-Magnitude (CLBP-M).

CLBP-Magnitude (CLBP-M) and CLBP-Sign (CLBP-S) are likewise computed using m_p and s_p. They are theoretically expressed by the following equations:

$$\boldsymbol{CLBP\ S_{(P,R)} = \sum_{p0}^{p-1} 2^p s\left(i_p - i_c\right) \ \ where\ s_p = \begin{cases} 1 \ , \left|i_p - i_c\right| \geq c \\ 0 \ , \left|i_p - i_c\right| < c \end{cases}} \quad \textbf{(11)}$$

$$\boldsymbol{CLBP\ M_{(P,R)} = \sum_{p0}^{p-1} 2^p t\left(m_p, c\right) \ \ where\ t(m_p, c) = \begin{cases} 1 \ , \left|i_p - i_c\right| \geq c \\ 0 \ , \left|i_p - i_c\right| < c \end{cases}} \textbf{(12)}$$

Where i_P is the neighboring pixel's intensity level, and i_c represents the central pixel's intensity level, and R is the radius of the neighborhood.

According to Guo et al. [85], the center pixel also contains distinguishing details. Thus, to acquire the local central information, an operator termed CLBP-Center (CLBP-C) is exhibited in Equation 13:

$$\boldsymbol{CLBP\ C_{(P,R)} = t(i_c, c_i)} \qquad \textbf{(13)}$$

Where i_c is the central pixel's gray level value, and c_i is the image's entire average gray level.

The three descriptors are combined to provide the final CLPB descriptor.

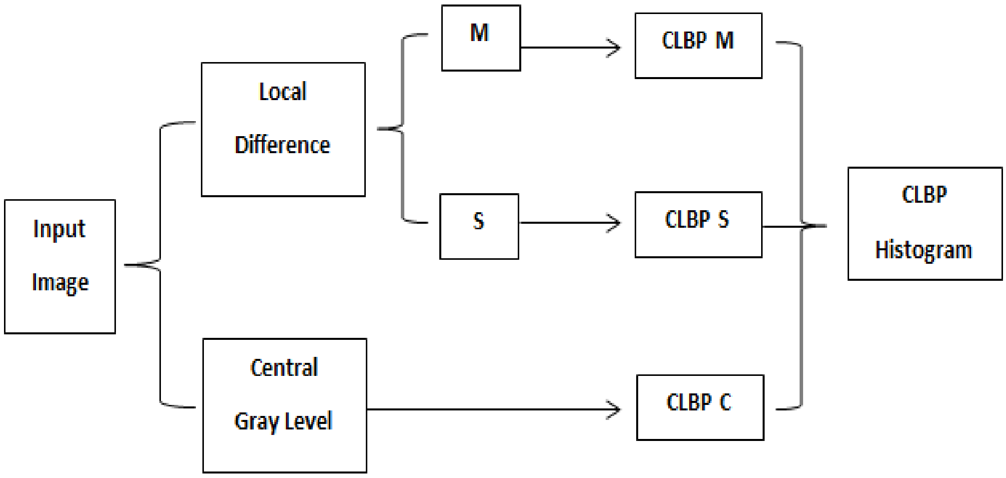


Figure 3.6 Framework of CLBP code

3.2.1.7 Completed Robust Local Binary Pattern (CRLBP)

The Completed Robust Local Binary Pattern (CRLBP) is the LBP method that uses ALG [87] as the threshold rather than the gray value. This ought to be hardy to noise and monotonically constant to a gray-scale transformation. The ALG can be written in equation 14 and the CRLBP in equation 15:

$$ALG = \frac{\sum_{1}^{8}(g_i + g)}{9} \qquad (14)$$

Where g is the gray value for the pixel that is found in the center, with g_i(i=0,1,...,8) representing the gray value for the neighboring pixel.

$$CRLBP_{P,R} = \sum_{p=0}^{p-1} s(g_p - ALG_c)2^p = \sum_{p=0}^{p-1} s\left(g_p - \frac{\sum_{i=1}^{8} g_{ci} + g_c}{9}\right)2^p \qquad (15)$$

Where g_{ci} (i=0,1,….,8) represents the pixel's gray value that is neighboring to g_c.

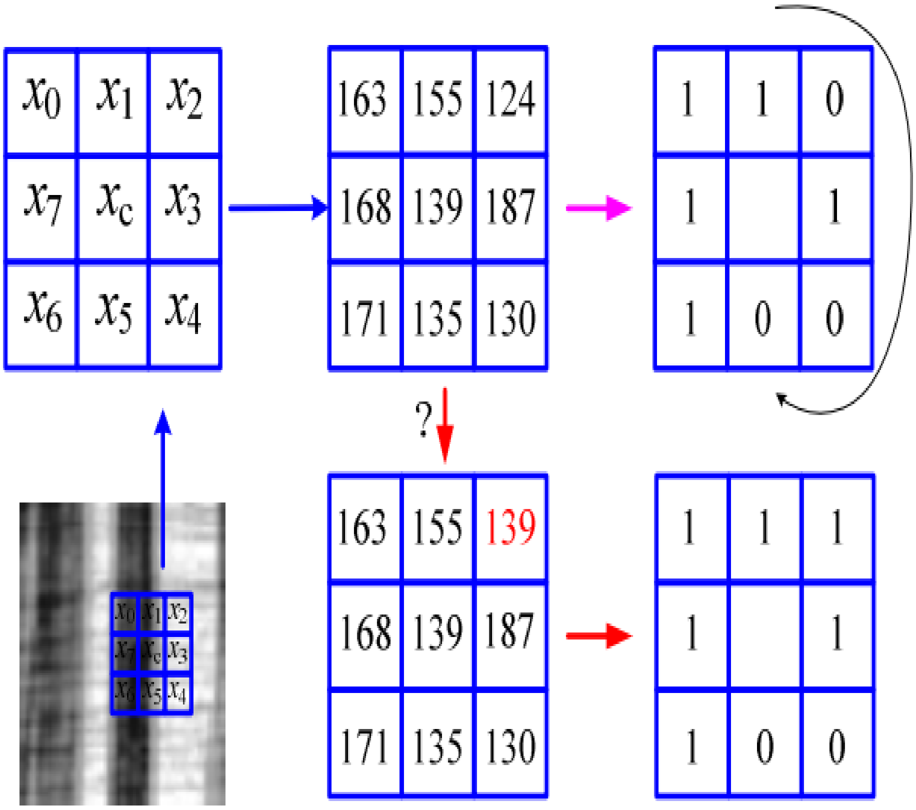


Figure 3.7 The computation of CRLBP code [88]

3.2.2 Image Matching

Once the word images in the reference base are represented by the textual descriptors, we can provide a query image and carry out the matching. For this purpose, we primarily employ the City-block distance [68], and the Euclidean distance [69] to compare the feature vectors of a pair of word images while several other metrics are also studied. These include the Chebychev distance, Cosine similarity, and Correlation coefficient.

3.3 Data Bases

On both databases from the ICFHR 2014 Handwritten Keyword Spotting Competition (H-KWS 2014) [64] for the segmentation-based track where the locations of the word images are provided in the document, the proposed methodology is tested.

3.3.1 Bentham database

A data collection was compiled by Bentham. It is made up of manuscripts in handwritten English that Bentham Jeremy and his secretarial staff authored [70]. This database contains 10,000 segmented word images derived from 50 document images, as well as 320 image queries for the competition's segmentation track.

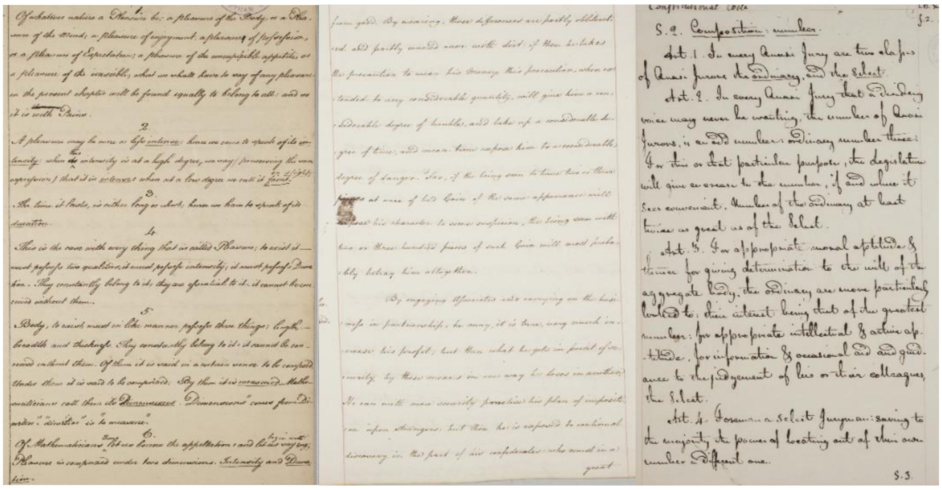


Figure 3.8 Examples of images from the Bentham database of handwritten documents

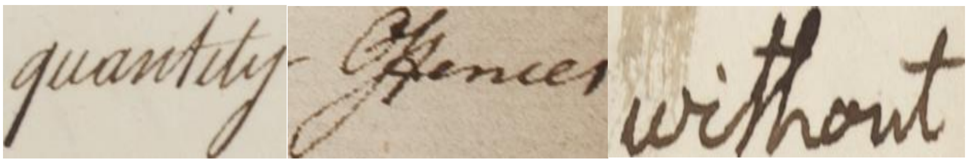


Figure 3.9 Examples of images from the Bentham database of handwritten documents

3.3.2 Modern database

It includes modern handwritten writings from the ICDAR 2009 Handwritten Segmentation Contest [71] in four languages: French, English, Greek, and German. For the competition's segmentation-based track, this database comprises 15,000 segmented word images and 300 image queries from 100 document images (25 for each language).

