الجمهورية الجزائرية الديمقراطية الشعبية
People's Democratic Republic of Algeria

Ministry Of Higher Education
and Scientific Research
Akli Mohand Oulhadj Bouira University
Tasdawit Akli Muḥend Ulḥağ - Tubirett -

وزارة التعليم العالي والبحث العلمي
جامعة أكلي محند أولحاج
-البويرة-

Faculty of Sciences and Applied Sciences
Department of computer science

THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR
IN
COMPUTER SCIENCE
BY

ABBAS Faycel

Writer Identification and Writer Retrieval of a handwritten document

This thesis was publicly defended on June 9th, 2022 in front of the examination committee composed of:

Pr. AMAD Mourad	President	Professor, Bouira University, Algeria
Dr. SAOUD Bilal	Examiner	Associate Professor, Bouira University, Algeria
Dr. MARIR Toufik	Examiner	Associate Professor, Oum El Bouaghi University, Algeria
Dr. MENASSEL Rafik	Examiner	Associate Professor, Tebessa University, Algeria
Dr. GATTAL Abdeljalil	Supervisor	Associate Professor, Tebessa University, Algeria
Dr. SAOUDI Kamel	Co-Supervisor	Associate Professor, Bouira University, Algeria

ACKNOWLEDGEMENT

First and foremost, praise and thank to Allah, the almighty, for his blessings throughout my research work to complete the research successfully.

The completion of this work would not have been possible without the support of many people whose names may not all be enumerated. Their contributions are sincerely appreciated and gratefully acknowledged.

Foremost, I would like to express my sincere gratitude to my supervisor Dr.Abdeljalil GATTAL, for their endless support, immense knowledge ,kind and understanding during our case thesis, My sincere thanks also to Dr.Kamal SAOUDI, who offered guidance and support.

I would like to thank my thesis committee: Pr. AMAD Mourad, Dr. SAOUD Bilal, Dr. MARIR Toufik and Dr. MENASSEL Rafik, for the honor they have giving me by accepting to be examiners of this thesis.

To all relatives, friends and others who in one way or another shared their support, either morally, thank you.

Table of Contents

Chapter 1: state of the art of writer identification and retrieval

1. Introduction 1
2. Writer identification system /writer retrieval system 1
3. Text Dependent / Text Independent 2
4. Online / Offline 4
5. Benchmarking Datasets 5
5.1. ICDAR-2013 dataset 5
5.2. ICFHR 2018 dataset 6
5.3. LAMIS-MSHD Dataset 6
5.4. CERUG Dataset 8
5.5. WDAD Dataset 9
5.6. ICDAR2017 Dataset (Historical-WI) 9
5.7. BFL Dataset 10
5.8. KHATT Dataset 11
6. Preprocessing 13
6.1. Noise reduction 14
6.2. binarization 14
6.3. Size normalization 15
7. Text extraction and segmentation 15
8. Features extractions 16
9. Classification techniques 16
9.1. Matching step 17
9.2. Support vector machines (SVM) 19
10. Conclusion 22

Chapter 2: Offline writer identification and retrieval based on single script

1. INTRODUCTION 24
2. RELATED WORK 26
3. FEATURE EXTRACTION 27
3.1 COMPLETE LOCAL BINARY PATTERNS (CLBP) 28
3.2 LOCAL BINARY PATTERN VARIANCE (LBPV) 29
4. PROBABILITY DENSITY FUNCTION (PDF) 30
5. DECISION 31
6. EXPERIMENTATION AND RESULTS 31
7. Conclusion 36

Chapter 3: offline writer identification based on multi script

1. Introduction 38
2. Related work 40
3. Proposed Approach 44
3.1. Texture Feature Column Scheme 44
3.1.1. LBP Column Histogram Scheme 44
3.1.2. oBIF column histogram scheme 44
4. Classification 48

5. System evaluation 50
5.1. Dataset and Experimental Protocol 50
5.4. Evaluations on Single-script protocol 53
5.5. Evaluations on Multi-script protocol 55
6. Conclusion 61
Chapter 4: offline writer identification for Historical Document
1. Introduction 63
2. Related work 66
3. Feature extraction 68
3.1 oBIF column histogram scheme 68
3.2 Hinge feature 69
4. New moment matching method 70
5. Experimental Results 72
6. Conclusion 77
Conclusion 78

List of figures

Figure 1.Illustration of the difference between writer identification and retrieval 2

Figure 2.sample text dependent from ICDAR 2013..........4

Figure 3.Handwritten document samples from ICDAR-2013 dataset. 6

Figure 4.image from LAMIS dataset contain Arabic handwritten text..........8

Figure 5. image from CERUG dataset contain two paragraphs in Chinese..........9

Figure 6.image from WDAD dataset contain Farsi handwritten text..........10

Figure 7. Binarized images from the ICDAR 2017 dataset...........11

Figure 8.sample of the BFL dataset..........12

Figure 9.sample of the KHATT dataset..........13

Figure 10.Overview of the Proposed System 25

Figure 11.Samples of two different writers 38

Figure 12. Overview of the Proposed System..........39

Figure 13. Circularly neighbor sets for two different values of P and R..........43

Figure 14. Different steps of the LBP Column scheme..........45

Figure 15. Different steps of the oBIF colunm histogram.......... 47

Figure. 16. Classification rates on KHATT and BFL dataset using the LBP column histograms..........51

Figure. 17. Classification rates on KHATT and BFL dataset using the OBIF column histograms..........51

Figure.18. Classification rates on three datasets - ICFHR 2018 Experimental settings..........54

Figure 19.Visually similar produced by the proposed method (misclassification) on three datasets..........57

Figure 20. Examples of the correct classification produced by the proposed approach on three datasets..........57

Figure 21. Overview of the Proposed System..........63

ABSTRACT

Identification and retrieval writer's focused on the recognition of his handwriting is highly complex and challenging task, these challenges included variations in handwriting styles, handwriting languages, character sets, layout or legibility; although Handwriting exhibits behavioral features of an individual and has been considered as unique. Where the letters style and shape of written entirely different for different writers and vary slightly for same writer. Also alphabets in the handwritten text may have junctions, crossings, loops, different directions, what Makes Identification /retrieval even more difficult is the fact that handwriting of the same writer may also present high variability in handwriting, such as writing letters of different sizes, changing writing instruments or document layout. Therefore, we intend in this work to identify specific characteristics from writers which can be robust against the aforementioned changes and proposes a model for discovering the writer's identity based on features extraction of handwriting using machine learning algorithms. In this thesis we used several datasets constrained and unconstrained, some of which contain single-script and multi-script handwritings written in different languages.

RÉSUMÉ

L'identification et la recherche de scripteur base sur son écriture manuscrite est une tâche très complexe et difficile. Ces défis comprenaient des variations dans les styles d'écriture manuscrite, les langues d'écriture manuscrite, les jeux de caractères, la mise en page ou la lisibilité ; bien que l'écriture manuscrite présente les caractéristiques comportementales d'un individu et ait été considérée comme unique. Où le style et la forme des lettres écrites varient légèrement pour le même scripteur et totalement différents pour différents scripteurs. De plus, les alphabets dans le texte manuscrit peuvent avoir des boucles, des croisements, des jonctions, des directions différentes, ce qui rend l'identification/recherche de scripteur une tache plus difficile, puisque l'écriture manuscrite du même scripteur peut également présenter une grande variabilité dans l'écriture manuscrite, comme l'écriture de lettres de différentes tailles, changer les instruments d'écriture ou la mise en page du document. Par conséquent, nous avons l'intention dans ce travail d'identifier les caractéristiques spécifiques des scripteurs qui peuvent être robustes contre les changements susmentionnés et propose un modèle pour découvrir l'identité de scripteur basé sur l'extraction de caractéristiques de l'écriture manuscrite à l'aide d'algorithmes d'apprentissage automatique. Dans cette thèse, nous avons utilisé plusieurs bases de données, dont certains contiennent des textes manuscrites écrits dans une seule langue et d'autres sont écrits en plusieurs langues.

ملخص

يعد التعرف على الكاتب والبحث عنه بناءً على خط يده مهمة معقدة وصعبة للغاية. وتشمل هذه التحديات اختلافات في أنماط الكتابة اليدوية ، ولغات الكتابة اليدوية ، ومجموعات الأحرف ، النسق والوضوح ؛ على الرغم من أن الكتابة اليدوية تعرض الخصائص السلوكية الكاتب وتعتبر فريدة من نوعها. حيث يختلف أسلوب وشكل الحروف المكتوبة اختلافًا طفيفًا بالنسبة لنفس الكاتب ويختلف تمامًا باختلاف الكاتب. بالإضافة إلى ذلك ، قد تحتوي الحروف الهجائية في النص المكتوب بخط اليد على حلقات وتقاطعات واتجاهات مختلفة ، وما يجعل التعرف / البحث أكثر صعوبة هو حقيقة أن الكتابة اليدوية من الكاتب نفسه قد تظهر أيضًا تنوعًا كبيرًا في الكتابة اليدوية ، مثل كتابة رسائل بأحجام مختلفة أو تغيير أدوات الكتابة أو تخطيط المستند. لذلك ، نعتزم في هذا العمل تحديد خصائص محددة للكاتب تكون قوية ضد التغييرات المذكورة أعلاه واقتراح نموذج التعرف / البحث عن هوية الكاتب بناءً على استخراج الميزات من الكتابة اليدوية باستخدام خوارزميات التعلم الآلي. في هذه الرسالة ، استخدمنا العديد من قواعد البيانات المقيدة وغير المقيدة ، بعضها يحتوي على نصوص مكتوبة بخط اليد بلغة واحدة والبعض الآخر مكتوب بلغات متعددة.

