

Word Spotting in Offline Handwritten Documents

Thesis submitted to the Kuvempu University for the award of the degree of Doctor of Philosophy in Computer Science under the Faculty of Science and Technology

Submitted by

Thontadari C.

Research Supervisor

Dr. Prabhakar C.J.



Department of P.G. Studies and Research in Computer Science
Kuvempu University, Shankaraghatta
Karnataka, India.

2018

R/E
004
THO

32(a)

t-3867

Kuvempu University Library
Jnanasahyadri, Shankaraghatta

Declaration

I hereby declare that, the research work embodied in this thesis entitled “**Word Spotting in Offline Handwritten Documents**” has been carried out by me in the Department of P.G. Studies and Research in Computer Science, Kuvempu University, Jnana Sahyadri, Shankaraghatta under the supervision of **Dr.Prabhakar C.J.**, for the award of the degree of Doctor of Philosophy in Computer Science. Further, I declare that, this thesis has not been submitted in part or full for the award of any other Degree or Diploma of this or any other University or Institute.

Place : Shankaraghatta

Date : 9/02/2018


Thontadari C.

Kuvempu



University

Department of P.G. Studies and Research in Computer Science

Certificate

This is to certify that, the thesis entitled “**Word Spotting in Offline Handwritten Documents**” submitted by Thontadari C. to the Kuvempu University, Jnana Sahyadri, Shankaraghatta for the award of the degree of Doctor of Philosophy (Ph.D.) in Computer Science during the year 2017 - 2018 is a record of original bonafide research work carried out by him under my supervision. The results embodied in this thesis have not been submitted fully or in part of any Degree or Diploma of this or any other University or Institute.

Jnana Sahyadri Campus

Date: 9/02/2018

Dr. Prabhakar C.J.

Research Supervisor

Department of Computer Science
Kuvempu University, Shankaraghatta

Acknowledgements

This is one of the most significant and delighting moments in my life. This thesis is the result of devoted work which would not have been possible without the support of many. Here, I would like to express my thanks and gratitude to all individuals who have been helped me lot during my research work.

First of all, I would like to take this opportunity to gratefully acknowledge the wholehearted supervision of my very learned supervisor **Dr. Prabhakar C.J.** for his invaluable scholarly advice and his inspiring excellent motivation throughout my research work. The patience, dedication, and constant encouragement of my supervisor has made it possible to me write a dissertation which is appreciable standard and quality. I was always led by his skillful guidance and helpful suggestions, which made it possible for me to go this far.

My cordial gratitude extends to Dr. Ravikumar M., Dr. Yogish Naik G.R., and Dr. Suresha M. faculty members of Department of Computer Science for administrative help, invaluable advices and worthwhile co-operation throughout my research work.

I am grateful to Prof. Bagewadi C.S., UGC Emeritus Fellow, Prof. Narasimhamurthy S.K., Professor & Chairman, Dr. Venkatesha and Dr. Gireesha B.J., faculty members of Department of Mathematics for extending their invaluable suggestion and advice.

I am very much appreciating the time and efforts of Dr. Jose Garcia-Rodriguez, University of Alicante, Spain, Editor-in-Chief of IJCVIP, Dr. Zhongyu (Joan) Lu, University of Huddersfield, UK, Editor-in-Chief of IJIRR, Dr. Jan Zizka, Mendel University, Czech Republic, Editor-in-Chief of IJCSIT and Dr. Divyendu K.Mishra, Editor-in-Chief IJARCSSE for their valuable comments and suggestions which enhance the quality of my research articles.

I express my sincere thanks to Prof. B. Jayadevappa, Principal, University College of Arts & Commerce, Jnanasahyadri, Shankarghatta, for his full support. This made me possible to pursue my research work in a prestigious institution. I am very thankful to my colleagues, teaching and non-teaching staffs, and students of B.C.A Department, for extending their continuous help and support during my research work.

I would also like to thank the staff members of Department of Computer Science, staff members of Examination, Academic, and Finance section of Kuvempu University for their help and encouragement during my research work.

My heartfelt thanks to my senior research scholars Dr. Mohana S.H., Dr. Arun Kumar H.D., Dr. Praveen Kumar P.U., Dr. Nagaraja S., Dr. Jyothi K., and Dr. Too Kipeyago Boaz, for their scholarly advice and invaluable constructive criticism during my research.

I am thankful to my fellow researchers Mr. Puneeth Kumar B.S., Mr. Srikanth K.N., Mr. Pradeep R., Mr. Harishnaik T., Mr. Madhusudhan S., Mr. Sandeep and Mr. Chandra Shekhar. I am sincerely grateful to them for sharing their truthful and illuminating views on number of issues related to the work.

I take this opportunity to express my gratitude to Mr. Mahesh M.C., Mr. Avinash B.S., Dr. Naveen Kumar R.T., Dr. Rudraswamy N.G., Dr. Ashok S.R., and Dr. Ramesh M., who have been the moral support throughout the course of this research.

My special thank to all my friends Dr. Krishna Murthy M.R., Dr. Vishnuvardan S.V., Mr. Ganesh Kumar K., Mr. Chandru K., Mr. Naveen Kumar and Mr. Mahenthesh B. not only for their friendly advices during my research days also for making life at Kuvempu University memorable one.

I would like to thank my M.Sc friends Mr. Ramesh A, Mr. Shivakumar Y. L. Mr. Raghunath Chowan, and Mr. Somanagowda, who have been a source of moral support to me and have extend their helping hands without fail.

I am much thankful to Mrs. Vani Prabhakar, Aishwarya, Darshan for the homely atmosphere I received and I feel privileged to be associated with the family.

I am indebted to my parents for providing me the right education and for inculcating dedication and discipline in my life. I am extremely thankful to my mother Smt. Kumaramma for her endless love, affection and blessings in every walk of my life. I am grateful to my father Sri. Chikkanna who supported me in all my endeavors and encouraged me all through the journey of my research.

Finally, I am forever indebted to my extremely loving brothers Mr. Padmanabha H.C., Mr. Chandra Shekhara H.C. for alleviating my family responsibilities and encouraging me to concentrate on my study and my sweet sisters Smt. Manjula Mahalingaiah., Smt. Supriya Shivaraju for affectionate words and smiling faces provided me additional strength and motivation always.

It will be endless to list each and every friend and relative of mine who have directly or indirectly wished my success in this endeavor. I thank one and all.

Thontadari C.

This thesis is bestowed to

My Beloved Parents

Kumaramma and Chikkanna.

Abstract

A huge amount of information in libraries and cultural institutions exist all over the world and need to be digitized so as to preserve it and protect it from frequent handling. In order to create digital libraries which allow efficient searching and browsing for users, thousands of digitized documents have to be transcribed. To achieve efficient transcription, Optical Character Recognition (OCR) is first used to convert image-based documents into ASCII format through automatic recognition. The automatic recognition by OCR system is suitable for modern high quality printed documents with simple layouts and known fonts. The performance of OCR is very poor for handwritten text due to various challenges posed by handwritten text such as unconstrained writing styles, open vocabulary and paper degradation such as stains, ancient fonts, and faded ink.

The Document Image Analysis (DIA) community has developed a method called word spotting aimed at locating and retrieving a particular word from a document image corpus without explicitly transcribing the whole corpus. In this thesis, we proposed the efficient and novel approaches for retrieving/spotting words in offline handwritten documents, that shows high accuracy, high speed with minimum preprocessing steps. The proposed techniques are learning-based where supervised learning techniques are used to train the documents of a corpus and standard matching technique is used to retrieve/spot word image instances (sub images) similar to a query image.

Generally, a typical word spotting system consists of three main modules: preprocessing, feature extraction and feature matching. Among them, feature

extraction is one of the most important steps for achieving high retrieval performance, because features with strong discriminative information can be well classified even using simplest classifier. Co-occurrence Histograms of Oriented Gradients (Co-HOG) descriptor, which captures the character shape information efficiently and encodes the local spatial information by counting the co-occurrence frequency of gradient orientation of neighboring pixel pairs. This motivated us to propose a segmentation based word spotting method for handwritten document images using Co-HOG descriptor. In order to construct Co-HOG descriptor, initially, we divide a word image into blocks, then, from each block, we extract Co-HOG descriptor and concatenate them to form a feature descriptor, which represent a word image in compact form. Finally, Dynamic Time Warping (DTW) technique is used to retrieve word image instances similar to a query image.

The literature survey reveals that the holistic shapes such as whole words and simple shapes such as characters are extracted more effectively at coarser scales. Hence, we proposed a segmentation based word spotting method for handwritten documents using Scale Space representation. Initially, we derive multi scale images using Gaussian convolution operation employed on original word image. Further, the Co-HOG descriptor is extracted from every multi scale image, and finally concatenated to form a scale space Co-HOG descriptor which yields scale space representation of a word image. The extracted scale space Co-HOG descriptor of the training set is matched with Co-HOG descriptor of a query word image in order to retrieve word image instances similar to query image using a DTW matching technique.

The Bag of Visual Words (BoVW) based image representation is holistic and fixed-length while keeping the discriminative power of local descriptor and that guarantees to reduce computational time by avoiding redundant information. Hence, we proposed a segmentation based word spotting method using BoVW for word image representation which represent the word image as histogram of visual words. The corner points are extracted from word image using Harris-Laplace Corner detector. Then, curvature features are extracted at each corner point. The codebook is used to quantize the visual words by clustering the curvature

features using K-means algorithm. Finally, each word image is represented by a vector that contains the frequency of visual words appeared in the image. For the word retrieval phase, Nearest Neighbor Search (NNS) algorithm is used to match the visual word vector of query image and the visual word vectors presented in the codebook.

One of the main limitations of the segmentation-based word spotting methods is that any segmentation errors can affect the subsequent word representations and matching steps. This motivates us to move towards segmentation-free word spotting method. We proposed a segmentation-free based word spotting method using BoVW based word image representation. The handwritten documents are divided into local patches and local patches are represented with the help of frequency of occurrence of visual words which are constructed using Co-HOG descriptor. Finally, for word retrieval, order the best patches yielded by Nearest Neighbor Search (NNS) similarity measure with respect to their minimum matching cost.

The proposed techniques are evaluated using standard metrics such Precision, Recall, and mean Average Precision (*mAP*). The experiments were conducted using popular offline datasets available for English text handwritten documents such as George Washington (GW), IAM and Bentham datasets. We employed five-fold cross-validation method to validate our approaches. The comparative study of the results obtained using our approaches conclude that the proposed segmentation-free based word spotting technique outperforms state-of-the-art word spotting techniques for offline handwritten documents.

Table of Contents

Abstract	i
List of Figures	ix
List of Tables	xi
1 Introduction	1
1.1 Document Image Processing	1
1.2 Optical Character Recognition (OCR)	2
1.2.1 Applications of OCR	4
1.2.2 Limitations of OCR	5
1.3 Word Spotting	7
1.3.1 Challenges Posed by Word Spotting Methods	8
1.3.2 Applications of Word Spotting	9
1.3.3 Query Format	10
1.4 Features for Word Image Representation	11
1.5 Related Work	13
1.5.1 Scripts Addressed so far	19
1.6 Datasets	21
1.7 Evaluation Metrics	25
1.8 Organization of the Thesis	26

2	Co-HOG Descriptor for Word Spotting	27
2.1	Introduction	27
2.2	Histogram of Oriented Gradients (HOG)	29
2.3	Co-HOG Descriptor	31
2.4	Proposed Method	33
2.4.1	Extraction of Co-HOG Descriptor	33
2.4.2	Word Image Matching	34
2.5	Experimental Results	36
2.5.1	Pre-Processing	36
2.5.2	Parameter Selection	37
2.5.3	Experiments on GW dataset	38
2.5.4	Experiments on IAM dataset	40
2.6	Chapter Summary	42
3	Scale Space Co-HOG Descriptor for Word Spotting	43
3.1	Introduction	43
3.2	Proposed Work	44
3.2.1	Scale Space Representation	44
3.2.2	Extraction of Scale Space Co-HOG Descriptors	45
3.3	Experimental Results	45
3.3.1	Pre-Processing	45
3.3.2	Parameter Selection	47
3.3.3	Experiments on GW dataset	49
3.3.4	Experiments on IAM dataset	50
3.4	Chapter Summary	52
4	Curvature Features Based BoVW for Word Spotting	53
4.1	Introduction	53
4.2	Bag of Visual Words (BoVW) Framework	55
4.3	Proposed Method	56

4.3.1	Corner Points Detection	58
4.3.2	Curvature Features	59
4.3.3	Codebook Generation	60
4.3.4	Word Retrieval	62
4.4	Experimental Results	62
4.4.1	Segmentation	63
4.4.2	Selection of Codebook Size	63
4.4.3	Experiments on GW Dataset	65
4.4.4	Experiments on IAM Dataset	67
4.4.5	Experiments on Bentham Dataset	69
4.5	Chapter Summary	71
5	Co-HOG Descriptor Based BoVW for Word Spotting	72
5.1	Introduction	72
5.2	Proposed Work	74
5.2.1	Document Image Representation	76
5.2.2	Local Patch Representation	78
5.2.3	Retrieval Stage	79
5.3	Experimental Results	81
5.3.1	Parameters Used	81
5.3.2	Selection of Codebook Size	82
5.4	Experiments on GW dataset	83
5.5	Experiments on IAM dataset	86
5.6	Comparative Study of Proposed Techniques	89
5.7	Chapter Summary	92
6	Conclusions and Future Directions	93
6.1	Conclusions	93
6.2	Future Directions	95
7	Author's Publications	96
	Bibliography	98

List of Figures

1.1	Steps involved in OCR system	3
1.2	Example images of (a) fraktur font and (b) gothik font	5
1.3	Image having different font size	6
1.4	Broken characters image	6
1.5	Examples of handwritten documents having (a) unconstrained writing style and (b) faded ink	6
1.6	General architecture of word spotting (Courtesy: Giotis et al., 2017)	8
1.7	Sample handwritten document image from GW dataset	22
1.8	Sample handwritten document image from IAM dataset	23
1.9	Sample handwritten document image from Bentham dataset	24
2.1	Intermediate results for extraction of HOG feature descriptor from word image	30
2.2	Lexis of gradient orientation:(a) single gradient orientation (b) a pair of gradient orientation. (Courtesy: Watanabe et al., 2009)	31
2.3	Extraction of Co-HOG feature descriptors: (a) offset used in Co-HOG, (b) Co-occurrence matrix for the given offset. (Courtesy: Tian et al., 2016)	32
2.4	Flow diagram of proposed method	34
2.5	Illustration of Co-HOG feature descriptor extraction process: (a) Sample word image (b) Word image divided in to 3×6 blocks and corresponding co-occurrence matrices of each block and (c) concatenated one after another to form a Co-HOG feature vector	35

2.6	Sample results of denoised and segmented word images	37
2.7	The comparison of mAP of our approach for different number of blocks	38
3.1	Illustration of Scale Space Co-HOG feature descriptor extraction process: (a) sample word image (b) scales at $\sigma = 0$ (original image), $\sigma = 1$ and $\sigma = 2$ respectively. (c) word image divided into blocks and corresponding co-occurrence matrices of each block and (d) concatenate one after another to form a Scale Space Co-HOG feature vector	46
3.2	Sample results of denoised and segmented word images	47
3.3	The comparison of mAP of our approach for different number of blocks	48
4.1	The pipeline of proposed word spotting method	57
4.2	The intermediate results for corner points detection: (a) segmented word image, (b) canny edge map, (c) detected corner points on canny edge map image using Harris-Laplace operator (d) detected corner points on original word image (e) corner point at the global level (f) sample corner point at the local level	58
4.3	Curvature of the curve C	60
4.4	BoVW framework for handwritten word images	61
4.5	Segmented word images (a) from GW database (b) from IAM database (c) from Bentham database	63
4.6	The performance comparison of our approach for varying codebook size on three different datasets based on mAP	64
5.1	The pipeline of proposed word spotting method	75
5.2	An overview of the document image representation (a) sample document image from GW database (b) dividing an image into regular grids of size and (c) Co-HOG histogram computed from each regular grid	77
5.3	The second level SPM configuration of proposed approach	78

5.4	Visualization of the word spotting. (a) query image, (b) sample document from GW dataset (c) spotted local patches where words similar to query word are found	80
5.5	mAP of our approach for varying codebook size for two datasets .	83
5.6	Example of spotting results for the query word “ <i>your</i> ” in one of documents of GW dataset	84
5.7	Example of spotting results for the query word “ <i>British</i> ” in one of the documents of IAM dataset	86
5.8	PrecisionRecall curves for different configurations for GW dataset	91
5.9	PrecisionRecall curves for different configurations for IAM dataset	91

List of Tables

1.1	Script addressed by existing word spotting methods	20
2.1	mAP of our approach for different number of blocks	37
2.2	Sample retrieval results: Query word images (first row) and their retrieval results of our approach for GW dataset	39
2.3	Performance comparison of our approach with existing word spotting methods for GW dataset	40
2.4	Sample retrieval results: Query word images (first row) and their retrieval results of our approach for IAM dataset	41
2.5	Performance comparison of our approach with existing methods for IAM dataset	41
3.1	mAP of our approach for different number of blocks	48
3.2	Sample retrieval results: Query word images (first row) and their retrieval results of our approach for GW dataset	49
3.3	Performance evaluation of our approach using GW dataset	50
3.4	Sample retrieval results: Query word images (first row) and their retrieval results of our approach for IAM dataset	51
3.5	Performance evaluation of our approach using IAM dataset	52
4.1	Performance evaluation of our approach for varying codebook size using three datasets	64
4.2	Sample retrieval results: Query word images (first column) and corresponding retrieved word instances from GW dataset	66
4.3	Performance evaluation of our approach using GW dataset	66

4.4	Sample retrieval results: Query word images (first column) and corresponding retrieved word instances from IAM dataset	68
4.5	Performance evaluation of our approach using IAM dataset	68
4.6	Sample retrieval results: Query word images (first column) and corresponding retrieved word instances from Bentham dataset	69
4.7	Performance evaluation of our approach using Bentham dataset	70
5.1	Performance evaluation of our approach for varying codebook size	82
5.2	Sample retrieval results: Query word images (first column) and corresponding retrieved word instance patches from GW dataset	84
5.3	The performance comparison of our approach with state-of-the-art segmentation free methods for GW dataset	85
5.4	Sample retrieval results: Query word image (first column) and corresponding retrieved word instance patches from IAM dataset	87
5.5	The performance comparison of our approach with state-of-the-art segmentation free word spotting methods for IAM dataset	88
5.6	Examples for some false positives	88
5.7	The performance comparison of proposed word spotting techniques	90

Chapter 1

Introduction

Analysis of document images for information extraction has become very prominent domain in recent days. In this chapter, we present a brief introduction to document image processing, steps involved in optical character recognition (OCR), its applications and limitations. Then, we present an overview of word spotting method and its techniques. Finally, we present an organization of the thesis.

1.1 Document Image Processing

In the phrase “**Document Image**”, the word “**Document**” is derived from the Latin word “**documentum**” which in Medieval Latin is referred to as a “**written instruction**” is normally used to communicate and store information (Javed 2016). Document image processing is a subfield of digital image processing. It mainly deals with the transformation of digitized document images into electronic form for storage, transmission, reprocess, and modification. The objective of document image processing is to recognize the text and graphics components in document images, and to extract the intended information. Document image processing can be categorized into two types such as Textual processing and Graphics processing.

Textual Processing: Textual processing deals with the text components of a document image. Some tasks here are: recognizing the text by OCR, determining the skew, finding columns, paragraphs, text lines, and words. Text document images are composed of plain or illustrated text (books, magazines, newspaper, archives) and structured text (forms, invoices, envelop).

Graphics Processing: Graphics processing deals with the non-textual line and symbol components that make up line diagrams, delimiting straight lines between

text sections, company logos, etc. Graphics document images are composed of maps, engineering drawings, and music sheets; their corresponding document image processing goals are Geographic Information Systems (GIS) representation for maps, Computer Aided Design (CAD) format for engineering drawing, and Musical Instrument Digital Interface (MIDI) representation for music scores.

1.2 Optical Character Recognition (OCR)

Optical character recognition (OCR) has become one of the most successful research areas in document image analysis. It involves many sub disciplines of computer science like Image Processing, Pattern Recognition (PR), Natural Language Processing (NLP), Artificial Intelligence (AI) and database systems. OCR is the process of converting a raster image representation of a document into an electronic file format, which can be edited. OCR system converts different types of documents such as scanned paper documents, portable document files (pdf) or images which are captured by a digital camera into editable and searchable data. OCR system has properties of a good document search engine, i.e., fast access, efficiency, low computational cost, and acceptable accuracy. OCR has been used recognition and retrieval of document images.

A typical OCR system consists of binarization of scanned image followed by applying some preprocessing steps such as separation of text and graphics, skew correction, noise removal. Then, the preprocessed image is segmented into characters. The segmented characters undergoes a feature extraction step. Finally extracted features are fed into a suitable classifier for recognition. Basic steps involved in OCR is shown in Figure 1.1.

Input: The input to the OCR is document image, which is usually captured by an image sensor such as an optical scanner or digital camera. The document images can be in any format such as color image or grayscale image or binary image.

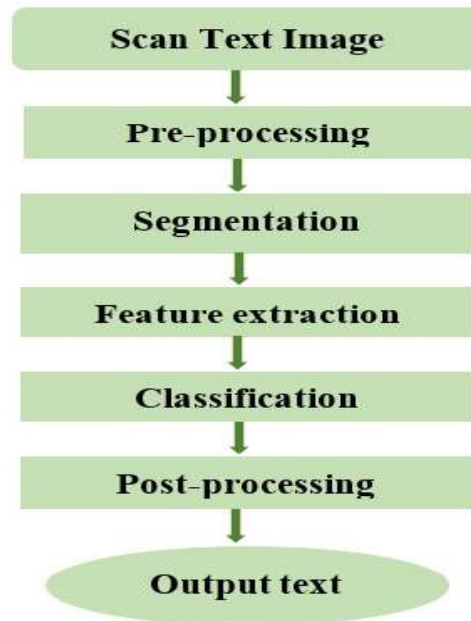


Figure 1.1: Steps involved in OCR system

Pre-processing: The initial step is binarization which includes a threshold for converting color image to grayscale or color image to binary image. After binarization, the other pre-processing techniques which are applied commonly such as skew detection, correction, normalization and image enhancement.

Segmentation: The pre-processing stage yields a clean image, which is having sufficient amount of shape information and low noise on a normalized image. Segmentation is an important step in OCR because the separation of individual characters directly affects the recognition rate. The segmentation step is divided into three categories (a) line segmentation, (b) word segmentation and (c) character segmentation.

Feature extraction: Extraction of feature plays an important role in recognition system. The features which are extracted from the whole image are known as the global features. A number of holes in the character, the number of concavities in its outer contour, and the relative protuberance of character extremities are global features. These features are extracted to make the document image invariant to rotation, translation and scaling. The features which are extracted

from segmented blocks or from the subdivision of the document are known as local features. Regions corresponding to individual character compactness, aspect ratio, black pixel density, asymmetry, number of line crossings and line endings are local features.

Classification: Extracted features are used to train a classifier. When a new document image is given as input, classifier recognizes it by the trained information. Classifiers such as Neural Networks, Hidden Markov Models (HMM), Support Vector Machines (SVM), Bayesian classifier and Template matching can be used.

Post-processing: Post-processing is used to improve the recognized document image results and thereby it increases the accuracy. Character n-gram, domain knowledge, dictionary and semantic knowledge can be used for this purpose.

Output: The output of OCR is text document which is in machine editable form.

1.2.1 Applications of OCR

The last few decades have seen a widespread appearance of OCR system satisfying requirements of different users. The below mentioned application areas are those in which OCR has been successfully used.

Aid for blind: OCR system is used to enable blind to understand documents through text information extraction and recognition.

Automatic number plate recognition: OCR system for automatic reading of number plates of vehicles. As opposed to other OCR applications, an input image is not natural bi-level image and must be captured by the very fast camera.

Automatic cartography: The character recognition from maps is one of the applications of OCR and called as automatic cartography. The challenges associated with automatic cartography is that symbols are intermixed with graphics, text is printed at different angles and characters are of several fonts or even handwritten.

Form readers: OCR systems are able to read specially designed forms. The characters are in printed or handwritten upper case letters or numerals in specified boxes. The processing speed is dependent on amount of data on each form but may be few hundred forms per minute.

Signature verification and identification: This application is useful for banking environment. Such system establishes the identity of a writer without attempting to read handwriting. The signature is simply considered as the pattern which is matched with signatures stored in the reference database.

1.2.2 Limitations of OCR

There are several limitations of OCR that makes inaccurate recognition result. Hence, classification and retrieval process becomes difficult.

Font type: OCR techniques usually recognize words by processing fonts independently and works well with machine printed fonts against a clean environment. OCR cannot recognize the characters in old font types such gothic, fraktur fonts shown in Figure 1.2 and also characters present in old documents that are not available in modern computer fonts.

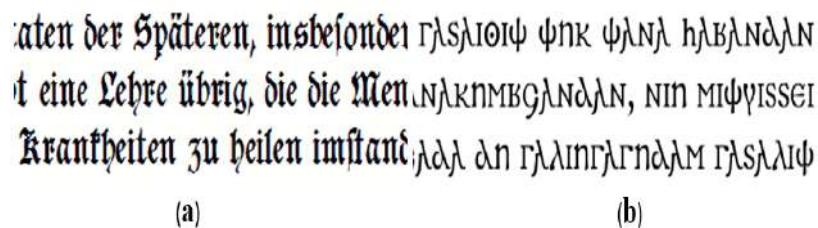


Figure 1.2: Example images of (a) fraktur font and (b) gothik font

Font Size: OCR may not convert document image contains characters which is having very large or very small font sizes as shown in Figure 1.3. This can make the most important characters and words unavailable for text-based systems.

Languages: Many languages have special characters those characters can be lost or incorrectly recognized by OCR system.

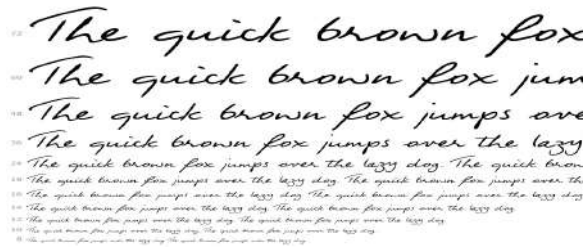


Figure 1.3: Image having different font size

Quality of the characters: Document images containing broken characters (Figure 1.4) and mixed with noise also affect the accuracy of OCR.

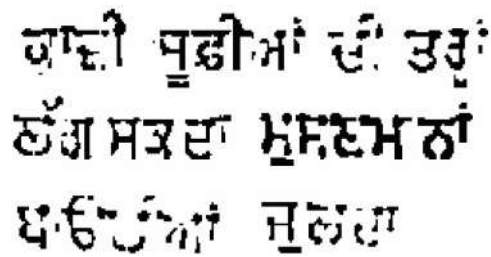


Figure 1.4: Broken characters image

Processing of documents through OCR requires high computation rate due to difficulty involved in understanding the page layout of digitized documents, dull ink, stained paper and other adverse factors. The performance of OCR is very poor for handwritten text due to various challenges posed by handwritten documents having unconstrained writing styles (Figure 1.5) causes still very high error rate.

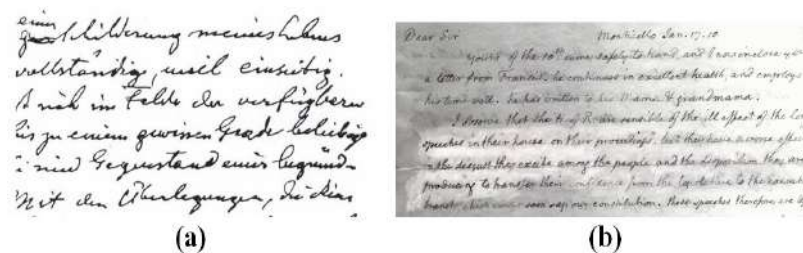


Figure 1.5: Examples of handwritten documents having (a) unconstrained writing style and (b) faded ink

1.3 Word Spotting

Retrieving information from a huge collection of historical and modern documents is useful for interpreting and understanding documents in various domains. Document digitization provides an inspiring alternative to preserve valuable historic and modern manuscripts. However, digitization solitary cannot be obliging until these collections of manuscripts can be indexed and made searchable. Spotting particular regions of interest in a digital document are easy owing to the possibility to search for words in huge sets of document images. The procedure of manual or semi-automatic transcription of the whole text of handwritten documents for searching any particular word is a tiresome and expensive job and automation is desirable in order to reduce costs.

In earlier research work, character recognition is used widely in order to search the required word in a document. Available OCR engines designed for different languages yield excellent recognition results on scanned images of good quality printed documents. However, the performance of OCR engines significantly degraded when applied to handwritten documents due to faded ink, stained paper, and other adverse factors of handwritten documents. Another factor is that OCR techniques that usually recognize words character by character. The performance of the available OCR engine is highly dependent on the burdensome process of learning. Moreover, the writing and font style variability, linguistics and script dependencies are the impediments of such systems.

To overcome the aforementioned limitations of OCR, the Document Image Analysis (DIA) community has developed a recognition free technique called as word spotting. Word spotting is a moderately new alternative technique for text recognition and retrieval of words in document images. Word spotting can be defined as the pattern recognition task aimed at locating and retrieving a particular word from a document image collection without explicitly transcribing the whole corpus. The optimum trend in word spotting systems is to propose methods that show high accuracy, high speed and work on any language with minimum preprocessing steps.

The components of a word spotting method is a collection of documents or document corpus and an input element denoted as a query. The output of word spotting method should be the localization in documents or sub images of this collection that are similar to the query, making it similar to the classical information retrieval system. Figure 1.6 illustrates a general architecture of word spotting method where the whole procedure is divided in an offline and an online phase. In the offline stage, a set of features are extracted from either word images, or text lines or whole document pages which are then represented by feature vectors. In the online phase, a user formulates a query either by selecting an actual example from the collection or by typing an ASCII text word. Then matching process is applied to these representations in order to obtain a similarity score which yields a ranking list of results according to their similarity with the query.