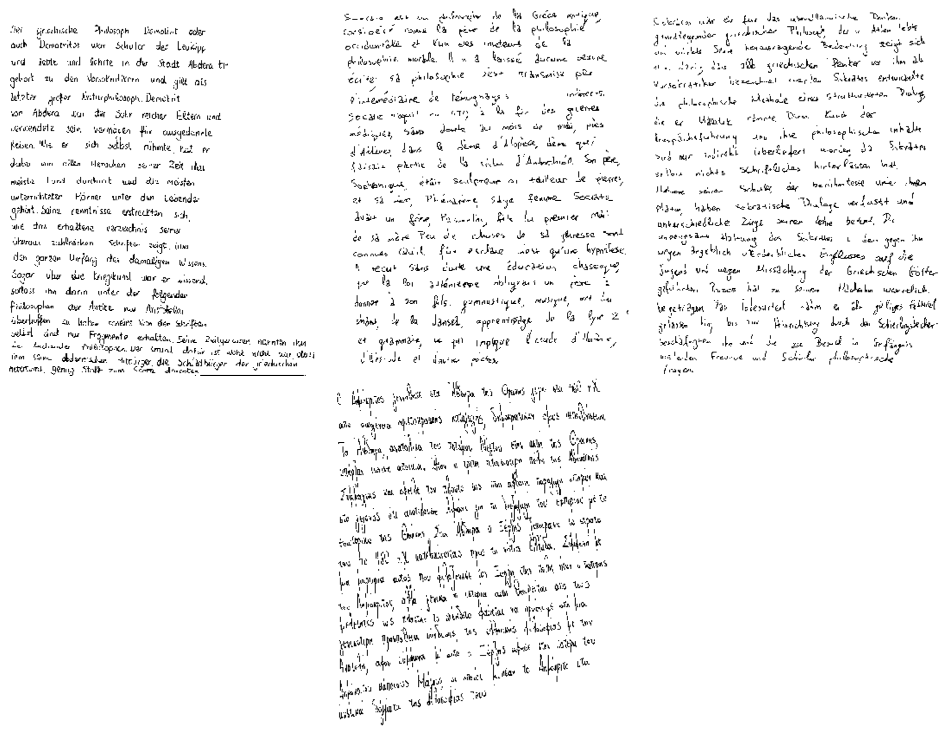


Figure 3.10 Examples of images from the Modern database of handwritten documents

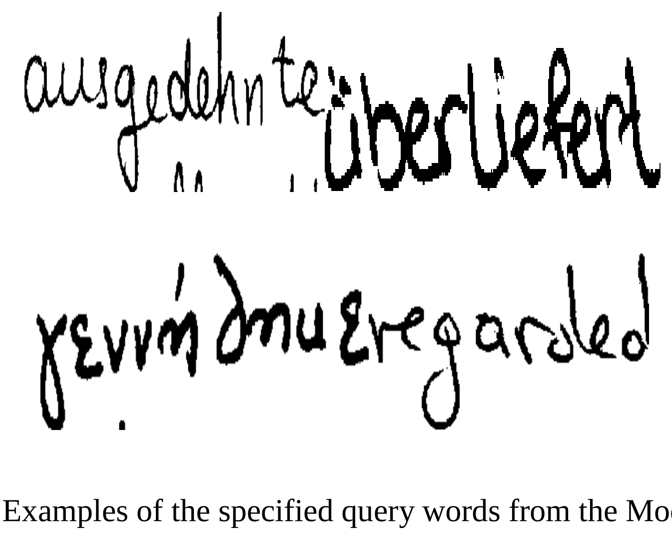


Figure 3.11 Examples of the specified query words from the Modern database

3.4 Metrics used

To measure the system's performance in various experiments, we use standard metrics such as Top retrieved words (P@k) and mean Average Precision (mAP). These metrics are further elaborated on in the next.

3.4.1 Top retrieved words (P@k)

The P@k, the precision at k, is calculated by calculating the precision at a specific cut-off rank while only considering the top k results returned by the system [72]. We report the results of our experiments for three various values of k. (1, 5, and 10). Formally, the P@k metric is acquainted as follows:

$$\mathbf{P@k} = \frac{|\{\mathbf{relevant\ words}\} \cap \{\mathbf{k\ retrieved\ words}\}|}{|\{\mathbf{k\ retrieved\ words}\}|} \qquad (\mathbf{16})$$

3.4.2 mean Average Precision (mAP)

The mAP for a specific set of query images is the mean of the average precision score for each query. A query's average precision is given as follows:

$$\mathbf{AP} = \frac{\sum_{\mathbf{k=1}}^{\mathbf{n}} (\mathbf{p(k)} \ * \ \mathbf{rel(k)})}{|\{\ \mathbf{retrieved\ words}\}|} \qquad (\mathbf{17})$$

3.5 Conclusion

This chapter was devoted to describing the different steps of our study. We presented the different textual descriptions used as well as the word matching method, the databases used in the tests conducted to validate our approach, and the performance metrics adopted for evaluation. In the next chapter, we will present the results of our experiments.

Chapter 04: Experiments and Discussions

4.1 Introduction

For the validation of the proposed solutions, we need to perform evaluations. In this chapter, we will present the tools and language used during the implementation of the proposed models. Then, we will analyze the results obtained and a comparative study with previous solutions will follow. To conclude this chapter we will summarize the results obtained.

4.2 Software and hardware used

The implementation of the solutions is done in MATLAB 2013a. The characteristics of the machine used are Intel(R) Core (TM) i5-9300H CPU @ 2.40 GHz 2.40 GHz and 8,00 Go of RAM under Windows 10.

4.3 Obtained results

Choosing feature extraction is the most necessary step in any matching problem. For this, we conducted many experiments on feature extraction. Once the textural descriptors for images in the reference base and the query word image have been assigned, we run the matching process. For this purpose, we primarily and in some experiments employ the City-block distance [68], and in others, we utilize the Euclidean distance [69] to compare the feature vectors of a pair of word images. While several other metrics are also studied. These include the correlation distance, Cosine distance, and the Chebychev distance.

A key highlight of the chosen features is their ability to encrypt the textual information as a function of neighboring pixels, resulting in a discriminative descriptor. Furthermore, combining multiple textual measures also makes the final feature set robust to noise and other artifacts.

Below, we will explain all the experiments and share the results received from each experiment:

4.3.1 Experiment 01: Using six-query images

- **<u>The oBIFs features</u>**

In the first series of experiments, we evaluated the effectiveness of the oBIFs features for spotting words in document images. The 100 document images for the Modern databases from the ICFHR 2014 competition database are employed as the reference base, while six-word images are employed as query samples.

The orientation is quantized by the parameter, which is fixed to 4, resulting in a dimensionality of 5n+3 for the oBIFs feature vector (which translates to 23 for n = 4), and we study the impact of the scale parameter σ (σ ∈ {1, 2, 4, 8, 16}) in computing the oBIFs features. City-block distance is employed as a distance metric for matching purposes in these experiments, while the parameter ε is fixed to a small value of 0.001. The performance in terms of Top1, Top5, Top10, and mAP is summarized in **Figure 4.1**.

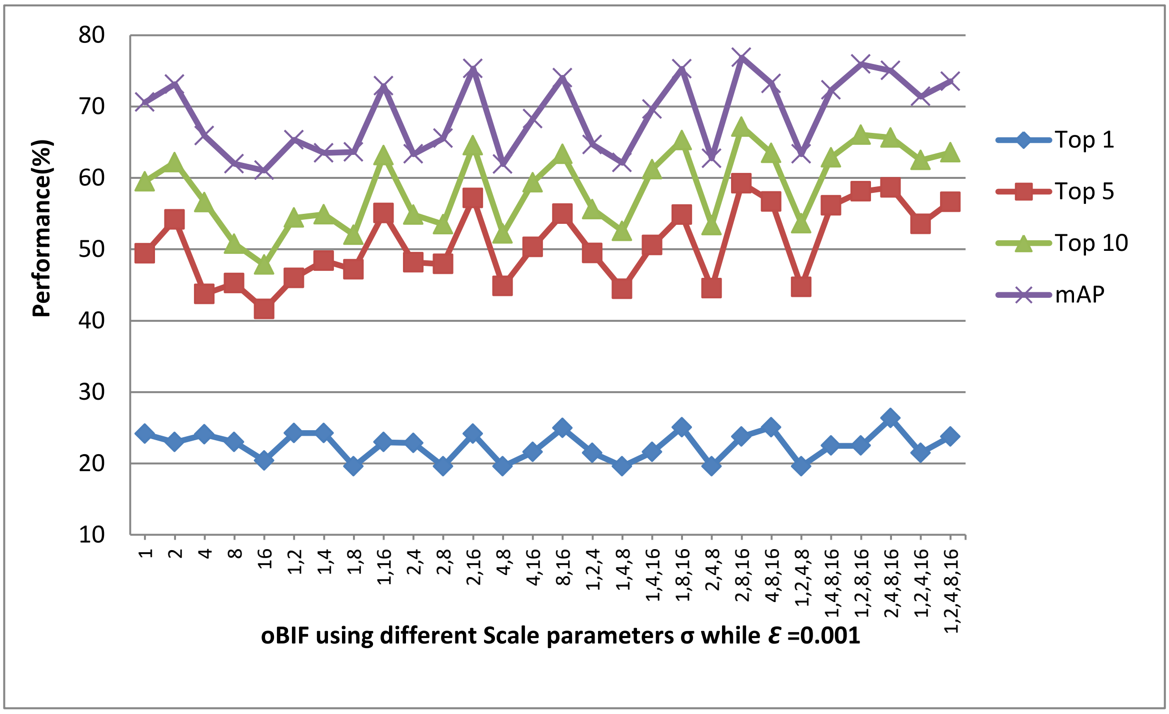


Figure 4.1 The performance of the proposed method using oBIFs

It can be seen from **Figure 4.1** that the matching performance varies as a function of the scale parameter in computing the oBIFs. The oBIFs features generated and combined using the values of the scale parameter σ= 2,8,16 outperform the other configurations, reporting a mAP of 76.86 %. The performance on other metrics reads a Top-1 precision of 23.77%, Top-5 of 59.23%, and Top-10 of 67.16%.

In addition to the city-block distance, we also evaluated the best configuration of oBIFs using different metrics, and the corresponding results are presented in **Table 4.1**. It can be observed from the reported results that, with few exceptions, the performance of different metrics is more or less similar. The highest mAP is reported using the correlation measure and reads 78.89%. These results are quite promising and validate the effectiveness of oBIFs for characterizing word images. It should also be noted that we do not carry out any pre-processing on the word images, and the features are directly extracted from the raw images.

oBIFs Parameter	Distance Type	*Top 1*	*Top 5*	*Top 10*	*mAP*
Ɛ=0.001* and *σ = 2,8,16	City-block	23.77	**59.23**	67.16	76.86
	Euclidean	**27.83**	56.37	65.51	76.26
	Cosine	26.44	59.07	68.44	78.28
	Hamming	10.47	26.40	30.71	40.41
	Correlation	26.44	59.06	**68.81**	**78.89**
	Spearman	23.08	49.62	59.18	69.65
	Chebychev	24.14	52.76	61.75	72.90

Table 4.1 Performance of the proposed method as a function of different distance metrics

4.3.2 Experiment 02: Using twelve-query

In the second experiment, we utilized textual features such as LDNP, CLBP, and CRLBP features, and incorporated the product (Prod) of diverse distances to enhance the results. The experimental study was conducted on the ICFHR 2014 dataset for the Handwritten Keyword Spotting Competition (H-KWS 2014), while twelve-word images were employed as query samples.

- <u>**The LDNP feature**</u>

We display the results with the LDNP in **Figure 4.2**, where we discuss the effect of the mask "Gaussian".

Observations suggest that the LDNP on the Bentham database performs better than that of the Modern database.