INTRODUCTION

Writer identification is the job of assigning a writer to a document who's the writer already does not know. For this task a dataset of handwritings by well- known writers must be available. Features are generated on the handwriting of all documents in the dataset, and when identifying the writer of the handwriting the same features are generated and the features comparison are accomplished. With a certain distance measurement, the most similar document in the database can be found and the writer of this particular document is then assigned to the new one. This allows for finding out the writer of a specific document and can be used for example by investigators to identify who wrote the threat, as well as to verify the authenticity of the document such as a will and whether the handwriting document was written by the authorized individual. In contrast to this, the task of writer retrieval is to find documents in a database which have the most similar handwriting and no need to know the writer of the document in the dataset. After the comparison of the features the documents in the dataset are ranked according to the distance and presented to the user as the most similar handwritings. Generally, the most similar documents should originate from the selfsame writer as the reference handwritings if possible. in addition, writer retrieval allows finding the most similar documents concerning the handwriting in a set of documents. It allows the users to look for documents which may have been written by the same writer. This can be used for example in the Museums for indexing the archives.

This thesis is giving an overview in the first chapter of the current state of the art of writer identification and retrieval followed by three chapters, which have been developed in the scope of this thesis, with the aim to propose a method out of the language and the seniority of the handwritten document which avoids unnecessary preprocessing steps. in more detail:

- In chapter 2: we propose two new LBPs based methods for extracting the features from handwritings: The Local Binary Pattern Variance (LBPV) and the Complete Local Binary Patterns (CLBP). These features are normalized using the Probability Density Function (PDF). the minimization of similarity criteria based on the distance between two feature vectors are applied as a classification step. The experimentations were conducted with different combinations of metrics' distances.

- In chapter 3: We propose a Writer Identification system that employs combinations of different configurations of both Texture Features such as Local Binary Patterns (LBP) column scheme and oriented Basic Image Features (oBIFs) column scheme to identify writer from handwriting. LBP and oBIFs columns histograms extracted from handwriting samples are used to train a support vector machine classifier (SVM).

- In chapter 4: we demonstrate how the combination of well-known Features using the proposed moment distance can be well employed to improve the performance of Historical Document Writer Identification systems. These features are calculated from whole binarized images. The classification step is based on minimizing a similarity criteria based on The New Moment distance proposed.

- In conclusion, we conclude by summarizing the essential contributions of this work.

CHAPTER 1

STATE OF THE ART OF WRITER IDENTIFICATION AND RETRIEVAL

1. Introduction:

Writer identification is the task of recognizing the writer of a document among of knowing writers form dataset. In contrast to identification, writer retrieval is the task of finding handwritings in a database according to the similarity of the handwriting to a reference handwriting.

This Chapter presents briefly overview the Identification and retrieval writer's research field, including the basic concepts associated in this area and the datasets used in experimental setup on evaluation protocols and in the performance of current identification/ retrieval systems, also the step of features extraction of specific styles of writer in the handwriting document and the machines learning used.

2. Writer identification system /writer retrieval system

In Writer identification system a dataset with the handwriting of known writers is available and for a new document image the system should return the identification of its writer. the retrieval system a set of documents, with known or unknown writers, is given and for a new document image a ranking according to the similarity of the handwriting of the document images in the dataset should be returned.

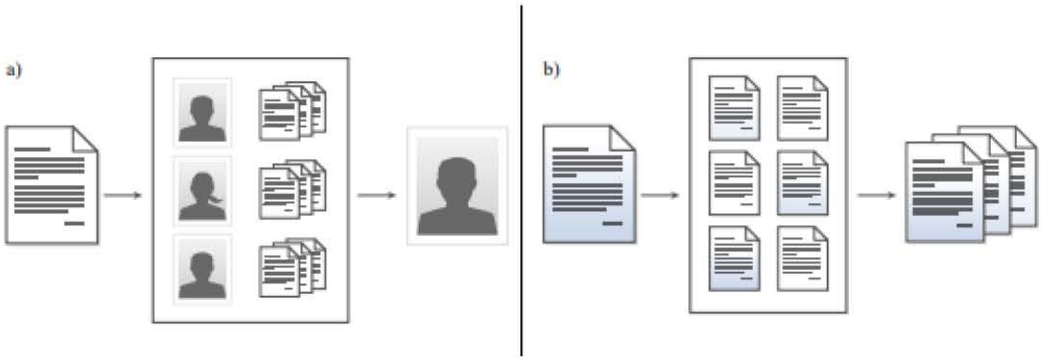


Figure 1.Illustration of the difference between writer identification and retrieval: a) Writer identification, b) writer retrieval.

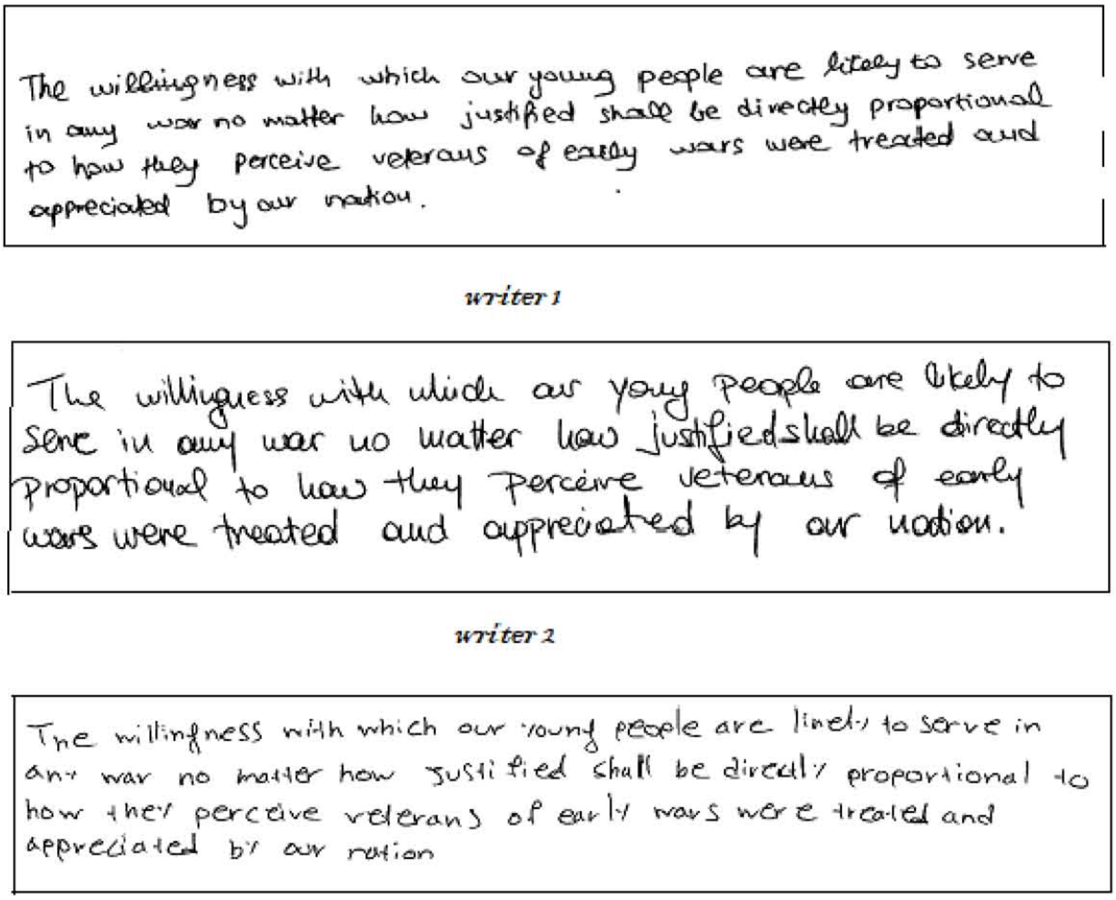
3. Text Dependent / Text Independent

The Classification systems for writers' identification/retrieval can be categorized into two large families, based on the level of text dependency: text-independent and text-dependent writer's identification [1,2,3,4].

Text-dependent writer identification/retrieval systems are more constrained and require a writer to write a particular predefined text that will be used to identify (verify) their identity or to recognize their gender. Text-dependent systems employ the comparison between different characters or words of known semantic content. These methods therefore need the prior location and segmentation of appropriate information. This is usually performed interactively manually by a human user or automatically by a segmentation algorithm, but as it requires the same writing content this method is not adapted for many practical situations. otherwise, the second method which called independent text that use any given text can be used in order to determine (verify) the identity or gender of the writer and it does not require comparison of same characters. Thus, it is very similar to uses the comparison between individual characters or words of known semantic (ASCII) content and signature verification techniques. This method considers as the universal model of hand writing text as the metric for comparison, because no restriction on the subject of the written text, this task is more general and thus more difficult to establish.

To extract stable characteristics, do not be influenced by the content of the text, from the writing sample, it is preferable to recommend writing a minimal amount of Text (for example, a paragraph containing a few lines of text). Even though it got a wider applicability, text-independent methods do not obtain the same high performance as text-dependent methods do. In case of notification, we cannot forget that both of these two systems (text dependent or independent) have advantages and disadvantages. Text-dependent systems achieve very high performance using less data, which is not possible for text-independent systems. in addition, they are more vulnerable to counterfeiting because the text on which the verification is done is known in advance. In the case of text-independent systems, forgery is not a significant problem because these systems extract, from a handwritten document, the less frequent properties that are difficult to forge [5]. The use of a text dependent or independent system is entirely dependent on the application area and the availability of data. For example, in the case of legal applications, we cannot ensure that the data that is available for verification matches the reference entry, so the only alternative is to use a text independent system [6].