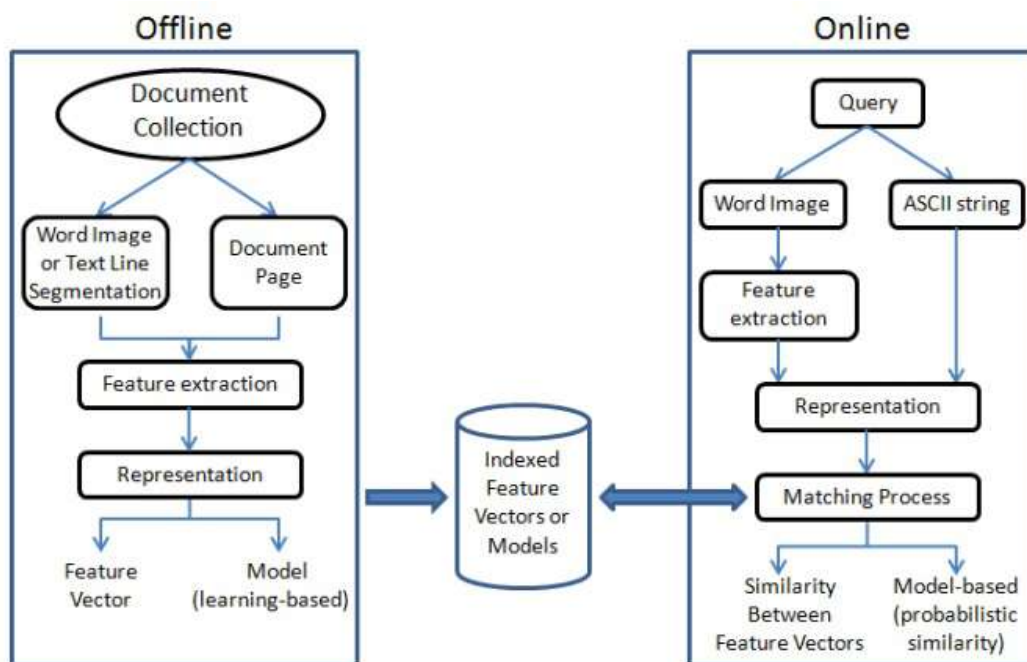


Figure 1.6: General architecture of word spotting (Courtesy: Giotis et al., 2017)

1.3.1 Challenges Posed by Word Spotting Methods

The word spotting in handwritten documents is not completely solved due to various challenges posed by handwritten documents and the challenges involved in handwritten documents are :

- Handwritten documents, either historical or modern, always suffer from variability in writing style, not only for different authors but also for documents of the same writer. But, in the case of machine printed text, it is mainly concern of variations in the font type
- In handwritten documents, the words may be skewed, characters may be slanted, non-text content such as symbols may be present and letters may be broken or connected in a cursive manner
- Historical printed documents also present challenges for word spotting because of degradations such as missing data, non-stationary noise due to illumination changes during the scanning process, low contrast, show through or warping effects
- Degradations involved in documents impede the overall performance of a word spotting method. Because low image quality directly affects the segmentation and feature extraction stages of a word spotting method
- A learning-based word spotting methods provide good results for different languages of a relevant script, but it is not suitable for a different script, unless new training data are used

Hence, in this research work, we focused on word spotting in offline handwritten documents rather than printed documents.

1.3.2 Applications of Word Spotting

For historical documents, a key application is given by integrating handwritten documents in digital libraries (Nagy and Lopresti, 2006). There are a variety of applications of word spotting for document indexing and retrieval including the following:

- Great assistance is provided in searching and browsing historical handwritten collections written by a single or several authors for researchers and the public

- Assisting human transcribers in identifying words in degraded documents especially those appearing for the first time
- Automatic sorting of handwritten mail containing significant words like urgent, cancellation and complain
- Identification of figures and their corresponding captions
- Retrieval of documents with a given word in company files
- Keyword retrieval in pre-hospital care reports (PCR forms)
- Word spotting in graphical documents such as maps

1.3.3 Query Format

For any word spotting methods, the query can be given in two ways. Such as:

- Query-By-Example(QBE) or Query-By-Image(QBI)
- Query-By-String (QBS)

Query-By-Example(QBE) or Query-By-Image(QBI): The input is an image of the word to search and the output is a set of the representative images in the database containing a similar to the query word. In this research work, we focused on the QBE paradigm, which consists of retrieving word image instances which are similar to a given query word image. In this context, words are represented as sequences of feature vectors. Therefore, the quality of the matching directly depends on the measure of similarity between such sequences.

Query-By-String (QBS): The input is text string. Character models are learned in advance and at runtime, the character models are combined to form words and probability of each word is evaluated. This approach is more complicated than query-by-example because word models are needed to be created even though no samples of that word exist in the training set and also faced similar drawbacks of OCR systems.

1.4 Features for Word Image Representation

In word spotting, the selection of proper and effective features plays a vital role, especially in historical and degraded handwritten documents. Features are broadly classified into two categories: global features and local features. In global features, an image is represented by single vector while in local features, the first keypoint is selected and based on keypoints neighborhood feature is designed. Some of the popular features are: profiles feature, moments, Gradient based binary feature, Scale Invariant Feature Transform (SIFT), Speeded Up Robust Feature (SURF) and Shape Context etc. Here we will describe popular features used in word spotting methods.

Projection Profiles : These features are calculated by summing the number of black pixels in each image column. In this way, the length of the feature vector is equal to the total number of columns, where each vector value represents a single column. The black pixels can either be counted for the entire image column or part of it, i.e., partial projection profile. For extracted partial projection profiles, three zones of the image are considered i.e., ascenders-zone, x-zone, and descenders zone. The projection profiles are calculated for each zone, which is called upper projection profile, middle projection profile and lower projection profile, respectively.

Word Profile : Word profile features can be calculated from four directions of a given word image, i.e., upper, lower, right and left. The first two are calculated by scanning the image column-wise whereas the latter two are calculated by scanning the image row-wise. The distance of first ink pixel, within each image column, from the upper word boundary is called the upper word profile, whereas that from the lower word boundary is called the lower word profile. Similarly, the distance of the first ink pixel, within each image row, from the left boundary is called the left word profile while that from the right is called the right word profile.

Mesh Features : The image is logically divided into a fixed number of zones, and for each zone the black pixel density is calculated, i.e., total number of ink

pixels. The length of the feature vector is equal to the total number of zones.

Cavity Features : These are features which represent the gaps between strokes of word. These features capture local variations and are used to distinguish different words with similar general shapes. A region point which is bounded by the character strokes on at least three directions/sides is called a cavity. Keaton et al. (1997) have used six types of cavities, i.e., east cavity, west cavity, north cavity, south cavity, center cavity and hole.

Gradient Orientation : A feature which describes the shapes of words by depicting the strokes' true local structure and the orientation of characters' contours. Usually, it is a feature vector, computed from the gradient of grey level, which points in the direction in which the grey level's value increases to maximum and whose value represents the rate of change.

GSC Features : These are the combination of three features, i.e., Gradient, Structure and Concavity (GSC) features. These features measure the characteristics of an image at local, intermediate and global ranges respectively. The Gradient features measure edge curvature in the neighborhood of a pixel and provide useful information about the stroke shape. They are then further extended to a longer distance by Structural features in order to achieve important information about stroke trajectories. The stroke relationships at a global scale are detected by Concavity features, i.e., across the entire image.

Angular Lines Features : An x-y plane is created, which has its origin at the centroid of the word image. Each quadrant of this plane is divided into two equal sections (45 degrees each), thus forming 8 equal regions. In each region, the number of pixels is counted which results in the creation of an 8-valued feature vector. The number of pixels in each region is divided by the total number of pixels in the word image for the purpose of normalization.

Concentric Circle Features : Several concentric circles are drawn by considering the centroid of the word image as the center. The number of pixels between the two consecutive circles are counted and used as feature values. For computing the number of pixels between two circles, the total number of pixels of the inner circle is subtracted from the total number of pixels enclosed by the outer circle.

Upper/Lower Contour : The upper contour is the distance from the upper outline of a word to the top of the bounding box. The distance from the lower outline of a word to the bottom of the bounding box is called its lower contour. If characters within the word are not touching each other, then the lower contour follows (i.e., touches) the top of the bounding box and the upper contour follows (i.e., touches) the bottom of the bounding box.

Ascenders/Descenders : These features depend upon the baseline and the x-line of the word image. The ascenders are the characters extended above the x-line while descenders are the characters extended below the baseline. These features have been largely used for word spotting in Latin scripts. Like cavity features, these features are also used in several ways, e.g., they can either be used as principal features or initially used to define a set of shape codes.

Scale Invariant Feature Transform (SIFT) : Scale Invariant Feature Transform (SIFT) is one of the most popular descriptor proposed in literature. SIFT is widely used in different computer vision applications like image retrieval and image classification, etc. consists of four major stages: (1) scale-space peak selection; (2) keypoint localization; (3) orientation assignment; (4) keypoint descriptor.

Speeded Up Robust Features (SURF) : Speeded up robust features (SURF) is a local feature detector and descriptor. It is partly inspired by the SIFT descriptor. To detect interest points, SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a precomputed integral image. SURF feature is based on the sum of the Haar wavelet response around the point of interest. These can also be computed with the aid of the integral image. SURF descriptors have been used to locate and recognize objects, people or faces, to reconstruct 3D scenes, to track objects and to extract points of interest.

1.5 Related Work

Word spotting algorithms intend to reduce tedious and time overwhelming manual annotation applied to the pictorial representation of input document words.

The literature survey reveals that a lot of techniques have been developed to spot the words in handwritten documents. In this section, we briefly discuss the word spotting methods proposed by various researchers for handwritten documents.

Word spotting was initially proposed by Jones et al. (1995) in the field of speech processing, while later this concept was agreed by several researchers in the field of printed and handwritten documents for the purpose of spotting and indexing. This approach enables to localize a user preferred word in a document without any syntactic restriction and without an explicit text recognition or training phase. The concept of word spotting (Manmatha et al., 1999) has been introduced as an alternative to OCR based results. The word spotting methods have followed a well-defined process. Initially, layout analysis is carried out to segment the words from document images. Then, the segmented word images are symbolized as sequences of features such as geometric features (Marti et al., 2001), profile based features (Rath et al., 2003; Larvenko et al., 2004), local gradient features (Rodriguez et al., 2008). Finally, similarity measure methods, such as XOR comparison, Euclidean distance, Scott and Longuet Higgins distance, Hausdorff distance of connected components, the sum of Euclidean distances of corresponding key points. More, recently Dynamic Time Warping and Hidden Markov Models are used to compare the feature of query word image and set of word images presented in the dataset. Finally, retrieved word images are ranked according to this similarity.

Rath et al. (2003) proposed an algorithm for matching handwritten words in noisy historical documents. The segmented word images are preprocessed to create 1-dimensional features, which are used to train the probabilistic classifier, which is then used to estimate similarity between word images. Rothfeder et al. (2003) presented an algorithm to draw correspondences between points of interest in two word images and utilizes these correspondences to measure the similarities between the images. Rath et al. (2003) proposed an approach which involves grouping similar word images into clusters of words by using both K-means and agglomerative clustering techniques. They constructed an index that links words to the locations of occurrence which helps to spot the words easily.

Zhang et al. (2004) have proposed word spotting technique based on gradient based binary features can offer higher accuracy and much faster speed than DTW matching of profile features. Srihari et al. (2005) indexed documents using global word image features such as stroke width and slant. Word gaps are used to measure the similarities between the spotted words and a set of prototypes from known writers. Srihari et al. (2006) developed a word spotting system that retrieves the candidate words from the documents and ranks them based on global word shape features. Rath et al. (2007) proposed an approach which involves grouping word images into clusters of similar words by using both K-means and agglomerative clustering techniques. They construct an index automatically that links words to the locations of occurrence which helps to spot the words easily. Rodriguez et al. (2008) proposed a method that relaxes the segmentation problem by requiring only segmentation at the text line level.

Louloudis et al. (2009) proposed a segmentation of the page to detect significant text regions. The represented queries are in the form of a descriptor based on the density of the image patches. Then, a sliding window search is performed over the significant regions of the documents using refined template-based matching. A probabilistic representation for learning based word spotting in a multi-writer text is proposed by Rodriguez et al. (2009). The query and the dataset word images are represented as sequences of feature vectors extracted using a sliding window in the writing direction and they are modeled using statistical models.

Unsupervised writer adaptation for unlabelled data has been successfully used for word spotting method proposed by Rodriguez-Serrano et al. (2010) based on statistical adaption of the initial universal codebook to each document. Rodriguez-Serrano et al. (2010) proposed an unsupervised handwritten word spotting using semi-continuous hidden Markov model to separate the word model parameters into a codebook of shapes and a set of word-specific parameters. An HMM-based method which learns character models for word spotting in handwritten documents is proposed by Fischer et al. (2010). Initially, text lines of the document are represented by a sequence of nine geometrical features which is obtained by a sliding window of one-pixel width from left to right over the

image. The character models are trained in offline using labeled text line images. Then, a text line model is created as a sequence of letter models according to the transcription.

Kesidis et al. (2011) extracted pixel density of zones and projection profile features and computes the Euclidean distance between word images, and refines the search procedure by user feedback. A template free word spotting method for handwritten documents was described by Frinken et al. (2011) which is derived from the neural network based system. The word spotting is done using a modification of the Connectionist Temporal Classification (CTC) token passing algorithm in conjunction with a recurrent neural network. Rusinol et al. (2011) have proposed a word spotting method for handwritten documents using SIFT descriptor. The drawback of this technique is that directly matching local key points is computationally expensive when dealing with large datasets. Frinken et al. (2011) have proposed word spotting using a modification of the Connectionist Temporal Classification (CTC) token passing algorithm in conjunction with a recurrent neural network.

Rodriguez-Serrano et al. (2012) proposed a model-based similarity between vector sequences of handwritten word images with semi-continuous Gaussian mixture HMMs. The work of Shekhar et al. (2012) avoids segmentation by representing regions with a fixed length SIFT descriptor. Khurshid et al. (2012) have proposed a word spotting method for scanned documents using a sequence of sub-patterns. The connected components algorithm is used to transform a word pattern into a sequence of sub-patterns. Each sub-pattern is represented by a sequence of feature vectors. Then, modified Edit distance is used to perform a segmentation-driven string matching and to compute the Segmentation Driven Edit (SDE) distance between the words.

The segmentation free method for word spotting in handwritten documents proposed by Zhang et al. (2013) based on heat kernel signatures (HKS) descriptors are extracted from a local patch centered at each keypoint. Kessentini et al. (2013) proposed a novel system for segmentation free and lexicon free word spotting and regular expression detection in handwritten documents using filler

model which allows accelerating the decoding process. Huang et al. (2013) proposed contextual word model for keyword spotting in off-line Chinese handwritten documents by combining a character classifier and the geometric context as well as linguistic context. They conducted experiments on handwriting database CASIA-HWDB demonstrate the effectiveness of the proposed method and justify the benefits of geometric and linguistic contexts. A template-free word spotting method for handwritten documents was proposed by Almazan et al. (2013) for multi-writer handwritten documents, which uses attributes-based approach for the pyramidal histogram of characters. This embeds the handwritten words in a more discriminative space, where the similarity between words is independent of the writing style. Rothacker et al. (2013) have proposed to combine SIFT descriptor with Hidden Markov Models (HMM) in a patch-based segmentation free word spotting. Zhang et al. (2013) have proposed segmentation free word spotting based on Heat Kernel Signature (HKS) descriptor. HKS descriptors are extracted from a local patch centered at each key point detected by the SIFT key point detector on the document pages and the query image.

Wshah et al. (2014) proposed a statistical script independent line based word spotting method for offline handwritten documents based on hidden Markov models by comparing filler models and background models for the representation of background and non-keyword text. An unsupervised Exemplar SVM framework for segmentation free word spotting method proposed by Almazn et al. (2014) using a grid of HOG descriptors for documents representation. Then, a sliding window is used to locate the document regions that are most similar to the query. Coherent learning segmentation based Arabic handwritten word spotting system proposed by Khayyat et al. (2014) which can adapt to the nature of Arabic handwriting and the system recognizes Pieces of Arabic Words (PAWs).

An efficient segmentation free word spotting method for the historical document is proposed by Rusinol et al. (2015). They used a patch-based framework of the bag-of-visual-words model powered by SIFT descriptors. By projecting the patch descriptors to a topic space with the latent semantic analysis technique and compressing the descriptors with the product quantization method efficiently

index the document information both in terms of memory and time. Based on inkball character models, Howe (2015) proposed a word spotting method using synthetic models composed of individual characters. Ghosh et al. (2015) proposed a segmentation-free query by string word spotting method based on a Pyramidal Histogram of Characters (PHOC) are learned using linear SVMs along with the PHOC labels of the corresponding strings.

Line-level keyword spotting method proposed by Toselli et al. (2016) on the basis of frame-level word posterior probabilities of a full-fledged handwritten text recognizer based on hidden Markov models and N-gram language models. A template-based learning-free word spotting method proposed by Dey et al. (2016) by combining the Local Binary Pattern (LBP) histograms and spatial sampling. One of the main advantages is its independence from the actual representation formalism as well as the underlying language of the document. However, template-based word spotting does not generalize well to different writing styles.

Sfikas et al. (2017) have proposed semi supervised segmentation based keyword spotting based on probabilistic interpretation of Canonical Correlation Analysis (CCA), using Expectation Maximization (EM). Segmentation-free word spotting method for multi-write handwritten documents based on Radial Line Fourier (RLF) descriptor is proposed by Hast et al. (2017). RLF is a short-length feature vector of 32 dimensions, that adheres to the property that the handwritten words across different documents are indeed similar.

1.5.1 Scripts Addressed so far

Regarding the nature of documents which have been addressed by the research community for word spotting, Table 1.1 illustrates the various scripts addressed by most of the representative works for word spotting. So far, word spotting has been applied to various scripts, such as Arabic, Chinese, Devanagari, Greek and Latin. These scripts differ from each other owing to factors such as the writing direction, size of the alphabet, number of characters and cursiveness. For example, documents in Arabic scripts are written from right to left, in the horizontal direction and are fully cursive. On the contrary, text in Latin script is written from left to right in the horizontal direction only, cursively in some cases. Chinese scripts contain thousands of characters and are written in two dimensions, either from left to right horizontally, or from top to bottom vertically. Devanagari scripts are written horizontally, from left to right in a complex cursive way, whereas Greek scripts are written from left to right without cursiveness. Furthermore, each separate character of the Chinese scripts has specific meanings or semantics, in contrast with the isolated characters of other scripts.

Table 1.1: Script addressed by existing word spotting methods

Publications	Context	Languages	Script	Type
Aldavert et al. (2013;2015), Zagoris et al. (2014)	Historical	English	Latin	Handwritten
Zhang et al. (2013), Fornes et al. (2011)	Historical	English	Latin	Handwritten
Roy et al. (2013), Rothacker et al. (2013;2015)	Historical	English	Latin	Handwritten
Mondal et al. (2013), Dovgalecs et al. (2013)	Historical	English	Latin	Handwritten
Rath et al. (2007), Zhong et al. (2016)	Historical	English	Latin	Handwritten
Cao et al. (2009), Wagan et al. (2010)	Modern	English	Latin	Handwritten
Kumar et al. (2012)	Modern	English	Latin	Handwritten
Retsinas et al. (2016), Krishnan et al. (2016)	Historical, modern	English	Latin	Handwritten
Almazan et al. (2014), Liang et al. (2012)	Historical, modern	English	Latin	Handwritten
Wilkinson et al. (2016), Fischer et al. (2013)	Historical, modern	English	Latin	Handwritten
Ghosh et al. (2015)	Historical, modern	English	Latin	Handwritten
Kessentini et al. (2013;2015), Choisy (2007)	Modern	French	Latin	Handwritten
Howe (2013;2015), Frinken et al. (2012)	Historical	English, German	Latin	Handwritten
Puigcerver et al. (2014), Riba et al. (2015)	Historical	Spanish	Latin	Handwritten
Fink et al. (2014)	Historical	German	Latin	Handwritten
Chatbri et al. (2014)	Modern	French	Latin	Handwritten
Mondal et al. (2014;2015)	Historical	English, French	Latin	Handwritten, machine-printed
Sfikas et al. (2015)	Historical	Greek	Greek	Handwritten, machine-printed
Rodriguez-Serrano et al. (2012)	Historical, modern	English, French, Arabic	Latin, Arabic	Handwritten
Sudholt et al. (2016)	Historical, modern	English, Spanish, Arabic	Latin, Arabic	Handwritten
Leydier et al. (2009)	Historical	Middle English, Semitic, Chinese	Latin, Arabic, Chinese	Handwritten
Terasawa et al. (2009)	Historical	English, Japanese	Latin, Chinese	Handwritten
Abidi et al. (2012), Sagheer et al. (2010)	Historical	Urdu	Arabic	Handwritten
Khayyat et al. (2014), Li et al. (2014)	Modern	Farsi	Arabic	Handwritten
Kumar et al. (2014), Wshah et al. (2014)	Modern	English, Urdu, Hindi	Latin, Arabic, Devanagari	Handwritten
Srihari et al. (2008)	Modern	English, Urdu, Hindi	Latin, Arabic, Devanagari	Handwritten
Huang et al. (2013)	Modern	Chinese	Chinese	Handwritten
Giotis et al. (2014)	Modern	Greek	Greek	Handwritten
Saabni et al. (2012)	Modern	Arabic	Arabic	Handwritten
Shah et al. (2010)	Modern	Pashto	Arabic	Handwritten
Can et al.(2011), Rusinol et al. (2011)	Historical	English, Ottoman	Latin, Arabic	Handwritten machine-printed
Wei et al. (2011;2015)	Historical	Kanjur	Mongolian	Woodblock-printed
Ranjan et al. (2014), Li et al. (2007)	Modern	English	Latin	Machine-printed
Zagoris et al. (2010), Bai et al. (2009)	Modern	English	Latin	Machine-printed
Louloudis et al. (2012), Roy et al. (2011)	Historical	French	Latin	Machine-printed
Papandreou et al. (2014)	Historical	French	Latin	Machine-printed
Gatos et al. (2009)	Historical	German	Latin	Machine-printed
Sousa et al. (2007)	Historical	Portuguese	Latin	Machine-printed
Marinai (2011)	Historical	Latin	Latin	Machine-printed
Konidaris et al. (2007), Kesidis et al. (2011)	Historical	Greek	Greek	Machine-printed
Xia et al. (2011)	Historical	Chinese	Chinese	Machine-printed
Hassan et al. (2013), Krishnan et al. (2013)	Modern	English, Indian, Gujarati	Latin, Bangla, Devanagari	Machine-printed
Shekhar et al. (2013), Yalniz et al. (2012)	Modern	English, Indian	Latin, Telugu	Machine-printed
Meshesha et al.(2008)	Modern	English, Amharic, Hindi	Latin, Amharic, Devanagari	Machine-printed

1.6 Datasets

In order to evaluate the proposed word spotting methods, we conducted the experiments using three standard offline datasets of handwritten documents, which are widely used in word spotting methods, such as:

George Washington (GW) dataset (Lavrenko et al., 2004) : The GW dataset is a historical dataset consists of 20 pages from a collection of English letters written by George Washington and his associates in the year 1755. An exemplary document image is shown in Figure 1.7.

IAM dataset (Marti et al., 2002) : The IAM dataset is modern handwritten dataset consists of 1539 pages of handwritten modern English text from the Lancaster-Oslo/Bergen corpus (LOB) (Johansson et al., 1978), written by 657 writers. The sample document image is shown in Figure 1.8.

Bentham dataset (Long, 1981) : The Bentham dataset is modern English handwritten dataset consists of 50 high qualities handwritten documents written by Jeremy Bentham as well as fair copies written by Bentham's secretarial staff. Figure 1.9 shows the sample document image of Bentham dataset.