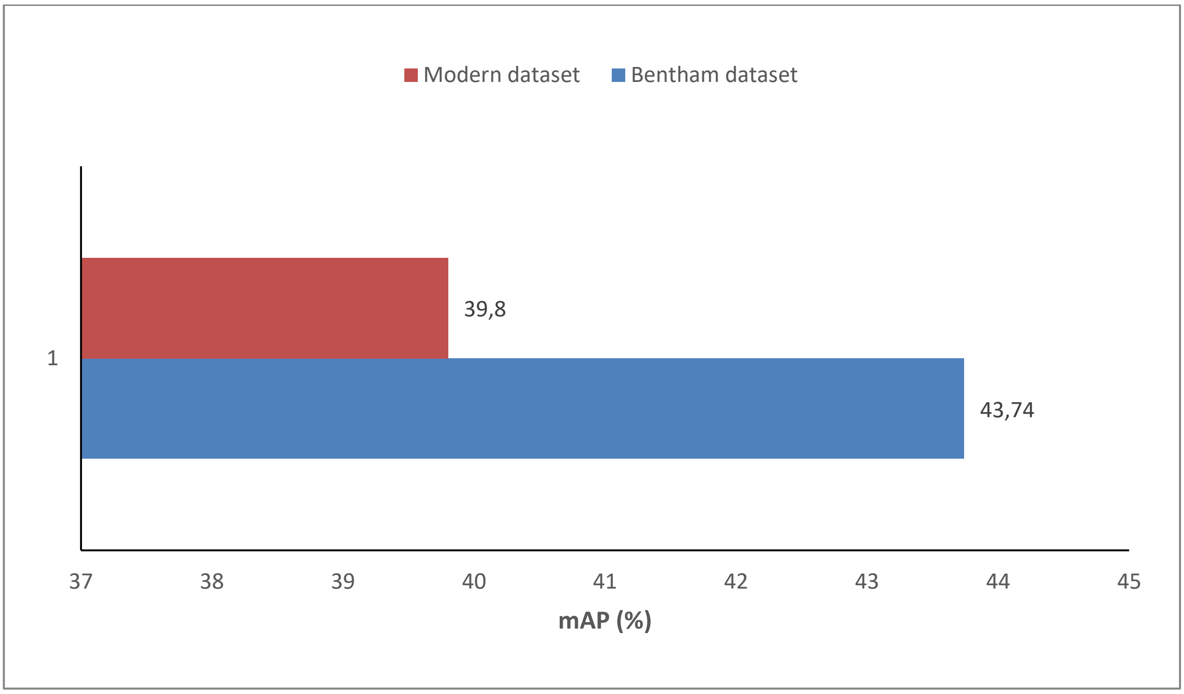


Figure 4.2 The performance of the LDNP features on the Modern and the Bentham databases

- **<u>The CLBP feature</u>**

For the CLBP feature, we change the neighborhood's radius (P) and number of nearby pixels (R) to see how performance changes. **Figure 4.3** demonstrates the effect of these parameters on the general system performance of the two databases. While the performance varies for various P and R values in the Bentham database, it is often more steady for texts from the Modern.

It can be seen that the P =8 and R =4 values outperform the other parameters in the Bentham database with a mAP of 53.47%, while the P = 4 and R =1 values exceed the other parameters in the Modern database with a mAP of 11.07%.

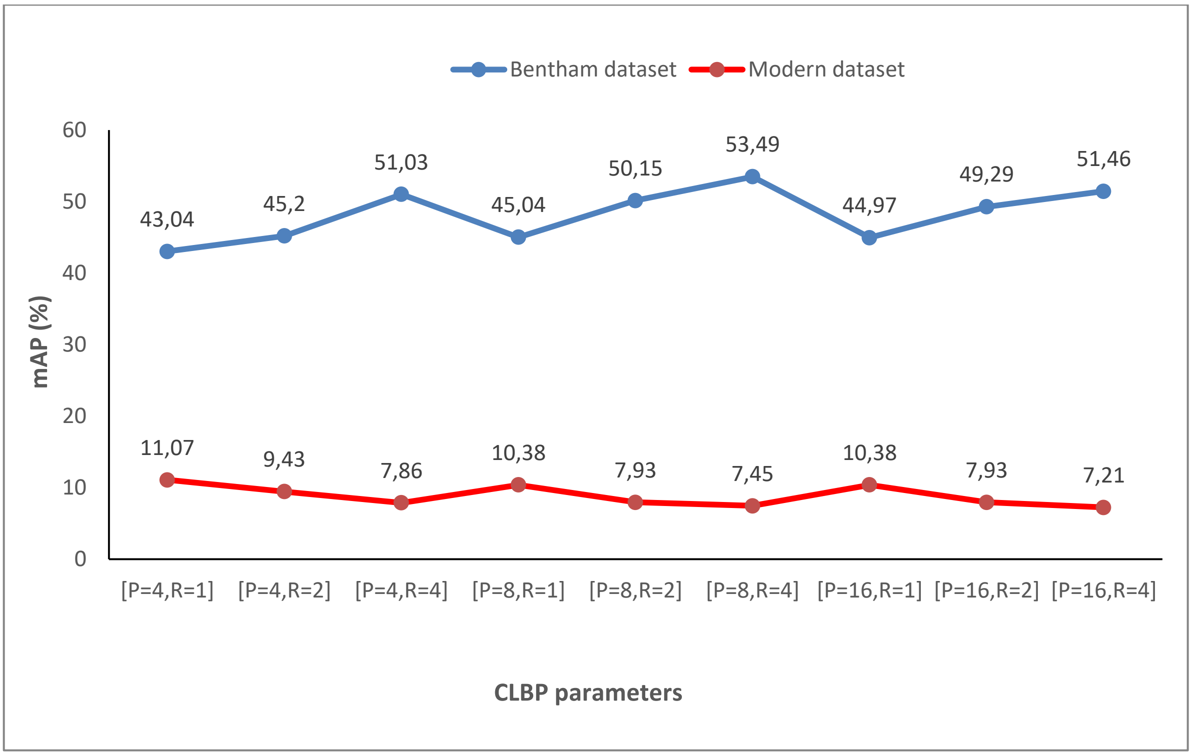


Figure 4.3 The performance of the CLBP features on the Modern and the Bentham databases for different combinations (P, R)

- **<u>The CRLBP feature</u>**

In implementing the CRLBP in our research and to examine how performance changes, we alter the neighborhood's radius (P) and several nearby pixels (R).

The performance in terms of a mAP of the results acquired in the Bentham and Modern databases is summarized in **Figure 4.4**. Observations indicate that the CRLBP for the (P = 8, R = 4) outperforms other parameters in the Bentham database, while the (P = 4, R = 1) exceeds other parameters in the Modern database.

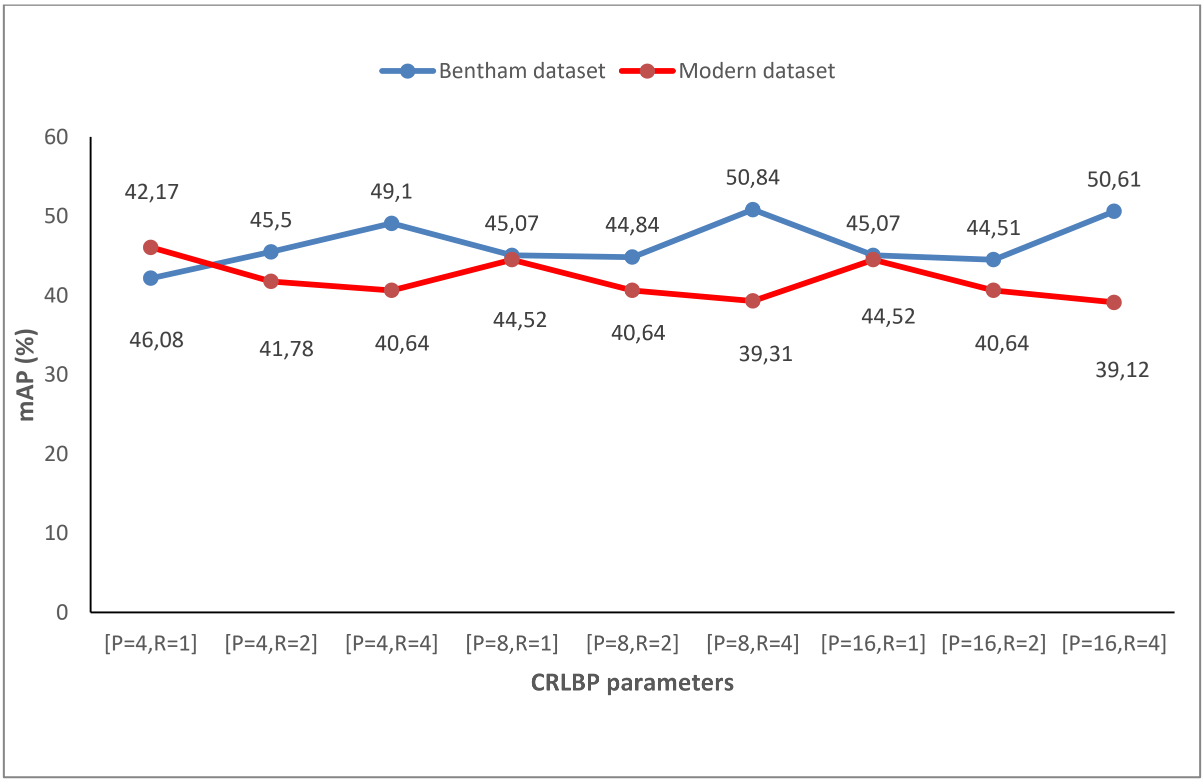


Figure 4.4 The performance of the CRLBP features on the Bentham and the Modern databases for different combinations (P, R)

To sweeten the matching rates, we also assessed the optimal features utilizing various distance metrics like the City-block distance and Chebychev distance, in addition to Euclidean distance.

Table 4.2 and **Table 4.3** recapitulate the performance of these features utilizing the various distance metrics. We see that the most elevated results were returned using the City-block distance on the Bentham database.

Feature	Dimension	distance metrics	Bentham Database	
			Top 1	**mAP**
CLBP (8,4)	486	Euclidean	72.22	53.49
		City-block	**79.17**	**57.41**
		Chebychev	68.06	52.48
CRLBP (8,4)	243	Euclidean	63.19	50.84
		City-block	**72.22**	**54.18**
		Chebychev	67.36	47.96
LDNP	56	Euclidean	37.5	43.74
		City-block	**57.64**	**49.31**
		Chebychev	39.58	43.5

Table 4.2 The performance of the propounded method using the best configuration with diverse distance metrics (The Bentham database)

Feature	Dimension	distance metrics	Modern Database	
			Top 1	**mAP**
CLBP (4,1)	486	Euclidean	**2.72**	**11.07**
		City-block	0	4.27
		Chebychev	0	4.31
CRLBP (4,1)	243	Euclidean	13.01	46.08
		City-block	13.01	46.08
		Chebychev	**16.21**	**48.49**
LDNP	56	Euclidean	11.73	39.8
		City-block	**11.73**	**39.8**
		Chebychev	11.73	39.8

Table 4.3 The performance of the propounded method using the best configuration with diverse distance metrics (The Modern database)

To grow the matching rate, we calculated the precision Top1 and the mAP for several combinations of distance metrics, where we used Prod (the product) of the diverse distance metrics with associated features. **Tables 4.4** to **Table 4.7** deliver an outline of the results.