In this work, the study involving text-dependent and text-independent writer`s identification and retrieval of offline handwritten text is focused.



The willingness with which our young people are likely to serve
in any war no matter how justified shall be directly proportional
to how they perceive veterans of early wars were treated and
appreciated by our nation.

writer 1

The willigness with which our young people are likely to
Serve in any war no matter how justifiedshall be directly
proportional to how they percive veterans of early
wars were treated and appreciated by our nation.

writer 2

The willingness with which our young people are linely to serve in
any war no matter how justified shall be directly proportional to
how they perceive veterans of early wars were treated and
appreciated by our nation

writer 3

Figure 2:sample text dependent from ICDAR 2013.

4. Online / Offline

Handwritten documents can be mainly divided throughout the means used for data acquisition into two classes: online and offline handwritings; With the ease of availability and enhance in use of Tablet PCs, Pocket PCs, and other pen enabled input devices, online documents are gaining popularity. Online handwritings contain the handwriting temporal information process in addition to the pen movements coordinates, along with pen-up and pen-down events. In contrast, Offline documents are digitalized images of handwritten documents.

Online handwriting allows us to use pressure, spatial information and velocity, which are not available with offline data. Online writer identification systems can be used in automated identity verification systems, as though online data contains useful information for writer identification (i.e., the order of the lines, the position of the pose and the raising of pencil, the pressure, speed, …… etc.) that is lost in offline data; offline writer identification and retrieval is considered more challenging than online writer identification and retrieval [7] The sample process shows the scanning of the documents from different writers, pre-processing done to the scanned documents, feature extraction that make major distinctions between the scanned images while at the same time reducing its dimensionality, and finally the classification step done writer recognition was carried out.

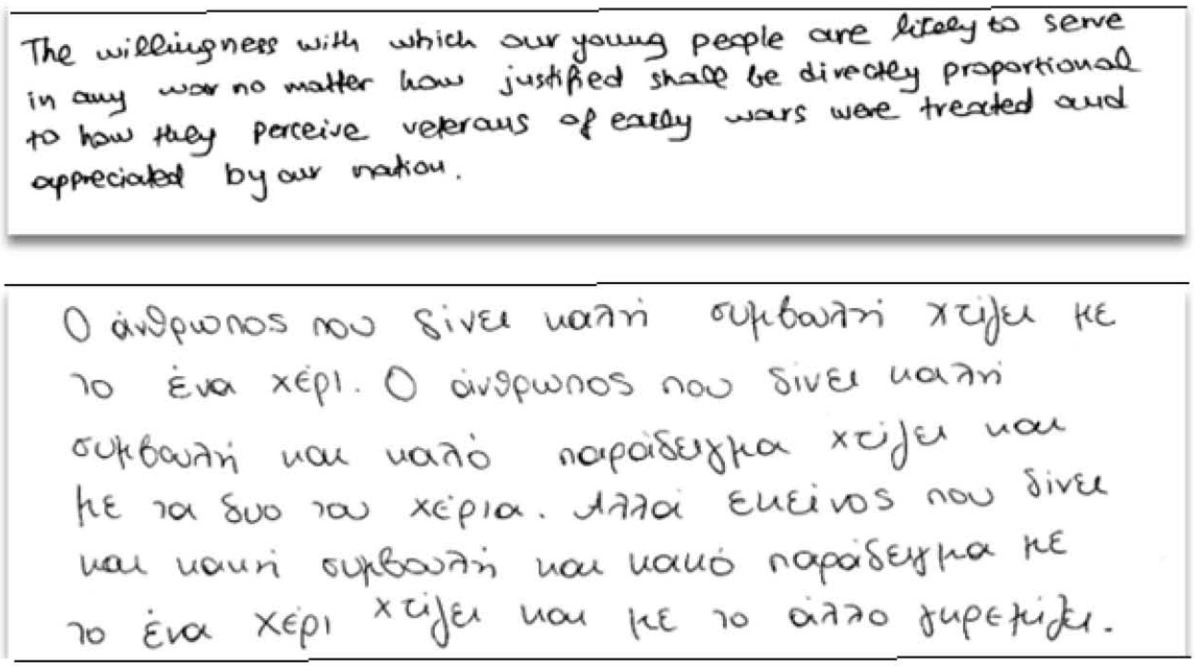
In our case, the handwriting documents acquired by the off-line mode includes capturing the text image using the sensors such as scanner or camera with a minimum degradation. Through this step, despite the good acquisition quality systems, noises might appear in scanned images. This is caused by the area, texture type, the document state and its lighting.

5. Benchmarking Datasets

the databases of images of Handwritten documents are in fact the main entrance in the evaluation of systems writer identification and verification. They provide an effective means for the unification and comparison of the work carried out within the different research teams around the world. In the work presented in this thesis we used several datasets, some of which contain single-script and multi-script documents written in different languages and collected under different constraints. In the following, we will review the most important of them.

5.1. ICDAR-2013 dataset

This dataset is proposed in Writer Identification Competition [8] of (ICDAR 2013) International Conference on Document Analysis and Recognition for the well-established evaluation methodology to enable research teams to make their contributions in order to find effective writer identification systems for Latin scripts. It consists of 1000 handwritten documents written in two different languages (English and Greek). Each writer filled 4 handwritings (two in Greek and two in English). The first 200 writers share the same texts while the remaining part of this dataset (50 writers) was acquired from data created for the testing phase of handwriting segmentation contest.



The willingness with which our young people are likely to serve in any war no matter how justified shall be directly proportional to how they perceive veterans of early wars were treated and appreciated by our nation.

Ο άνθρωπος που δίνει καλή συμβουλή χτίζει με το ένα χέρι. Ο άνθρωπος που δίνει καλή συμβουλή και καλό παράδειγμα χτίζει και με τα δυο του χέρια. Αλλά εκείνος που δίνει και καλή συμβουλή και κακό παράδειγμα με το ένα χέρι χτίζει και με το άλλο γκρεμίζει.

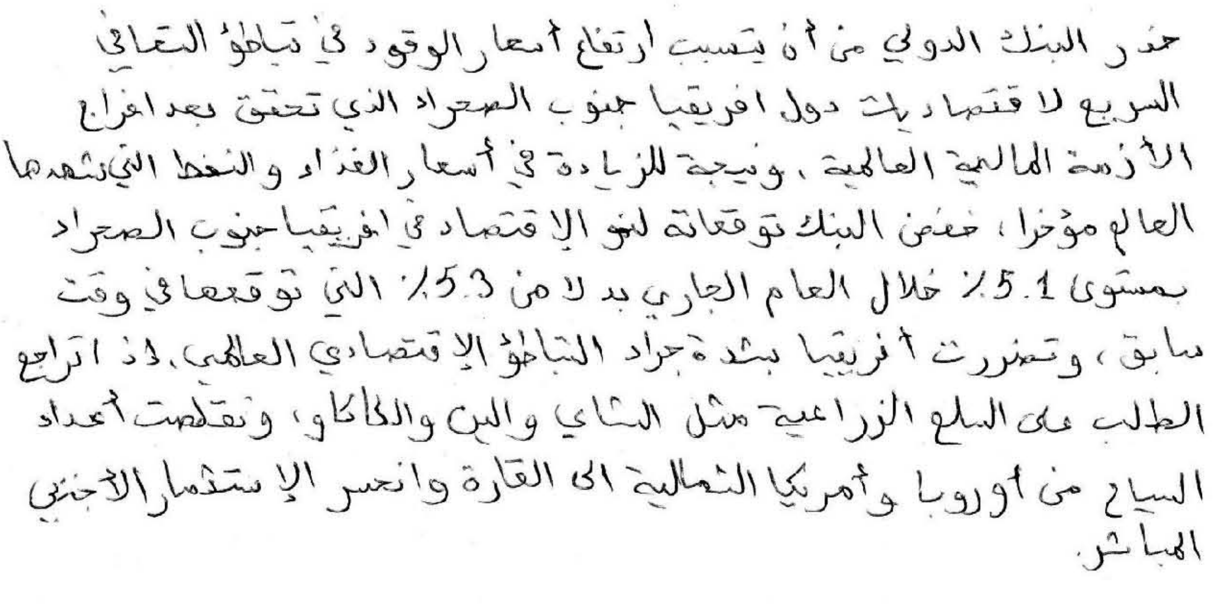
Figure 3:Handwritten document samples from ICDAR-2013 dataset.

5.2. ICFHR 2018 dataset

The 16th International Conference on Frontiers in Handwriting Recognition (ICFHR 2018) propose in this Competition of Multi-script Writer Identification to using CERUG Data sets, WDAD , and LAMIS-MSHD, In the context of the 16th International Conference on Frontiers in Handwriting Recognition In order to examine the overall performance of new Contributions in multi-script writer identification and to analyze the overall performance of well-known script-dependent writer identification systems in a multi-script environment [9].