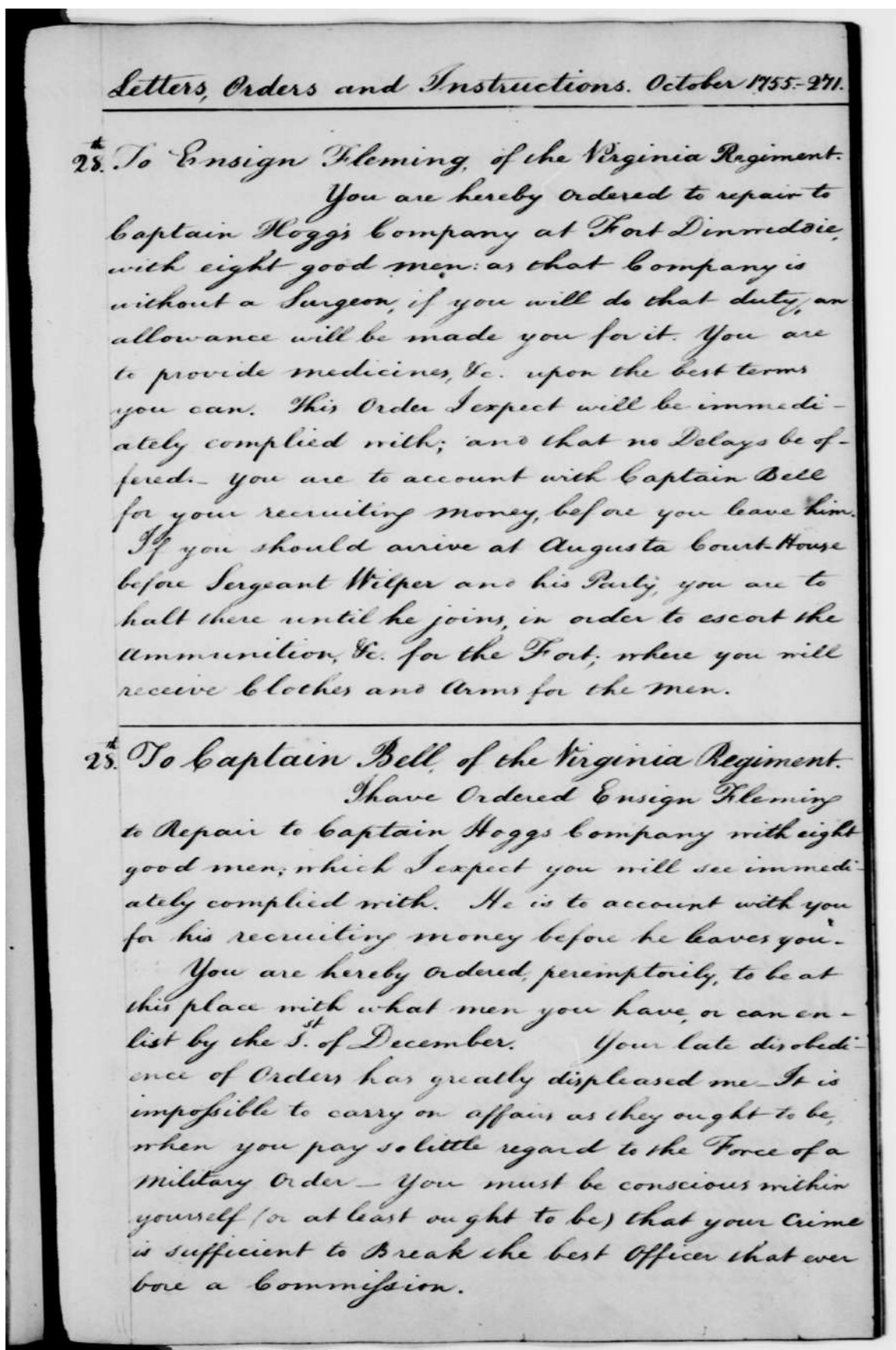


Figure 1.7: Sample handwritten document image from GW dataset

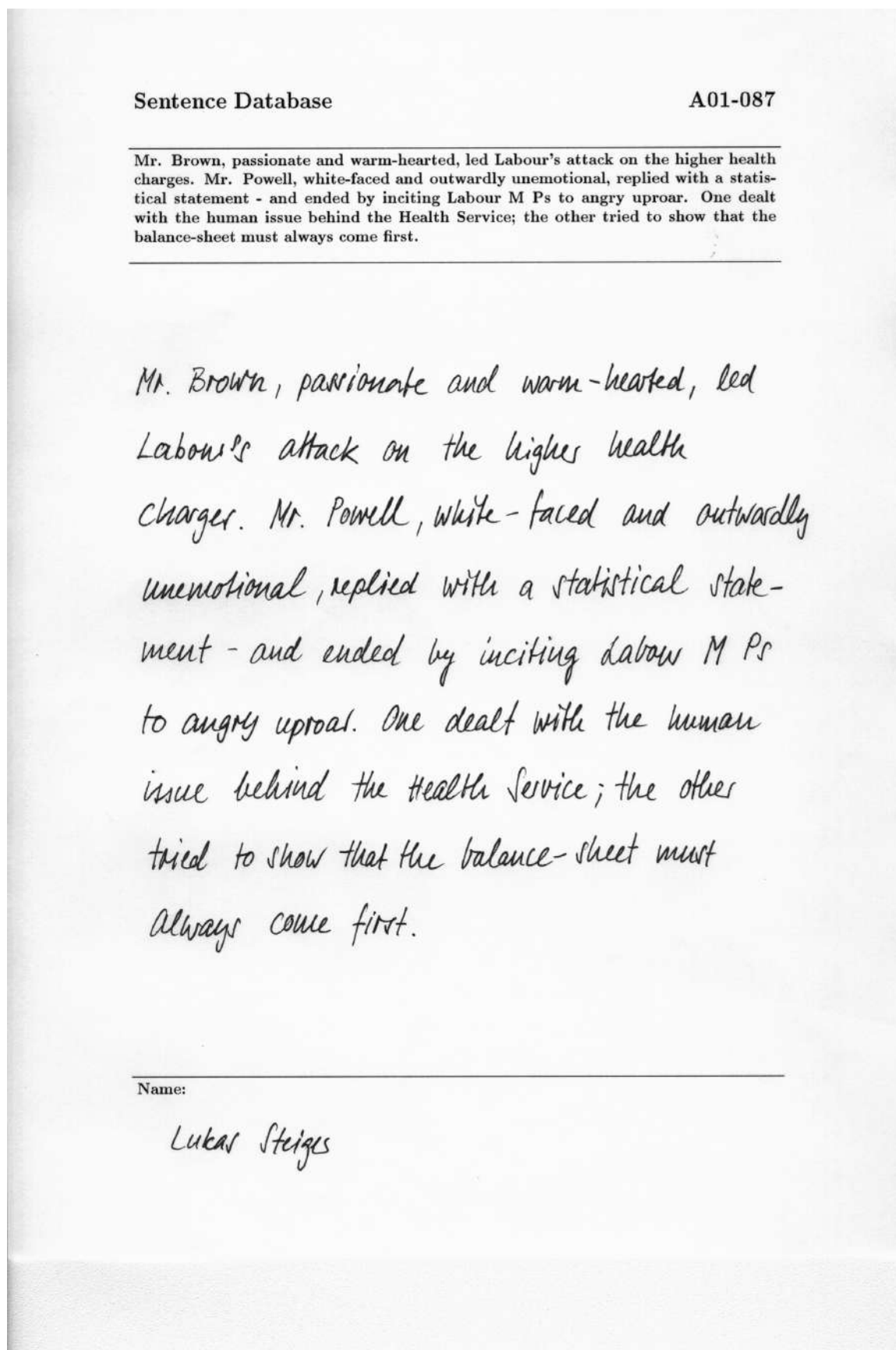


Figure 1.8: Sample handwritten document image from IAM dataset

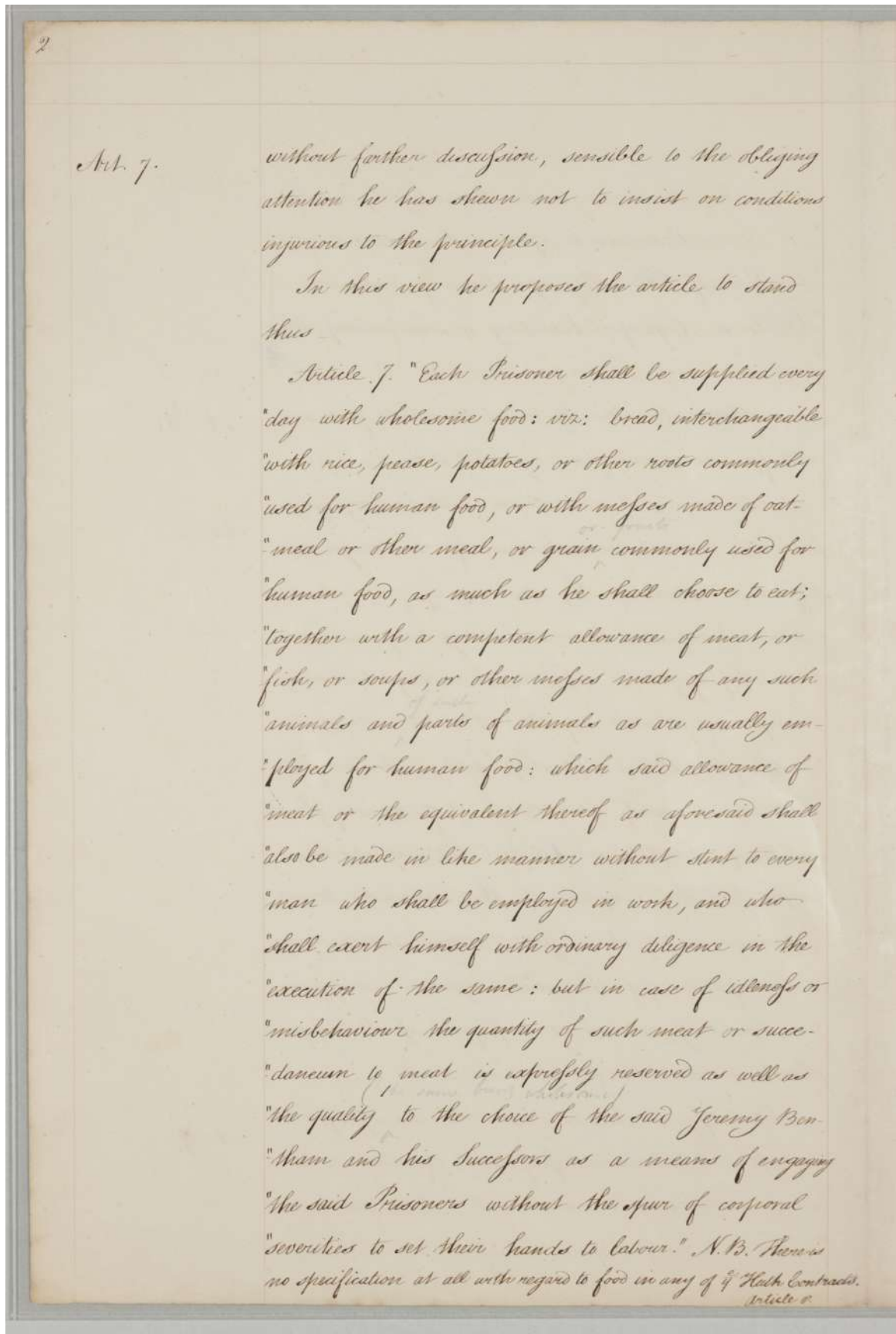


Figure 1.9: Sample handwritten document image from Bentham dataset

1.7 Evaluation Metrics

In order to measure the accuracy of proposed word spotting methods, we used popular metrics such as, Precision (P) and Recall (R). Precision gives the fraction of retrieved word instances that are relevant to the query. Recall gives the fraction of relevant word instances that are successfully retrieved. The Precision (P) and Recall (R) are defined as follows :

$$P = \frac{|\{\text{relavant instances}\} \cap |\{\text{retrieved instances}\}|}{|\{\text{retrieved instances}\}|} \quad (1.1)$$

$$R = \frac{|\{\text{relavant instances}\} \cap |\{\text{retrieved instances}\}|}{|\{\text{relavant instances}\}|} \quad (1.2)$$

The mean Average Precision (mAP) provides a mean of the average precision for all the query word images and it can be computed as

$$mAP = \frac{\sum_{q=1}^Q Ave P(q)}{Q}, \quad (1.3)$$

where, Q is the number of query word images.

We employed five-fold cross-validation method to validate our approaches. Each dataset is partitioned into five complementary subsets, among five subsets, four subsets are used for training and remaining one subset is used as validation (test) set. The cross-validation process is repeated five times, with each subset used exactly once for validation. To estimate the overall evaluation results, we compute the mean Average Precision (mAP) of the five validation results.

1.8 Organization of the Thesis

The thesis is organized into six chapters.

- Chapter 1 provides an overview of the general background and the problem setting. The motivation behind the research work and its methods are briefly described.

- In chapter 2, we present a segmentation based word spotting method for handwritten document images using Co-occurrence Histograms of Oriented Gradients (Co-HOG) descriptor.

- In chapter 3, we present a segmentation based word spotting method for handwritten documents using Scale Space Co-HOG descriptor. .

- In chapter 4, we present a segmentation based word spotting method for handwritten documents using Bag of Visual Words (BoVW) framework based on curvature features.

- In chapter 5, we present a segmentation free word spotting method for handwritten documents using Bag of Visual Words (BoVW) framework based on Co-HOG descriptor.

- In chapter 6, Conclusions and possible scopes for the work in future are presented.

Chapter 2

Co-HOG Descriptor for Word Spotting

A part of this chapter is published in International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Advance Academic Publisher, vol. 7, issue 6, page 35-40, 2017.

Chapter 2

Co-HOG Descriptor for Word Spotting

In this chapter, we present a segmentation based word spotting method for handwritten document images using Co-occurrence Histograms of Oriented Gradients (Co-HOG) descriptor. In order to construct Co-HOG descriptor, initially, we divide a word image into blocks, then, from each block, we extract Co-HOG descriptor and concatenate them to form a feature descriptor, which represents a word image in the compact form. Finally, Dynamic Time Warping(DTW) technique is used to retrieve word image instances similar to a query image.

2.1 Introduction

The state-of-the-art word spotting methods for handwritten documents can be categorized into two groups, segmentation based methods and segmentation free methods. In segmentation based methods, the sequences of operations are applied to the document images. First, the document is pre-processed based on text layout analysis; the handwritten document is segmented into word images. Then, extract the feature descriptor from the segmented word image. Based on this feature descriptor, a distance measure is used to measure the similarity between the query word image and the segmented word image.

In segmentation based word spotting methods, extraction of feature descriptors is one of the most important steps for achieving high retrieval performance, because of feature descriptor with strong discriminative information can be well classified even using with the simplest classifier. The literature survey reveals that, Histogram of Oriented Gradients (HOG) descriptor is widely used in several recognition applications because of its discriminating ability compared to other existing feature descriptors.

Rodriguez et al. (2008) have proposed local gradient histogram features for word spotting in unconstrained handwritten documents. A sliding window moves from left to right over a word image. At each position, the window is subdivided into cells, and in each cell, a histogram of orientations is accumulated. Slit style HOG features for handwritten document image word spotting is proposed by Terasawa et al. (2009). Newell et al. (2011) have extended the HOG descriptor to include features at multiple scales for character recognition. Saidani et al. (2015) have proposed a novel approach for Arabic and Latin script identification based on HOG feature descriptors. HOG is first applied at word level based on writing orientation analysis. Then, they are extended to word image partitions to capture fine and discriminating details. The unsupervised segmentation-free HOG based word spotting method was proposed by Almazan et al. (2014). Documents are represented by a grid of HOG descriptors, and a sliding-window approach is used to locate the document regions that are most similar to the query.

The appearance and shape of the local object in an image is represented through the distribution of the local intensity gradient orientation and edge direction without requiring equivalent gradient and edge positions (Carcagn et al., 2015). This orientation analysis is robust to lighting changes since the histogram provides translational invariance. The HOG feature descriptor summarizes the distribution of measurements within the image regions. When extracting HOG features, the orientations of gradients are usually quantized into histogram bins and each bin has an orientation range. A histogram of oriented gradients falling into each bin is computed and then normalized to overcome the illumination variation. The orientation of gradients from all blocks are then concatenated together to form a feature descriptor of the whole image.

HOG feature descriptor captures orientation of only isolated pixels, whereas spatial information of neighboring pixels is ignored. Co-occurrence Histogram of Oriented Gradients (Co-HOG) (Watanabe et al., 2009) feature descriptor is an extension of the original HOG feature descriptor that captures the spatial information of neighboring pixels. Instead of counting the occurrence of the gradient orientation of a single pixel, gradient orientations of two or more neighboring

pixels are considered. For each pixel in an image block, the gradient orientations of the pixel pair formed by its neighbor and itself are examined.

Co-occurrence Histogram of Oriented Gradients is dominant feature descriptor widely used in object detection because Co-HOG feature descriptor accurately represents significant characteristics of the object structure. At the same time, it is more efficient compared with HOG and therefore more suitable for real-time applications. In this chapter, we propose a segmentation based word spotting method for handwritten document images using Co-HOG feature descriptors. Initially, we divide a word image into the number of blocks, then, Co-HOG feature descriptors are extracted from each block. Then, we concatenate Co-HOG feature descriptors of each block, which represent for word image in the compact form. Finally, using DTW matching technique, we retrieve word images that are similar to given query image.

2.2 Histogram of Oriented Gradients (HOG)

The Histogram of Oriented Gradients descriptor is developed by Dalal et al. (2005) for human detection. The HOG has been successfully applied in many research fields such as word spotting task (Rodriguez et al., 2008; Terasawa et al., 2009), body parts detection (Corvee et al., 2010), face recognition (Deniz et al., 2011; Shu et al., 2011), character recognition (Newell et al., 2011), text/non-text classification problem (Minetto et al., 2013) and detection of vehicles in traffic videos (Arrospide et al., 2013).

The HOG computes a histogram of gradient orientation in a local region of an image I . One of the significant differences between SIFT and HOG is that HOG normalizes the histograms in overlapping local blocks and makes the redundant expression. And, another difference is that SIFT describes the scale and orientation normalized image patch around the detected key point, while HOG is computed in a grid of regular window without scale or normalization.

In order to extract HOG feature, initially, gradient orientation of every pixel is calculated as:

$$\theta = \arctan\left(\frac{I_y}{I_x}\right), \quad (2.1)$$

where, $\arctan(I_y/I_x)$ returns the inverse tangent of the elements in degrees. I_y and I_x are vertical and horizontal gradient respectively calculated by Gaussian filter. Then, histogram of each orientation in a small rectangular region is calculated. The orientation of pixels is quantized to N bins (N_{bin}) and a histogram of orientation is calculated at each bin as follows:

$$H(i) = \sum_{x,y \in I, G_m(x,y)} G_g(x,y), \quad i = 1, 2, 3, \dots, n, \quad (2.2)$$

where, G_m and G_g is magnitude and gradient orientation at (x, y) respectively. Finally, HOG feature descriptor is constructed by concatenating the histogram $H(i)$ for all small regions. The Figure 2.1 shows intermediate results for extraction of HOG feature descriptors from a word image.

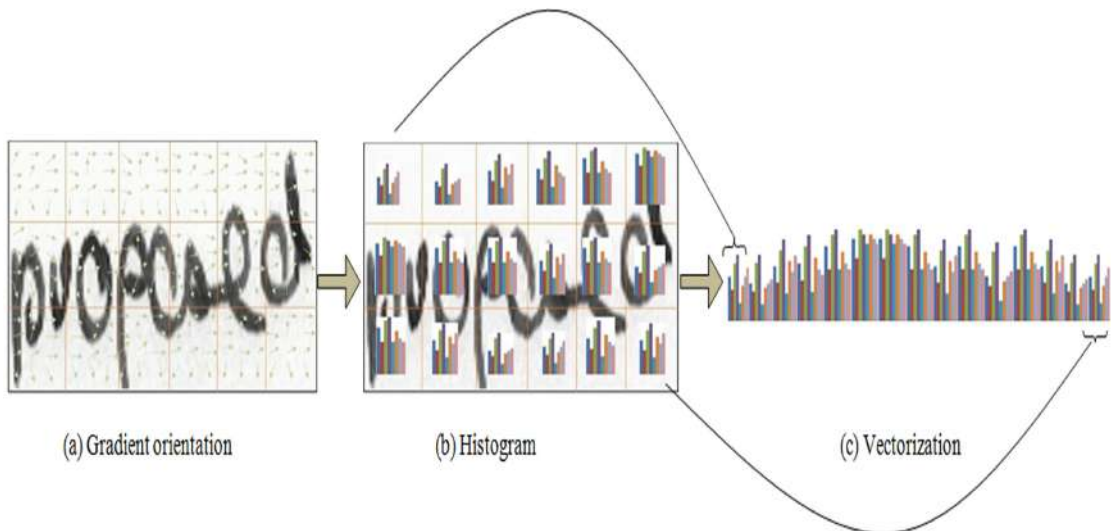


Figure 2.1: Intermediate results for extraction of HOG feature descriptor from word image

2.3 Co-HOG Descriptor

Co-HOG feature descriptor is robust against deformation and illumination variation because it is gradient based histogram feature descriptor. The pedestrian detection method is proposed by Watanabe et al. (2009) based on extraction of Co-HOG feature descriptors. Ren et al. (2010) have proposed object detection method using Co-HOG features with variable location and variable size blocks which captures the characteristics of object structure. Face recognition using weighted Co-HOG feature descriptor is proposed by Do (2012). The weighted voting Co-HOG features for scene text recognition are proposed by Tian et al. (2013). The character recognition in natural scenes using Convolutional Co-HOG feature descriptors are proposed by Su et al. (2014) and multilingual scene character recognition using Co-HOG and Convolutional Co-HOG feature descriptors are proposed by Tian et al.(2015).

Co-HOG capture spatial information by counting frequency of co-occurrence of oriented gradients between pixel pairs. For each pixel in an image block, the gradient orientations of the pixel pair formed by its neighbor and itself are examined. Since, the pair of gradient orientations has more lexis than single one because single gradient orientation has only eight varieties (Figure 2.2(a)) and in a pair of them has much more than single orientation as shown in Figure 2.2(b).

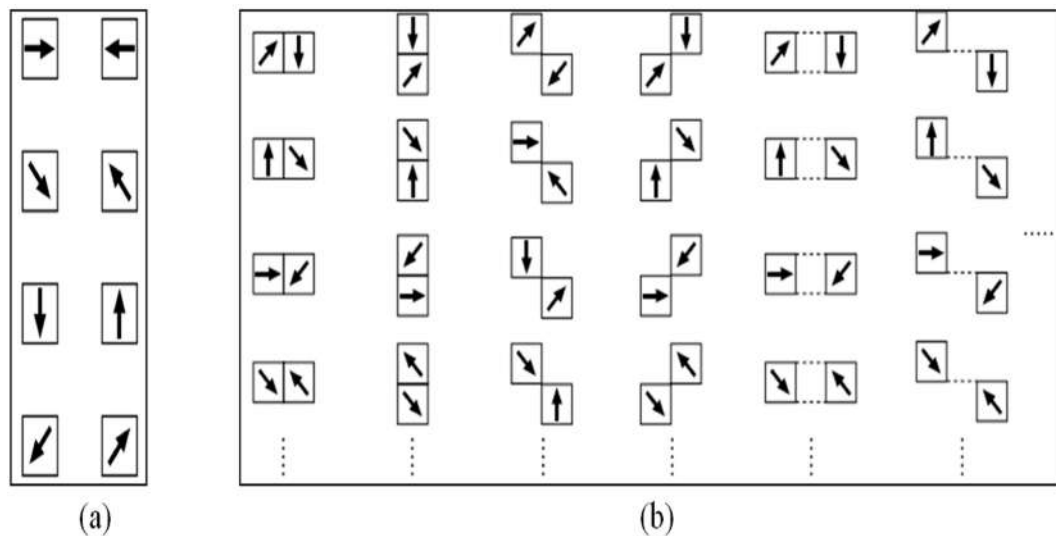


Figure 2.2: Lexis of gradient orientation:(a) single gradient orientation (b) a pair of gradient orientation. (Courtesy: Watanabe et al., 2009)

In order to extract Co-HOG feature descriptors from an image, firstly, gradient orientations at every pixel are calculated using Eq.(2.1). In particular, the gradient image G of an input image I is computed. Then, each pixel in a gradient image, the gradient orientations of the pixel pair formed with its neighbor, and itself are examined. The relative locations are reflected by the offset between two pixels as shown in Figure 2.3(a). The pixel at the center is the pixel under study and the neighboring ones are pixels with different offsets. Each neighboring pixel forms an orientation pair with the center pixel and accordingly votes to the co-occurrence matrix as illustrated in Figure 2.3(b).

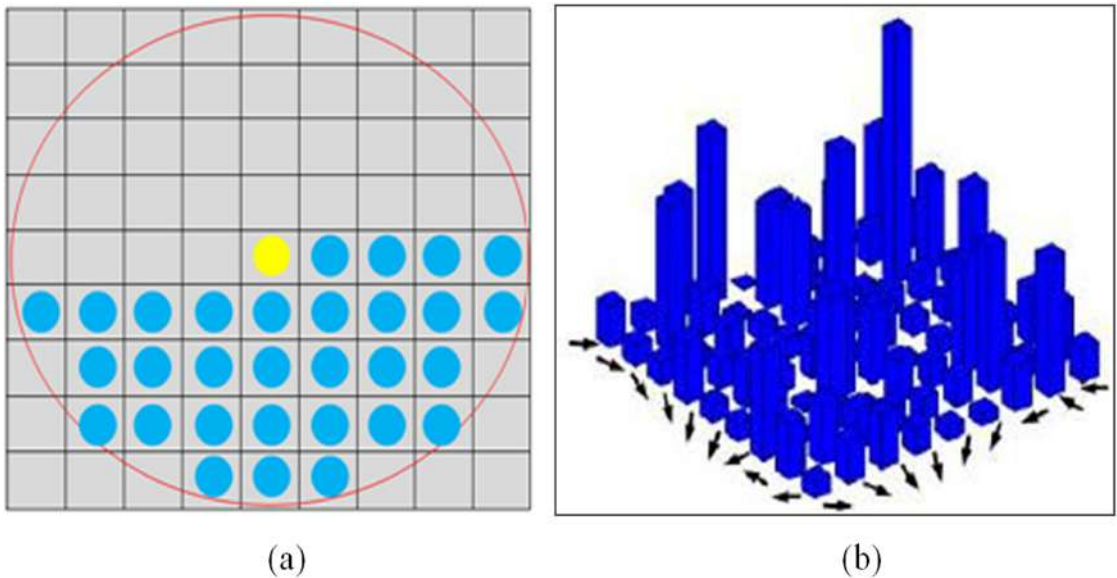


Figure 2.3: Extraction of Co-HOG feature descriptors: (a) offset used in Co-HOG, (b) Co-occurrence matrix for the given offset. (Courtesy: Tian et al., 2016)

We compute co-occurrence matrix K over an gradient image G of size $N \times M$ at an offset (x, y) as follows:

$$K_{x,y}(i, j) = \sum_{p=1}^N \sum_{q=1}^M \begin{cases} 1 & \text{if } G(p, q) = i \text{ and } G(p + x, q + y) = j, \\ 0 & \text{otherwise.} \end{cases} \quad (2.3)$$

where, $K_{x,y}$ is the co-occurrence matrix at an offset (x, y) , which is a square matrix and its dimension is decided by the number of orientation bins (N_{bin}).

In our experiments, the gradient orientation interval $[0, 360^0]$ is divided into 8 orientations per 45^0 . Therefore, the dimension of co-occurrence matrix is 8×8 . The maximum offset is set to 4 as in Figure 2.3(a), it will give rise to 31 co-occurrence matrices. For zero offset, co-occurrence matrix size is 1×8 (only eight effective values) because non-diagonal components are zero and the co-occurrence matrix size for other offsets is 8×8 . One such co-occurrence matrix is shown in Figure 2.3(b).

2.4 Proposed Method

The handwritten word image is having more local variation within the class when compared to variation between other classes because of variation in the writing style of different writers. Hence, in order to capture distinct local shape information within the class, we divide the word image into a number of blocks of equal size. The advantages of dividing an image into blocks are that the feature descriptor can express local and global shapes in detail and decreases the space computation complexities for feature extraction, and to utilize the distinct location information from each block. Therefore, we divide an image into blocks and extract Co-HOG features from each block. Then, extracted Co-HOG features are concatenated to form a feature descriptor of a word image. In the testing phase, the extracted Co-HOG feature descriptors of the training set are matched with feature descriptor of a query word image in order to retrieve word images similar to query image using a DTW matching technique. The Figure 2.4 shows flow diagram of the proposed method.

2.4.1 Extraction of Co-HOG Descriptor

In order to extract Co-HOG feature descriptors, we divide an image into non-overlapping blocks of dimension $Block_h \times Block_w$ and co-occurrence matrices are computed for each block using Eq.(2.3). The Figure 2.5(b) shows word image which is divided into non-overlapping blocks and corresponding co-occurrence

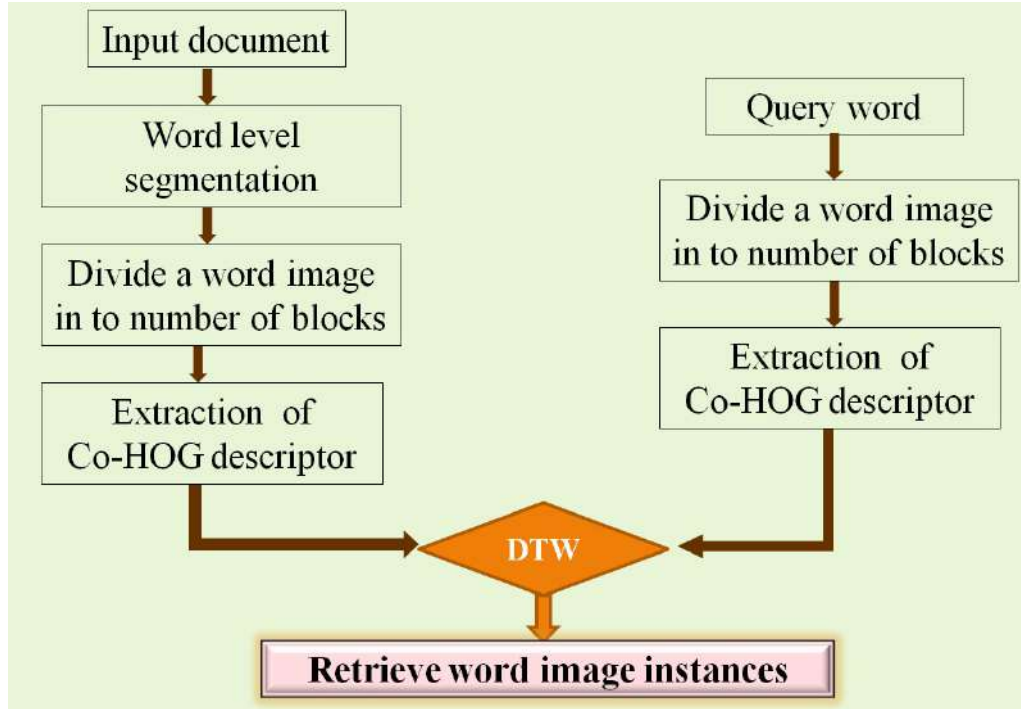


Figure 2.4: Flow diagram of proposed method

matrices. Then, Co-HOG feature descriptor of the scale image is constructed by vectorizing and concatenating the co-occurrence matrices of all blocks. Finally, the components of all the co-occurrence matrices are concatenated to form Co-HOG feature descriptor which represents the word image (Figure 2.5(c)). For example, when the word image is divided into (3×6) blocks, the dimension of Co-HOG feature vector is computed by $(N_{bin} \times N_{bin} \times 31)(Block_h \times Block_w)$ i.e. $(8 \times 8 \times 30 + 8)(3 \times 6) = 34,704$.

2.4.2 Word Image Matching

The literature survey and experimental based verification reveals that Dynamic Time Warping (DTW) (Rath et al., 2007) matching algorithm yields highest accuracy for handwritten word spotting compared to other matching algorithms. Hence, for the word spotting task, we compare Co-HOG feature vectors of the training set with the Co-HOG feature vector of a query word image using a DTW matching algorithm to retrieve the word instances similar to the query word. Consider two word images X and Y of each dimension m . They are represented

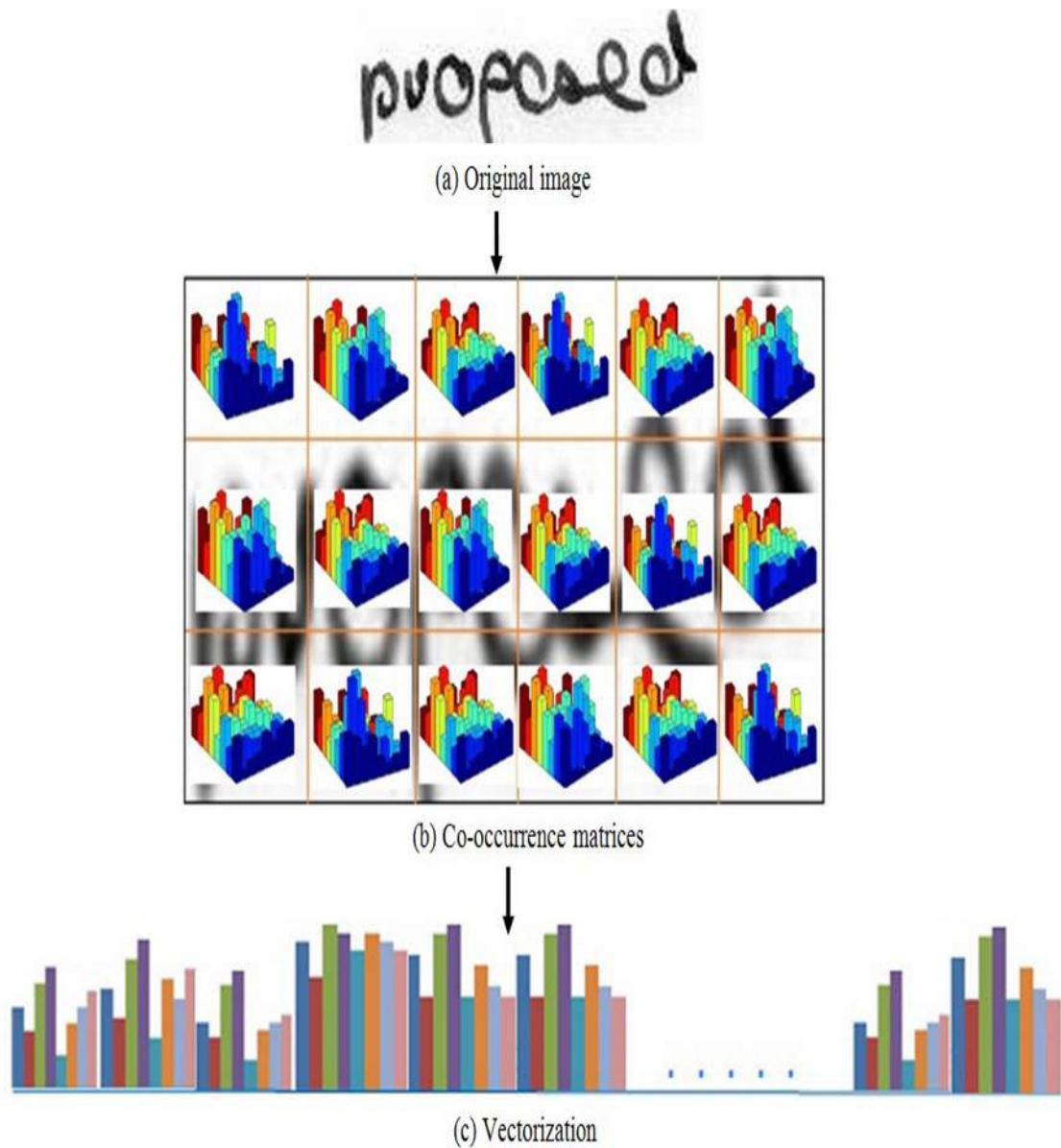


Figure 2.5: Illustration of Co-HOG feature descriptor extraction process: (a) Sample word image (b) Word image divided in to 3×6 blocks and corresponding co-occurrence matrices of each block and (c) concatenated one after another to form a Co-HOG feature vector

by vector $X = (x_1, \dots, x_m)$ and $Y = (y_1, \dots, y_m)$. To determine the DTW distance between these two vectors, a matrix D of dimension $m \times m$ is built where each element $D(i, j)$ is the cost of matching the two vectors (x_1, \dots, x_m) and (y_1, \dots, y_m) . Element $D(i, j)$ in the matrix is calculated in the following manner:

$$D(i, j) = \min \left\{ \begin{array}{l} D(i, j - 1) \\ D(i - 1, j) \\ D(i - 1, j - 1) \end{array} \right\} + d(x_i, y_j), \quad (2.4)$$

where, $1 < i \leq m$ and $1 < j \leq m$.