F1: $CLBP_{8,2}$	F2: $CRLBP_{8,4}$	F3: $LDNP$	Bentham Database			
			Prod (F1,F2)		**Prod (F1,F3)**	
			Top 1	**mAP**	**Top 1**	**mAP**
City-block	City-block	City-block	75.00	53.83	**72.22**	**52.60**
	Euclidean	Euclidean	75.00	53.84	57.64	49.15
	Chebychev	Chebychev	72.22	53.27	46.53	46.83
Euclidean	City-block	City-block	75.00	54.49	59.72	51.04
	Euclidean	Euclidean	72.22	52.45	46.53	46.74
	Chebychev	Chebychev	69.44	50.85	43.75	44.96
Chebychev	City-block	City-block	**75.00**	**55.47**	59.72	50.57
	Euclidean	Euclidean	63.89	52.11	46.53	45.83
	Chebychev	Chebychev	56.94	49.43	43.75	44.67

Table 4.4 The performance of keyword-spotting for different combination features (The Bentham database)

F2: $CRLBP_{8,4}$	**F3:$LDNP$**	**Bentham Database**	
		Prod (F2,F3)	
		Top 1	**mAP**
City-block	City-block	**72.22**	**53.79**
	Euclidean	70.14	51.44
	Chebychev	54.86	47.43
Euclidean	City-block	72.22	52.67
	Euclidean	46.53	46.73
	Chebychev	43.75	45.10
Chebychev	City-block	67.36	50.78
	Euclidean	42.36	44.88
	Chebychev	37.50	43.18

Table 4.5 The performance of keyword-spotting for different combination features (The Bentham database)

F1: $CLBP_{4,1}$	F2: $CRLBP_{4,1}$	F3:$LDNP$	Modern Database			
			Prod (F1,F2)		Prod (F1,F3)	
			Top 1	mAP	Top 1	mAP
City-block	City-block	City-block	13.01	45.57	11.73	39.17
	Euclidean	Euclidean	12.37	45.24	11.73	39.39
	Chebychev	Chebychev	**14.29**	**46.23**	11.67	**39.64**
Euclidean	City-block	City-block	12.37	45.30	11.73	38.79
	Euclidean	Euclidean	12.37	45.22	**12.37**	39.58
	Chebychev	Chebychev	13.01	45.39	11.03	39.21
Chebychev	City-block	City-block	12.37	45.01	11.73	38.92
	Euclidean	Euclidean	12.37	42.73	11.73	39.15
	Chebychev	Chebychev	10.87	45.13	11.67	39.39

Table 4.6 The performance of keyword-spotting for different combination features (The Modern database)

F2: $CRLBP_{4,1}$	F3:$LDNP$	Modern Database	
		Prod (F2,F3)	
		Top 1	mAP
City-block	City-block	**13.37**	43.97
	Euclidean	10.34	40.49
	Chebychev	11.03	41.36
Euclidean	City-block	12.37	**44.36**
	Euclidean	12.37	42.11
	Chebychev	10.98	41.62
Chebychev	City-block	13.01	44.22
	Euclidean	11.08	41.84
	Chebychev	10.98	41.61

Table 4.7 The performance of keyword-spotting for different combination features (The Modern database)

Tables 4.4 to **4.7** illustrate that the suggested method by using the product (Prod) of the Chebychev-Cityblock distance metrics and the City-block-Chebychev distance metrics with CLBP and CRLBP features on the Bentham and Modern databases, respectively, exceeds the product (Prod) with Top 1 of 75.00% and mAP of 55.47% on the Bentham database and Top 1 of 14.29 % and mAP of 46.23% on the Modern database. These results support the efficacy of textural features with suitable combination schemes for keyword spotting in historical handwritten manuscripts.

- **Discussion**

Table 4.8 compares the proposed twelve-word query image method with the best systems offered to the ICFHR 2014 Keyword Spotting Competition.

Methods	Features	mAP (%)	
		Bentham Database	Modern Database
Proposed Method	CLBP and CRLBP	55.47	46.23
Best system in [64]	HOG and LBP	52.40	33.80

Table 4.8 Comparing the proposed method's performance to the top-performing systems in the ICFHR 2014

Table 4.8 shows how the proposed method performs better than the best system in [64] when utilizing additional textural features like HOG and LBP. It should nonetheless be mentioned that the features of the proposed method are taken from document images and do not require any preprocessing.

4.3.3 Experiment 03: Applied on the H-KWS 2014 database

In the third and last experiment, we decided to check the effectiveness of the textual features, which are the oBIFs columns, the LBP, the LPQ features, and combined of the diverse features. The experimental study of the proposed technique is carried out on the ICFHR 2014 dataset for the Competition on Handwritten Keyword Spotting (H-KWS 2014). Since we target a word-level approach, we have chosen the segmentation-based tracks from the competition, where two datasets (Bentham and Modern) are employed.

- **<u>The oBIFs column feature</u>**

When orientation is quantized by the parameter, which is fixed to 4, the oBIFs columns are produced with dimensions $(5n+2)^2$, i.e., 484. The oBIFs column features are created utilizing various values of the scale parameter $\sigma \in \{1, 2, 4, 8, 16\}$, and the parameter ε is specified with a small value of 0.001. Finally, the generated feature vector is normalized.

Figure 4.5 and **Figure 4.6** present a summary of the findings from these experiments in the Bentham database and the modern handwriting database, respectively. It can be observed that the oBIFs column histograms for the scale parameter combinations σ = [8, 16] and σ = [4, 16] outperform other configurations on the two databases, respectively.

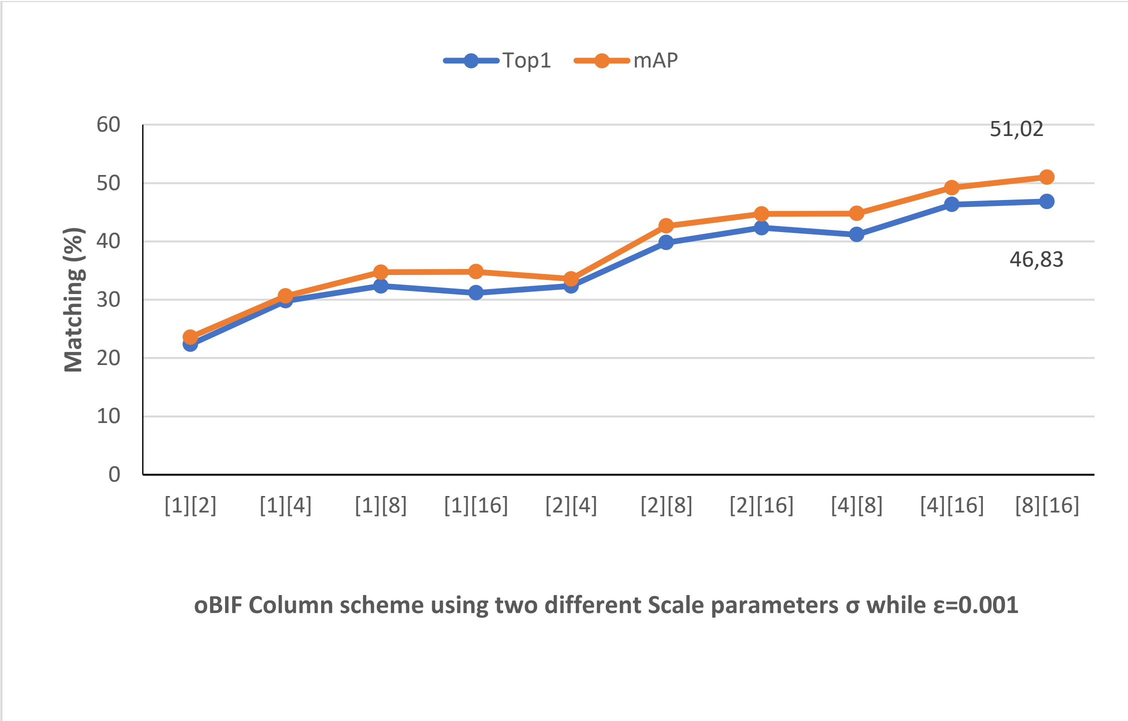


Figure 4.5 The Bentham database's matching rates for the proposed approach employing oBIFs column

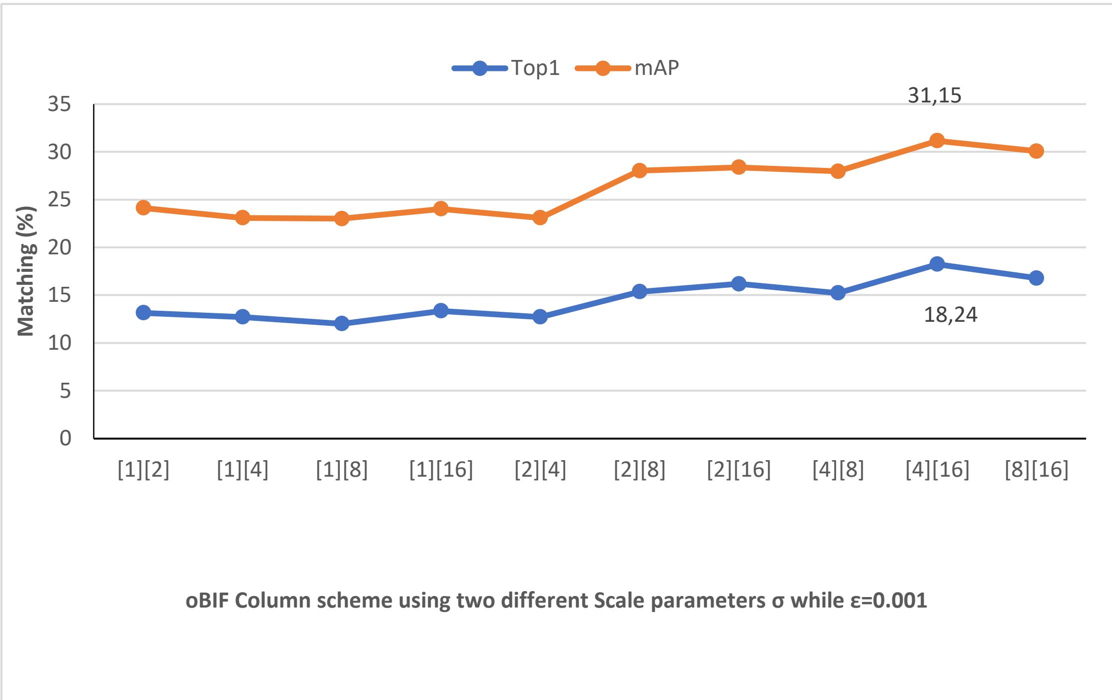


Figure 4.6 The Modern database's matching rates for the proposed approach employing oBIFs column

- **The LPB feature**

For the LBP feature, we vary the number of neighboring pixels (R) and the radius of the neighborhood (P) and study the evolution of performance. **Figure 4.7** and **Figure 4.8** illustrate the effect of these parameters on the overall system performance for the two databases. While the performance varies for various combinations of P and R values for the Bentham database, it is relatively more stable for contemporary writings.