5.3. LAMIS-MSHD Dataset

A Multi-script Offline Handwriting Dataset (LAMIS-MSHD) [9] this database is mainly targeting writer identification and verification in a multi-script environment and can also be effectively used to evaluate systems Such as signature verification, script recognition and handwriting recognition…., Contains constrained handwritten samples forms with content from 13 different sources including handwritten text, signatures and digits. Each form was subsequently digitized at 300 dpi. The forms have been replenished by 100 randomly selected volunteers in Tebessa (Algeria) from different age groups, gender (Arabic and French writers), level of hands and educational backgrounds. Each writer filled 13 forms making a total of 1300 forms on each of the 13 forms six in French, six in Arabic and one form with isolated digits and numerical strings., the individuals provided their personal information including education level, age group, name, gender. The volunteers were also asked to provide their signatures in the specified place. The writing samples were collected by asking the writers to copy the filled text on each form, six in French, six in Arabic and one form with isolated digits and numerical strings. All writings were produced with a blue or a black pen.

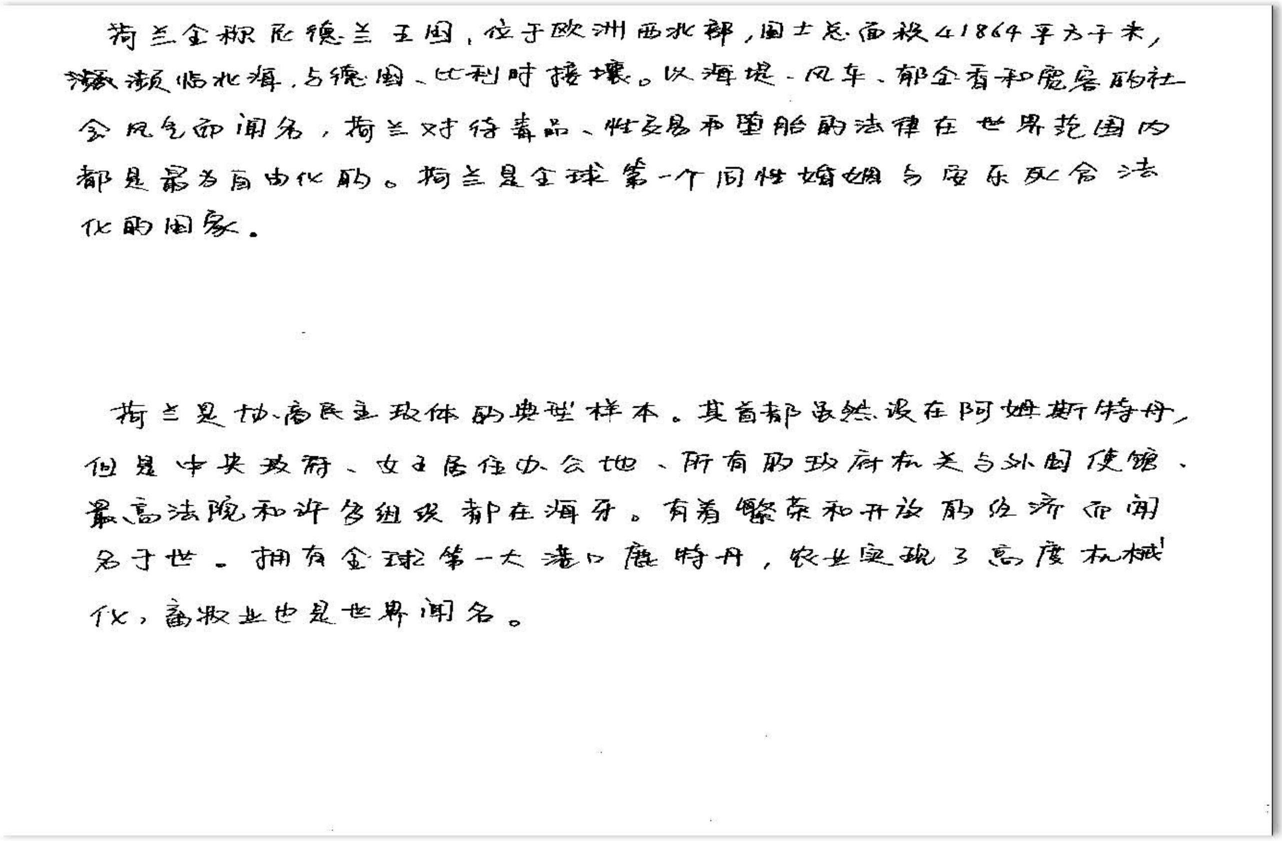


حذر البنك الدولي من أن يتسبب ارتفاع أسعار الوقود في تباطؤ التعافي
السريع لاقتصاديات دول افريقيا جنوب الصحراء الذي تحقق بعد افراج
الأزمة المالية العالمية، ونتيجة للزيادة في أسعار الغذاء والنفط التي تشهدها
العالم مؤخرا، خفض البنك توقعاته لنمو الاقتصاد في افريقيا جنوب الصحراء
بمستوى 1.5% خلال العام الجاري بدلا من 5.3% التي توقعها في وقت
سابق، وتضررت أفريقيا بشدة جراء التباطؤ الاقتصادي العالمي. إذ تراجع
الطلب على السلع الزراعية مثل الشاي والبن والكاكاو، وتقلصت أعداد
السياح من أوروبا وأمريكا الشمالية الى القارة وانحسر الاستثمار الأجنبي
المباشر.

Figure 4:image from LAMIS dataset contain Arabic handwritten text.

5.4. *CERUG Dataset*

The CERUG dataset [10] comprises handwritings collected from 105 Chinese subjects, mainly students from China. Some of them live in China and the others study in the Netherlands. Every subject is required to write four different A4 pages. On page 1, the contributors were asked to copy a text of two paragraphs in Chinese. On page 2, the subjects described certain fields they liked in their own words in Chinese. Page 3 includes English text copied from two paragraphs. This page is split into two subpages, and each subpage contains one paragraph. Overall, there are four handwritten samples from each writer, two in English language and two in Chinese language. All the handwritings were scanned at 300 dpi, 8 bits/pixel and gray-scale.



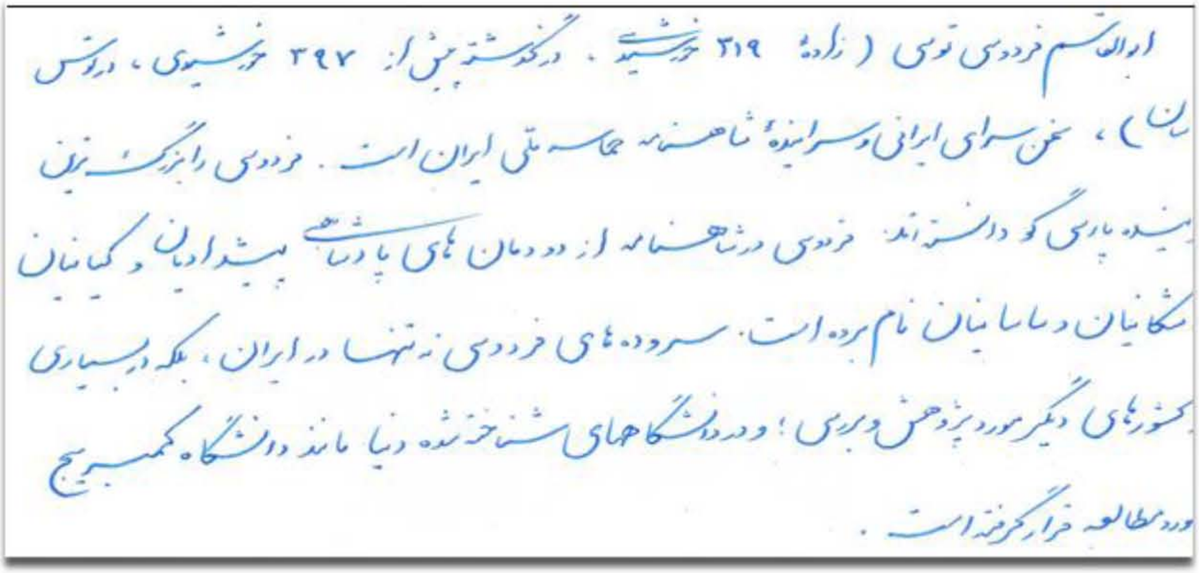
荷兰全称尼德兰王国，位于欧洲西北部，国土总面积41864平方千米，濒临北海，与德国、比利时接壤。以海堤、风车、郁金香和宽容的社会风气而闻名，荷兰对待毒品、性交易和堕胎的法律在世界范围内都是最为自由化的。荷兰是全球第一个同性婚姻与安乐死合法化的国家。

荷兰是协商民主政体的典型样本。其首都虽然设在阿姆斯特丹，但是中央政府、女王居住办公地、所有的政府机关与外国使馆、最高法院和许多组织都在海牙。有着繁荣和开放的经济而闻名于世。拥有全球第一大港口鹿特丹，农业实现了高度机械化，畜牧业也是世界闻名。

Figure 5: image from CERUG dataset contain two paragraphs in Chinese.

5.5. WDAD Dataset

The WDAD Dataset [11] contains 800 handwritings images written by 200 different writers. All the forms in the dataset are filled by volunteers selected from different of genders, ages, educational levels and handedness. Each volunteers wrote 4 handwritten documents. The specificities of the dataset are as follows: The first 2 pages comprise Farsi while the remaining 2 pages comprise English handwritten text. The text on each of the 12 pages is different and all volunteers copied the same text.