In order to use DTW to match such feature vectors X and Y , we need to define a distance measure d for the local distance between two sequence samples, x_i and y_j . We used the Euclidean distance measure and it is defined as:

$$d(x_i, y_j) = \sqrt{\sum_{k=1}^m (x_{i,k} - y_{j,k})^2}, \quad (2.5)$$

where, index k is used to access the elements from the vectors x_i and y_j .

2.5 Experimental Results

For evaluating the proposed word spotting method, two popular datasets are used such as George Washington (GW) dataset (Lavrenko et al., 2004) and the IAM dataset (Marti et al., 2002). In order to evaluate the performance of our approach, we conducted the experiments and results are evaluated based on popular metric such as *mAP*.

2.5.1 Pre-Processing

The document images of GW and IAM datasets are unconstrained and therefore documents contain different writing styles, artifacts and other types of noise.

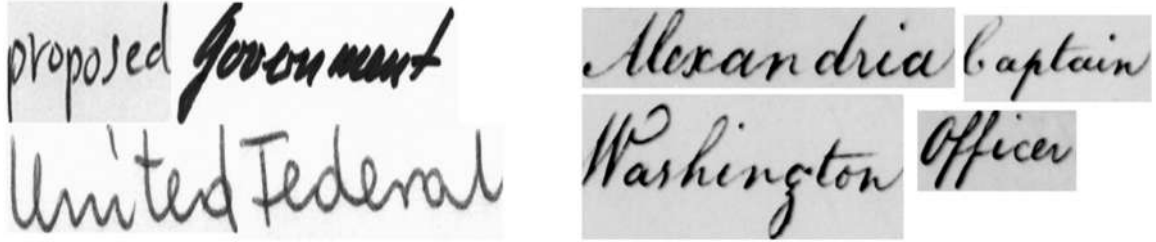


Figure 2.6: Sample results of denoised and segmented word images

In preprocessing step, the handwritten document images are denoised and segmented into individual text lines, then words using the adjacent connected components technique (Papavassiliou et al., 2010). The sample results of denoised and segmented word images are shown in Figure 2.6.

2.5.2 Parameter Selection

One of the important factors to achieve the highest accuracy is an optimal number of blocks of an image. Hence, in order to find an optimal number of blocks, we conducted the experiments using GW (336 word instances from all 20 pages) and IAM dataset (200 word instances from all 1539 pages) for a different number of blocks such as 1×2 (2 blocks), 2×4 (8 blocks), 3×6 (18 blocks) and 4×8 (32 blocks).

Table 2.1: *mAP* of our approach for different number of blocks

Number of blocks	mAP(%)	
	For GW dataset	For IAM dataset
1×2 (2 blocks)	86.42	79.64
2×4 (8 blocks)	88.67	82.52
3×6 (18 blocks)	91.03	88.41
4×8 (32 blocks)	87.22	86.43

The *mAP* obtained on two datasets using our approach for a different number blocks is presented in Table 2.1 and its pictorial representation is shown in the Figure 2.7. It is observed that when the word image is divided into 3×6 blocks (18 blocks), the accuracy of our approach is significantly better than for the other

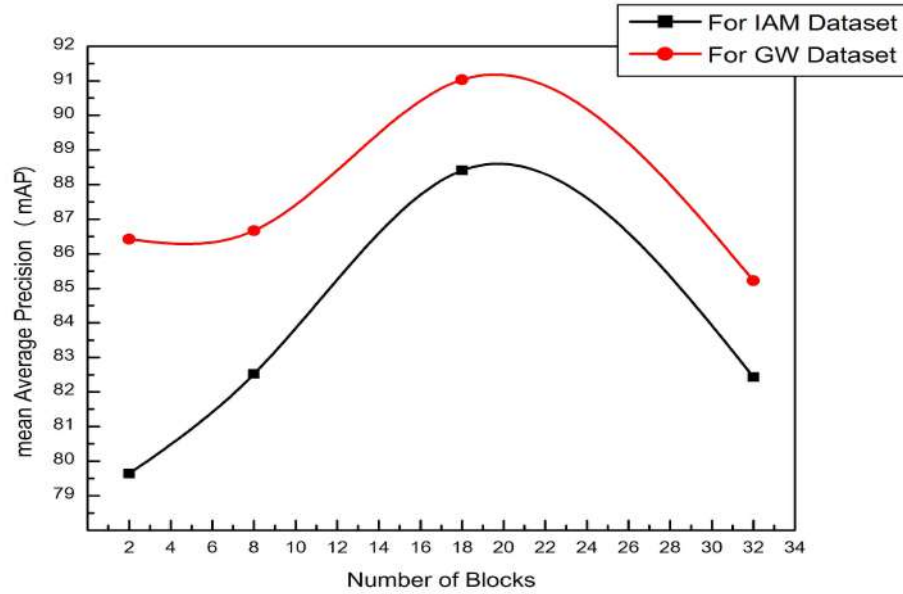


Figure 2.7: The comparison of mAP of our approach for different number of blocks


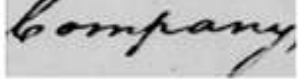
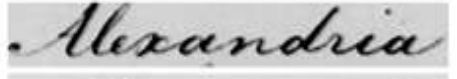
number of blocks. Hence, in all the experiments, we divide the word image into 3×6 blocks (18 blocks) and co-occurrence matrices are extracted from each block.

2.5.3 Experiments on GW dataset

The most important challenge for word spotting using this dataset is to deal with a small amount of training data, because this dataset consists of twenty pages. From this dataset, we have taken manually labeled 1680 word instances of 28 classes of words from all the 20 pages of scanned handwritten documents ($28 \text{ classes} \times 20 \text{ pages} \times 3 \text{ instances} = 1680$). Each class of words may occur at least three times per document. While performing five-fold cross-validation, we partitioned the dataset into five disjoint subsets and each subset consists of 336 word images (from each class of words we have taken 12 word images). In the training stage, 1344 word images are used and remaining 336 word images are used for testing. Hence, for every experiment, we have 336 word images as test samples.

It is noticed that our approach achieves the highest accuracy (%) with mAP is 91.03 using GW dataset. Table 2.2 shows the sample retrieved word instances for a given query. The first row in Table 2.2 shows query word images and each column displays corresponding retrieved word instances.

Table 2.2: Sample retrieval results: Query word images (first row) and their retrieval results of our approach for GW dataset

We used a GW dataset to compare the performance of our approach with existing word spotting methods such as (a) word profile features based approach proposed by Rath et al. (2007), (b) HOG based approach by extracting local gradient histogram features through the sliding window proposed by Rodriguez et al. (2008) and (c) Fisher vector representation computed over SIFT descriptors from the word image proposed by Almazan et al. (2013). We implemented existing methods and validated using five-fold cross-validation method. To estimate the overall evaluation results, we have taken the mAP . Table 2.3 shows the comparison results of our approach with existing methods. It is observed that, compared to existing methods, our approach yields the highest accuracy (mAP) of 91.03%.

Table 2.3: Performance comparison of our approach with existing word spotting methods for GW dataset

Method	Features	Segmentation	mAP (%)
Rath et al. (2007)	Word profile	Word	73.71
Rodriguez et al. (2008)	HOG	Word	71.70
Almazan et al. (2013)	SIFT	Word	85.63
Proposed method	Co-HOG	Word	91.03

2.5.4 Experiments on IAM dataset

From the IAM dataset, we have taken manually labeled 1000 word images of 20 different classes i.e., each class of word image has 50 instances of different writing styles. While performing five-fold cross-validation, we partitioned the dataset into five disjoint subsets and each subset consists of 200 word images (from each class of words we have taken 10 word images). In the training stage, 800 word images are used and remaining 200 word images are used for testing. Hence, for every experiment, we have 200 word images as test samples.

Table 2.4 shows the sample retrieved word instances of our approach for the IAM dataset. The first row shows query word and corresponding retrieved word instances of a given query word are shown. The results show the robustness of our approach for different kinds of words with variation in writing style. To estimate the overall evaluation results of our approach, we have taken the mAP of the five validation results. It is observed that our approach yields the highest accuracy (%) with mAP is 88.41.

We compared the performance of our approach with existing word spotting methods using IAM dataset. The existing methods considered for comparison are (a) HOG based approach by extracting local gradient histogram features through the sliding window proposed by Rodriguez et al. (2008) and (b) word profile based approach proposed by Rath et al. (2007). The existing methods are implemented and five-fold cross-validation method is used to validate based on mAP . Table 2.5 presents results comparison of our approach with existing methods based on mAP . From the experimental results, it is notified that our approach outperforms

Table 2.4: Sample retrieval results: Query word images (first row) and their retrieval results of our approach for IAM dataset

existing methods for IAM dataset.

Table 2.5: Performance comparison of our approach with existing methods for IAM dataset

Method	Features	Segmentation	mAP (%)
Rath et al. (2007)	Word profile	Word	73.71
Rodriguez et al. (2008)	HOG	Word	71.70
Almazan et al. (2013)	SIFT	Word	85.63
Proposed method	Co-HOG	Word	88.41

Based on the experimental results obtained on GW and IAM dataset, we can conclude that our approach efficiently retrieves the handwritten words which are having non uniform illumination, suffering from the noise and written by different writers. The highest accuracy of our approach is due to the extraction of illumination and scale invariant Co-HOG feature descriptor at multiple blocks of an image. Another factor for highest accuracy is a division of the word image into blocks which capture the local variations within the class. It is identified through evaluation of experimental results on GW and IAM dataset that our

approach outperforms existing word profile, HOG and SIFT based word spotting methods because Co-HOG is more robust and discriminating than word profile, SIFT, HOG because it considers gradient orientations of two or more neighboring pixels.

2.6 Chapter Summary

In this chapter, we proposed a segmentation based word spotting method for handwritten document images using Co-HOG feature descriptors. The Co-HOG feature descriptor is designed to capture the local spatial information by counting the frequency of co-occurrence of gradient orientation of neighboring pixel pairs. Since the existing HOG feature descriptor captures orientation of each pixel in an isolated manner and does not take into account the spatial information described by their neighboring pixels. Co-HOG captures the character shape information more precisely and robustness to illumination variation and invariance to local geometric transformation.

Chapter 3

Scale Space Co-HOG Descriptor for Word Spotting

A part of this chapter is published in International Journal of Computer Vision and Image Processing (IJCVIP), IGI Global Publisher, vol. 6, issue 2, pages 71-86, 2016.

Chapter 3

Scale Space Co-HOG Descriptor for Word Spotting

In this chapter, we proposed a Scale Space Co-occurrence Histograms of Oriented Gradients method (SS Co-HOG) for retrieving words from handwritten documents. We employed three scale representation of an image and at each scale, we divide the word image into blocks and Co-HOG features are extracted from each block and finally concatenate them to form a feature descriptor. Experiments on GW and IAM handwritten datasets demonstrates the effectiveness of the proposed method.

3.1 Introduction

In the preceding chapter, we proposed word spotting method using Co-HOG feature descriptor which is extracted at a fixed scale or uni scale. The holistic shapes such as whole words and simple shapes such as characters are extracted more effectively at coarser scales. The extracted shape features of the whole word at coarser scales are potentially more resistant to variation occurring in a different writing style. At finer scales, it is convenient to extract the local information about image gradient and stroke orientation of the words which is more convenient for word image contour description. Hence, in order to represent the word image in multi scale, we derive images from original word image, which yields scale space representation of a word image. Integrating scale space representation with Co-HOG feature descriptor helps to extract sufficient information from handwritten word image contour. Hence, in this chapter, we proposed to extract Co-HOG feature descriptor of the word image at multiple scales and these feature descriptors are called as Scale Space Co-HOG (SS Co-HOG) feature descriptor.

3.2 Proposed Work

In order to represent the word image in multi scale, we derive images from the original word image based on Gaussian convolution operation based on Gaussian convolution operation. We employed three scale representation of an image and at each scale, we divide the word image into blocks and Co-HOG features are extracted from each block and finally concatenate them to form a feature descriptor. For the word spotting task, we compare Co-HOG feature vectors of the training set with the Co-HOG feature vector of a query word image using a DTW matching algorithm to retrieve the word instances similar to the query word.

3.2.1 Scale Space Representation

The scale space representation of an image is an embedding of the original image into a family of the derived images constructed by convolution with kernels of increasing width. Scale space representation is a special kind of multiscale that comprises a continuous scale parameter and preserves the similar spatial sampling at all scales (Lindeberg et al., 1994).

Let I represent an original image. Then, the scale space representation of I can be defined as

$$L(\sigma, I) = I * G(\sigma), \quad (3.1)$$

where, $G(\sigma)$ is a Gaussian kernel with variance σ and $(*)$ is convolution operation. The width of the Gaussian kernel is controlled by σ . Larger the value of σ makes the Gaussian kernel G wider and smooth the original image more significant. So using a larger value of σ , the convolve operation gets a coarser scale version of the original image. Therefore, different values of σ indicate different scales. It is verified that the width of the Gaussian kernel σ preserves the similar spatial sampling at three scale parameters such as $\sigma=0$, $\sigma=1$ and $\sigma=2$. Hence, in all the experiments, we derive three images from original word image based on Gaussian convolution operation using these three different values of σ . The Figure 3.1(b) shows a scale space representation of a word image.

3.2.2 Extraction of Scale Space Co-HOG Descriptors

After representing word image into multiple scales, for each scale, we divide the word image into a number of blocks of equal size. For each scale, we divide an image into $Block_h \times Block_w$ non-overlapping blocks and co-occurrence matrices are computed for each block using Eq.(2.3). The Figure 3.1(c) shows word image which is divided into non-overlapping blocks and corresponding co-occurrence matrices. Then, Co-HOG feature vector of the scale image is constructed by vectorizing and concatenating the co-occurrence matrices of all blocks. For example, when the word image is divided into 3×6 blocks, the dimension of Co-HOG feature vector is $(8 \times 8 \times 30 + 8)(3 \times 6) = 34,704$. Finally, Scale space Co-HOG feature vector of the word image is constructed by consecutively arranging Co-HOG feature vectors of all three scales (Figure 3.1 (d)). Thus, the dimension of the Scale space Co-HOG feature vector of a word image is $34,704 \times 3 = 104112$.

3.3 Experimental Results

For evaluating the proposed word spotting algorithm, two datasets are used such as GW dataset (Lavrenko et al., 2004) and the IAM dataset (Marti et al., 2002). In order to evaluate the performance of our approach, we conducted the experiments and results are evaluated based on popular metrics such as Precision (P). We employed five-fold cross-validation method to validate our approach, each dataset is partitioned into five complementary subsets i.e. four subsets are used for training and remaining one subset is used as validation (test) set. The cross-validation process is repeated five times, with each subset used exactly once for validation. To estimate the overall evaluation results of our approach, we compute the mAP of the five validation results.

3.3.1 Pre-Processing

The document images of GW and IAM datasets are unconstrained and therefore documents contain different writing styles, artifacts and other types of noise.

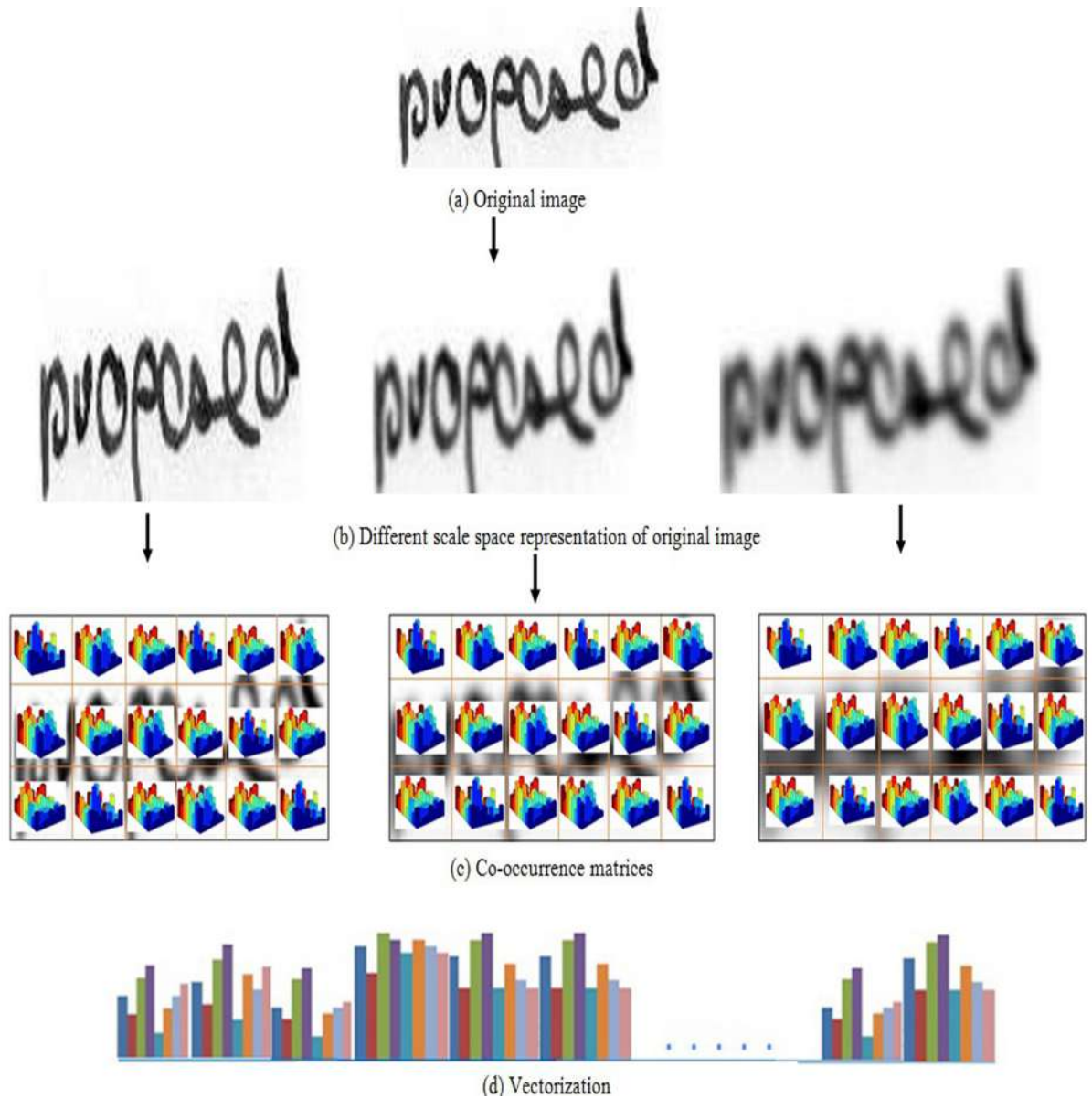


Figure 3.1: Illustration of Scale Space Co-HOG feature descriptor extraction process: (a) sample word image (b) scales at $\sigma = 0$ (original image), $\sigma = 1$ and $\sigma = 2$ respectively. (c) word image divided into blocks and corresponding co-occurrence matrices of each block and (d) concatenate one after another to form a Scale Space Co-HOG feature vector

In preprocessing step, the handwritten document images are denoised and segmented into individual text lines, then words using the adjacent connected components technique (Papavassiliou et al., 2010). The sample results of denoised and segmented word images are shown in Figure 3.2.

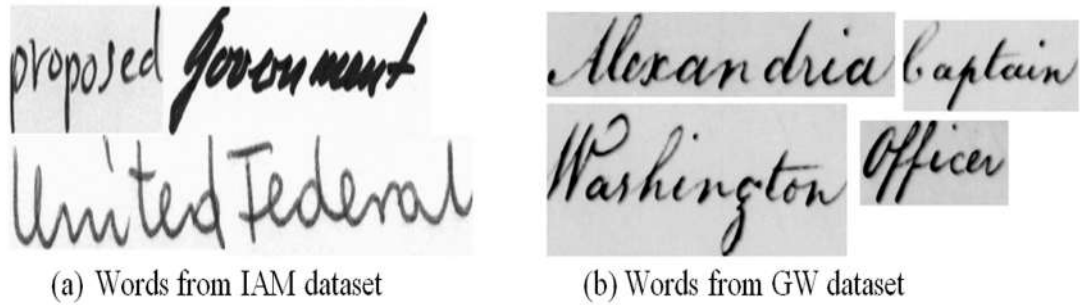


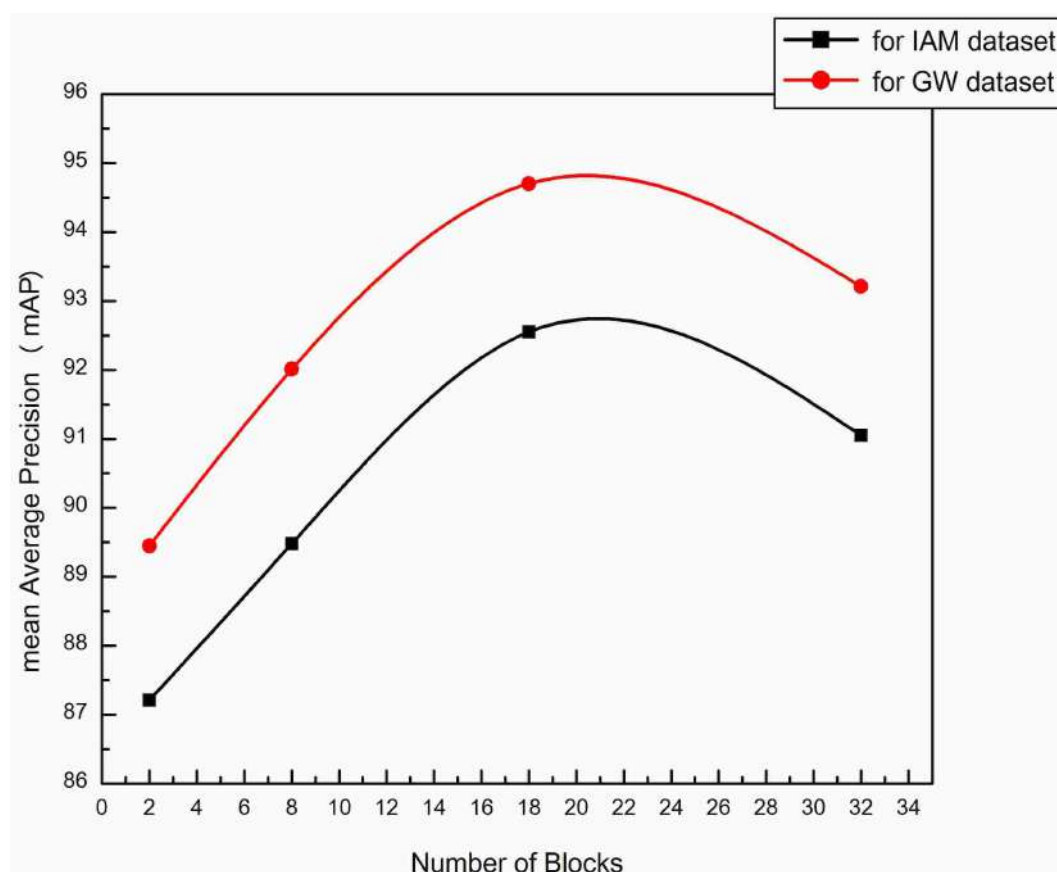
Figure 3.2: Sample results of denoised and segmented word images

3.3.2 Parameter Selection

For proposed approach, one of the important factors to achieve the highest accuracy is an optimal number of blocks of an image. Hence, in order to find an optimal number of blocks, we conducted the experiments using GW (336 word instances from 20 pages) and IAM (200 word instances from 1539 pages) dataset for a different number of blocks such as 1×2 (2 blocks), 2×4 (8 blocks), 3×6 (18 blocks) and 4×8 (32 blocks). The *mAP* obtained on two datasets using our approach for a different number blocks is presented in Table 3.1 and its pictorial representation is shown in the Figure 3.3. It is observed that when the word image is divided into 3×6 blocks (18 blocks), the accuracy of our approach is significantly better than for the other number of blocks. Hence, in all the experiments, we divide the word image into 3×6 blocks (18 blocks) and co-occurrence matrices are extracted from each block.

Table 3.1: mAP of our approach for different number of blocks

Number of blocks	mAP(%)	
	For GW dataset	For IAM dataset
1×2 (2 blocks)	89.45	87.21
2×4 (8 blocks)	92.01	89.48
3×6 (18 blocks)	94.70	92.55
4×8 (32 blocks)	93.21	91.05

Figure 3.3: The comparison of mAP of our approach for different number of blocks

3.3.3 Experiments on GW dataset

From this dataset, we have taken manually labeled 1680 word instances of 28 classes of words from all the 20 pages of scanned handwritten documents ($28 \text{ classes} \times 20 \text{ pages} \times 3 \text{ instances} = 1680 \text{ words}$). Each class of words may occur at least three times per document. While performing five-fold cross-validation, we partitioned the dataset into five disjoint subsets and each subset consists of 336 word images (from each class of words we have taken 12 word images). In the training stage, 1344 word images are used and remaining 336 word images are used for testing. Hence, for every experiment, we have 336 word images as test samples. It is noticed that our approach achieves the highest accuracy (%) with mAP is 94.70. Table 3.2 shows the retrieved word instances for a given query. The first row in Table 3.2 shows query word images and each column displays corresponding retrieved word instances.

Table 3.2: Sample retrieval results: Query word images (first row) and their retrieval results of our approach for GW dataset

We used a GW dataset to compare the performance of our approach with existing word spotting methods such as Rath et al. (2007), Rodriguez et al. (2008), and Almazan et al. (2013). We also compared the proposed method with preceding chapter results, where word image is represented as Co-HOG feature descriptor alone. The existing methods are implemented and five-fold

cross-validation method is used to validate based on *mAP*. Table 3.3 shows the comparison results of our approach with existing methods. It is observed that, compared to existing methods, our approach yields the highest accuracy of 94.70%. Another observation we found that, the scale space Co-HOG feature descriptor yields highest accuracy compared to Co-HOG feature descriptor (Chapter 2).

Table 3.3: Performance evaluation of our approach using GW dataset

Method	Features	Segmentation	mAP (%)
Rath et al. (2007)	Word profile	Word	73.71
Rodriguez et al. (2008)	HOG	Word	71.70
Almazan et al. (2013)	SIFT	Word	85.63
Our Approach presented in Chapter 2	Co-HOG alone	Word	91.03
Proposed method	Scale space Co-HOG	Word	94.70

3.3.4 Experiments on IAM dataset

From the IAM dataset, we have taken manually labeled 1000 word images of 20 different classes i.e., each class of word image has 50 instances of different writing styles. While performing five-fold cross-validation, we partitioned the dataset into five disjoint subsets and each subset consists of 200 word images (from each class of words we have taken 10 word images). In the training stage, 800 word images are used and remaining 200 word images are used for testing. Hence, for every experiment, we have 200 word images as test samples. The cross-validation process is repeated five times, with each subset used exactly once for validation.

Table 3.4 shows the retrieved word images for given query image using our approach for the IAM dataset. The first row shows query word and each column shows corresponding retrieved word instances of a given query word are shown. The results show the robustness of our approach for different kinds of words with variation in writing style. To estimate the overall evaluation results of our approach, we have taken the *mAP* of the five validation results. It is observed

that our approach yields the highest accuracy (%) with mAP is 92.55.

Table 3.4: Sample retrieval results: Query word images (first row) and their retrieval results of our approach for IAM dataset

We compared the performance of our approach with existing word spotting methods using IAM dataset. The existing word spotting methods considered for comparison are (a) HOG based approach by extracting local gradient histogram features through the sliding window proposed by Rodriguez et al. (2008) and (b) word profile based approach proposed by Rath et al. (2007). The existing methods are implemented and five-fold cross-validation method is used to validate based on mAP . Table 3.5 presents results comparison of our approach with existing methods based on mAP . From the experimental results, it is notified that our approach outperforms existing methods as well as the method presented in the preceding chapter for IAM dataset.

Based on the experimental results obtained on GW and IAM dataset, we can conclude that our approach efficiently retrieves the handwritten words which are having non uniform illumination, suffering from the noise and written by different writers. The highest accuracy of our approach is due to the extraction of illumination and scale invariant Co-HOG feature descriptor at multi scale representation of word images. Complex objects such as handwritten words are perceived at

Table 3.5: Performance evaluation of our approach using IAM dataset

Method	Features	Segmentation	mAP (%)
Rath et al. (2007)	Word profile	Word	73.71
Rodriguez et al. (2008)	HOG	Word	71.70
Almazan et al. (2013)	SIFT	Word	85.63
Our Approach presented in Chapter 2	Co-HOG alone	Word	88.41
Proposed method	Scale space Co-HOG	Word	92.55

multiple scales, which encodes the information of the word image at multiple scales. Another factor for highest accuracy is a division of the word image into blocks which capture the local variations within the class.

3.4 Chapter Summary

In this chapter, we proposed a word spotting method using scale space extended Co-HOG feature descriptor. The experimental results obtained on publicly available datasets such as GW and IAM dataset demonstrate that the proposed Scale Space Co-HOG feature descriptor achieves consistently excellent for word spotting in handwritten document images. Scale space representation dominates more discriminating ability because Co-HOG feature descriptor captures the co-occurrence features of the gradient, edge and stroke orientation of a word image at multiple scales. The Scale Space Co-HOG feature descriptors are robust to illumination variation and invariance to local geometric transformation compared to HOG feature descriptors. The drawback of our approach is its high dimensionality, which has a negative impact on the computational cost when matching by the DTW algorithm. However, it provides the highest accuracy compared to existing methods.