We observe that the R = 8 and P = 4 values outperform the other configurations in the two databases.

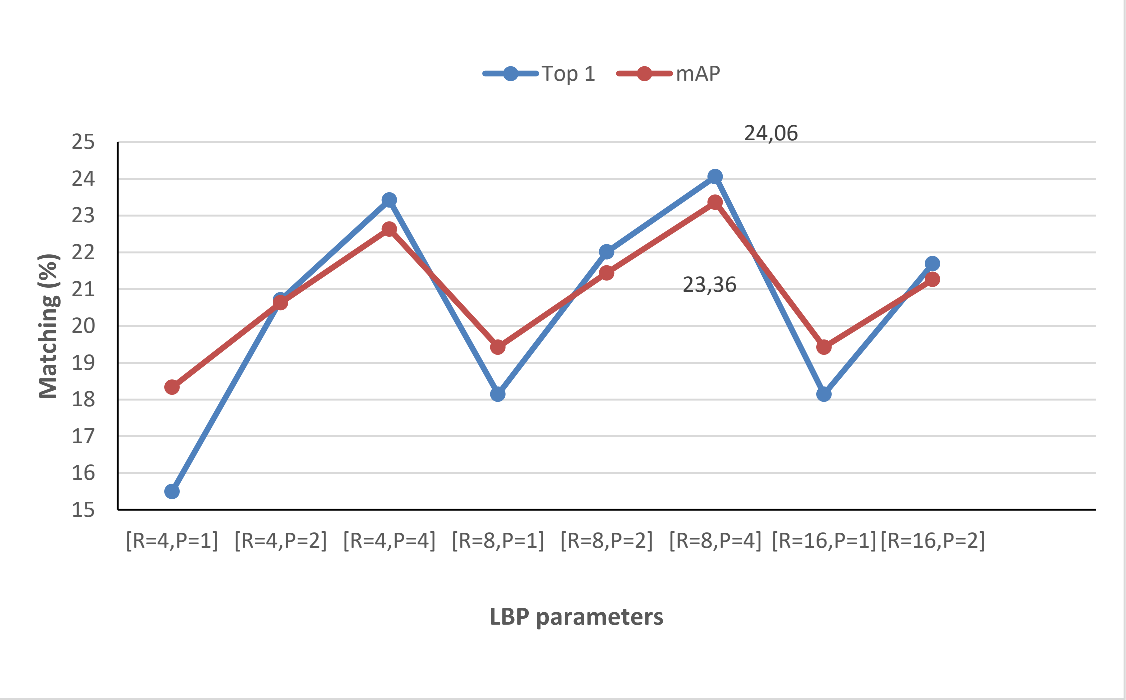


Figure 4.7 Performance of LBP features on the Bentham database for different combinations (P, R)

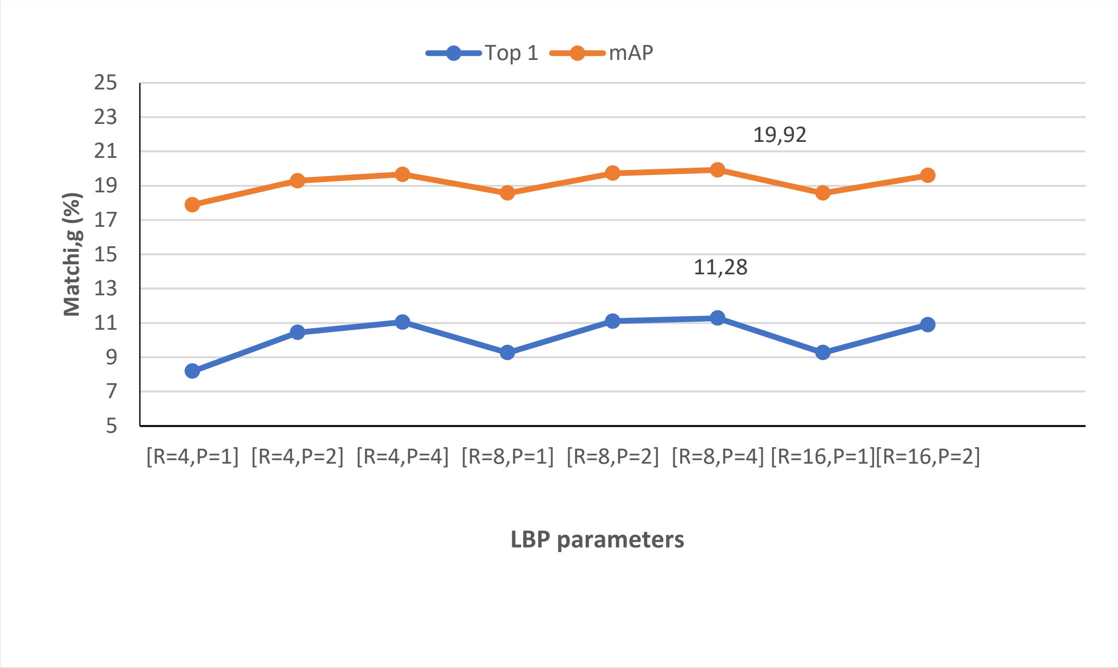


Figure 4.8 Performance of LBP features on the Modern database for different combinations (P, R)

- <u>**The LPQ feature**</u>

For the LPQ features, we calculate the descriptor by varying the local size of the window and evaluate the system using the query word images. The performance as a function of the window size is exhibited in **Figure 4.9** and **Figure 4.10** for the historical and modern samples, respectively. It is interesting to note that a similar trend is observed for both databases, where the performance gradually improves with the increase in the size of the window and then stabilizes, and the largest value in the window is w=37*37 in both databases.

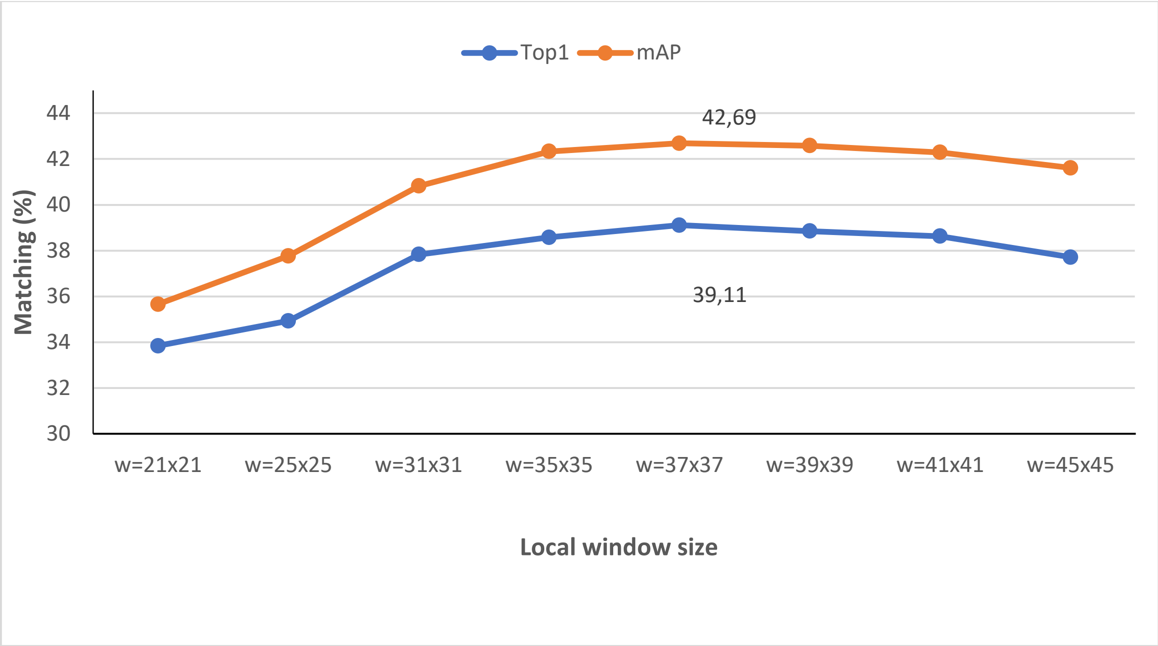


Figure 4.9 Performance of LPQ features on the Bentham database as a function of window size

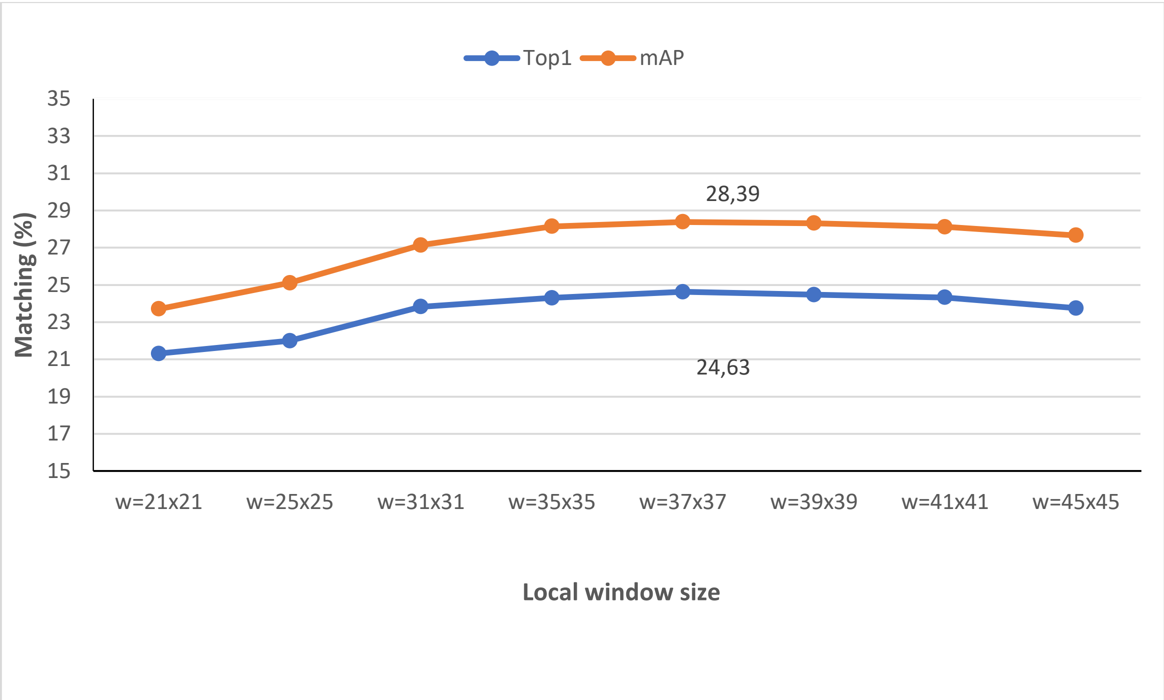


Figure 4.10 Performance of LPQ features on the Modern writing samples as a function of window size

For matching the feature vectors of the two words being compared, in addition to the City-block distance, we also evaluate the best-performing feature (f1) of the oBIFs column using various distance metrics. These contain the Euclidean distance, the Cosine distance, and the Correlation distance.