ابوالقاسم فردوسی توسی (زادهٔ ۳۱۹ خورشیدی ، درگذشته پیش از ۳۹۷ خورشیدی ، در توس
سان) ، سخن‌سرای ایرانی و سرایندهٔ شاهنامه حماسه ملی ایران است . فردوسی را بزرگ‌ترین
ینده پارسی‌گو دانسته‌اند . فردوسی در شاهنامه از دودمان‌های پادشاهی پیشدادیان و کیانیان
شکانیان و ساسانیان نام برده است . سروده‌های فردوسی نه تنها در ایران ، بلکه در بسیاری
کشورهای دیگر مورد پژوهش و بررسی ؛ و در دانشگاه‌های شناخته‌شده دنیا مانند دانشگاه کمبریج
مورد مطالعه قرار گرفته است .

Figure 6:image from WDAD dataset contain Farsi handwritten text

5.6. ICDAR2017 Dataset (Historical-WI)

This dataset used in the ICDAR2017 competition on the identification of a historical document writer (Historical-WI) [12] consists of 3,600 handwritten text pages selected from the current 13th-20th century electronic library of the University of Basel that includes 140,000 images. this dataset contains handwritings from 720 different writers, each writer produced five pages for the test sets, and in the training sets, 394 writers remained, or a total of 1182 pages. All documents' images are digitalized at 300 dpi, saved in *jpeg* format and *png* format for the binarized images respectively.

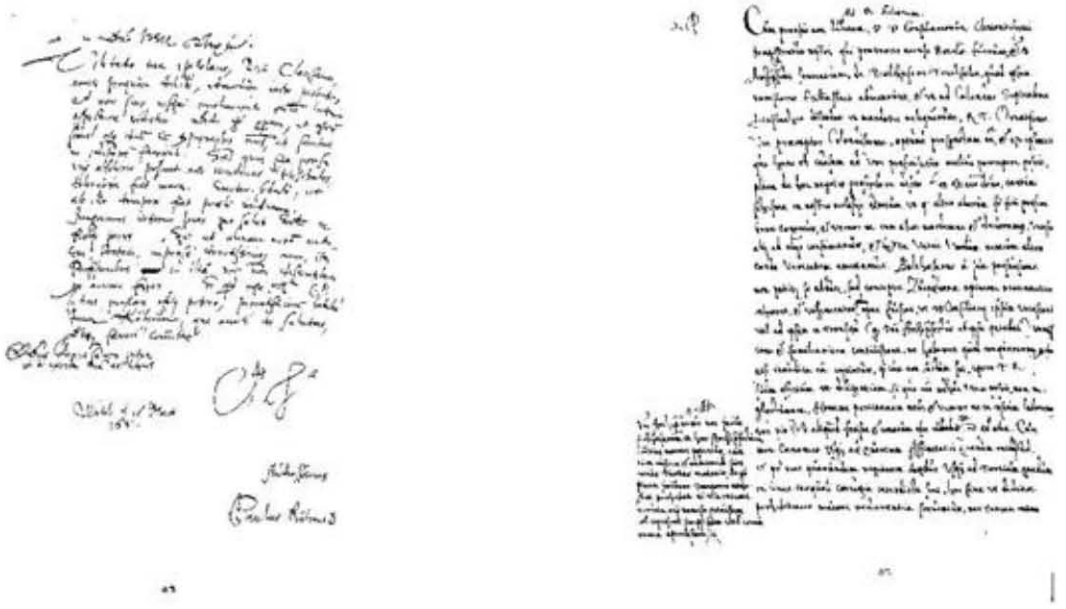
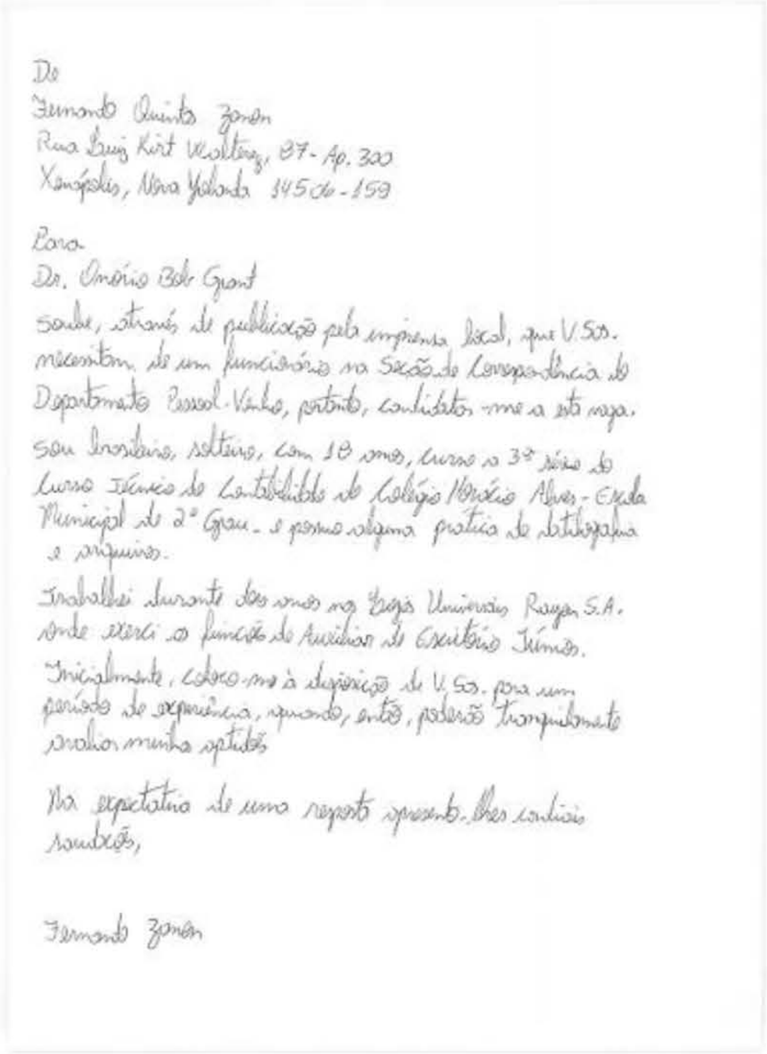


Figure 7: Binarized images from the ICDAR 2017 dataset.

5.7. BFL Dataset

The Brazilian Forensic Letter (BFL) dataset [13] was created for use by Brazilian forensic experts as well as the Brazilian Federal Police for writer identification and retrieval by 315 undergraduate students in three various courses during one month, each of whom wrote three samples, totaling 945 images. The texts were collected on A4 white paper without a pencil line and then digitalized in grayscale at 300 dpi (3760 x 2448). Each writer was used his own pen, which meant that several various pens were used. The text is short (131 words in Portuguese) and complete in the sense that it covers all the characters (letters and numbers) and certain characters combinations. This gives it suitable for text-dependent writer identification and retrieval as well.



De
Fernando Quinto Zonen
Rua Luiz Kirt Weltniz, 87 - Ap. 300
Xenópolis, Nova Yolanda 14506-159

Para
Dr. Onório Bob Grant

Soube, através de publicação pela imprensa local, que V.Sas. necessitam de um funcionário na Seção de Correspondência do Departamento Pessoal-Vendas, portanto, candidato-me a esta vaga.

Sou brasileiro, solteiro, com 18 anos, curso a 3ª série do Curso Técnico de Contabilidade do Colégio Horácio Alves - Escola Municipal de 2º Grau e possuo alguma prática de datilografia e arquivo.

Trabalhei durante dois anos na Lojas Universais Rayon S.A. onde exerci a função de Auxiliar de Escritório Júnior.

Inicialmente, coloco-me à disposição de V.Sas. para um período de experiência, quando, então, poderão tranquilamente avaliar minhas aptidões

Na expectativa de uma resposta apresento-lhes cordiais saudações,

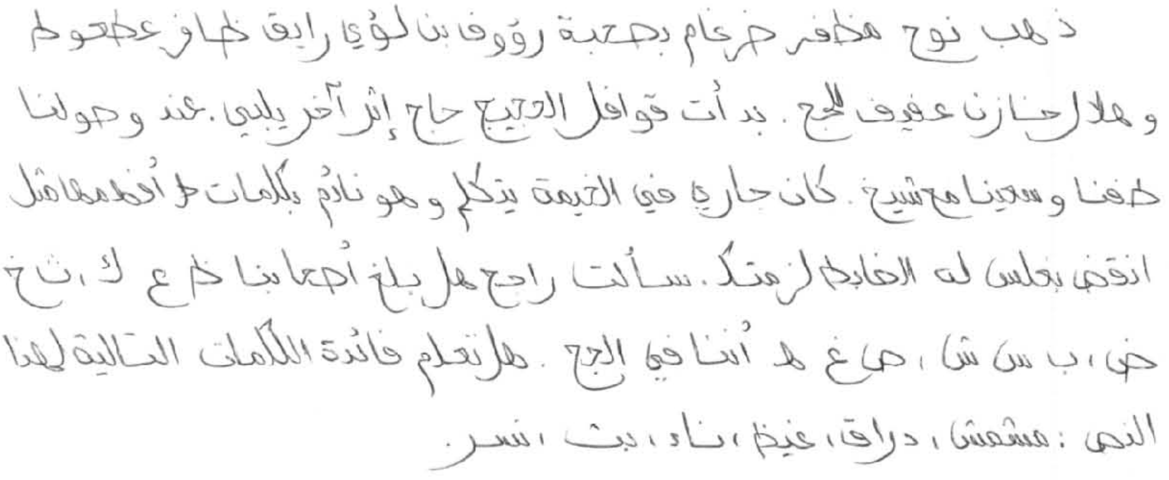
Fernando Zonen

Figure 8:sample of the BFL dataset

5.8. KHATT Dataset

KHATT (KFUPM Handwritten Arabic Text) dataset [14]: A comprehensive Arabic handwritten text dataset is made freely available from 2012 onwards to researchers world-wide ,It make use for research in different handwritten text related problems such as writer identification and verification, text recognition, segmentation, pre-processing ,forms analysis ,It contains unrestricted writing styles of Arabic texts written by 1000 different contributors from various countries, gender ,age groups , level of hands and education, digitized at various resolutions

(200, 300 ,600 dpi), the half of 4,000 images that make up the database contain a similar text covering all Arabic characters and numbers while the remaining 2000 images contain free paragraphs written by writers on any topic of their choice, the verified ground truth dataset including meta-data describing the written text at the page, paragraph, and line levels in text and XML formats.