Chapter 4

Curvature Features Based BoVW for Word Spotting

A part of this chapter is published in International Journal of Computer Science & Information Technology (IJCSIT), AIRCC Publisher, vol. 9 no. 4 page 77-92, 2017.

Chapter 4

Curvature Features Based BoVW for Word Spotting

In this chapter, we present a segmentation-based word spotting method for handwritten documents using Bag of Visual Words (BoVW) framework based on curvature features. We proposed to extract curvature feature at each detected corner points of word image in a BoVW framework. The curvature feature is scalar value describes the geometrical shape of the strokes and requires less memory space to store.

4.1 Introduction

In Holistic word representations, a good description of the word image is a key issue. The researchers have developed word spotting techniques using different feature representations. The shape descriptors are widely used in word spotting and shape descriptors can be classified into statistical and structural. The statistical descriptor represents the image as an n-dimensional feature vector, whereas, the structurally based techniques represent the image as a set of geometric and topological primitives and relationships among them. Statistical descriptors are the most frequent and they can be classified as global and local features. Global features are computed from the image as a whole, for example, widths, height, aspect ratio, the number of pixels.

In contrast, local features are those which refer independently to different regions of the image or primitives extracted from it. For example, position/number of holes, valleys, dots or crosses, gradient orientation based SIFT (Lowe 2004) and SURF (Bay et al., 2008) features. These features have been proved useful due to invariance to scale and rotation as well as the robustness across the considerable range of distortion, noise, and change in intensity. Hence, these features

are extensively used in different computer vision applications such as image retrieval (Sivic et al., 2003) and image classification (Wang et al., 2010) and sign board detection (Schroth et al., 2011). The SIFT and SURF feature descriptors have recently achieved a great success in the document image analysis domain. Hence, the researchers are adapting these local features for development of many applications in document analysis such as word image retrieval (Rusinol et al. 2011; Yalniz et al., 2012), page retrieval (Smith et al. 2011), and logo retrieval (Jain et al., 2012).

Many authors have proposed handwritten word spotting techniques based on the matching of keypoints extracted using SIFT or SURF. The techniques have been used to directly estimate similarities between word images, or by searching the query model image within complete pages in segmentation free circumstances. However, the key points matching framework presents the disadvantage that such methods are memory intensive and requires alignment between the keypoint sets. In order to avoid matching all the keypoints among them, the Bag of Visual Words (BoVW) technique has been used for word spotting in handwritten documents.

The BoVW based word spotting methods yield holistic and fixed-length image representation while keeping the discriminative power of local descriptor. Rusinol et al. (2011) have proposed a segmentation free word spotting technique using BoVW model based on SIFT descriptors. Shekhar et al. (2012) used query by example model where the local patches expressed by a bag of visual words model powered by SIFT descriptors. Rothacker et al. (2013) have proposed to combine the SIFT descriptors based on BoVW representation with Hidden Markov Models in a patch-based segmentation-free framework in handwritten documents. The drawback of SIFT-based word spotting in BoVW technique is that they are memory intensive; window size cannot be adapted to the length of the query, relatively slow to compute and match. The performances of these methods are dependent on the length of the query with respect to the fixed size of the window.

In order to overcome the drawbacks of SIFT based word spotting using BoVW, in this chapter, we proposed segmentation based word spotting technique using a BoVW framework powered by curvature features. For the purpose of shape description of the handwritten word, curvature features at corner points are extracted. Then, we construct a Bag of Visual Words based on curvature features. After construction of Bag of Visual Words of the training set, we represent the training word images and query word image using visual word vector. Finally, Nearest Neighbor Search (NNS) similarity measure algorithm is used to retrieve word images similar to a query image.

4.2 Bag of Visual Words (BoVW) Framework

The BoVW framework consists of three main steps: in the first step, a certain number of image local interest points are extracted from the image. These keypoints are significant image points having rich of information content. In the second step, feature descriptors are extracted from these keypoints, and these feature descriptors are clustered. Each cluster corresponds to a visual word that is a description of the features shared by the descriptors belongs to that cluster. The cluster set can be interpreted as a visual word vocabulary. Finally, each image is represented by a vector, which contains occurrences of each visual word that appears in the image. Based on these feature vectors, a similarity measure is used to measure the likeness between given query image and the set of images in the dataset.

In the BoVW framework, an image is represented by an unordered set of non-distinctive discrete visual words. In retrieval phase, an image is retrieved by computing the histogram of visual word frequencies, and returning the image, with the closest histogram computed by the cosine of the angles. This can also be used to rank the returned images. An advantage of this approach is that, matches can be efficiently computed. Therefore, images can be retrieved with no delay.

The Bag of Visual Words (Fei-Fei et al., 2007) is an extension of Bag of Words (BoW) (Csurka et al., 2004) to the container of digitized documents and it is an alternative approach for word image retrieval in documents. This is inspired by several reasons (i) BoVW framework is best general representation for document (text) retrieval, (ii) Being a loose representation, the BoVW framework can retrieve sub-words, which is challenging in vector space models, (iii) the BoVW framework has revealed to achieve outstanding performance in recognition and retrieval tasks in images and videos (Sivic et al., 2003; Lazebnik et al., 2006).

4.3 Proposed Method

To the best of the author's knowledge, there is no approach using curvature feature in the BoVW framework for word spotting in handwritten document images. The proposed method based on the BoVW framework for word spotting is illustrated in Figure 4.1. The proposed method composed of four stages: (i) corner points detection (ii) extraction of curvature features (iii) codebook generation and (iv) word retrieval.

In the first stage, the document image is segmented into text lines and each text lines are segmented into primitive segments i.e. words using directional local profile technique (Papavassiliou et al., 2010). The corner points are extracted from word image using Harris-Laplace corner detector (Harris et al., 1988). In the second stage, curvature features are extracted from detected corner points. In the third stage, the codebook is used to quantize the visual words by clustering the curvature features using K-means algorithm. Finally, each word image is represented by a vector that contains the frequency of visual words appeared in the word image. In the word retrieval phase, for a given query image, we construct the visual word vector. Then, Nearest Neighbor Search is used to match the visual word vector of the query word image and the visual word vectors presented in the codebook. Finally, based on the ranking list, retrieved word images are presented.

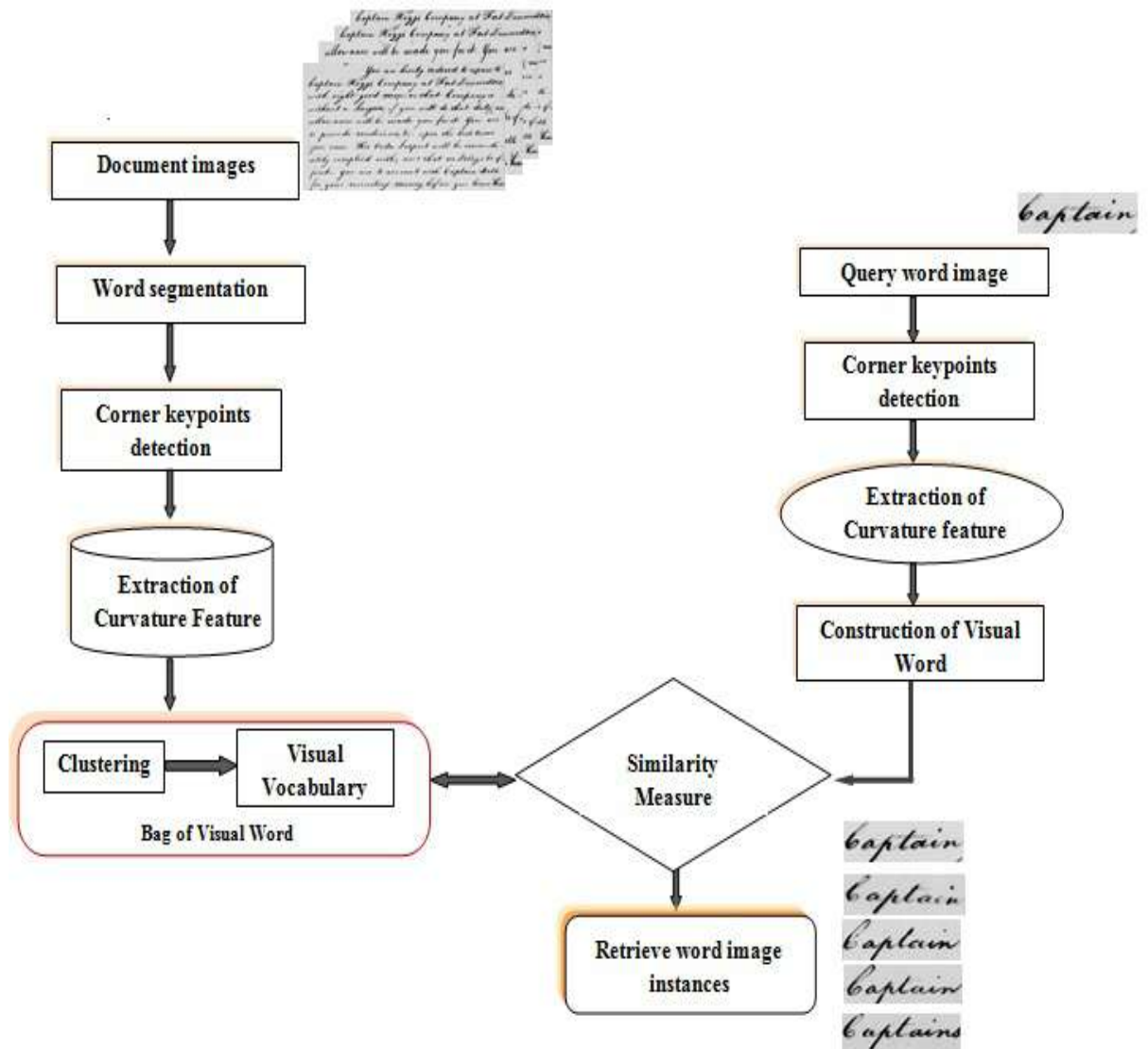


Figure 4.1: The pipeline of proposed word spotting method

4.3.1 Corner Points Detection

Once the document image is segmented into words, next step is the detection of corner points (keypoints). In this work, we detect the corner points of the word images using Harris-Laplace corner detector (Harris et al., 1988). Harris-Laplace corner detector is an accepted interest point detector due to its strong invariance to scale, rotation, and image noise. It is believed that most of the information on a contour is concentrated at its corner points. These corner points, which have high the curvature on the contour, play a very important role in shape analysis of handwritten word images. The Figure 4.2(d-f) shows the intermediate result for detection of corner points in the word image.

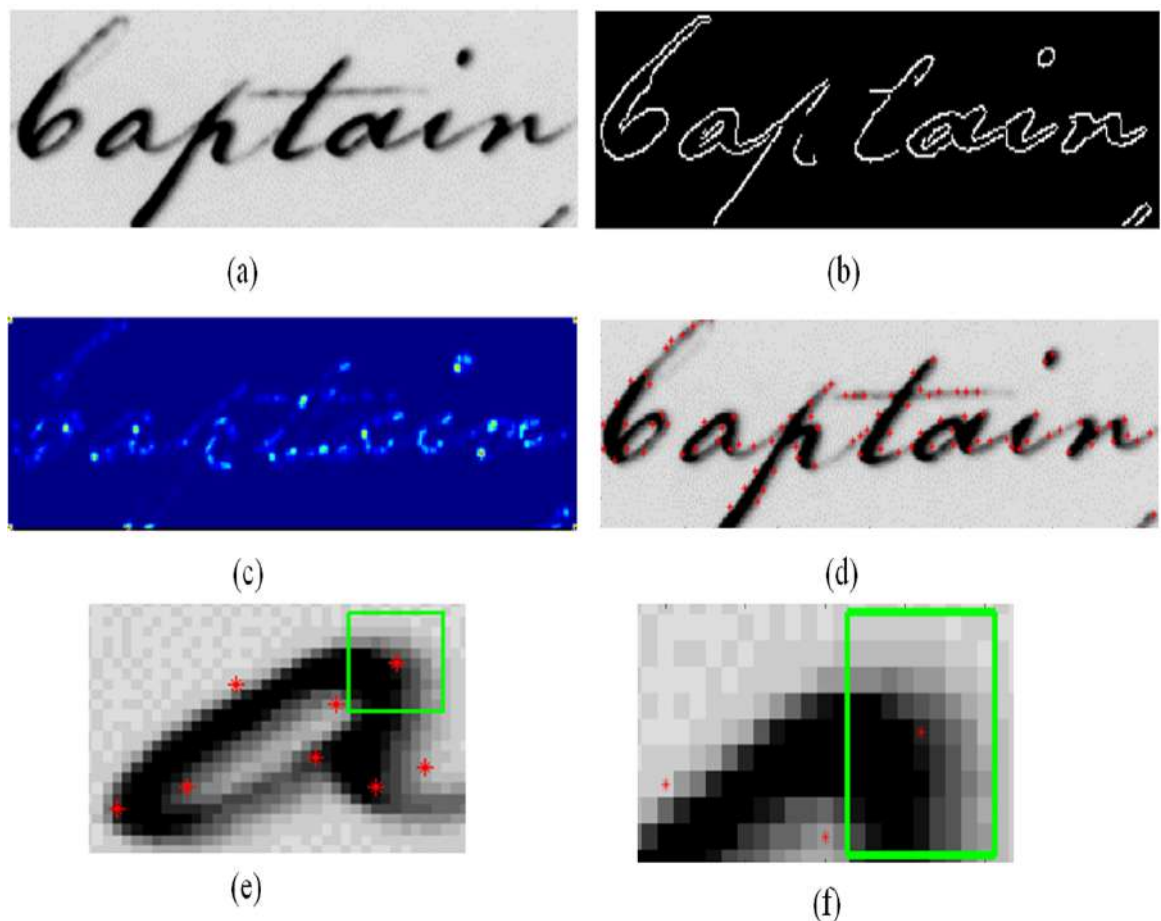


Figure 4.2: The intermediate results for corner points detection: (a) segmented word image, (b) canny edge map, (c) detected corner points on canny edge map image using Harris-Laplace operator (d) detected corner points on original word image (e) corner point at the global level (f) sample corner point at the local level

4.3.2 Curvature Features

The literature survey on theories of vision reveals that curvature is important in shape perception and measures of curvature play an important role in shape analysis algorithms. High curvature points are the best place to break the lines where a maximal amount of information can be extracted which are necessary for successful shape recognition. This is based on the information that the corners are points of high curvature. Asada et al. (1986) proposed an approach for representing the significant changes in curvature along the bounding of planar shape and this representation called as curvature primal sketch.

Mokhtarian et al. (1992) applied the scale space approach to the description of planar shapes using the shape boundary. The curvature of the contour was computed and the scale space image of the curvature function was used as a hierarchical shape descriptor that is invariant to translation, scale, and rotation. Recognition of handwritten numerals based on gradient and curvature features of the gray scale character proposed by Shi et al. (2002). Kannan et al. (2013) proposed an approach for offline Malayalam recognition using gradient and curvature feature. An advantage of using curvature feature is that it reduces the dimension of the feature descriptor when compared to other local features.

We extract curvature features from each detected corner points of word image. The curvature is a measure of how geometric object surface deviates from being a flat plane, or a curve is turning as it is traversed. The curvature features are the salient points along the contour of word image. At a particular point (M) along the curve, a tangent line can be drawn; this line making an angle θ with the positive x-axis (Figure 4.3). The curvature at the point is defined as the magnitude of the rate of change of θ with respect to the measure of length on the curve. Mathematically, the curvature of a point M in the curve C is defined as follows:

$$K(M) = \lim_{\Delta S \rightarrow 0} \left| \frac{\Delta\theta(M)}{\Delta S} \right|, \quad (4.1)$$

where, $\theta(M)$ is the tangential angle of the point M and S is the arc length.

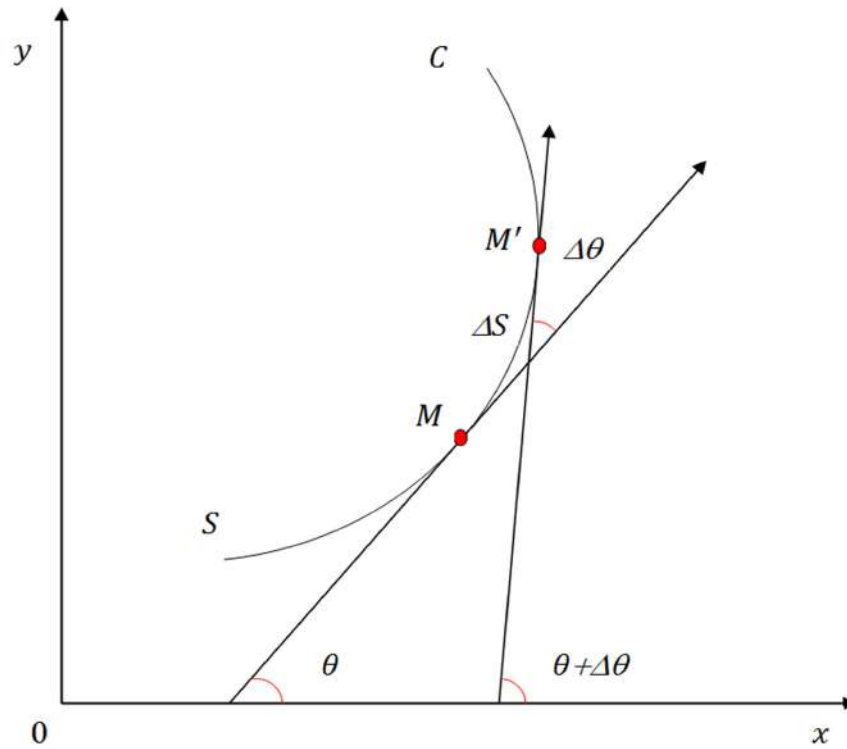


Figure 4.3: Curvature of the curve C

4.3.3 Codebook Generation

The single codebook is the house of all possible visual words that correspond to spotting the word image in handwritten document images. Generally, the codebook must hold the following constraints, the codebook should be small, that guarantees to minimum computational rate through minimum dimensionality, redundant visual words minimization, and provide high discrimination performance. The Figure 4.4 shows the pipeline of codebook generation process.

In order to generate the codebook for given training samples, we cluster curvature features using K-means algorithm. The curvature feature computed from detected corner point is allotted to a cluster through minimum distance from the center of the corresponding cluster. The clusters are treated as visual words. The number of clusters characterizes the size of the codebook. Finally, word image is formally represented as visual word vector with the help of frequency of occurrences of the visual words in the codebook.

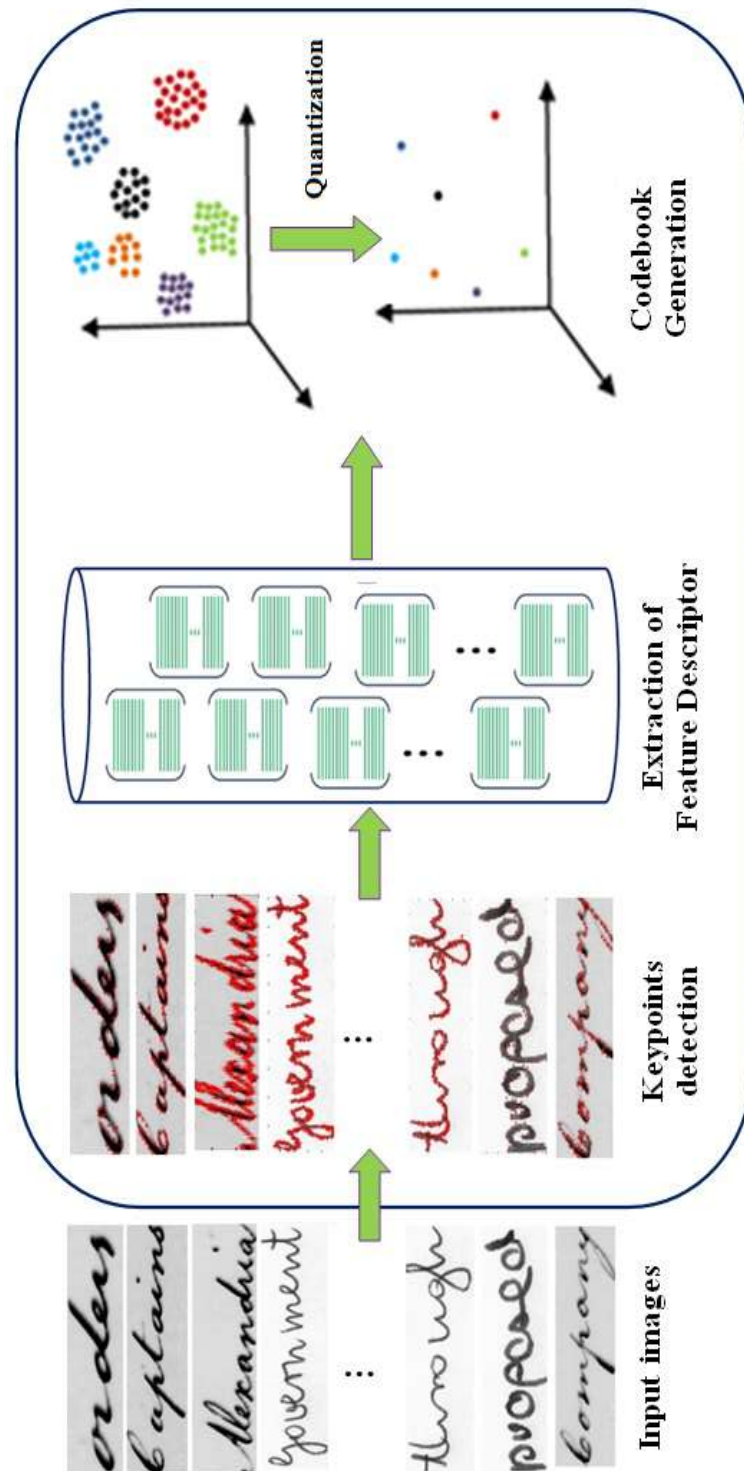


Figure 4.4: BoVW framework for handwritten word images

For example, consider a codebook with size 7 visual words and a word image that contains 16 local curvature features, which are assigned as follows: 3 local curvature features for the first visual word, 5 local curvature features for the second, 2 local curvature features for the third, 3 local curvature features for the fourth visual word, 2 local curvature features for the fifth visual word and 1 local curvature feature for the seventh visual word. Then, the visual word vector which represents the Bag of Visual Words of the word image is [3, 5, 2, 3, 2, 0, 1]. The dimension of this visual word vector is equal to the number of visual words in the codebook.

4.3.4 Word Retrieval

In order to retrieve the word images similar to query word, we compute the similarity between visual word vector of the query word image and visual word vector of word images present in the dataset. The most commonly used approach is Nearest Neighbor Search (*NNS*) using the Euclidean distance between visual word vector of a given query image and visual word vector of dataset images by considering appropriate threshold T .

$$NNS = \sqrt{\sum_{i=1}^n (d_j(i) - q(i))^2} < T, \quad (4.2)$$

where, n is the dimension of visual word vector, d_j and q are visual word vectors of j^{th} training sample and query word image respectively.

4.4 Experimental Results

In this Section, we present the experimental results of the proposed word spotting method in comparison with the state of the art word spotting methods. The proposed method is evaluated on three handwritten datasets of different nature, such as GW dataset, (Lavrenko et al., 2004), IAM English dataset (Marti et al., 2002) and Bentham dataset (Long 1981). In order to evaluate the performance of our approach, we used *mAP* metric, which provides a single value measure of precision for all the query word images.

4.4.1 Segmentation

Before the extracting the corner points from word images, the document must be segmented into words. We segment document images into lines and then into words using robust algorithm proposed by Papavassiliou et al. (2010). The sample results of segmented word images are shown in Figure 4.5.

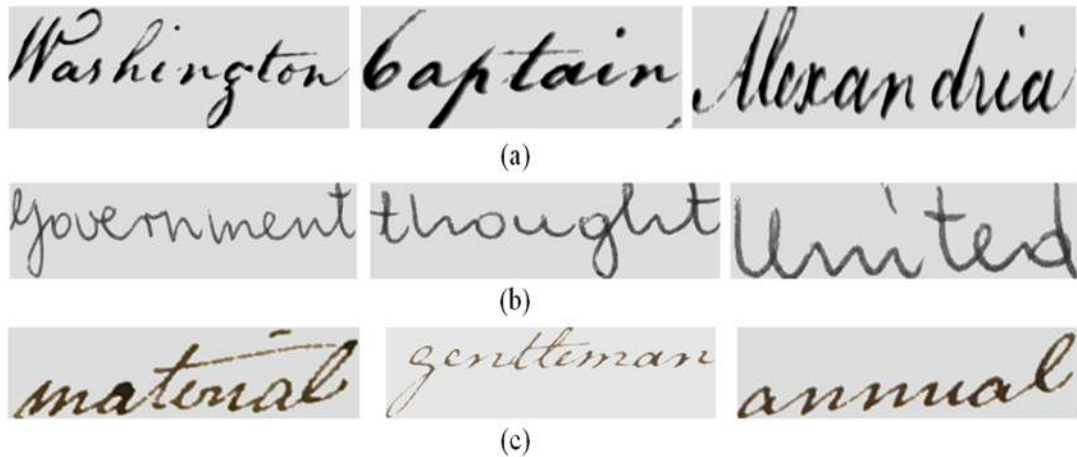


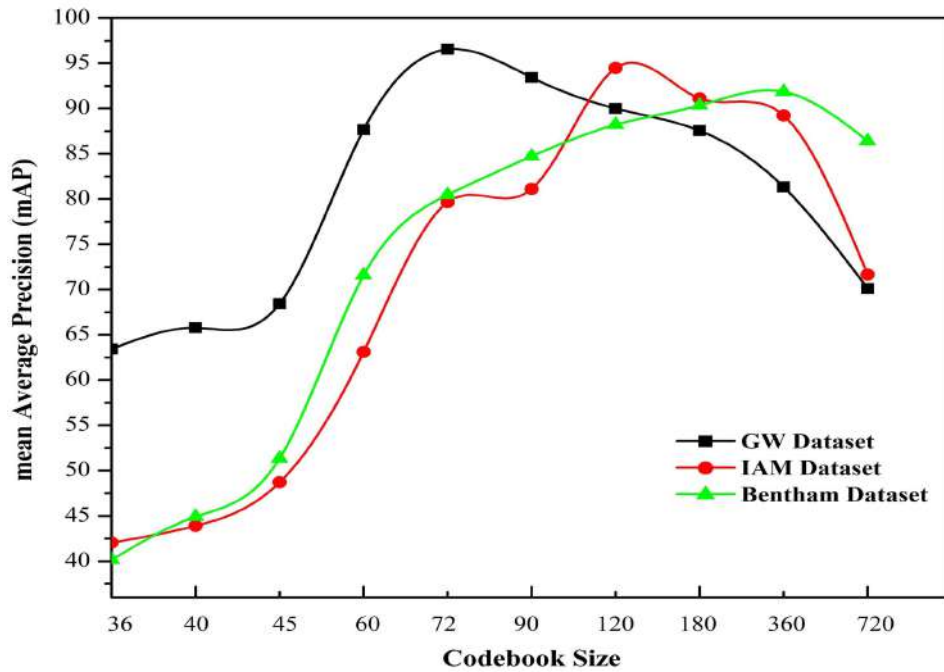
Figure 4.5: Segmented word images (a) from GW database (b) from IAM database (c) from Bentham database

4.4.2 Selection of Codebook Size

The size of the codebook is pre-computed and selection of the optimal size of a codebook is one of the important factors in achieving highest accuracy. Predicting the desirable clusters and optimal codebook size is non-straight forward and it is dataset-dependent. Hence, in order to find an optimal size, we conducted the experiments using GW (336 word instances from 20 pages), IAM (414 word instances from 1539 pages) and Bentham (240 word instances from 50 pages) datasets by varying codebook size. For each dataset, we employed five-fold cross validation process. To estimate the overall evaluation results of our approach, we compute the average of the five validation results. The mAP obtained on three datasets using our approach for a different codebook size is presented in Table 4.1 and its pictorial representation is shown in the Figure 4.6.

Table 4.1: Performance evaluation of our approach for varying codebook size using three datasets

Codebook size	(mAP %)		
	GW dataset	IAM dataset	<i>Bentham dataset</i>
36	63.43	42.06	40.17
40	65.78	43.89	44.91
45	68.43	48.72	51.36
60	87.65	63.12	71.59
72	96.72	79.64	80.64
90	93.42	81.12	84.72
120	90.01	94.46	88.21
180	87.57	91.13	90.34
360	81.32	89.24	91.84
720	70.13	71.64	86.42

Figure 4.6: The performance comparison of our approach for varying codebook size on three different datasets based on mAP

From the Table 4.1 and Figure 4.6, it is observed that, when codebook size is large the accuracy of our approach decreases. It is concluded that for GW dataset, the performance of our approach is significantly better when a size of the codebook is set to 72. Similarly, for IAM dataset and Bentham dataset, the performance of our approach yields good accuracy when a size of the codebook is set to 120 and 360 respectively. Hence, in all the experiments, the codebook size is 72, 120 and 360 for GW, IAM and Bentham dataset respectively.

4.4.3 Experiments on GW Dataset

From this dataset, we have taken 1680 word instances of 42 different classes of words from all the 20 pages of scanned handwritten documents. Each class of words may occur at least two times per document. While performing five-fold cross-validation, we partitioned the dataset into five disjoint subsets and each subset consists of 336 word images (from each class of words we have considered 8 word instances).