The matching results are exhibited in **Figure 4.11** and **Figure 4.12**, where observations indicate that the performance using the City-block distance outperforms other metrics.

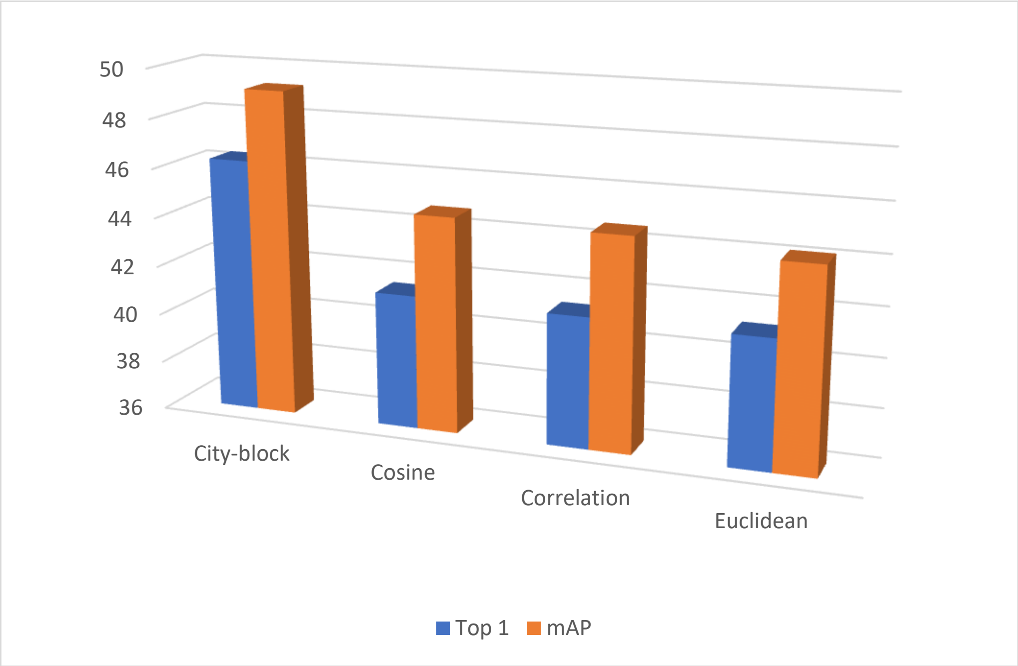


Figure 4.11 Performance of the proposed method using the best configuration of oBIFs column scheme with different distance metrics (The Bentham database)

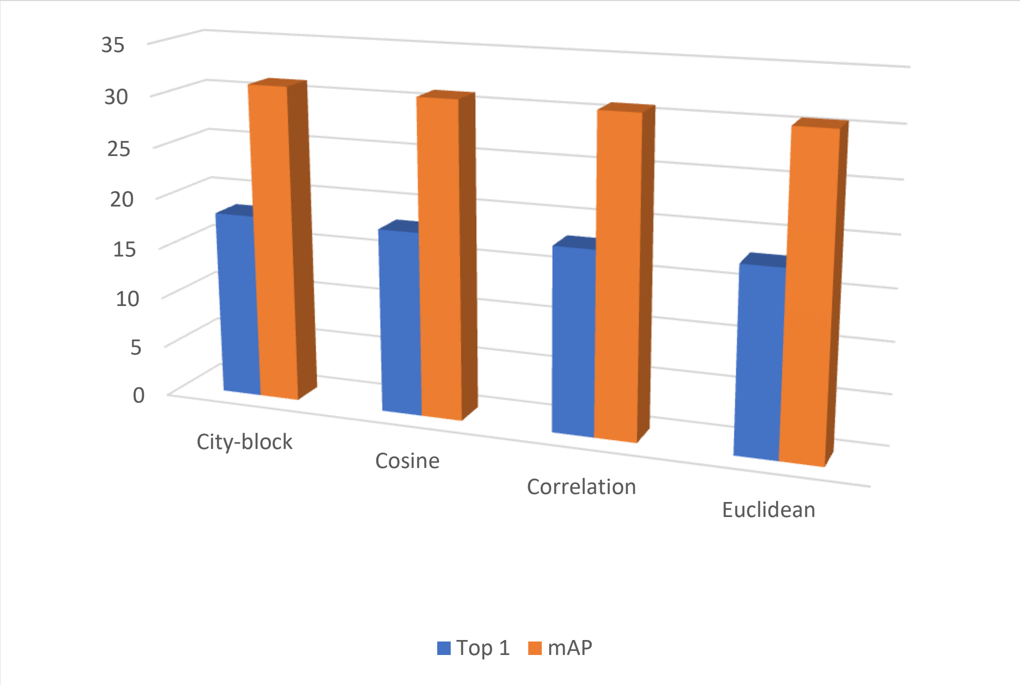


Figure 4.12 Performance of the proposed method using the best configuration of oBIFs column scheme with different distance metrics (The Modern database)

- <u>**The combination of oBIFs column and LPQ**</u>

To raise the caliber of the outcomes achieved, we decided to combine the best features.

A summary of the results of the top three best-performing features in addition to their combinations is presented in **Table 4.9**. The combination of the oBIFs column at σ =4 and σ = 16, the oBIFs column at σ =8 and σ =16, and LPQ at w = 37x37 reports the best performance, with Top 1 of 48.52% and mAP of 52.74% on the Bentham database and Top 1 of 25.82 % and mAP of 34.01% on the Modern database.

Features	Parameters	Size	Bentham Database				Modern Database			
			Matching rate (%)				Matching rate (%)			
			Top 1	**Top 5**	**Top10**	**mAP**	**Top 1**	**Top 5**	**Top 10**	**mAP**
f1	oBIFs column at σ =4 and σ =16, ε = 0.001	484	46.32	46.55	47.21	49.20	18.24	25.91	28.24	31.15
f2	oBIFs column at σ =8 and σ =16, ε = 0.001	484	46.83	47.16	49.12	51.02	16.78	24.82	27.10	30.07
f3	LPQ at w = 37x37	256	39.11	40.10	41.34	42.69	24.63	27.88	28.21	28.39
Combination (f1, f2)		968	48.34	48.51	50.41	52.19	20.87	26.08	29.01	31.33
Combination (f1, f2, f3)		1224	**48.52**	**48.67**	**50.87**	**52.74**	**25.82**	**28.04**	**30.11**	**34.01**

Table 4.9 Results using the top three best-performing features in addition to their combinations

- **Discussion**

This section compares the proposed technique's performance to the most advanced approaches tested on the ICFHR 2014 Competition on Handwritten Keyword Spotting (H-KWS 2014) database. We employ the same experimental protocol as that of the competition for an objective comparison.

Table 4.10 presents the comparison results. It is evident that, out of all the methods mentioned, the proposed technique yields the greatest mAP value of 52.74% of the historical documents in the Bentham database. For the modern writing samples, though the performance is relatively low with respect to the best-reported results of the competition, it is comparable to other methods. However, it is pertinent to mention that our method does not require any size normalization.

Methods	Features	mAP (%)	
		Bentham Database	**Modern Database**
Proposed Method	oBIFs column scheme and LPQ	**52.74**	34.01
Best system in [64]	HOG and LBP	52.40	33.80
Almazan et al. [64]	HOG descriptors	51,30	**52.30**
How. [64]	Gaussian random-walk deformation	46,20	27.80

Table 4.10 Performance comparison of the proposed method with the participating systems in the ICFHR 2014 Competition on Handwritten Keyword-Spotting

4.3.4 Summarize the results

This section summarizes our previous experiments in both the Bentham database and the Modern database. The following table presents the best results of the experiments (**Table 4.11**).

Experiments	Features	Database	mAP (%)
Experiment 01	oBIFs	Six-words queries in the Modern database	76.86
Experiment 02	CLBP and CRLBP	Twelve-words queries in the Modern database	46.23
		Twelve-words queries in the Bentham database	55.47
Experiment 03	oBIFs column scheme and LPQ	The Modern database	34.01
		The Bentham database	52.74

Table 4.11 Our best experiment results

We observe in **Table 4.11** that the best results were obtained by the oBIFs column scheme and LPQ compared to other descriptors, whereas when checking oBIFs, CLBP, and CRLBP in a small database, they gave satisfactory results.

4.4 Conclusion

This chapter contains the tools that were utilized to execute the solutions that were submitted, as well as the databases that were utilized in the tests that were conducted to validate our approach. Subsequently, we showcased the approaches we had supplied and the outcomes of assessments of our approach's solutions. Ultimately, a comparative study between our approach and the related works showed that the proposed models concerning keyword spotting proved to be better suited.

Conclusion

In recent years, the recovery of old documents has become an important issue for libraries and archive services, and new technologies have helped them through digitization. Therefore, it is necessary to develop applications to facilitate the use of these images. Keyword spotting in document images is among the most interesting applications, as it allows you to locate a query word in document images. It has become a broad research field that has resulted in a large number of research studies. There is an improvement in results, but there is no perfect solution to the problem of word detection in handwritten documents that achieves a 100% matching rate. For this reason, keyword spotting remains an open topic of research.

A keyword spotting system undergoes various stages: preprocessing, segmentation, feature extraction, and matching.

Throughout my years of research in this field, I have discovered that feature extraction and matching are two important steps for any keyword spotting system in documents. Therefore, in this thesis, we decided to concentrate on the feature extraction stage and determine the extent of the impact of textual features such as oBIFs, oBIFs column, LBP, LPQ, LDNP, CRLBP, and CLBP in this field. To match the words, we used several different metrics, including the City-block distance, the Correlation distance, the Euclidean distance, the Spearman distance, and the Cosine distance.

In order to validate our system, an experimental study is carried out using the ICFHR 2014 word spotting competition database, and promising results are reported in Precision at k (P@k), where the three values of k (1, 5, and 10) and the mean Average Precision (mAP).

The first results obtained are encouraging and validate the effectiveness of textual features in the characterization of word images. The proposed method, when we combine the different texture features, gives the greatest mAP value of 52.74% of the historical documents in the Bentham database. For the modern database, though the performance is relatively low for the best-reported results of the competition, it is comparable to other methods. However, note that we do not carry out any processing on the word images and that the features are directly extracted from the raw images.