ذهب نوح مظفر ضرغام بصحبة رؤوف بن لؤي رايق ظمآن عصفور
وهلال خازن عفيف للحج. بدأت قوافل الحجيج حاج إثر آخر يلبي. عند وصولنا
طفنا وسعينا مع شيخ. كان جاره في الخيمة يتكلم وهو نائم بكلمات لم أفهمها مثل
انقض يجلس له الغائط زمتك. سألت راجح هل بلغ أصحابنا خ ع ك، ث خ
ض، ب س ش، ص غ هـ أننا في الحج. هل تعلم فائدة الكلمات التالية لهذا
النص: مشمش، دراق، غيظ، نشاء، بث، نسر.

Figure 9:sample of the KHATT dataset.

6. Pre-processing

Pre-processing steps are the technique applied to format images and improves their quality or reducing the amount of information to process to keep only the most significant information before they are used by training template and inference. This includes, but is not limited to, the skew/slant detection and correction, resizing and colour corrections for text document, and that are a segmentation challenge. These tasks generally include noise reduction, normalization and binarization.

In this thesis, we aim to create a system that works on the original document without pre-processing step.

6.1. Noise reduction

Noise in a Document image is random variation brightness or colour information in the images captured due to the document age, the human handling of papers or the low quality of the scanning devices. Generally, noise is degradation in image signal caused by external sources. It can have various forms and appearances within an image and is, in several cases, disturbing or an unwanted artifact that reduces the subjective quality of image. Basically, to reduce the diminish spurious points and noises in the image, various filtering techniques were proposed [15]. The filter choice is carried out according to the noise importance. A low additive noise can be perfectly removed by low-pass filtering such as the Gaussian filter. For more complicated noises, more powerful filtering techniques are used such as the “Transform Domain Filtering” [16].

6.2. Binarization

Image binarization it is an important step pre-processing in document analysis by converts a multicolor image to grayscale image. For the handwriting images, it transforms a gray image coded on 256 values to a binary image which affects 0 or 1 for the background and 1 or 0 for the handwritten text by using a threshold. Thresholding techniques are generally classified into global and local (adaptive) methods. For global thresholding, the Otsu binarization is the most popular technique (Otsu, [17]). It consists to find a single cut-off level at which pixels in a gray scale image can be classified into two groups, one for foreground handwritten text and the other for the background. On the other hand, local methods use adaptive techniques to find multiple threshold levels each for a local area of an image. In this respect, there are several effective techniques such as “Bernsen thresholding” [18],“Sauvola thresholding” [19], and “Niblack thresholding” [20], in this work The proposed method also assumes that the writing is a texture image, thus a binarization step is not needed. The benefit is that the writer identification or writer retrieval does not require a binarization step.

6.3. Size normalization

Size normalization is employed to correct the shape, size (dimension) and position of the handwritten image. This step is required to reduce the shape variability between class images in order to facilitate feature extraction and enhance its classification.

Normalizing document images to the same size is not always necessary in handwriting recognition, since various feature generation schemes, like histogram-based techniques (including Histogram of Oriented Gradients, Histogram of Templates, Local Binary Patterns, etc) provide feature vectors with the same size whatever the initial image size. Nevertheless, various experimental studies reveal that handling images with the same size can produce more homogeneous features and accelerates the document processing [21]. Besides, for some particular features such as convolutional neural network-based features, the size normalization is mandatory for the network training. In this aim,[21]; [22] utilized a normalization of text images to facilitate the network training.

7. Text extraction and segmentation

Text extraction is extensively employed in handwritten document processing, because the text is often situated in the middle of document pages. So, extracting the text can obviously, accelerate the feature generation process, and remove useless information brought by the background. Text is extracted by using histograms of horizontal and vertical projections The horizontal projection corresponds to the number of text pixel in this line. Similarly, each column is used to compute the vertical projection by calculating the occurrences of text pixels along this column. The horizontal projection is used to extract lines, while the vertical projection is used for words extraction.

Note that, generating features from the full handwritten document, provides a coarse description of the writer traits. For this reason, a segmentation into smaller handwritten image patches such as lines, words, or fragments can help to get more effective features. In fact, the use of full documents has been employed in some applications, where there is no need to highlight the writing style of each individual. For instance, in soft biometrics prediction, which aims to predict the writer’s gender, handedness and age range, similar performance is obtained by using full document or segmented lines ([23]; [24]). For the writer identification and retrieval, researchers cope the large size of the handwritten text by using features that are locally computed [25].

8. Features extractions

Feature extraction includes the translation of an input image to vectors contains of numerical feature sets that allows them to shown as an image. The information extracted from the image during the feature extraction step is called a feature. The local features are the important fragments of interconnected objects; instead, the global features are all properties which can be attributed to an object [26, 27…,31]. Best features are those that satisfy the two conditions of having small intra-class invariance and large inter-class invariance. In a right system, the extraction and selection of features is so paramount and still the study focus in last few years [26-31]. The methods used for extracting features in Writer identification and retrieval can be clustered into two main groups.

A) Global Features: it attained from the handwritten text includes manipulation of the text images instead of handwritten document. These comprise co-occurrence matrices and Gabor filter.

B) Local Features: it Includes manipulation of the handwritten document in the shape of text images so that numerous statistical properties can be found such as averages of width, height, and characters legibility.

There are many methods for feature extraction, this thesis covers Texture-based approaches that extract textural information as features for writer identification such us oriented Basic Image Features (oBIFs), Complete Local Binary Patterns (CLBP), HINGE, Local Binary Patterns (LBP), The Local Binary Pattern Variance (LBPV). Each method of them will be detailed in the chapter in which he was used.

9. Classification techniques

The classification is a step which makes it possible an unknown pattern to be assigned to a known class [32]. Two techniques of classification approach are used to categorize handwritings features into several classes, unsupervised classification that can make without producing training dataset and generated the classes according to the similarity between samples. Overwise, the second technique of classification which is supervised needs training dataset and identify the method of classification like the minimum distance, Spectral Angle Mapper (SAM), the maximum likelihood, or Mahalanobis distance.

For writer identification, the nearest neighbour classifier is one of the basic techniques and is widely used to classify handwritten text image. Hidden Markov Model (HMM) is a current approach to this problem which uses the structural and statistical information included in handwritten text image. Artificial Neural Networks (ANN) are widely used for image recognition and also extra well studied classification method [33]. Hence, Support Vector Machines (SVM) which is developed at the beginning to classify two class and then extended to problems of multi-class and perform to writer identification [33-36].

In the case of the writer's identification, supervised classification has been approved since writer's classes are well definite. It can be performed in two stages:

Training stage: The goal of the training stage is to train the classifier with the known writer dataset for further recognition with unknown writer dataset.

Recognition and decision stage: This stage classifies the input pattern by comparing them to a list of reference patterns. This stage also uses classification techniques in the form of decision. The decision stage is strongly influenced by the feature extraction step, and a successful of the classifier.

9.1. Matching step

in writer identification and Retrieval system, the matching is a way of measuring how handwritten samples are related or closed to each other. On the other hand, the dissimilarity measure is to tell how much the handwriting document from one writer to another writer are distinct.in another meaning, the matching measures the similarity between the feature vector of a query document and documents feature vectors available in a reference dataset. Thereby, all similarity or dissimilarity measures can be used in this step. Most of research works report the use of standard metrics like Euclidean, Manhattan distance, city block, correlation and spearman distance. Table (1) summarizes measures employed to perform writer identification and Retrieval task in this thesis.

Table 1 : Adopted similarity/dissimilarity measures

Distance	Equation
Euclidean	$D_{eucl}(x,y) = \sqrt[2]{\sum_{i}^{n}(x_i - y_i)^2}$
Correlation	$D_{corr}(x,y) = 1 - \frac{cov\ (x,y)}{\sqrt{var(x)}.\sqrt{var(y)}}$
Cosine	$D_{cos}(x,y) = \frac{x.y}{\lVert x\rVert\lVert y\rVert}$
spearman	$D_{sp}(x,y) = r_s = \frac{6\sum d_i^2}{n(n^2-1)}$ where $d_i = R(x_i) - R(y_i)$, n number observations
city block	$D_{cityblock}(x,y) = \sum_{i=1}^{n}\lvert x_i - y_i\rvert$

Where:

x and y are feature vectors of two documents.

n: is the size of each feature vector.

9.2. Support vector machines (SVM)

Support vector machine (SVM) is an attractive machine learning algorithms and regression rules from dataset including non-linear and linear, not only does it have a solid theoretical foundation, although it is also more performance and efficient in Hight dimensionality feature spaces than other classifiers in numerous applications areas for dataset classification.

Additionally, it covers two main negative factors in machine learning:

- The Vapnik-Chervonenkis (VC) dimension provides a general measure of parametrically controlling the SVM capacity.
- Avoid underfitting and overfitting.