In the training stage, 1344 word instances are used and remaining 336 word images are used for testing. Hence, for every experiment, we have 336 word images as test samples. The cross-validation process is repeated five times, with each subset used exactly once for validation. It is noticed that our approach achieves the highest accuracy (%) with mAP is 96.72. Table 4.2 shows the qualitative results of our approach for the GW dataset. The first column shows query words and corresponding retrieved word instances of a given query word are shown in subsequent columns. The result confirms that the robustness of our approach for different kinds of words with a small variation in writing style.

We compared the performance of our approach with other existing word spotting methods using GW dataset. The work of Rath et al. (2003) is considered as a baseline for our experiment. The second and third method used for comparison with our approach results are Bag of Visual Words based methods powered by dense SIFT feature descriptor, presented in two different papers proposed by Rusinol et al. (2011) and Rothacker et al. (2013). The fourth method used for comparison is unsupervised word spotting proposed by Almazn et al. (2012)

Table 4.2: Sample retrieval results: Query word images (first column) and corresponding retrieved word instances from GW dataset

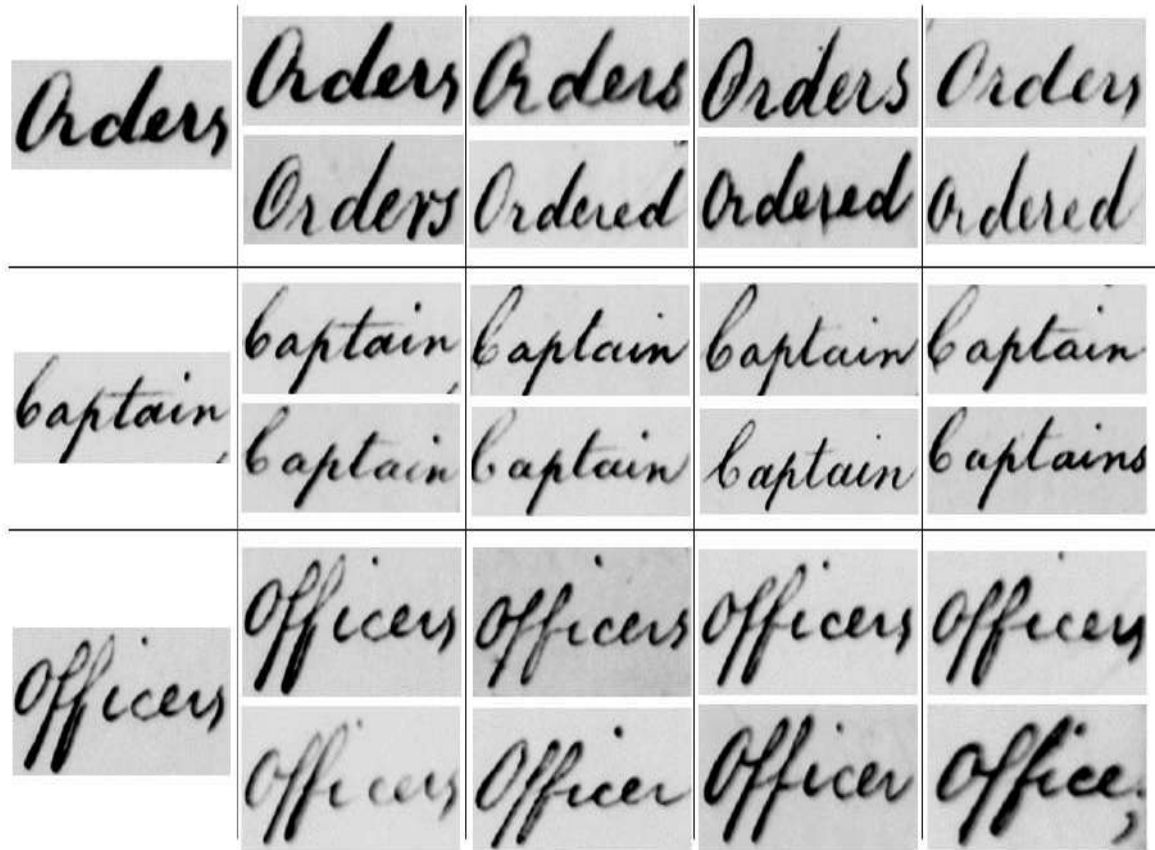


Table 4.3: Performance evaluation of our approach using GW dataset

Methods	Features	Experimental setup	mAP (%)
Rath et al. (2003)	Projection profile, Background/Ink transitions	10 pages, 1680 queries	40.90
Rusinol et al. (2011)	SIFT	20 pages, 1680 queries	30.40
Almazn et al. (2012)	HOG	20 pages, 1680 queries	54.40
Rothacker et al. (2013)	SIFT	20 pages, 1680 queries	61.10
Our Approach presented in Chapter 3	Co-HOG	20 pages, 1680 queries	94.70
Proposed method	Curvature	20 pages, 1680 queries	96.72

using a grid of HOG descriptors by sliding window framework. And also we compared the performance of proposed method with the approach presented by the author in Chapter 3 based on Co-HOG feature descriptor extracted in scale space representation.

The results of existing methods are extracted from their papers where results are reported as shown in Table 4.3. In Table 4.3, the size of the dataset, the feature descriptors are used and comparison results of our approach with existing methods are shown. It is observed that, compared to existing methods as well as the method presented preceding chapter, the proposed word spotting method yields the highest accuracy of 96.72%.

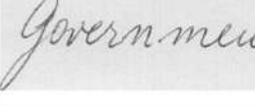
4.4.4 Experiments on IAM Dataset

From this dataset, we have taken 2070 segmented non-stop word images of 46 different classes i.e., each class of word image has 45 instances of different writing styles. For five-fold cross-validation process, each subset consists of 414 word images (from each class of words we have taken 9 word images). In the training stage, 1656 word images are used and remaining 414 word images are used for testing. It is observed that our approach achieves the highest accuracy (%) with *mAP* is 94.46. Table 4.4 shows the word retrieval results of our approach for the IAM dataset.

We compared the performance of our approach with existing word spotting method which is based on SIFT descriptor proposed by Almazan et al. (2014) and second method which is based on scale space Co-HOG feature descriptor presented by the author in Chapter 3. is SIFT descriptors based method proposed by Almazan et al. (2014) using Fisher vector representation computed over extracted densely from the word image and second method is Co-HOG feature descriptor based word spotting method presented in Chapter 3.

The results of existing methods are extracted from their papers where results are reported as shown in Table 4.5. Table 4.5 shows the comparison results of our approach with the existing method, it is observed that our approach yields the highest accuracy of 94.46% compared to popular existing method as well as

Table 4.4: Sample retrieval results: Query word images (first column) and corresponding retrieved word instances from IAM dataset

method presented by the author in preceding chapter.

Table 4.5: Performance evaluation of our approach using IAM dataset

Methods	Features	Experimental setup	mAP (%)
Almazn et al.(2014)	SIFT	1539 pages, queries1539	54.78
Our Approach presented in Chapter 3	Co-HOG	1539 pages, 2070 queries	92.55
Proposed method	Curvature	1539 pages, 2070 queries	94.46

4.4.5 Experiments on Bentham Dataset

From this dataset, we have taken correctly segmented 1200 word instances of 12 different classes of words from all the 50 pages of scanned handwritten documents. Each class of words may occur at least two times per document. While performing five-fold cross-validation, we partitioned dataset into five disjoint subsets and each subset consists of 240 word images (from each class of words we have consider 20 word instances). In the training stage, 960 word instances are used and remaining 240 word images are used for testing. It is noticed that our approach achieves the highest accuracy (%) with mAP is 91.84. Table 4.6 shows the qualitative results of our approach for the Bentham dataset. The result confirms that the robustness of our approach for a different variation of the same word involves, different writing style, font size, noise as well as their combination.

Table 4.6: Sample retrieval results: Query word images (first column) and corresponding retrieved word instances from Bentham dataset

Table 4.7: Performance evaluation of our approach using Bentham dataset

Methods	Features	Experimental setup	mAP (%)
Zagoris et al. (2013)	Texture	50 pages, 1200 queries	68.01
Proposed method	Curvature	50 pages, 1200 queries	91.84

We compared the performance of our approach with existing word spotting method proposed by Zagoris et al. (2014) based on texture feature descriptors extracted around the selected keypoints. Existing method is implemented, the training set and testing set is maintained same as the proposed method. The Table 4.7 shows the comparison results of our approach with the existing method. It is observed that, compared to the existing method, our approach achieves promising results with *mAP* is 91.84.

Based on the experimental results, we can conclude that our approach efficiently retrieves the handwritten words which are having non-uniform illumination, suffering from the noise and written by different writers. The highest accuracy of our approach is due to the extraction of curvature at corner point describes the geometrical shape of strokes present in handwritten words. The construction of BoVW using curvature features is simple when compared to SIFT or SURF feature descriptors because it has scalar value. It is identified through evaluation of experimental results on GW, IAM and Bentham dataset that, our approach outperforms existing SIFT or SURF based word spotting methods because the curvature is robust with respect to noise, scale, orientation and it preserves the local spatial information of the word shape.

The advantage of our approach is that it uses less memory space because of scalar value extracted at every corner point when compared to SIFT or SURF feature descriptors where feature vector is extracted at every keypoint. Another advantage involved in our approach is that the codebook size is very small compared to a size of codebook generated using SIFT or SURF feature descriptors. A benefit of using BoVW representation is that, word retrieval can be effectively

carried out because word image is retrieved by computing the histogram of visual word frequencies, and returning the word image, with the closest histogram. Therefore, word images can be retrieved with no delay.

4.5 Chapter Summary

In this chapter, we proposed a novel approach for word spotting in handwritten documents using curvature features in Bag of Visual Words framework. The curvature feature is significant in word shape perception and preserves the advantage of holistic representation of word image. From the experimental results, it is observed that curvature features are more suitable for handwritten word representation and which can improve the performance of proposed word spotting method as compared to existing word spotting methods. The use of BoVW framework has gained attention as a way to represent segmented handwritten words and can lead to a great boost in performance of our approach.

Chapter 5

Co-HOG Descriptor Based BoVW for Word Spotting

Chapter 5

Co-HOG Descriptor Based BoVW for Word Spotting

In this chapter, we propose a segmentation free word spotting in handwritten document images using a BoVW framework based on Co-HOG descriptor. Initially, the handwritten document is represented using visual word vectors which are obtained based on the frequency of occurrence of Co-HOG descriptor within local patches of the document. In order to add spatial distribution information of visual words into the unstructured BoVW framework, we adopted Spatial Pyramid Matching (SPM) technique.

5.1 Introduction

One of the limitations of the segmentation-based word spotting methods presented in the preceding chapters is that they usually need a layout analysis step for segments the document into words and they also need to perform segmentation step for select the candidate words. But this segmentation step is not always straightforward and any segmentation errors can affect the subsequent word representations and matching steps. This motivates us to move towards segmentation-free word spotting method.

The segmentation free approaches overcome the problems associated with poor segmentation results by considering the document image as a whole. The gradient information and local image features have been used in segmentation free approaches trying to benefit from the scale, rotation invariance and they offer robustness to noise. These methods usually involve a voting scheme in order to detect and localize potential word matches in the document image.

In segmentation free methods (Leydier et al., 2005; Gatos et al., 2009), the document images are represented by feature descriptor such as SIFT. Then, sliding window or patch based approaches are used to locate the document regions that are most similar to the query word (Rusinol et al., 2015; Shekhar et al., 2012; Rothacker et al., 2013 and Zhang et al., 2013). The drawback of SIFT-based word spotting is that they are memory intensive; window size cannot be adapted to the length of the query, relatively slow to compute and match. In order to avoid matching all the key points among them, the BoVW technique has been used for word spotting in handwritten documents (Rusinol et al., 2011; Shekhar et al., 2012). The BoVW based word spotting methods yield holistic and fixed-length image representation while keeping the discriminative power of local descriptor.

Almazan et al. (2014) have proposed unsupervised segmentation free word spotting method based on HOG descriptor. Documents images are represented through a grid of HOG descriptor, and a sliding-window approach is used to locate the document regions that are most similar to the query. HOG feature descriptor captures orientation of only isolated pixels, whereas spatial information of neighboring pixels is ignored. In order to capture the spatial information of neighboring pixels, we propose a Co-HOG descriptor for word spotting (Chapter 2) in handwritten documents. The Co-HOG is an extension of HOG descriptor, which encodes gradient orientation of neighboring pixel pairs and accordingly captures more spatial and relative information, making it more dominant to represent the characters shape precisely and effectively.

The benefit of a BoVW framework for effective matching of feature vectors motivated us to use visual word representation in conjunction with Co-HOG feature descriptor which represents the character shape precisely and effectively through capturing more spatial and relative information. Hence, in this chapter, we propose a segmentation free word spotting technique using a Bag of Visual Words framework powered by Co-HOG feature descriptor. Initially, the handwritten document is represented using visual word vectors which are obtained based on the frequency of occurrence of Co-HOG descriptor within the local patches of the document. When we represent local patch using visual word vector, it does

not consider their spatial location. Hence, visual word vector suffers from spatial information.

In word spotting methods, spatial information is essential because it helps to determine a location exclusively with visual information when the different location can be perceived as the same. This is one of the limitations of the BoVW framework for representing a patch using visual word vector. In order to add spatial distribution information of visual words into the unstructured BoVW framework, we adopted Spatial Pyramid Matching (SPM) technique (Lazebnik et al., 2006). This technique takes into account the visual word distribution over the image by creating a pyramid of spatial bins. This SPM representation sacrifices the geometric invariance properties of BoVW framework. It more than recompenses for this loss with increased discriminative power derived from the global spatial information. Therefore, the SPM technique significantly outperforms BoVW on handwritten word spotting tasks.

5.2 Proposed Work

To the best of the author's knowledge, there is no approach which uses segmentation free word spotting using Co-HOG feature descriptor in the BoVW framework for handwritten document images. The proposed approach is made up of three steps (Figure 5.1). Initially, all the document images of the corpus are represented using codebook, which is obtained by clustering the Co-HOG descriptor feature space into k-different clusters by using the k-means algorithm. In the second step, query model is constructed through patch based approach, in which it accumulates the frequencies of each visual word within the local patch.

The feature descriptor of local patch consists of a frequency of visual words and does not take into account the spatial distribution of the visual words within the patch. In order to include spatial distribution information to the BoVW framework, we adopted spatial pyramid matching method (Lazebnik et al., 2006). Finally, for word retrieval, order the best patches yielded by Nearest Neighbor Search (NNS) similarity measure with respect to their minimum matching cost.

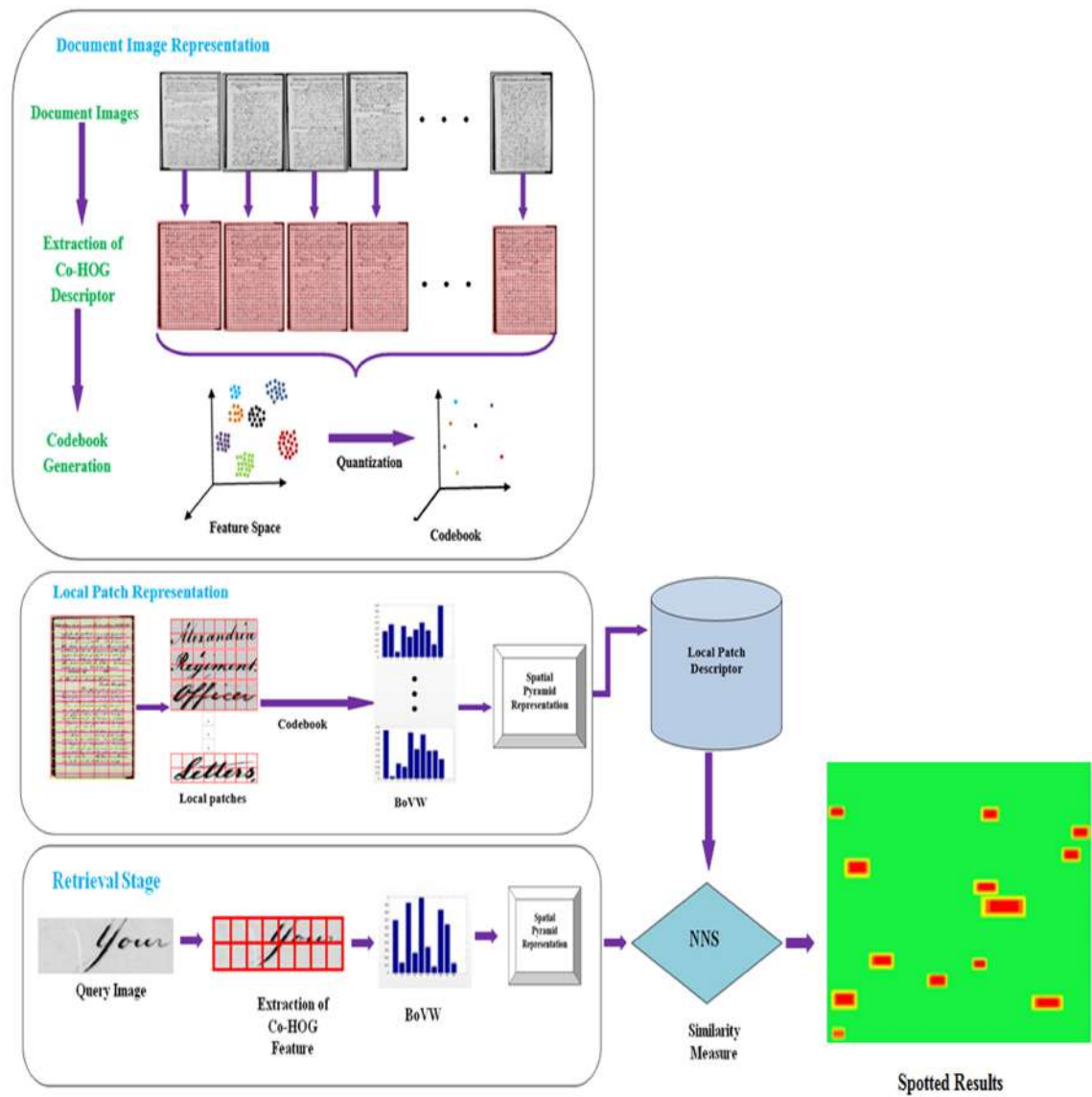


Figure 5.1: The pipeline of proposed word spotting method

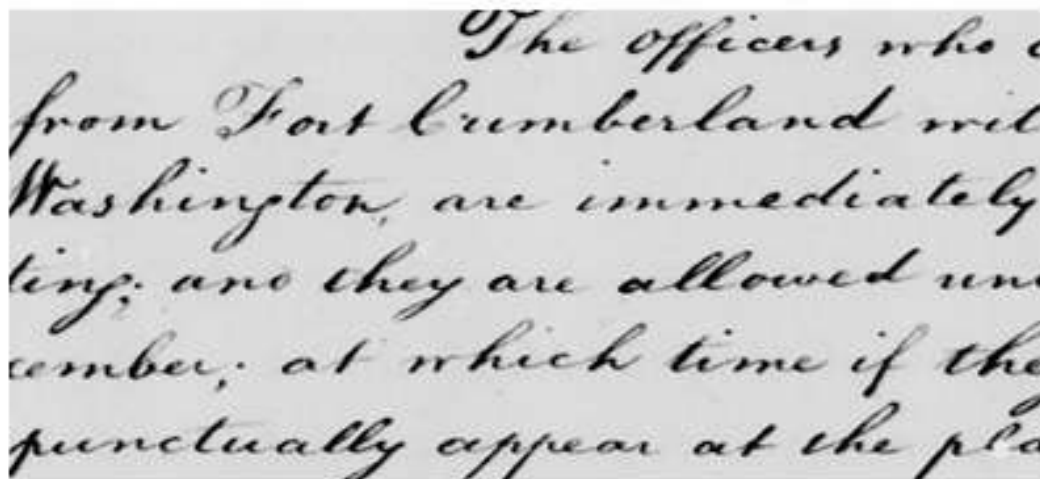
5.2.1 Document Image Representation

In this section, we present a construction of visual words using Co-HOG feature descriptor to represent the document image. Initially, we extract the Co-HOG feature descriptor from the handwritten document images of the corpus. In order to extract Co-HOG feature descriptor, we divide a document image into an equal sized regular grid of dimension (Figure 5.2(b)). This parameter is related to the font size, and in our method, experimentally set. For each regular grid in the document, we calculate the co-occurrence matrix. We compute co-occurrence matrix over a grid G of size at an offset as follows:

$$K_{x,y}(i, j) = \sum_{p=1}^N \sum_{q=1}^M \begin{cases} 1 & \text{if } G(p, q) = i \text{ and } G(p + x, q + y) = j, \\ 0 & \text{otherwise,} \end{cases} \quad (5.1)$$

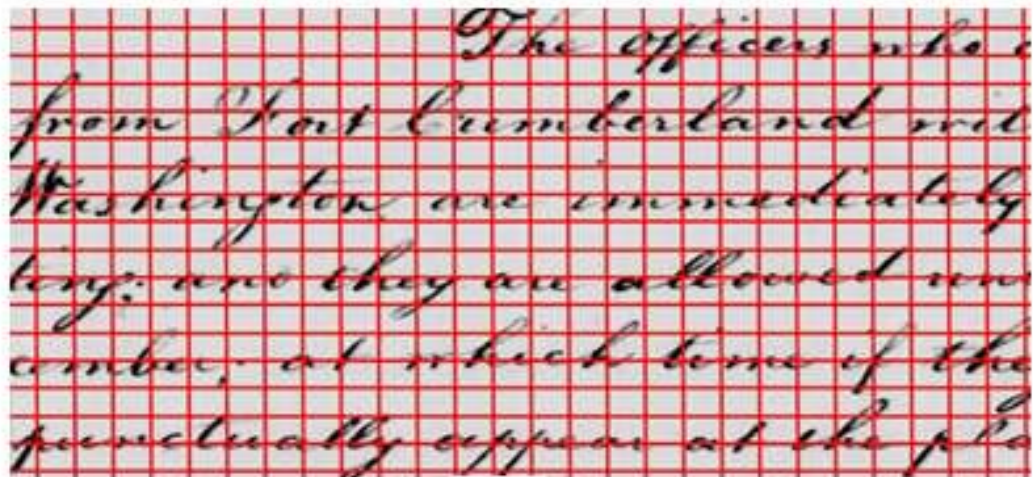
where, $K_{x,y}$ is a square matrix of dimension 8×8 . Gradient orientation interval $[0^0, 360^0]$ is divided into $N_{bin} = 8$ orientations per 45^0 . For example, the document has dimension 2120×3360 . When divide a document into a regular grid of 40×40 pixels, we have 4452 grids (Figure 5.2(b)). The computed co-occurrence matrix from each grid is represented using Co-HOG histogram is called as Co-HOG descriptor is shown in Figure 5.2(c). The co-occurrence matrix expresses the distribution of gradient orientations at a given offset over a grid and combinations of neighbor gradient orientations can express shapes in detail.

After extracting the Co-HOG feature descriptor from document images, we generate codebook which is a collection of visual words. The computed Co-HOG descriptors are clustered by the k-means clustering algorithm. These clusters are treated as visual words. The number of visual words characterizes the dimension of a codebook.



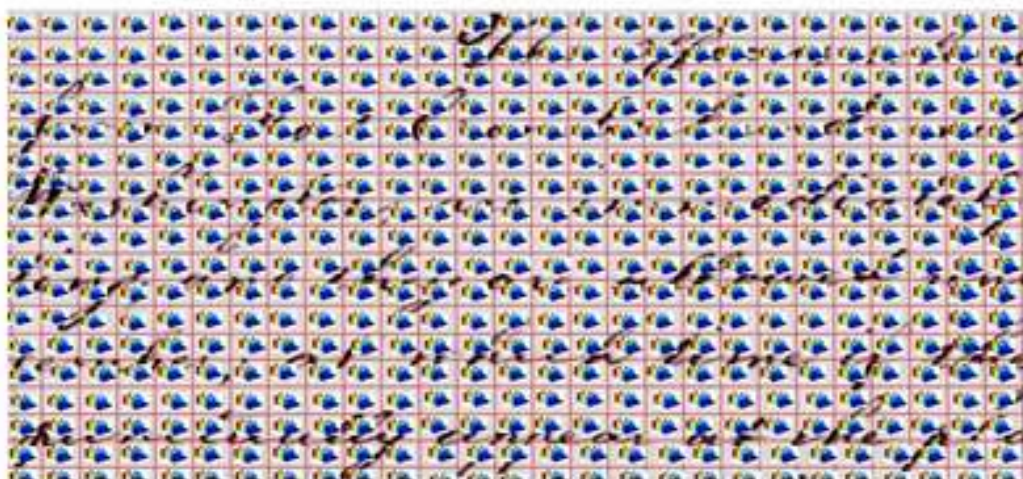
The officers who
from Fort Cumberland with
Washington, are immediately
ting; and they are allowed un
ember; at which time if the
punctually appear at the pla

(a)



The officers who
from Fort Cumberland with
Washington are immediately
ting; and they are allowed un
ember; at which time if the
punctually appear at the pla

(b)



The officers who
from Fort Cumberland with
Washington are immediately
ting; and they are allowed un
ember; at which time if the
punctually appear at the pla

(c)

Figure 5.2: An overview of the document image representation (a) sample document image from GW database (b) dividing an image into regular grids of size and (c) Co-HOG histogram computed from each regular grid

5.2.2 Local Patch Representation

Once we construct the codebook for a corpus of documents, we split every document image into a set of overlapping local patches. The local patch guarantees that almost all the words in the documents from the database will fit in a single local patch. Based on experimental observation on the document, resolution of a document, height and width of characters present in the document, we consider the local patch has a dimension of $LP_w \times LP_h$ pixels and patch is densely sampled at each S pixels. This patch displacement promises that all the words in a document page are represented by at least one patch. The visual word vector of every local patch is formally represented with the help of frequency of occurrence in each visual word which lies within the local patch. Therefore, the dimension of visual word vector of the local patch is same as the number of visual words present in the codebook.

The spatial pyramid is constructed by splitting the local patch LP into $LP_x \times LP_y$ spatial bins, where LP_x and LP_y represents the number of partitions along horizontally X and vertically Y directions, respectively. After splitting the local patch into $LP_x \times LP_y$ spatial bins, we compute descriptor of each spatial bin using codebook, which consists of a frequency of visual words present within the spatial bin. Finally, the resultant descriptor of the local patch LP is obtained by concatenating all the descriptor of each spatial bin.

In our experiments, we split the patch into two levels of spatial pyramids presented in Figure 5.3. In the first level, $LP_x = LP_y = 1$; i.e. the whole patch

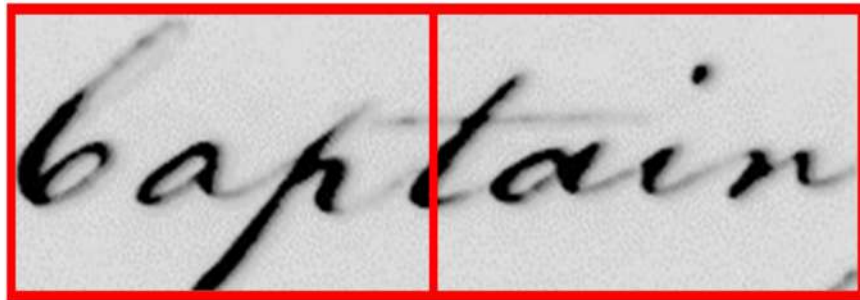


Figure 5.3: The second level SPM configuration of proposed approach

is considered and in the second level, $LP_x = 1$ and $LP_y = 2$; i.e. left and right parts of the patch. Therefore, three spatial bins are obtained from two level representation of spatial pyramid. With this configuration, we aim to capture information about the left and right parts of the words. Since, we used a two level SPM with 3 spatial bins such whole, left and right part of the patch. Using this configuration, the descriptors are encoded separately from the left and right side of the word. Since, the frequency of visual words computed from each spatial bin is minimum when considering at higher levels of the pyramid, because of the spatial bins are smaller and visual words contribution is weighted according to the spatial coverage. Finally, descriptor of local patch F_{LP} is represented by concatenating all the three spatial bins and is described by

$$F_{LP} = [f_w^{LP}, f_l^{LP}, f_r^{LP}], \quad (5.2)$$

where, f_w^{LP} , f_l^{LP} , f_r^{LP} represents descriptor corresponds to the whole, left and right spatial bins of the local patch respectively. Therefore, the dimension of each local patch D_{LP} is calculated by using Eq. 5.3. This spatial distribution information increases the performance of our method.

$$D_{LP} = \sum_{l=1}^L LP_x^l \times LP_y^l, \quad (5.3)$$

where, L is number of levels and l is corresponding pyramid level.

5.2.3 Retrieval Stage

At the retrieval stage, Co-HOG feature descriptors are extracted from the query patch which is cropped from the document corresponding to a single patch within a document. Then, Co-HOG feature descriptors are quantized by using the codebook, which yields visual vector of the query patch. Then, apply the spatial pyramid configuration to the query patch to construct the query patch descriptor. Finally, we retrieve the local patches using Nearest Neighbor Search (NNS)

similarity measure algorithm by computing the similarity between the query patch descriptor (f_q) and the document patch descriptors (f_P) of training set by considering appropriate threshold T . The NNS similarity measure between query descriptor and document patch descriptor is calculated using following equation:

$$NNS = \sqrt{\sum_{i=1}^{D_{LP}} (f_P(i) - f_q(i))^2} < T, \quad (5.4)$$

where, NNS is the similarity distance computed between two patch descriptors i.e. f_P and f_q and D_{LP} is the dimension of a patch descriptor.

The returned region from the documents will be considered as relevant if it overlaps with a query patch by at least 60%. Once the most similar local patches have been retrieved, the regions of the document found which is most similar to patch selected. For each document page image, a 2D voting space is constructed where each retrieved local patch will cast its votes. In our case, we consider each grid dimension of the voting space is same as the local patch. Then, each retrieved local patch casts a vote to the location of the document where its geometric center falls and weighted by the approximate threshold T . Figure 5.4 shows an example of the word spotting using local patches voting procedure obtained for a given query image.