In our further research on this problem, we plan to study other textural features to determine the most proper textural features for this problem, and we also intend to employ feature extraction using deep convolutional neural networks. Furthermore, the current study's matching step is quite traditional, and we intend to improve the matching method by using a new distance metric.

SCIENTIFIC CONTRIBUTIONS

- Douaa, Y., Abdeljalil, G., & Chawki, D. (2024). Keyword Spotting from Historical Handwritten Manuscripts using CLBP and CRLBP. *International Journal of Performability Engineering*, *20*(2). https://doi.org/10.23940/ijpe.24.02.p4.9198

- Yousfi, D., Gattal, A., Djeddi, C., Siddiqi, I., & Bensefia, A. (2021, December). Keyword Spotting in Modern Handwritten Documents Using oBIFs. In *Mediterranean Conference on Pattern Recognition and Artificial Intelligence* (pp. 240-250). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-04112-9_18

BIBLIOGRAPHY

[1] Retsinas, G., Louloudis, G., Stamatopoulos, N., & Gatos, B. (2018). Efficient learning-free keyword spotting. IEEE transactions on pattern analysis and machine intelligence, 41(7), 1587-1600.

[2] Rath, T. M., & Manmatha, R. (2003, June). Word image matching using dynamic time warping. In 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. (Vol. 2, pp. II-II). IEEE.

[3] Kundu, S., Malakar, S., Geem, Z. W., Moon, Y. Y., Singh, P. K., & Sarkar, R. (2021). Hough transform-based angular features for learning-free handwritten keyword spotting. Sensors, 21(14), 4648.

[4] Krishnan, P., & Jawahar, C. V. (2019). HWNet v2: an efficient word image representation for handwritten documents. International Journal on Document Analysis and Recognition (IJDAR), 22(4), 387-405.

[5] Sudholt, S., & Fink, G. A. (2018). Attribute CNNs for word spotting in handwritten documents. International journal on document analysis and recognition (ijdar), 21(3), 199-218.

[6] Rohlicek, J. R., Russell, W., Roukos, S., & Gish, H. (1989, May). Continuous hidden Markov modeling for speaker-independent word spotting. In International Conference on Acoustics, Speech, and Signal Processing, (pp. 627-630). IEEE.

[7] Manmatha, R., Han, C., & Riseman, E. M. (1996, June). Word spotting: A new approach to indexing handwriting. In Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 631-637). IEEE.

[8] Khoubyari, S., & Hull, J. J. (1993, April). Keyword location in noisy document image. In Second Annual Symposium on Document Analysis and Information Retrieval, UNLV, Las Vegas (pp. 217-231).

[9] Chen, F. R., Wilcox, L. D., & Bloomberg, D. S. (1993, April). Word spotting in scanned images using hidden Markov models. In 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 5, pp. 1-4). IEEE.

[10] Rath, T. M., & Manmatha, R. (2007). Word spotting for historical documents. International Journal of Document Analysis and Recognition (IJDAR), 9, 139-152.

[11] Haitaamar, S. (2007). Segmentation de textes en caractères pour la reconnaissance optique de l'écriture arabe (Doctoral dissertation, Université de Batna 2).

[12] Tarafdar, A., Pal, U., Ramel, J. Y., Ragot, N., & Chaudhuri, B. B. (2014, August). Word spotting in Bangla and English graphical documents. In 2014 22nd International Conference on Pattern Recognition (pp. 3044-3049). IEEE.

[13] Cao, H., Bhardwaj, A., & Govindaraju, V. (2009). A probabilistic method for keyword retrieval in handwritten document images. Pattern Recognition, 42(12), 3374-3382.

[14] Rodríguez-Serrano, J. A., & Perronnin, F. (2009). Handwritten word-spotting using hidden Markov models and universal vocabularies. Pattern Recognition, 42(9), 2106-2116.

[15] Khurshid, K., Faure, C., & Vincent, N. (2009, July). Fusion of word spotting and spatial information for figure caption retrieval in historical document images. In 2009 10th International Conference on Document Analysis and Recognition (pp. 266-270). IEEE.

[16] Rothacker, L., Fisseler, D., Müller, G. G., Weichert, F., & Fink, G. A. (2015, August). Retrieving cuneiform structures in a segmentation-free word spotting framework. In Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing (pp. 129-136).

[17] Belaïd, A. (2001). Reconnaissance automatique de l'écriture et du document. Campus scientifique, Vandoeuvre-Lès-nancy.

[18] Leydier, Y., Ouji, A., LeBourgeois, F., & Emptoz, H. (2009). Towards an omnilingual word retrieval system for ancient manuscripts. Pattern Recognition, 42(9), 2089-2105.

[19] Terasawa, K., & Tanaka, Y. (2007, September). Locality sensitive pseudo-code for document images. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) (Vol. 1, pp. 73-77). IEEE.

[20] Konidaris, T., Gatos, B., Ntzios, K., Pratikakis, I., Theodoridis, S., & Perantonis, S. J. (2007). Keyword-guided word spotting in historical printed documents using synthetic data and user feedback. International Journal of Document Analysis and Recognition (IJDAR), 9(2), 167-177.

[21] Rothacker, L., & Fink, G. A. (2015, August). Segmentation-free query-by-string word spotting with bag-of-features HMMs. In 2015 13th International conference on document analysis and recognition (ICDAR) (pp. 661-665). IEEE.

[22] Farahmand, A., Sarrafzadeh, A., & Shanbehzadeh, J. (2013). Document Image Noises and Removal Methods.Proceedings of the International MultiConference of Engineers and Computer Scientists 2013, 1, 436-440.

[23] Niblack, W. (1985). An introduction to digital image processing. Strandberg Publishing Company.

[24] Sauvola, J., Seppanen, T., Haapakoski, S., & Pietikainen, M. (1997, August). Adaptive document binarization. In Proceedings of the fourth international conference on document analysis and recognition (Vol. 1, pp. 147-152). IEEE.

[25] Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.

[26] Fiel, S., & Sablatnig, R. (2015). Writer identification and retrieval using a convolutional neural network. In Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II 16 (pp. 26-37). Springer International Publishing.

[27] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). Ieee.

[28] Fei-Fei, L., Fergus, R., & Perona, P. (2004, June). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In 2004 conference on computer vision and pattern recognition workshop (pp. 178-178). IEEE.

[29] Csurka, G., Dance, C., Fan, L., Willamowski, J., & Bray, C. (2004, May). Visual categorization with bags of keypoints. In Workshop on statistical learning in computer vision, ECCV (Vol. 1, No. 1-22, pp. 1-2

[30] Lowe, D. G. (1999, September). Object recognition from local scale-invariant features. In Proceedings of the seventh IEEE international conference on computer vision (Vol. 2, pp. 1150-1157). Ieee.

[31] Rusiñol, M., Aldavert, D., Toledo, R., & Lladós, J. (2015). Efficient segmentation-free keyword spotting in historical document collections. Pattern recognition, 48(2), 545-555.

[32] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60, 91-110.

[33] Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. Lecture notes in computer science, 3951, 404-417.

[34] Almazán, J., Gordo, A., Fornés, A., & Valveny, E. (2014). Word spotting and recognition with embedded attributes. IEEE transactions on pattern analysis and machine intelligence, 36(12), 2552-2566.

[35] Senin, P. (2008). Dynamic time warping algorithm review. Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA, 855(1-23), 40.

[36] Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. The annals of mathematical statistics, 37(6), 1554-1563.

[37] Rath, T. M., Manmatha, R., & Lavrenko, V. (2004, July). A search engine for historical manuscript images. In Proceedings of the 27th annual international ACM SIGIR conference on research and development in information retrieval (pp. 369-376).

[38] Marti, U. V., & Bunke, H. (2002). The IAM-database: an English sentence database for offline handwriting recognition. International Journal on Document Analysis and Recognition, 5, 39-46.

[39] Pechwitz, M., Maddouri, S. S., Märgner, V., Ellouze, N., & Amiri, H. (2002, October). IFN/ENIT-database of handwritten Arabic words. In Proc. of CIFED (Vol. 2, pp. 127-136). Citeseer.

[40] Al Maadeed, S., Ayouby, W., Hassaine, A., & Aljaam, J. M. (2012, September). QUWI: an Arabic and English handwriting dataset for offline writer identification. In 2012 International Conference on Frontiers in Handwriting Recognition (pp. 746-751). IEEE.

[41] Grosicki[1], E., Carre, M., Brodin, J. M., & Geoffrois[1], E. (2008). RIMES evaluation campaign for handwritten mail processing.

[42] Fernández-Mota, D., Almazán, J., Cirera, N., Fornés, A., & Lladós, J. (2014, August). Bh2m: The barcelona historical, handwritten marriages database. In 2014 22nd International Conference on Pattern Recognition (pp. 256-261). IEEE.

[43] Alamri, H., Sadri, J., Suen, C. Y., & Nobile, N. (2008). A novel comprehensive database for Arabic off-line handwriting recognition. In Proceedings of 11th

international conference on frontiers in handwriting recognition, ICFHR (Vol. 8, pp. 664-669).

[44] Pratikakis, I., Zagoris, K., Gatos, B., Puigcerver, J., Toselli, A. H., & Vidal, E. (2016, October). ICFHR2016 handwritten keyword spotting competition (H-KWS 2016). In 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR) (pp. 613-618). IEEE.

[45] Rothfeder, J. L., Feng, S., & Rath, T. M. (2003, June). Using corner feature correspondences to rank word images by similarity. In 2003 Conference on Computer Vision and Pattern Recognition Workshop (Vol. 3, pp. 30-30). IEEE.

[46] Rodriguez, J. A., & Perronnin, F. (2008). Local gradient histogram features for word spotting in unconstrained handwritten documents. Proc. 1st ICFHR, 7-12.

[47] Terasawa, K., & Tanaka, Y. (2009, July). Slit style HOG feature for document image word spotting. In 2009 10th international conference on document analysis and recognition (pp. 116-120). IEEE.

[48] Rusinol, M., Aldavert, D., Toledo, R., & Lladós, J. (2011, September). Browsing heterogeneous document collections by a segmentation-free word spotting method. In 2011 International Conference on Document Analysis and Recognition (pp. 63-67). IEEE.

[49] Aldavert, D., Rusinol, M., Toledo, R., & Lladós, J. (2013, August). Integrating visual and textual cues for query-by-string word spotting. In 2013 12th International conference on document analysis and recognition (pp. 511-515). IEEE.

[50] Zagoris, K., Pratikakis, I., & Gatos, B. (2014, September). Segmentation-based historical handwritten word spotting using document-specific local features. In 2014 14th International Conference on Frontiers in Handwriting Recognition (pp. 9-14). IEEE.

[51] Kovalchuk, A., Wolf, L., & Dershowitz, N. (2014, September). A simple and fast word spotting method. In 2014 14th international conference on frontiers in handwriting recognition (pp. 3-8). IEEE.