Two other benefits are especially imparted by SVMs. First, the SVM can play well even with a suit kernel, if datasets are a non-linearly separable in feature space. Another, particularly popular for images classification problems where very high-dimensional feature spaces are the standard. These benefits allow us to choose Support vector machine as top candidates for effecting writer's identification systems.

SVMs are a type of machine learning algorithm that effecting regression and classification using a hyperplane in a high feature space. The SVMs classifier is presented with an examples of training set (x_i,y_i) where the x_i are the data samples and the y_i are the labels suggesting which class the sample belongs to. For the two-class problem of pattern recognition, $y_i = -1$ or $y_i = +1$.

Where, example of a training (x_i,y_i) is called positive if $y_i = +1$ and negative otherwise.

SVM has its skill to select the representative training dataset, commonly referred to as a "Support vectors" and tries to find the optimal line is called hyperplane which maximizes the margin or distance between the nearby points from two classes. When learned, a decision function is built in order to classify data according to the area separated by the hyperplane.

The errors are minimized by maximizing the margins controlled by the classifier's VC dimension.

SVMs built a hyperplane which separates two classes and attempts to perform maximum separation between the classes. It allows to separate the classes with a large margin minimizes a bound on the predicted generalization error. Maximal Margin classifier is called simplest model of SVM which is constructs an optimal hyperplane using linear separator given by $w^T x - \gamma = 0$ between two examples classes. The free parameters are a vector of weights *w*, which is orthogonal to the hyperplane and a threshold value *γ*. These parameters are obtained by solving the following optimization problem using Lagrangian duality.

Minimize $\frac{1}{2}\|\boldsymbol{w}\|^2$

subject to $$\boldsymbol{D}_{\mathrm{ii}}(\boldsymbol{w}^T\boldsymbol{x}_i - \boldsymbol{\gamma}) \geq \mathbf{1}; \boldsymbol{i} = \mathbf{1}, \dots \dots, \boldsymbol{l}.$$

Where $\boldsymbol{D}_{ii}$ corresponds to class labels +1 and –1. The samples with non-null weights are called support vectors. In the presence of wrongly classified and outliers training examples it may be suitable to allow some training errors in order to avoid overfitting. A slack variable vector ξi that measure the violation amount of the constraints is introduced and the optimization problem mentioned to as soft margin is given below.

$$\underset{\boldsymbol{w},\boldsymbol{\gamma}}{\boldsymbol{minimize}} = \boldsymbol{c}\sum_{i=1}^{l} \boldsymbol{\varepsilon}_i + \frac{\mathbf{1}}{\mathbf{2}}\|\boldsymbol{w}\|^2$$

Subject to $$\boldsymbol{D}_{\mathrm{ii}}(\boldsymbol{w}^T\boldsymbol{x}_i - \boldsymbol{\gamma}) + \boldsymbol{\gamma}_i \geq \mathbf{1}; \boldsymbol{i} = \mathbf{1}, \dots \dots, \boldsymbol{l}.\ \boldsymbol{\varepsilon}_i \geq 0$$

In this equation the contribution to the training errors and objective function of margin maximization can be balanced through the use of regularization parameter **c**. The following decision function is used to correctly predict the class of new sample with a minimum error.

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{sgn}[\boldsymbol{w}^T\boldsymbol{x} - \boldsymbol{\gamma}]$$

The benefit of the dual equation is that it allows an efficient learning of non–linear SVM separators, by giving kernel functions. Theoretically, a non-linear kernel function computes a dot product between two vectors that have been mapped into a high dimensional feature space. Then there is no require to perform this mapping explicitly, the training is still possible although the real dimensional feature space can be very high or even infinite. The parameters are obtained by solving the following non-linear SVM equation.

$$\boldsymbol{Minimize\ L_D(u) = \frac{1}{2}u^T Q u - e^T u}$$

$$\boldsymbol{d^T u = 0, 0 \leq u \leq Ce}$$

Where $\boldsymbol{Q = DKD}$ and $\boldsymbol{K}$ the Kernel Matrix. The polynomial or Gaussian of kernel function $\boldsymbol{K}$ $\boldsymbol{(AA^T)}$ is used to construct hyperplane in the feature space, which splits two classes linearly, by achieving computations in the input space. The decision function is specified by

$$\boldsymbol{f(x) = sgn\left[K\left(x, x_i^T\right) * u - \gamma\right]}$$

where $\boldsymbol{u}$ is the Lagrangian multipliers.

When the class labels number is further than two, the binary SVM can be extended to multi class SVM. One of the classical methods for multiclass SVM is method of one versus rest. For each class a binary SVM classifier is created, separating the data points of that class against the rest. Therefore, in case of N classes, N binary SVM classifiers are constructed. Through testing, each classifier produces a decision value for the test data point and the classifier with the maximum positive decision value assigns its label to the data point. The comparison between the values of decision produced by different SVMs is still valid since the training parameters and the dataset remain the same.

10. Conclusion

In this chapter, firstly we presented an overview of the main modules composed a writer's identification and retrieval systems that the work of this thesis is based, where we focus on extracting the features characteristics of specific styles of the writer in the handwriting document, these features and their differences are required to create an effective system of writer's identification and retrieval based on the support vector machine (SVM) classifier or similarity or dissimilarity measures .

in the following chapter, we will present our contributions to the area of writer identification using handwritings document based on single script.

CHAPTER 02

OFFLINE WRITER IDENTIFICATION AND RETRIEVAL BASED ON SINGLE SCRIPT

1. Introduction

Identifying and authenticating people based on human samples, physical or behavioral, have a wide a range of applications, starting from a voice (behavioral) to communicate with a mobile phone to a DNA (physical) to recognize a homicide's writer.
Many researches proposed efficient approaches to identify people using different types of human samples. Handwriting is one of these samples that attracted a lot of interest in the last decades due to its applications in document retrievals, court of justices and also in electronic devices.
In this chapter, we propose an approach for writer identification, where the objective is to assign one of the pre-registered identities in the system to the handwritten sample read as input. Different approaches have been considered in the literature for this task, where they diverge mainly on the set of features used to describe the handwriting samples. Various features, such as LBP, LTP [37], LPQ [28], curvature features [38], RootSIFT descriptors [40], can be divided into two categories: Text-independent methods that classify the writer independently of the sample's semantic content. instead, the text-dependent methods require the samples to be of same fixed content. The text-dependent researches are mainly motivated by forensic applications; for this approaches a non-exhaustive list of some interesting studies is presented in [29], [40-44].

Plentiful competitions with the aim of identifying the writers have been held at well-known conferences such as the International Conference on Frontiers in Handwriting Recognition (ICFHR) and the International Conference on Document Analysis and Recognition (ICDAR), in addition in conjunction with ICDAR 2013 [8], the English and Greek handwriting samples has been orderly. In this chapter, our goal is to analyze the current state of Writer Identification methods based on the handwriting datasets of the competition ICDAR 2013 [8]. One of these methods, labelled as "CS-UMD-a" uses the gradients taken from the contour of the segments of words spliced by sewing cuts to form a feature vector. Features are then grouped together to identify the best representative character set. These character sets are used to calculate the similarity between images.

The proposed system ranked first used different configurations of Complete Local Binary Patterns (CLBP) and Local Binary Pattern Variance (LBPV), and various combinations of distances metrics to identify writer from handwriting document. Hence, the features vectors normalized using Probability Density Function (PDF). The proposed system uses a leave-one-out strategy for ranking according to the similarity between two handwritings. Figure 1 gives an overview of the proposed approach.

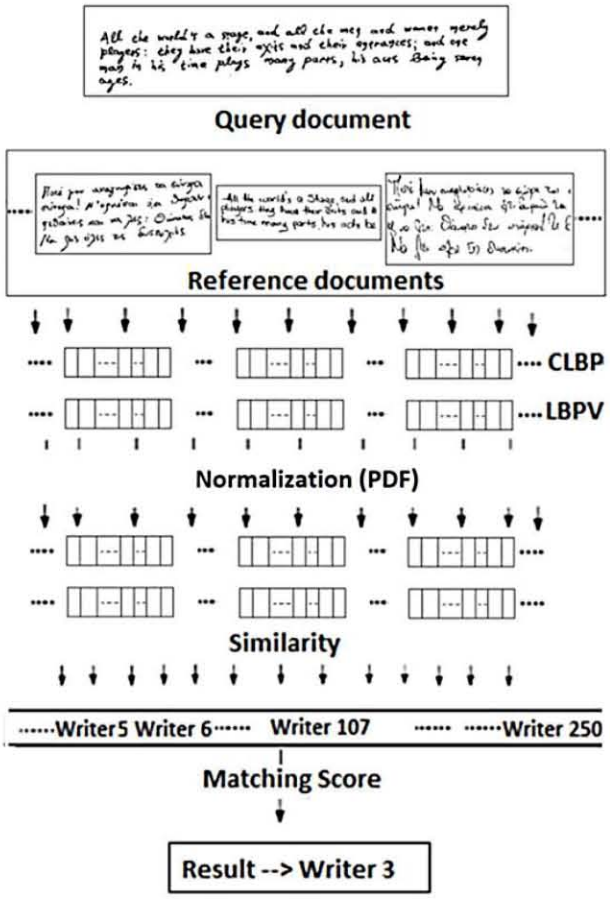


Figure 10: Overview of the Proposed System

In this chapter, first we present some of the relevant works in the writer identification and retrieval. Second, the choice and the motivation of features used in our approach are discussed, in the followed section we present the obtained results, where a comparative discussion is conducted between our approach and other present in the literature. Finally, we conclude with a discussion.