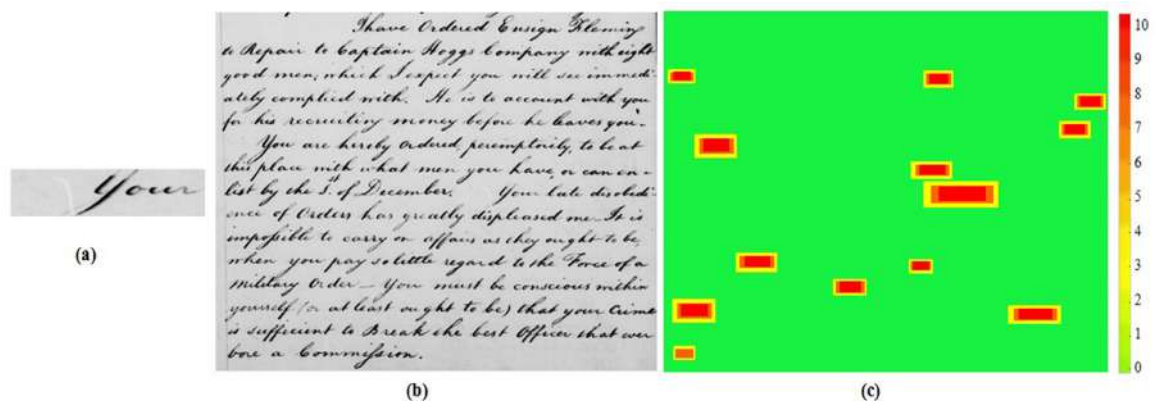


Figure 5.4: Visualization of the word spotting. (a) query image, (b) sample document from GW dataset (c) spotted local patches where words similar to query word are found

5.3 Experimental Results

The proposed method is evaluated on two popular handwritten datasets which are widely used for evaluation of word spotting methods, such as George Washington (GW) dataset (Lavrenko et al., 2004), and IAM dataset (Marti et al., 2002). In order to evaluate the performance of proposed method, we used popular metric such as *mAP*. In order to create query set for each dataset, we cropped and extracted the patches from all the documents of both datasets. From GW dataset, we extracted 4680 query patches and from IAM dataset, 5789 query patches are extracted which are having dimension similar to local patches of training set.

5.3.1 Parameters Used

For each database, we need to use the line height parameter in order to define the geometry parameters of the grid size and local patch size. The line height H has been estimated by projection profile technique over documents of each database. To compute the Co-HOG feature descriptors, the selection of grid size for each dataset is based on the criteria that a descriptor either covers part of a character, or a complete character or a character and its surroundings. We used optimal grid size 40×40 pixels for GW dataset.

Similarly, for IAM dataset we used 30×30 pixels. These selected grid sizes affect the number of grids per document image; therefore, each document image contains approximately 4450 grids per document when using grid size 40×40 pixels for GW dataset. Similarly, for IAM dataset, approximately 19422 grids per document image when using grid size 30×30 pixels. Similarly, the selection of local patch size is carried out based on criteria that almost all the words in the documents fit in a single local patch. To compute the local patch feature descriptors for documents of each database, we selected optimal local patch dimension 320×80 pixels for GW dataset and similarly 300×60 pixels for IAM dataset.

5.3.2 Selection of Codebook Size

The size of the codebook is pre-defined and selection of the optimal size of a codebook is one of the important factors in achieving highest accuracy. Predicting the desirable clusters and optimal codebook size is non-straightforward and it is dataset-dependent. Hence, in order to find an optimal size for each dataset, we conducted the experiments using GW (4860 query patches from 20 pages) and IAM (5789 query patches from 1539 pages) datasets by varying codebook size. In all the experiments, optimal grid size and optimal local patch size for each dataset are used. The *mAP* obtained on two datasets using our approach for a

Table 5.1: Performance evaluation of our approach for varying codebook size

Codebook size	(mAP %)	
	GW dataset	IAM dataset
32	32.19	28.64
64	46.23	37.31
128	57.01	42.58
256	68.33	56.42
512	76.14	62.39
1024	86.39	68.72
2048	84.62	74.38
4096	83.11	84.57
8192	82.67	82.68
16384	82.03	81.21

different codebook size is presented in Table 5.1 and its pictorial representation is shown in Figure 5.5. We can see from the Figure 5.5, the evolution of the *mAP* for varying codebook size from 2^5 to 2^{13} visual words for the two datasets.

The system tends to perform better with large codebook size in terms of accuracy. The increase in performance as the patch geometry and codebook size grows due to the perceptual aliasing. From the Table 5.1 and Figure 5.5, it is observed that, for GW dataset, the performance of our approach is significantly

better when a size of the codebook is set to 1024. Similarly, for IAM dataset, the performance of our approach yields good accuracy when a size of the codebook is set to 4096. Hence, in all the experiments, the codebook size is 1024 for GW and 4096 for IAM dataset.

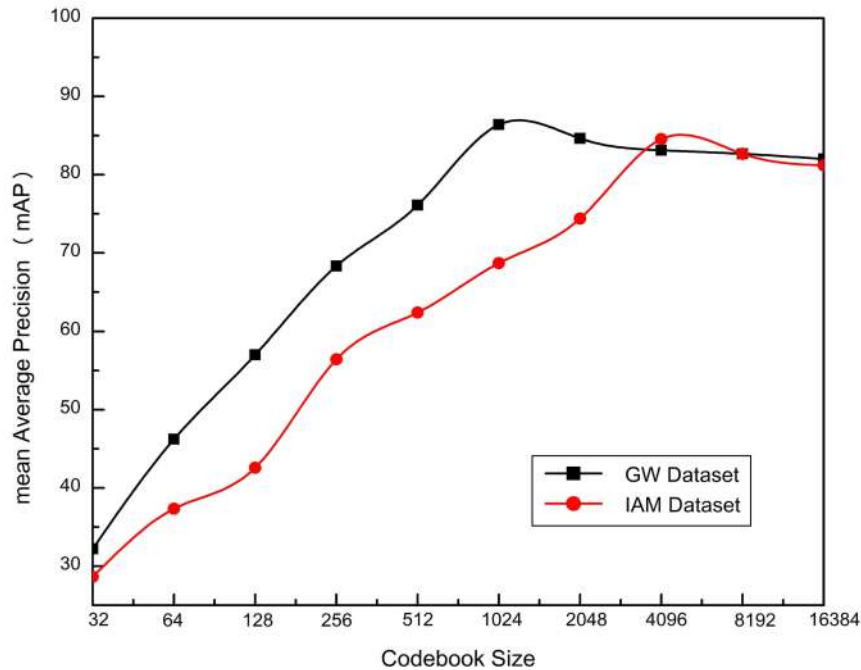


Figure 5.5: *mAP* of our approach for varying codebook size for two datasets

5.4 Experiments on GW dataset

The optimal grid size for calculating Co-HOG is 40×40 pixels, local patch size is 320×80 pixels and codebook size is 1024 visual words. We have conducted the experiments using 20 pages of GW dataset for 4860 query words. Figure 5.6 demonstrates the qualitative results obtained on one of the documents of GW dataset using our approach for the query word “*your*”. The spotted regions are shown using the red color. Similarly, Table 5.2 shows qualitative results of our approach obtained for documents of GW dataset. The first column shows query words and subsequent columns display spotted local patches where words similar to query word are found.

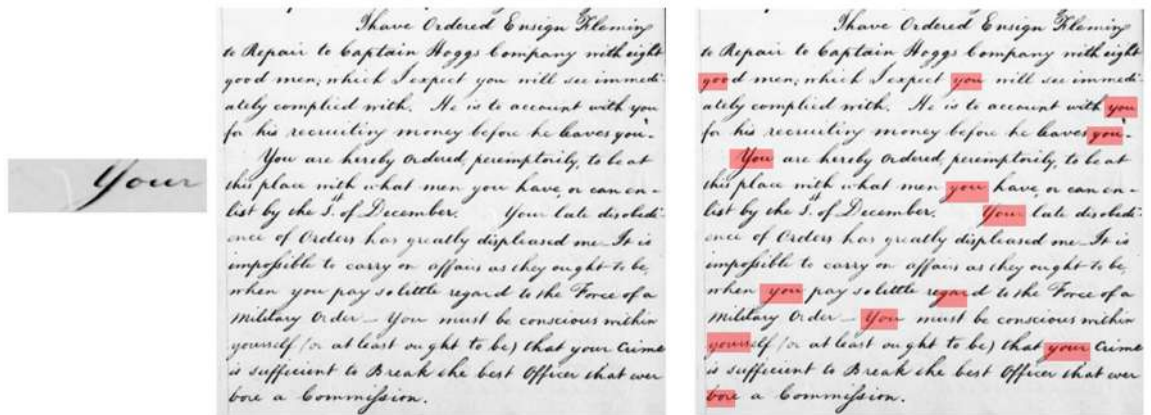


Figure 5.6: Example of spotting results for the query word “your” in one of documents of GW dataset

Table 5.2: Sample retrieval results: Query word images (first column) and corresponding retrieved word instance patches from GW dataset

Table 5.3 demonstrates quantitative results of our approach obtained on GW dataset and also the review of the achieved performances of four state-of-the-art segmentation free methods. We compared the performance of proposed method with existing segmentation-free methods that report experiments using GW dataset, such as Rothacker et al. (2013), Zhang et al. (2013), Almanzan et al. (2014) and Rusinol et al. (2015).

While comparing the performance of proposed method with these existing segmentation free word spotting methods, the training set size and number of query words are maintained same for all the methods. All these methods include the query image in retrieved results and in the computation of *mAP*. Moreover, Rothacker et al. (2013) use an overlap over a union of 20% with the ground truth to classify a window as positive, instead of the more traditional 50%, and Zhang et al. (2013) have only considered as queries the words with more than 5 characters. From the Table 5.3, we can see that proposed system obtains quite reasonable results and considerably outperform the existing segmentation free word spotting methods for GW dataset.

Table 5.3: The performance comparison of our approach with state-of-the-art segmentation free methods for GW dataset

Methods	Features	Experimental setup	Segmentation	mAP (%)
Zhang et al. (2013)	SIFT	20 pages, 4860 queries	None	79.35
Rothacker et al. (2013)	SIFT	20 pages, 4860 queries	None	61.10
Almanzan et al. (2014)	HOG	20 pages, 4860 queries	None	80.29
Rusinol et al. (2015)	SIFT	20 pages, 4860 queries	None	61.35
Proposed method	Co-HOG	20 pages, 4860 queries	None	86.39

5.5 Experiments on IAM dataset

From this dataset, we have used 5784 query words extracted from the pages of IAM dataset in order to evaluate our approach. In all the experiments, we used the optimal grid size for CO-HOG is 30×30 pixels, local patch size 300×60 pixels and codebook size is 4096 visual words. The Figure 5.7 shows word spotting results visually obtained on one of the documents of IAM dataset using our approach for the query word “British”. The spotted regions are shown using the red color.

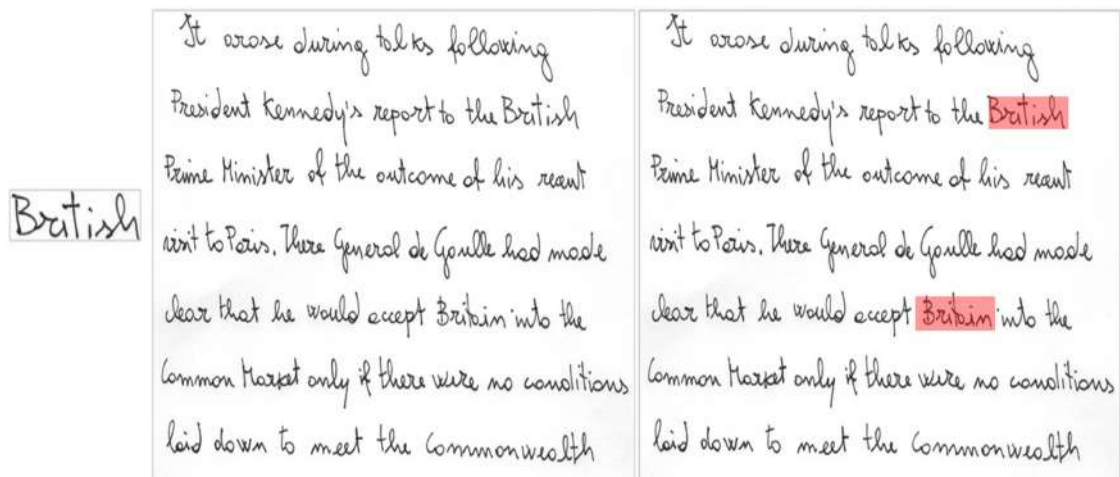


Figure 5.7: Example of spotting results for the query word “British” in one of the documents of IAM dataset

Table 5.4 shows the qualitative results of our approach to documents of IAM dataset. The first column shows query words and subsequent columns display spotted local patches where words similar to query word are found. As shown in Table 5.4, most of the positive matching of word instances and partial words are spotted. For example, local patches containing “Federation” are spotted as the positive matching for the query word “Federal”.

Table 5.5 shows the comparison of the performance of proposed method with existing segmentation free methods proposed by Rothacker et al. (2013), Zhang et al. (2013), Almazan et al. (2014) and Rusinol et al. (2015) for IAM dataset. The dataset size, number of queries, experimental setup and evaluation measures are

Table 5.4: Sample retrieval results: Query word image (first column) and corresponding retrieved word instance patches from IAM dataset

Federal				
British				
talles				

maintained same for all the methods including proposed and existing methods. It is observed that proposed approach yields the highest accuracy of 84.57 compared to existing segmentation free methods.

For the proposed segmentation free word spotting method, we can observe that most of the relevant word instances are spotted and retrieved, even if local patches overlap with a query patch by at least 60%. The short queries and substring from longer words are more possible to obtain false positives since we avoided matching the substrings in our method. Table 5.6 illustrates some false positives obtained using our approach for the queries which are short and substrings. The first column shows query words and subsequent column shows spotted and retrieved regions.

Based on the experimental results obtained on GW and IAM dataset, we can conclude that our approach efficiently retrieves and localize the handwritten words which are having non uniform illumination, suffering from the noise and written by different writers. The highest accuracy of our approach is due to many factors and one of the important factors is Co-HOG descriptor which captures the character shape information more precisely and encodes the local

Table 5.5: The performance comparison of our approach with state-of-the-art segmentation free word spotting methods for IAM dataset

Methods	Features	Experimental setup	Segmentation	mAP (%)
Zhang et al. (2013)	SIFT	1539 pages, 5784 queries	None	80.21
Rothacker et al. (2013)	SIFT	1539 pages, 5784 queries	None	63.18
Almanza et al. (2014)	HOG	1539 pages, 5784 queries	None	72.16
Rusinol et al. (2015)	SIFT	1539 pages, 5784 queries	None	64.29
Proposed method	Co-HOG	1539 pages, 5784 queries	None	84.57

Table 5.6: Examples for some false positives



spatial information by counting the frequency of co-occurrence of gradient orientation of neighboring pixel pairs. Moreover, Co-HOG descriptor is more robust and discriminative, because it captures the occurrence of edge characteristics of text strokes by exhaustively considering every grid within a word image patch. Another important factor for highest accuracy is the construction of visual words using Co-HOG descriptor.

In addition, the experiment results demonstrate that the performance of the approach is improved when the spatial information incorporated into the BoVW framework through SPM technique. The spatial pyramid matching representation yields the geometric invariance properties which are not provided by BoVW framework alone. An enhanced BoVW framework integrates the global and local information which can produce more discriminative visual words.

5.6 Comparative Study of Proposed Techniques

In this section, we present a comparative study of proposed techniques for word spotting in handwritten documents presented in preceding chapters and also the technique presented in this chapter. The first technique which is segmentation based extracts Co-HOG feature descriptors alone (Chapter 2), the second technique which also segmentation based and extracts Co-HOG descriptor in scale space representation (Chapter 3), and the third technique is segmentation based and uses curvature features in BoVW framework and the fourth technique (current chapter) is segmentation-free technique which is based on extraction of Co-HOG feature descriptors in BOVW framework.

The experiments are conducted using the same setup for all the proposed methods for GW and IAM datasets. The Table 5.7 shows the comparison of the performance of proposed word spotting methods. Figure 5.8 and Figure 5.9 shows Precision versus Recall plot obtained on GW and IAM dataset for the different configurations of the four-word spotting techniques proposed by the author. It is observed that among all the configurations, segmentation free based Co-HOG + BoVW + SPM configuration (proposed in this chapter) achieves

Table 5.7: The performance comparison of proposed word spotting techniques

Method	Experimental setup for		Segmentation	mAP (%)	
	GW	IAM		GW	IAM
Co-HOG alone (Chapter 2)	20 pages, 336 queries	1539 pages, 200 queries	Word	91.03	88.41
Co-HOG + SS (Chapter 3)	20 pages, 336 queries	1539 pages, 200 queries	Word	94.70	92.55
Curvature + BoVW (Chapter 4)	20 pages, 336 queries	1539 pages, 414 queries	Word	96.72	94.46
Co-HOG + BoVW + SPM (Chapter 5)	20 pages, 336 queries	1539 pages, 414 queries	None	98.43	96.53

the highest accuracy compared to segmentation based techniques presented in preceding chapters for both the datasets. The increase in accuracy of Co-HOG + BoVW + SPM configuration compared to other configurations is due to the incorporation of SPM method in a BoVW framework which provides a spatial distribution of constructed visual words.

The Table 5.3 and Table 5.5 shows the comparative study of proposed segmentation free technique with existing segmentation free word spotting techniques using GW and IAM datasets respectively where the accuracy of proposed technique is 86.39 for GW (4860 query patches) and 84.57 for IAM dataset (5784 query patches). Similarly, Table 5.7 shows the comparative study of proposed segmentation free technique with segmentation based techniques proposed by the author in preceding chapters, where the proposed technique performance increases (98.43 for GW and 96.53 for IAM dataset) compared to its results presented in Table 5.3 and Table 5.5.

The increase in the performance purely depends upon nature of query words. While comparing the performance of proposed technique with existing segmentation free techniques, we have used query patches which may or may not fit actual words and its instances. However, while comparing the performance of

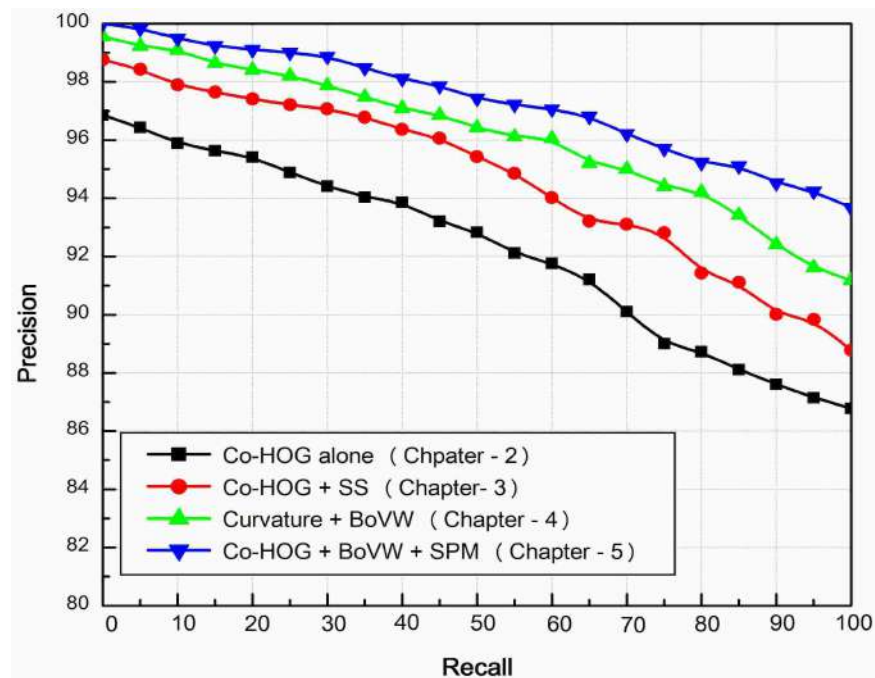


Figure 5.8: PrecisionRecall curves for different configurations for GW dataset

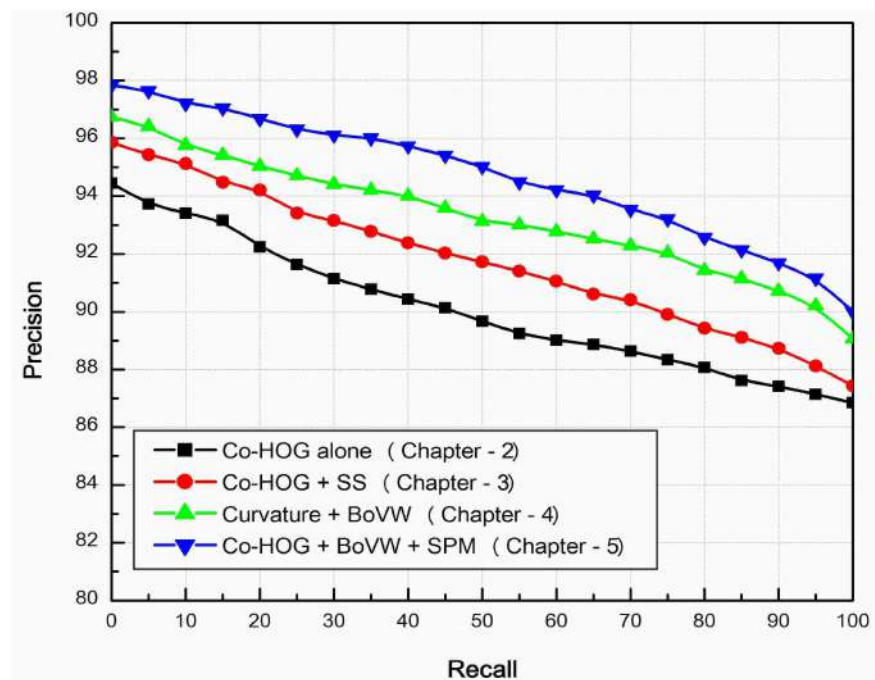


Figure 5.9: PrecisionRecall curves for different configurations for IAM dataset

proposed technique with segmentation based techniques presented in preceding chapters, we have considered query words which represent actual words contained in the documents. Hence, we have observed empirically that the performance of proposed word spotting method is quite dependent on the available number of training sets. This limitation can be handled by larger sets which provide learning of more number of short and substrings could improve the overall performance of proposed approach.

5.7 Chapter Summary

In this chapter, a segmentation free word spotting method for handwritten documents has been proposed. We have presented a systematic analysis of the proposed method using the popular datasets and evaluation of all the parameters, in order to assess the configuration which maximizes the word spotting performance. We offered a comprehensive comparison with state-of-the-art segmentation free word spotting methods, it is observed that, the proposed method achieves the highest accuracy compared to existing segmentation free methods for both types of handwritten documents such as historical documents (GW) and modern documents (IAM) which are a single writer and multi writer documents. We also conducted performance comparison of proposed segmentation free technique with segmentation based word spotting techniques presented in preceding chapters for both the datasets. It is observed from the experimental results, the proposed segmentation free based word spotting technique outperforms segmentation based techniques proposed by the author.

Chapter 6

Conclusions and Future Directions

Chapter 6

Conclusions and Future Directions

6.1 Conclusions

In this thesis, we addressed and investigated segmentation based as well as segmentation free based word spotting methods for handwritten documents. Through various experiments conducted using standard benchmark offline handwritten English text datasets, such as GW (historical documents), IAM (modern documents) and Bentham (modern documents), it is demonstrated that the proposed approaches achieve the highest accuracy compared to existing state-of-the-art word spotting methods and more robust against handwritten documents which are having non-uniform illumination, suffering from the noise and written by different writers. The main contributions of our research work presented in this thesis are summarized as follows:

We proposed a segmentation based word spotting method for handwritten document images using Co-HOG descriptor. The proposed approach efficiently retrieves the handwritten words and outperforms the existing segmentation based techniques. This is due to the fact that Co-HOG descriptor keeps robustness to illumination variation and invariance to local geometric transformation. Another factor for highest accuracy is a division of the word image into blocks which capture the local variations within the class. The main advantage of this approach is that there is minimum learning involved.

We proposed a segmentation based word spotting method using Scale Space Co-HOG descriptor for handwritten documents. The experimental results demonstrate that considerable improvement of accuracy compared to existing segmentation based word spotting techniques. Scale space representation dominates more discriminating ability than uni scale, this is due to properties of coarser and finer

scale. The coarser scale provides more resistant to variation occurring in a different writing style, and the finer scale provides the local information about image gradient and stroke orientation.

We proposed a segmentation based word spotting method using the BoVW framework powered by curvature features. The curvature feature describes the geometrical shape of the stroke. The experimental results show that proposed approach accurately retrieve words and its instances similar to query words from a large collection of handwritten documents. And also the proposed method outperforms existing SIFT or SURF based word spotting methods because the curvature feature is robust with respect to noise, scale, orientation and it preserves the local spatial information of the word shape. The highest accuracy is due to the fact that adaptation of BoVW framework which reduces redundancy of information occurred in features. The advantage of our approach is that it uses less memory space because of scalar value extracted at every corner point when compared to SIFT or SURF descriptors where feature vector is extracted at every keypoint.

We proposed a segmentation free based word spotting method using the BoVW framework in conjunction with Co-HOG feature descriptor and SPM technique. The performance of the approach is improved when the spatial information incorporated into the BoVW framework through Spatial Pyramid Matching technique. The spatial pyramid matching representation yields the geometric invariance properties which are not provided by BoVW framework alone. An enhanced BoVW framework integrates the global and local information which can produce more discriminative visual words. The qualitative results show that, such representation is able to retrieve/spot the document information efficiently in segmentation free scenarios.

Based on comparative study of results of state-of-the-art word spotting techniques with results of proposed word spotting techniques presented in this thesis, we can conclude that the segmentation-free based word spotting technique (Co-HOG + BoVW + SPM) outperforms state-of-the-art segmentation based as well

as segmentation free based word spotting techniques for offline handwritten documents. The high performance and efficiency of the proposed segmentation free based word spotting method can be employed for industrial applications, such as automatic mail sorting, retrieval of handwritten documents from digital libraries and etc.

6.2 Future Directions

It is observable from the experimental results that further research is needed to improve the efficiency and effectiveness of the proposed word spotting methods.

- The proposed word spotting method can be extended in order to reduce false positive rate through combining kernel functions with a BoVW framework which improves the bag of visual words model
- The second possibility is combination of Pruned Dictionary (codebook) with BoVW framework based on the Latent Semantic Topic (PD-LST) method provides discriminative power of a visual word in the codebook so that less meaningful words are removed to enable better similarity computation between word images and PD-LST method can largely reduce the number of required words, leading to higher retrieval efficiency
- The effectiveness of different low-level features, such as texture, SURF and Hashing features can be adopted for constructing the BoVW model
- The proposed word spotting methods can be extended by incorporating other newly emerging techniques of pattern recognition and machine learning in order to increase the accuracy and efficiency to overcome the limitations of the proposed techniques
- There is a scope to extend proposed word spotting techniques for other scripts with slight modifications.

Chapter 7

Author's Publications

Refereed International Journals

1. Thontadari C. and Prabhakar C.J.: Scale Space Co-Occurrence HOG Features for Word Spotting in Handwritten Document Images, *International Journal of Computer Vision and Image Processing (IJCVIP)*, vol. 6, issue 2, page 71-86, 2016.
2. Thontadari C. and Prabhakar C.J.: Segmentation Based Word Spotting Method for Handwritten Documents, *International Journals of Advanced Research in Computer Science and Software Engineering (IJARCSSE)*, vol. 7, issue 6, page 34-40, 2017.
3. Thontadari C. and Prabhakar C.J.: Bag of Visual Words for Word Spotting in Handwritten Documents Based on Curvature Features, *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 9, issue 4, page 77-92, 2017.
4. Thontadari C. and Prabhakar C.J.: Segmentation free Word Spotting for Handwritten Documents Using Bag of Visual Words Based on Co-HOG Features, *International Journal of Information Retrieval Research (IJIRR)*, vol. 9, issue 1, 2018.

Refereed Book Chapter

1. Thontadari C. and Prabhakar C.J.: Bag of Visual Words Based on Co-HOG Features for Word Spotting in Handwritten Documents, *Book entitled Advancements in Computer Vision and Image Processing (ACVIP)*, Editor-Jose Garcia-Rodriguez, IGI Global Inc., USA, 2018.

Bibliography

- Abidi, A., Jamil, A., Siddiqi, I., and Khurshid, K.: Word spotting based retrieval of urdu handwritten documents. *Proc. International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 331-336, 2012.
- Aldavert, D., Rusinol, M., Toledo, R., and Lladós, J.: Integrating visual and textual cues for query-by-string word spotting. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 511-515, 2013.
- Aldavert, D., Rusinol, M., Toledo, R., and Lladós, J.: A study of Bag-of-Visual-Words representations for handwritten keyword spotting. *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 18, issue 3, page 223-234, 2015.
- Almazan, J., Gordo, A., Fornes, A., and Valveny, E.: Efficient exemplar word spotting. *Proc. of the British Machine Vision Conf, Chicago*, vol. 1, issue 2, page 67.1-67.11, 2012.
- Almazan, J., Gordo, A., Fornes, A., and Valveny, E.: Handwritten word spotting with corrected attributes. *Proc. of International Conf. on Computer Vision*, page 1017-1024, 2013.
- Almazan, J., Gordo, A., Fornes, A., and Valveny, E.: Segmentation-free word spotting with exemplar SVMs. *Pattern Recognition*, vol. 47, issue 12, page 3967-3978, 2014.