[52] Yao, S., Wen, Y., & Lu, Y. (2015, August). Hog based two-directional dynamic time warping for handwritten word spotting. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 161-165). IEEE.

[53] Rabaev, I., Kedem, K., & El-Sana, J. (2016, April). Keyword retrieval using scale-space pyramid. In 2016 12th IAPR Workshop on Document Analysis Systems (DAS) (pp. 144-149). IEEE.

[54] Zagoris, K., Pratikakis, I., & Gatos, B. (2017). Unsupervised word spotting in historical handwritten document images using document-oriented local features. IEEE Transactions on Image Processing, 26(8), 4032-4041.

[55] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. American statistician, 175-185.

[56] Stauffer, M., Fischer, A., & Riesen, K. (2018). Keyword spotting in historical handwritten documents based on graph matching. Pattern Recognition, 81, 240-253.

[57] Serdouk, Y., Eglin, V., Bres, S., & Pardoen, M. (2019, September). KeyWord spotting using siamese triplet deep neural networks. In 2019 International Conference on Document Analysis and Recognition (ICDAR) (pp. 1157-1162). IEEE.

[58] Yousfi, D., Gattal, A., Djeddi, C., Siddiqi, I., & Bensefia, A. (2021, December). Keyword Spotting in Modern Handwritten Documents Using oBIFs. In Mediterranean Conference on Pattern Recognition and Artificial Intelligence (pp. 240-250). Cham: Springer International Publishing.

[59] Daraee, F., Mozaffari, S., & Razavi, S. M. (2021). Handwritten keyword spotting using deep neural networks and certainty prediction. Computers & Electrical Engineering, 92, 107111.

[60] Mohammed, H. H., Subramanian, N., & Al-Madeed, S. (2021). Learning-free handwritten word spotting method for historical handwritten documents. IET Image Processing, 15(10), 2332-2341.

[61] Scius-Bertrand, A., Studer, L., Fischer, A., & Bui, M. (2022, August). Annotation-free keyword spotting in historical Vietnamese manuscripts using graph matching. In Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR) (pp. 22-32). Cham: Springer International Publishing.

[62] Fischer, A., Suen, C. Y., Frinken, V., Riesen, K., & Bunke, H. (2015). Approximation of graph edit distance based on Hausdorff matching. Pattern Recognition, 48(2), 331-343.

[63] Retsinas, G., Sfikas, G., & Nikou, C. (2023, August). Keyword Spotting Simplified: A Segmentation-Free Approach Using Character Counting and CTC Re-

scoring. In International Conference on Document Analysis and Recognition (pp. 446-464). Cham: Springer Nature Switzerland.

[64] Pratikakis, I., Zagoris, K., Gatos, B., Louloudis, G., & Stamatopoulos, N. (2014, September). ICFHR 2014 competition on handwritten keyword spotting (H-KWS 2014). In 2014 14th International Conference on Frontiers in Handwriting Recognition (pp. 814-819). IEEE.

[65] Puigcerver, J., Toselli, A. H., & Vidal, E. (2015, August). ICDAR2015 competition on keyword spotting for handwritten documents. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 1176-1180). IEEE.

[66] Sánchez, J. A., Mühlberger, G., Gatos, B., Schofield, P., Depuydt, K., Davis, R. M., … & De Does, J. (2013, September). tranScriptorium: a 111iamese111 project on handwritten text recognition. In Proceedings of the 2013 ACM symposium on Document engineering (pp. 227-228).

[67] Pratikakis, I., Zagoris, K., Gatos, B., Puigcerver, J., Toselli, A. H., & Vidal, E. (2016, October). ICFHR2016 handwritten keyword spotting competition (H-KWS 2016). In 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR) (pp. 613-618). IEEE.

[68] Mitra, D., Sarkar, P., & Roy, P. (2019). Face recognition by city-block distance classifier in supervised machine learning. IJRAR DOI: http://doi. one/10.1729/Journal, 21653.

[69] Danielsson, P. E. (1980). Euclidean distance mapping. Computer Graphics and image processing, 14(3), 227-248.

[70] Long, D. G. (1981). The Manuscripts of Jeremy Bentham a Chronological Index to the Collection in the Library of University College, London: Based on the Catalogue by A. Taylor Milne.

[71] Gatos, B., Stamatopoulos, N., & Louloudis, G. (2011). ICDAR2009 handwriting segmentation contest. International Journal on Document Analysis and Recognition (IJDAR), 14(1), 25-33.

[72] Wang P., Eglin V., Garcia C., Largeron C., Llados J., Fornes A. A novel learning-free word spotting approach based on graph representation. 11th IAPR International Workshop on Document Analysis Systems (DAS), pp. 207–211 (2014).

[73] Newell, A.J., Griffin, L.D., Morgan, R.M., Bull, P.A.: Texture-based estimation of physical characteristics of sand grains. In: 2010 International Conference on Digital Image Computing: Techniques and Applications, pp. 504–509 (2010).

[74] Newell, A.J., Griffin, L.D.: Writer identification using oriented basic image features and the delta encoding. Pattern Recogn. 47(6), 2255–2265 (2013).

[75] Gattal, A., Djeddi, C., Chibani, Y., Siddiqi, I. Isolated Handwritten Digit Recognition Using oBIFs and Background Features. in Proceedings - 12th IAPR International Workshop on Document Analysis Systems, DAS 2016 (2016). doi:10.1109/DAS.2016.10

[76] Gattal, A., Djeddi, C., Siddiqi, I., Al-Maadeed, S. Writer identification on historical documents using oriented basic image features. in Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR 2018-August, 369–373 (Institute of Electrical and Electronics Engineers Inc., 2018).

[77] Abbas F., Gattal A., Djeddi C., Siddiqi I., Bensefia A., Saoudi K. Texture feature column scheme for single and multi-script writer identification. IET Biometrics ,10(2).179-193.(2021). https://doi.org/10.1049/bme2.12010

[78] Gattal, A., Djeddi, C., Siddiqi, I. & Chibani, Y. Gender classification from offline multiscript handwriting images using oriented Basic Image Features (oBIFs). Expert Syst. Appl. 99, (2018).

[79] Gahmousse A, Gattal A, Djeddi C and Siddiqi I. Handwriting based Personality Identification using Textu ral Features, 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain, 2020, pp. 1-6, https://10.1109/ICDABI51230.2020.9325664

[80] Abbas, F., Gattal, A., Djeddi, C., Siddiqi, I., Bensefia, A. and Saoudi, K. (2021), Texture feature column scheme for single- and multi-script writer identification. IET Biome, 10: 179-193. https://doi.org/10.1049/bme2.12010

[81] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24(7), 971-987.

[82] Ojansivu, V., & Heikkilä, J. (2008, July). Blur insensitive texture classification using local phase quantization. In International conference on image and signal processing (pp. 236-243). Springer, Berlin, Heidelberg.

[83] Belahcene, M., Laid, M., Chouchane, A., Ouamane, A., & Bourennane, S. (2016, October). Local descriptors and tensor local preserving projection in face recognition. In 2016 6th European workshop on visual information processing (EUVIP) (pp. 1-6). IEEE.

[84] Rivera, A. R., Castillo, J. R., & Chae, O. O. (2012). Local directional number pattern for face analysis: Face and expression recognition. IEEE transactions on image processing, 22(5), 1740-1752.

[85] Guo, Z., Zhang, L., & Zhang, D. (2010). A completed modeling of local binary pattern operator for texture classification. IEEE transactions on image processing, 19(6), 1657-1663.

[86] Thajeel, S. A. N., & Sulong, G. (2015). A Novel Approach for Detection of Copy Move Forgery using Completed Robust Local Binary Pattern. J. Inf. Hiding Multim. Signal Process., 6(2), 351-364.

[87] Zhao, Y., Jia, W., Hu, R. X., & Min, H. (2013). Completed robust local binary pattern for texture classification. Neurocomputing, 106, 68-76.

[88] Chen, J., Kellokumpu, V., Zhao, G., & Pietikäinen, M. (2013). RLBP: Robust Local Binary Pattern. In BMVC.

[89] Hannad, Y., Siddiqi, I., & El Kettani, M. E. Y. (2016). Writer identification using texture descriptors of handwritten fragments. *Expert Systems with Applications*, *47*, 14-22.

[90] Abbas, F., Gattal, A., Laouar, M. R., Saoudi, K., & Hadjadj, I. (2018, December). Local Binary Patterns for Gender Classification. In *Proceedings of the 7th International Conference on Software Engineering and New Technologies* (pp. 1-5).

[91] Gattal, A., & Abbas, F. (2020, June). Isolated handwritten digit recognition using LPQ and LBP features. In *Proceedings of the 10th International Conference on Information Systems and Technologies* (pp. 1-5).

[92] Douaa, Y., Abdeljalil, G., & Chawki, D. (2024). Keyword Spotting from Historical Handwritten Manuscripts using CLBP and CRLBP. *International Journal of Performability Engineering*, *20*(2).

ACRONYMS

A

Local Gray Level (ALG)

Artificial Neural Network (ANN)

Arabic Manuscripts from Harvard University (AH)

Alvermann Konzilsprotokolle (AK)

B

Bag of Visual Words (BoVW)

Botany (BOT)

Barcelona Historical Handwritten Marriages databases (BH2M)

C

Complete Local Binary Patterns **(**CLBP)

CLBP-Center (CLBP-C)

CLBP-Sign (CLBP-S)

CLBP-Magnitude (CLBP-M)

Completed Robust Local Binary Pattern (CRLBP)

Cairo Genizah (CG)

D

Derivative of Gaussian (DoG)

Discrete Fourier Transform (DFT)

Dynamic time warping (DTW)

Document Specific Local Features (DSLF)

Document-oriented local features (DoLF)

G

George Washington (GW)

Gaussian mixture model (GMM)

H

Hidden Markov model (HMM)

Histogram of Oriented Gradient (HOG)

Hausdorff Edit Distance (HED)

I

International Conference on the Frontiers of Handwriting Recognition (ICFHR)

K

Keyword Spotting (KWS)

L

Local Binary Patterns (LBP)

Local Phase Quantization (LPQ)

Local Directional Number Pattern (LDNP)

Local Gradient Histogram (LGH)

Local Proximity Nearest Neighbor (LPNN)

Lord Byron (LB)

M

mean Average Precision (mAP)

Multi instance Slective Matching (MISM)

N

Nearest Neighbor Search (NNS)

O

Optical Character Recognition techniques (OCR)

oriented Basic Image Features (oBIFs)

P

Precision at k (P@k)

Pyramidal Histogram of Characters (PHOC)

Parzival (PAR)

Q

Query-by-Example (QbE)

Query-by-String (QbS)

Qatar University Writer Identification (QUWI)

R

regions of interest (ROIs)

S

Short-Term Fourier Transform (STFT)

Scale Invariant Feature Transform (SIFT)

Speeded Up Robust Features (SURF)

Support vector machine (SVM)