- Almazan, J., Gordo, A., Fornes, A., and Valveny, E.: Word spotting and recognition with embedded attributes. *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 36, issue 12, page 2552-2566, 2014.
- Arrospeide, J., Salgado, L., and Marinas, J.: HOG-like gradient-based descriptor for visual vehicle detection. *Proc. of IEEE Intelligent Vehicles Symposium (IV)*, page 223-228, 2013.
- Asada, H., and Brady, M.: The curvature primal sketch. *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 1, page 2-14, 1986.
- Bai, S., Li, L., and Tan, C. L.: Keyword spotting in document images through word shape coding. *Proc. of International Conf. Document Analysis and Recognition (ICDAR)*, page 331-335, 2009.
- Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L.: Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, vol. 110, issue 3, page 346-359, 2008.
- Can, E. F., and Duygulu, P.: A line-based representation for matching words in historical manuscripts. *Pattern Recognition Letters*, vol. 32, issue 8, page 1126-1138, 2011.
- Cao, H., Bhardwaj, A., and Govindaraju, V.: A probabilistic method for keyword retrieval in handwritten document images. *Journal of Pattern Recognition*, vol. 42, issue 12, page 3374-3382, 2009.
- Carcagni, P., Del Coco, M., Leo, M., and Distanto, C.: Facial expression recognition and histograms of oriented gradients: a comprehensive study. *Springer-Plus*, vol.4, issue 1, page 645, 2015.
- Chatbri, H., Kwan, P., and Kameyama, K.: An application-independent and segmentation-free approach for spotting queries in document images. *Proc. of International Conf. on Pattern Recognition (ICPR)*, page 2891-2896, 2014.

- Chan, J., Ziftci, C., and Forsyth, D.: Searching off-line arabic documents. *Proc. of IEEE International Conf. on Computer Vision and Pattern Recognition*, vol. 2, page 1455-1462, 2006.
- Chen, F. R., Bloomberg, D. S., and Wilcox, L. D.: Spotting phrases in lines of imaged text. *Proc. of International Society for Optics and Photonics Document Recognition II*, vol. 2422, page 256-270, 1995.
- Chen, F. R., Wilcox, L. D., and Bloomberg, D. S.: Word spotting in scanned images using hidden Markov models. *Proc. of IEEE International Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, page 1-4, 1993.
- Choisy, C. Dynamic handwritten keyword spotting based on the NSHP-HMM. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, vol. 1, page 242-246, 2007.
- Corvee, E., and Bremond, F.: Body parts detection for people tracking using trees of histogram of oriented gradient descriptors. *Proc. of Seventh IEEE International Conf. on Advanced Video and Signal Based Surveillance*, page 469-475, 2010.
- Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C.: Visual categorization with bags of keypoints. *Proc. of ECCV International Workshop on Statistical Learning in Computer Vision*, vol. 1, page 1-22, 2004.
- Dalal, N., and Triggs, B.: Histograms of oriented gradients for human detection. *Proc. of IEEE International Conf. on Computer Vision and Pattern Recognition*, vol. 1, page 886-893, 2005.
- Deniz, O., Bueno, G., Salido, J., and De la Torre, F.: Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, vol. 32, issue 12, page 1598-1603, 2011.
- Dey, S., Nicolaou, A., Llados, J., and Pal, U.: Local binary pattern for word spotting in handwritten historical document. *Proc. of Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR)*

-
- and Structural and Syntactic Pattern Recognition (SSPR)*, page 574-583, 2016.
- Do, T. T., and Kijak, E.: Face recognition using co-occurrence histograms of oriented gradients. *Proc. of IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, page 1301-1304, 2012.
- Dovgalecs, V., Burnett, A., Tranouez, P., Nicolas, S., and Heutte, L.: Spot it! Finding words and patterns in historical documents. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 1039-1043, 2013.
- Fei-Fei, L., Fergus, R., and Perona, P.: Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer vision and Image understanding*, vol. 106, issue 1, pp. 59-70, 2007.
- Fernandez, S., Graves, A., and Schmidhuber, J.: An application of recurrent neural networks to discriminative keyword spotting. *Artificial Neural Networks*, page 220-229, 2007.
- Fink, G. A., Rothacker, L., and Grzeszick, R.: Grouping historical postcards using query-by-example word spotting. *Proc. International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 470-475, 2014.
- Fischer, A., Keller, A., Frinken, V., and Bunke, H.: HMM-based word spotting in handwritten documents using subword models. *Proc. 20th International Conf. on Pattern Recognition*, page 3416-3419, 2010.
- Fischer, A., Frinken, V., Bunke, H., and Suen, C. Y.: Improving HMM-based keyword spotting with character language models. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 506-510, 2013.
- Fornes, A., Frinken, V., Fischer, A., Almazan, J., Jackson, G., and Bunke, H.: A keyword spotting approach using blurred shape model-based descriptors.

-
- Proc. of International workshop on Historical Document Imaging and Processing* page 83-90, 2011.
- Frinken, V., Fischer, A., and Bunke, H.: A novel word spotting algorithm using bidirectional long short-term memory neural networks. *Artificial Neural Networks in Pattern Recognition*, page 185-196, 2011.
- Frinken, V., Fischer, A., Manmatha, R., and Bunke, H.: A novel word spotting method based on recurrent neural networks. *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 34, issue 2, page 211-224, 2012.
- Gatos, B., and Pratikakis, I.: Segmentation-free word spotting in historical printed documents. *Proc. of International Conf. on Document Analysis and Recognition (ICADR)*, page 271-275, 2009.
- Ghosh, S. K., and Valveny, E.: Query by string word spotting based on character bi-gram indexing. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 881-885, 2015.
- Giotis, A. P., Gerogiannis, D. P., and Nikou, C.: Word spotting in handwritten text using contour-based models. *Proc. International Conf. Frontiers in Handwriting Recognition (ICFHR)*, page 399-404, 2014.
- Giotis, A. P., Sfikas, G., Gatos, B., and Nikou, C.: A survey of document image word spotting techniques. *Pattern Recognition*, vol. 68, page 310-332, 2017.
- Harris, C., and Stephens, M.: A combined corner and edge detector. *Proc. of International Conf. on Alvey vision*, vol. 15, issue 50, page 10-5244, 1988.
- Hassan, E., Chaudhury, S., and Gopal, M.: Word shape descriptor-based document image indexing: a new DBH-based approach. *International Journal on Document Analysis and Recognition (IJ DAR)*, vol. 16, issue 3, page 227-246, 2013.
- Hast, A., and Vats, E.: Radial Line Fourier Descriptor for Handwritten Word Representation, arXiv preprint, 2017.

- Howe, N. R.: Part-structured inkball models for one-shot handwritten word spotting. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 582-586, 2013.
- Howe, N. R.: Inkball models for character localization and out-of-vocabulary word spotting. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 381-385, 2015.
- Huang, L., Yin, F., Chen, Q. H., and Liu, C. L.: Keyword spotting in unconstrained handwritten Chinese documents using contextual word model. *Image and Vision Computing*, vol. 31, issue 12, page 958-968, 2013.
- Jain, R., and Doermann, D.: Logo retrieval in document images. *Proc. of IAPR International Workshop on Document Analysis Systems (DAS)*, page 135-139, 2012.
- Javed, M.: On the possibility of processing document images in compressed domain. *PhD thesis, Department of Studies in Computer Science, University of Mysore*, 2016.
- Johansson, S., Leech, G., and Goodluck, H.: Manual of Information to Accompany the Lancaster-Olds/Bergen Corpus of British English, for Use with Digital Computers, 1978.
- Jones, G. J., Foote, J. T., Jones, K. S., and Young, S. J.: Video mail retrieval: The effect of word spotting accuracy on precision. *Proc. of International Conf. on Acoustics, Speech, and Signal Processing*, vol. 1, page 309-312, 1995.
- Kannan, B., Jomy, J., and Pramod, K. V.: A system for offline recognition of handwritten characters in Malayalam script, 2013.
- Keaton, P., Greenspan, H., and Goodman, R.: Keyword spotting for cursive document retrieval. *Proc. of International Workshop on Document Image Analysis*, page 74-81, 1997.

- Kesidis, A. L., Galiotou, E., Gatos, B., and Pratikakis, I.: A word spotting framework for historical machine-printed documents. *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 14, issue 2, page 131-144, 2011.
- Kessentini, Y., Chatelain, C., and Paquet, T.: Word spotting and regular expression detection in handwritten documents. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 516-520, 2013.
- Kessentini, Y., and Paquet, T.: Keyword spotting in handwritten documents based on a generic text line HMM and a SVM verification. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 41-45, 2015.
- Khayyat, M., Lam, L., and Suen, C. Y.: Learning-based word spotting system for Arabic handwritten documents. *Pattern Recognition*, vol. 47, issue 3, page 1021-1030, 2014.
- Khoubyari, S., and Hull J. J.: Keyword location in noisy document images. *Proc. of 2nd Annual Symposium on Document Analysis and Information Retrieval*, page 217-231, 1993.
- Khurshid, K., Faure, C., and Vincent, N.: Word spotting in historical printed documents using shape and sequence comparisons. *Pattern Recognition*, vol. 45, issue 7, page 2598-2609, 2012.
- Kim, S. H., Park, S. C., Jeong, C. B., Kim, J. S., Park, H. R., and Lee, G. S.: Keyword spotting on Korean document images by matching the keyword image. *Proc. on International Conf. on Asian Digital Libraries (ICADL)*, page 158-166, 2005.
- Kolcz, A., Alspector, J., Augusteijn, M., Carlson, R., and Popescu, G. V.: A line-oriented approach to word spotting in handwritten documents. *Pattern Analysis & Applications*, vol. 3, issue 2, page 153-168, 2000.

- Konidakis, T., Gatos, B., Ntzios, K., Pratikakis, I., Theodoridis, S., and Perantonis, S. J.: Keyword-guided word spotting in historical printed documents using synthetic data and user feedback. *International Journal of Document Analysis and Recognition (IJ DAR)* vol.9 issue 2-4, page 167-177, 2007.
- Kumar, G., Shi, Z., Setlur, S., Govindaraju, V., and Ramachandrupa, S.: Keyword spotting framework using dynamic background model. *Proc. of International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 582-587, 2012.
- Kumar, G., and Govindaraju, V.: A Bayesian approach to script independent multilingual keyword spotting. *Proc. of International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 357-362, 2014.
- Kuo, S. S., and Agazzi, O. E.: Keyword spotting in poorly printed documents using pseudo 2-D hidden Markov models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, issue 8, 842-848, 1994.
- Krishnan, P., and Jawahar, C. V.: Bringing semantics in word image retrieval. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 733-737, 2013.
- Krishnan, P., Dutta, K., and Jawahar, C. V.: Deep feature embedding for accurate recognition and retrieval of handwritten text. *Proc. of International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 289-294, 2016.
- Lavrenko, V., Rath, T. M., and Manmatha, R.: Holistic word recognition for handwritten historical documents. *Proc. of First International Workshop on Document Image Analysis for Libraries*, page 278-287, 2004.
- Lazebnik, S., Schmid, C., and Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. *Proc. of IEEE International Conf. on Computer Vision and Pattern Recognition*, vol. 2, page 2169-2178, 2006.

- Le Bourgeois, F., and Emptoz, H.: Omnilingual segmentation-free word spotting for ancient manuscripts indexation. *Proc. of IEEE International Conf. on Document Analysis and Recognition (ICDAR)*, page 533-537, 2005.
- Leydier, Y., Le Bourgeois, F., and Emptoz, H.: Omnilingual segmentation-free word spotting for ancient manuscripts indexation. *Proc. of IEEE International Conf. on Document Analysis and Recognition (ICDAR)*, page 533-537, 2005.
- Leydier, Y., Lebourgeois, F., and Emptoz, H.: Text search for medieval manuscript images. *Pattern Recognition*, vol. 40, page 3552-3567, 2007.
- Leydier, Y., Ouji, A., LeBourgeois, F., and Emptoz, H.: Towards an omnilingual word retrieval system for ancient manuscripts. *Pattern Recognition*, vol. 42, issue 9, page 2089-2105, 2009.
- Li, N., Chen, J., Cao, H., Zhang, B., and Natarajan, P.: Applications of recurrent neural network language model in offline handwriting recognition and word spotting. *Proc. of International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 134-139, 2014.
- Li, L., Lu, S. J., and Tan, C. L.: A fast keyword-spotting technique. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, vol. 1, page 68-72, 2007.
- Liang, Y., Fairhurst, M. C., and Guest, R. M.: A synthesised word approach to word retrieval in handwritten documents. *Pattern Recognition*, vol. 45, issue 12, page 4225-4236, 2012.
- Lindeberg, T.: Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, vol. 21, vol.1-2, page 225-270, 1994.
- Lladós, J., Rusinol, M., Fornes, A., Fernandez, D., and Dutta, A.: On the influence of word representations for handwritten word spotting in historical

- documents. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, vol. 26, issue 5, 2012.
- Long, D. G.: 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, 1981.
- Louloudis, G., Gatos, B., Pratikakis, I., and Halatsis, C.: Text line and word segmentation of handwritten documents. *Pattern Recognition*, vol. 42, issue 12, page 3169-3183, 2009.
- Louloudis, G., Kesidis, A. L., and Gatos, B.: Efficient word retrieval using a multiple ranking combination scheme. *Proc. of International Workshop on Document Analysis Systems (DAS)*, page 379-383, 2012.
- Lowe, D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, vol. 60, issue 2, page 91-110, 2004.
- Lu, Y., and Tan, C. L.: Word spotting in Chinese document images without layout analysis. *Proc. of International Conf. on Pattern Recognition*, vol. 3, page 57-60, 2002.
- Madhvanath, S., and Govindaraju, V.: The role of holistic paradigms in handwritten word recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, issue 2, page 149-164, 2001.
- Manmatha, R., and Croft, W. B.: Word spotting: Indexing handwritten archives. *Intelligent Multimedia Information Retrieval Collection*, page 43-64, 1997.
- Manmatha, R., and Rath, T.: Indexing of handwritten historical documents-recent progress. *Proc. of Symposium on Document Image Understanding Technology*, page 77-86, 2003.

- Manmatha, R., Han, C., and Riseman, E. M.: Word spotting: A new approach to indexing handwriting. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, page 631- 637, 1996.
- Marinai, S.: Text retrieval from early printed books. *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 14, issue 2, page 117-129, 2011.
- Marti, U. V., and Bunke, H.: Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system. *International Journal of Pattern Recognition and Artificial intelligence*, vol. 15, issue 01, page 65-90, 2001.
- Marti, U. V., and Bunke, H.: The IAM-database: An English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 5, issue 1, page 39-46, 2002.
- Meshesha, M., Jawahar, C. V.: Matching word images for content-based retrieval from printed document images. *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 11, issue 1, page 29-38, 2008.
- Minetto, R., Thome, N., Cord, M., Leite, N. J., and Stolfi, J.: T-HOG: An effective gradient-based descriptor for single line text regions. *Pattern Recognition*, vol. 46, issue 3, page 1078-1090, 2013.
- Mokhtarian, F., and Mackworth, A. K.: A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, issue 8, page 789-805, 1992.
- Mondal, T., Ragot, N., Ramel, J. Y., and Pal, U.: A fast word retrieval technique based on kernelized locality sensitive hashing. *Proc. International Conf. on Document Analysis and Recognition (ICDAR)*, page 1195-1199, 2013.
- Mondal, T., Ragot, N., Ramel, J. Y., and Pal, U.: Flexible sequence matching

- technique: Application to word spotting in degraded documents. *Proc. International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 210-215, 2014.
- Mondal, T., Ragot, N., Ramel, J. Y., and Pal, U.: Performance evaluation of DTW and its variants for word spotting in degraded documents. *Proc. International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 1141-1145, 2015.
- Nagy, G., and Lopresti, D.: Interactive document processing and digital libraries. *Proc. of Second International Conf. on Document Image Analysis for Libraries (DIAL)*, page 8, 2006.
- Newell, A. J., and Griffin, L. D.: Multiscale histogram of oriented gradient descriptors for robust character recognition. *Proc. of International Conf. on Document Analysis and Recognition*, page 1085-1089, 2011.
- Papandreou, A., Gatos, B., and Louloudis, G.: An adaptive zoning technique for efficient word retrieval using dynamic time warping. *Proc. of International Conf. on Digital Access to Textual Cultural Heritage*, page 147-152, 2014.
- Papavassiliou, V., Stafylakis, T., Katsouros, V., and Carayannis, G.: Handwritten document image segmentation into text lines and words. *Pattern Recognition*, vol. 43, issue 1, page 369-377, 2010.
- Perronnin, F., and Rodriguez-Serrano, J.: Fisher kernels for handwritten wordspotting. *Proc. International Conf. on Document Analysis and Recognition*, vol. 1, page 106-110, 2009.
- Plamondon, R., and Srihari, S.: Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, issue 1, page 63-84, 2000.
- Puigcerver, J., Toselli, A. H., and Vidal, E.: Word-graph-based handwriting keyword spotting of out-of-vocabulary queries. *Proc. International Conf. on Pattern Recognition (ICPR)*, page 2035-2040, 2014.

- Rath, T. M., and Manmatha, R.: Features for word spotting in historical manuscripts. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 218-222, 2003.
- Rath, T. M., and Manmatha, R.: Word image matching using dynamic time warping. *Proc. of IEEE International Conf. on Computer Vision and Pattern Recognition*, vol. 2, page II-II, 2003.
- Rath, T. M., and Manmatha, R.: Word spotting for historical documents. *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 9, issue 2, page 139-152, 2007.
- Rath, T. M., Lavrenko, V., and Manmatha, R.: A statistical approach to retrieving historical manuscript images without recognition (*No. CIIR-MM-42*). *Space and Naval Warfare Systems Center San Diego CA*, 2003.
- Rath, T. M., Lavrenko, V., and Manmatha, R.: Retrieving historical manuscripts using shape. *Massachusetts Univ. Amherst Center For Intelligent Information Retrieval*, 2003.
- Ranjan, V., Harit, G., and Jawahar, C. V.: Enhancing word image retrieval in presence of font variations. *Proc. International Conf. on Pattern Recognition (ICPR)*, page 2709-2714, 2014.
- Ren, H., Heng, C. K., Zheng, W., Liang, L., and Chen, X.: Fast object detection using boosted co-occurrence histograms of oriented gradients. *Proc. of International Conf. on Image Processing (ICIP)*, page 2705-2708, 2010.
- Riba, P., Lladas, J., and Fornes, A.: Handwritten word spotting by inexact matching of grapheme graphs. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 781-785, 2015.
- Retsinas, G., Louloudis, G., Stamatopoulos, N., and Gatos, B.: Keyword spotting in handwritten documents using projections of oriented gradients. *Proc. of IAPR Workshop on Document Analysis Systems (DAS)*, page 411-416, 2016.

- Rodriguez, J. A., and Perronnin, F.: Local gradient histogram features for word spotting in unconstrained handwritten documents. *Proc. of First International Conf. on Frontiers in Handwriting Recognition*, page 7-12, 2008.
- Rodriguez-Serrano, J. A., and Perronnin, F.: Handwritten word-spotting using hidden Markov models and universal vocabularies. *Pattern Recognition*, vol. 42, issue 9, 2106-2116, 2009.
- Rodriguez-Serrano, J. A., and Perronnin, F.: A model-based sequence similarity with application to handwritten word spotting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, issue 11, page 2108-2120, 2012.
- Rodriguez-Serrano, J. A., Perronnin, F., Snchez, G., and Llads, J.: Unsupervised writer adaptation of whole-word HMMs with application to word-spotting. *Pattern Recognition Letters*, vol. 31, issue 8, page 742-749, 2010.
- Rodriguez-Serrano, Saidani, A., Kacem Echi, A., and Belaid, A.: Arabic/Latin and Machine-printed/Handwritten Word Discrimination using HOG-based Shape Descriptor. *Electronic Letters on Computer Vision and Image Analysis (ELCVIA)*, page 0001-23, 2015.
- Rodriguez-Serrano, Shekhar, R., and Jawahar, C. V.: Word image retrieval using bag of visual words. *Proc. International Workshop on Document Analysis Systems (DAS)*, page 297-301, 2012.
- Roy, P. P., Ramel, J. Y., and Ragot, N.: Word retrieval in historical document using character-primitives. *Proc. International Conf. on Document Analysis and Recognition (ICDAR)*, page 678-682, 2011.
- Roy, U., Sankaran, N., Sankar, K. P., and Jawahar, C. V.: Character n-gram spotting on handwritten documents using weakly-supervised segmentation. *Proc. of International Conf. Document Analysis and Recognition (ICDAR)*, page 577-581, 2013.

- Rothacker, L., Fisseler, D., Muller, G. G., Weichert, F., and Fink, G. A.: Retrieving cuneiform structures in a segmentation-free word spotting framework. *Proc. of International Workshop on Historical Document Imaging and Processing*, page 129-136, 2015.
- Rothacker, L., Rusinol, M., and Fink, G.: Bag-of-features HMMs for segmentation-free word spotting in handwritten documents, *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 1305-1309, 2013.
- Rothfeder, J. L., Feng, S., and Rath, T. M.: Using corner feature correspondences to rank word images by similarity. *Proc. International Workshop on Computer Vision and Pattern Recognition*, vol. 3, page 30-30, 2003.
- Rusinol, M., Rodriguez-Serrano, Aldavert, D., Toledo, R., and Lladós, J.: Browsing heterogeneous document collections by a segmentation-free word spotting method. *Proc. International Conf. on Document Analysis and Recognition (ICDAR)*, page 63-67, 2011.
- Rusinol, M., Aldavert, D., Toledo, R., and Lladós, J.: Efficient segmentation-free keyword spotting in historical document collections. *Pattern Recognition*, vol. 48, issue 2, page 545-555, 2015.
- Saabni, R., and Bronstein, A.: Fast keyword searching using boostmap based embedding. *Proc. International Conf. on In Frontiers in Handwriting Recognition (ICFHR)*, page 734-739, 2012.
- Sagheer, M. W., Nobile, N., He, C. L., and Suen, C. Y.: A novel handwritten urdu word spotting based on connected components analysis. *Proc. of International Conf. on Pattern Recognition (ICPR)*, page 2013-2016, 2010.
- Schroth, G., Hilsenbeck, S., Huitl, R., Schweiger, F., and Steinbach, E.: Exploiting text-related features for content-based image retrieval. *Proc. of International Symposium on Multimedia (ISM)*, page 77-84, 2011.

- Scot, G. L., and Loguet-Higgins, H. C.: An algorithm for associating the features of two patterns. *Proc. of Royal Society of London*, vol. 224, page 21-26, 1991.
- Sfikas, G., Giotis, A. P., Louloudis, G., and Gatos, B.: Using attributes for word spotting and recognition in polytonic greek documents. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 686-690, 2015.
- Sfikas, G., Gatos, B., and Nikou, C.: SEMICCA: A New Semi-Supervised Probabilistic CCA Model For Keyword Spotting. *Proc. of International Conf. Image Processing*, page 1107-1111, 2017.
- Shah, M. I., and Suen, C. Y.: Word spotting in gray scale handwritten Pashto documents. *Proc. International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 136-141, 2010.
- Shahab, S. A., Al-Khatib, W. G., and Mahmoud, S. A.: Computer aided indexing of historical manuscripts. *Proc. International Conf. on Computer Graphics, Imaging and Visualisation*, page 287-295, 2006.
- Saidani, A., and Echi, A. K.: Pyramid histogram of oriented gradient for machine-printed/handwritten and arabic/latin word discrimination. *Proc. of International Conf. on Soft Computing and Pattern Recognition (SoC-PaR)*, page 267-272, 2015.
- Shekhar, R., and Jawahar, C. V.: Document specific sparse coding for word retrieval. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 643-647, 2013.
- Shi, M., Fujisawa, Y., Wakabayashi, T., and Kimura, F.: Handwritten numeral recognition using gradient and curvature of gray scale image. *Pattern Recognition*, vol. 35, issue 10, page 2051-2059, 2002.
- Shi, Z., Setlur, S., and Govindaraju, V.: A steerable directional local profile

- technique for extraction of handwritten arabic text lines. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 176-180, 2009.
- Shu, C., Ding, X., and Fang, C.: Histogram of the oriented gradient for face recognition. *Tsinghua Science & Technology*, vol. 16, issue 2, page 216-224, 2011.
- Sivic, J., and Zisserman, A.: Video google: A text retrieval approach to object matching in videos. *Proc. of International Conf. on Computer Vision (ICCV)*, vol. 2, issue 1470, page 1470-1477, 2003.
- Smith, D. J., and Harvey, R. W.: Document Retrieval Using SIFT Image Features. *Journal of Universal Computer Science (JUICS)*, vol. 17, issue 1, page 3-15, 2011.
- Sousa, J. M., Gil, J. M., and Pinto, J. R. C.: Word indexing of ancient documents using fuzzy classification. *IEEE Transactions on Fuzzy Systems*, vol. 15, issue 5, page 852-862, 2007.
- Srihari, S. N., Huang, C., and Srinivasan, H.: Search engine for handwritten documents, *Proc. of Society of Photographic Instrumentation Engineers (SPIE) 5676*, page 66-75, 2005.
- Srihari, S. N., Srinivasan, H., Babu, P., and Bhole, C.: Handwritten arabic word spotting using the cedarabic document analysis system. *Proc. of Symposium on Document Image Understanding Technology*, page 123-132, 2005.
- Srihari, S. N., Srinivasan H., Haung C., and Shetty, S.: Spotting words in Latin, Devanagari and Arabic scripts. *Indian Journal of Artificial Intelligence*, vol. 16, issue 3, page 2-9, 2006.
- Srihari, S. N., and Ball, G. R.: Language independent word spotting in scanned documents. *Proc. of International Conf. on Asian Digital Libraries*, page 134-143, 2008.

- Su, B., Lu, S., Tian, S., Lim, J. H., and Tan, C. L.: Character recognition in natural scenes using convolutional co-occurrence hog. *Proc. of 22nd International Conf. on Pattern Recognition*, page 2926-2931, 2014.
- Sudholt, S., and Fink, G. A.: PHOCNet: A deep convolutional neural network for word spotting in handwritten documents. *Proc. International Conf. Frontiers in Handwriting Recognition (ICFHR)*, page 277-282, 2016.
- Terasawa, K., and Tanaka, Y.: Slit style HOG feature for document image word spotting. *Proc. of International Conf. on Document Analysis and Recognition*, page 116-120, 2009.
- Thontadari, C., and Prabhakar, C. J.: Bag of visual words for word spotting in handwritten documents based on curvature features. *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 9, issue 4, page 77-92, 2017.
- Thontadari, C., and Prabhakar, C. J.: Scale space Co-occurrence HOG features for word spotting in handwritten document images. *International Journal of Computer Vision and Image Processing (IJCVIP)*, vol. 6, issue 2, page 71-86, 2016.
- Thontadari, C., and Prabhakar, C. J.: Segmentation based word spotting method for handwritten documents. *International Journals of Advanced Research in Computer Science and Software Engineering (IJARCSSE)*, vol. 7, issue 6, page 35-40, 2017.
- Tian, S., Bhattacharya, U., Lu, S., Su, B., Wang, Q., Wei, X., and Tan, C. L.: Multilingual scene character recognition with co-occurrence of histogram of oriented gradients. *Pattern Recognition*, vol. 51, page 125-134, 2015.
- Tian, S., Lu, S., Su, B., and Tan, C. L.: Scene text recognition using co-occurrence of histogram of oriented gradients. *Proc. of International Conf. on Document Analysis and Recognition*, page 912-916, 2013.

- Toselli, A. H., Vidal, E., Romero, V., and Frinken, V.: HMM word graph based keyword spotting in handwritten document images. *Information Sciences*, vol. 370, page 497-518, 2016.
- Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., and Gong, Y.: Locality-constrained linear coding for image classification. *Proc. of International Conf. on Computer Vision and Pattern Recognition (CVPR)*, page 3360-3367, 2010.
- Watanabe, T., Ito, S., and Yokoi, K.: Co-occurrence histograms of oriented gradients for pedestrian detection. *In Advances in Image and Video Technology, Springer Berlin Heidelberg*, page 37-47, 2009.
- Wagan, A. I., Bres, S., and Emptoz, H.: Word spotting in Alices adventures underground using multi scale integral orientation features. *Proc. of International Workshop on Document Analysis Systems (DAS)*, page 417-424, 2010.
- Wei, H., Gao, G., and Bao, Y.: A method for removing inflectional suffixes in word spotting of Mongolian Kanjur. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 88-92, 2011.
- Wei, H., Gao, G., and Su, X.: A multiple instances approach to improving keyword spotting on historical Mongolian document images. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 121-125, 2015.
- Wilkinson, T., and Brun, A.: Semantic and verbatim word spotting using deep neural networks. *Proc. of International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 307-312, 2016.
- Wshah, S., Kumar, G., and Govindaraju, V.: Statistical script independent word spotting in offline handwritten documents. *Pattern Recognition*, vol. 47, issue 3, page 1039-1050, 2014.

- Xia, Y., Wang, K., and Li, M.: Chinese keyword spotting using knowledge-based clustering. *Proc. of International Conf. on Document Analysis and Recognition (ICDAR)*, page 789-793, 2011.
- Yalniz, I. Z., and Manmatha, R.: An efficient framework for searching text in noisy document images. *Proc. of International Workshop on Document Analysis Systems (DAS)*, page 48-52, 2012.
- Yan, H.: Skew correction of document images using interline cross-correlation. *Computer Vision, Graphics, and Image Processing : Graphical Model and Image Processing (CVGIP:GMIP)*, vol. 55, issue 6, page 538-543, 1993.
- Zagoris, K., Pratikakis, I., and Gatos, B.: Segmentation-based historical handwritten word spotting using document-specific local features. *Proc. of 14th International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 9-14, 2014.
- Zagoris, K., Ergina, K., and Papamarkos, N.: A document image retrieval system. *Engineering Applications of Artificial Intelligence*, vol. 23, issue 6, page 872-879, 2010.
- Zhang, B., Srihari, S. N., and Huang, C.: Word image retrieval using binary features. In *Document Recognition and Retrieval XI, International Society for Optics and Photonics. 5296*, page 45-54, 2004.
- Zhang, X., and Tan, C. L.: Segmentation-free keyword spotting for handwritten documents based on heat kernel signature. *Proc. International Conf. on Document Analysis and Recognition (ICDAR)*, page 827-831, 2013.
- Zhong, Z., Pan, W., Jin, L., Mouchere, H., and Viard-Gaudin, C.: SpottingNet: learning the similarity of word images with convolutional neural network for word spotting in handwritten historical documents. *Proc. International Conf. on Frontiers in Handwriting Recognition (ICFHR)*, page 295-300, 2016.