2. Related Work

In many different literature reviews which was conducted in the topic of writer identification [48], where the handwriting variability and its impact on both the writer recognition task and text recognition task have been discussed [49]. For the case of the writer recognition, two approaches determined: the verification (retrieval) task and Identification task. The identification system must take a handwritten sample as input and match it with the ID of what is registered in the system.; at the outset at the beginning of the verification task, the system needs to take two handwritten samples and determine if they were written in the same handwriting. This section describes some of the work done as part of the writer identification task.

All writer identification approaches reported in the literature converge to the same structure, which is clearly represented around the feature extraction module and the decision module. The role of the preprocessing module, if present, is limited to the connected component extraction and image binarization.

Feature extraction is a central module that allows you to observe divergences between different writer identification approaches and can divide the features used into global features (textures) and local features (structures).

Textural features were examined in [37] proposed and evaluated with two IFN/ENIT datasets and IAM datasets, the same method focused on textural features and use three methods to extract features from fragments considered as a texture which come out of the handwriting images division, Local Phase Quantization features (LPQ); which relies on the local phase information extracted from the short-term Fourier transform. LBP features were extracted in addition to a variant of LBP less sensitive to noise and distortion mainly called Local Ternary Patterns features (LTP). With a performances rate of 89.5% and 94.9%, respectively. In [28], the authors extracted features from handwritten blocks based on local phase quantization (LPQ) and local binary patterns (LBP), This approach was evaluated on the Brazilian Forensic

Letter (BFL) dataset and the IAM dataset, the SVM classifier was used, and the authors had been announced that a good identification rate which is 99 %.

As already mentioned, other approaches to identifying authors are based on local features such as grapheme [39] resulting from the process of segmentation words handwriting, and the codebook [48] which are widely used in the literature. In fact, the author used it in [10] to characterize the handwritten junctions. This approach was evaluated on two different datasets, Fire Maker and IAM, with a high identification rate of 94%. using a Fire Maker dataset for evaluation and codebook as a module of characterize features character fragments, that containing fragmented connected component contours (FCO3). the same author [10] obtained a high identification rate of 97% in last method.

Due to their ability to describe the main characteristics of writers based on handwriting text, textural features extraction considered as the most important module in analysis of the writer identification stated in the literature. For this, we investigate in this chapter the relevance of using a couple of features: the VLBP and the CLBP, discussed in the next section.

3. Feature extraction

In this proposed approach, where we aim to identify an individual based on a sample of his handwritten text, we decided to capture the handwriting function using a combination of the LBPV feature and the CLBP feature. These allow to capture and distinguish curvature and texture information in handwritings documents. These descriptors have been successfully used to solve various tasks related to document analysis. These features are described in the following subsections.

3.1 Complete Local Binary Patterns (CLBP)

Local Binary Patterns consider the local structure of the image only discarding the difference of magnitude between the central pixel and its neighboring pixels. The authors of [49] noted that because LBP only evaluates the difference between two gray values, it frequently generates inconsistency codes. The LBP operator generates a binary code for a center pixel with an intensity value. The LPB code generated relates to a dark spot that does not exist in this instance. To address this issue, [49] introduced CLBP, a completed LBP. Each pattern's core gray level data were merged with the pattern's local magnitude and sign information. The sign and magnitude differences are both captured with two bits each. The calculation is summed up in (Equation 1).

$$s_P = s(i_P - i_c) \qquad m_P = |i_P - i_c| \qquad (1)$$

Where s_p denotes the sign difference between the central and nearby pixels' intensity levels, and m_p is the magnitude difference; i_p denotes the intensity level of the neighboring pixel, and i_c denotes the intensity level of the center pixel.

m_p and s_p are further used to calculate CLBP-Magnitude (*CLBP_M*) and CLBP-Sign (*CLBP_S*) .

CLBP Magnitude and CLBP Sign are mathematically expressed in (Equation 2) and (Equation 3) respectively.

$$CLBP\_M_{P,R} = \sum_{P=0}^{P-1} 2^P t(m_P, c), \quad \begin{Bmatrix} t(m_P, \mathrm{c}) = 1, |i_P - i_c| \geq c \\ \mathrm{t}(m_P, \mathrm{c}) = 0, |i_P - i_c| < c \end{Bmatrix} \qquad (2)$$

$$CLBP\_S_{P,R} = \sum_{P=0}^{P-1} 2^P s(i_P - i_c), \quad \begin{Bmatrix} s_P = 1, i_P \geq i_c \\ s_P = 0, i_P < i_c \end{Bmatrix} \qquad (3)$$

Where i_c is the intensity level of center pixel, i_p is the intensity level of neighboring pixel, R is the radius of neighborhood and P is the value of center pixel.

Furthermore, Guo et al. [49] introduce a new operator CLBP Center (*CLBP_C*) based on the gray level of each pattern (Equation 4).

$$CLBP\_C_{P,R} = t(i_c, c_i) \quad (2)$$

Where c_i is the average gray level of whole image and i_c is the gray level value of central pixel. The final CLPB descriptor is created by concatenating the three descriptors and it outdone the traditional LBP [18] for texture classification problems.

3.2 Local Binary Pattern Variance (LBPV)

Local Binary Pattern Variance (LBPV) is used to exploits the complementary information of local contrast into the one-dimensional LBP histogram [50]. A rotation invariant measure of the local variance (VAR) is quantized using the threshold values from the test images. The feature distribution, calculated from all the training images thanks to the threshold values, is defined as:

$$\mathrm{VAR}_{\mathrm{P,R}} = \frac{1}{\mathrm{P}} \sum_{\mathrm{P}=0}^{\mathrm{P}-1} (\mathrm{i}_{\mathrm{P}} - \mathrm{u})^2 \text{ Where } \mathrm{u} = \frac{1}{\mathrm{p}} \sum_{\mathrm{p}=0}^{\mathrm{p}-1} \mathrm{i}_{\mathrm{p}} \quad (3)$$

In order to partition the total distribution into N bins, with the same number of entries, some of the previous threshold values may be used.

The LBPV is a simple but effective method joint LBP and contrast distribution method.

To weight the contribution of the LBP code in the histogram calculation, we can consider the usage of the variance VAR. Furthermore, LBPV does not need any quantization and it is totally training-free. The LBPV histogram is computed as:

$$\mathrm{LBPV}_{\mathrm{P,R}}(\mathrm{k}) = \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{\mathrm{j}=0}^{\mathrm{M}} \mathrm{w}(\mathrm{LBP}_{\mathrm{P,R}}(\mathrm{i,j}), \mathrm{k}) \text{ , } \mathrm{k} \in [0, \mathrm{k}] \quad (4)$$

Where:

$$\mathrm{w}\big(\mathrm{LBP}_{\mathrm{P,R}}(\mathrm{i,j}), \mathrm{k}\big) = \begin{cases} \mathrm{VAR}_{\mathrm{P,R}}(\mathrm{i,j}), & \mathrm{LBP}_{\mathrm{P,R}}(\mathrm{i,j}) = \mathrm{k} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

In addition, these feature vectors are normalized using Probability Density Function (PDF) of an exponential distribution for providing a significant improvement for writer classification.

Indeed, to model the randomness of elapsed times between events, the PDF of an exponential distribution is frequently used. Thus, if the times between events follow an exponential distribution, then the number of events in a specific interval of time follows a so-called Poisson distribution.

The exponential distribution has mean parameter μ which must be greater than zero and evaluated at the values *x* in the vector *X*.

The exponential PDF is

$$f_X(x|\mu) = \begin{cases} \frac{1}{\mu} e^{-\frac{1}{\mu}x} & for\ x > 0 \\ 0 & for\ x \leq 0 \end{cases} \quad \text{Where} \quad \mu > 0 \qquad (6)$$

4. Probability Density Function (PDF)

The normalization of the data before evaluation is often necessary so that the inputs are uniform, in order to normalize the vectors resulting from the feature extraction phase, we used the probability density function (pdf) because it is the most fundamental building block for the statistical description of a variable. The probability density function (PDF), denoted f, of a continuous random variable X satisfies the following:

1. $f(x) \geq 0, for\ all\ x \in R$
2. f is piecewise continuous
3. $\int_{-\infty}^{\infty} f(d)dx = 1$
4. $P(a \leq X \leq b) = \int_b^a f(d)dx = 1$.

5. Decision

The last module in this approach is that which measures the distance between the features' vector of the query document and the features vectors of the reference documents as a result of the training phase. various distances metrics were considered, such as the Euclidean distance, city block distance, correlation distance, cosine distance and Spearman distance. in the step of matching, the final result of matching score reports the select minimum distance that most closely matches the query handwriting document.

The basic statistical reasoning measures based on the least product distance from the best distance metrics are utilized to arrive at a final conclusion by taking into account the decision of many distance metrics for combination distance metrics. We took for the objective of increasing the categorization rate in our study.

6. Experimentation and Results

The experiments aim to study the effect of the mean parameter μ of exponential distribution in normalizing the Complete Local Binary Patterns (CLBP) and Local Binary Pattern Variance (LBPV) features from the binarized image. In addition, the Euclidean distance measure is used for classifying each document. The realized Top1 is illustrated in Table 2 and Table 3.

As can be shown, the $LBPV_{16,8}$, $CLBP_{16,4}$, $CLBP_{16,8}$ features outperform the others features configuration when μ= 84 is used. As a result, these features are extracted from the entire handwriting image.

Table 1 : Top1 rates on ICDAR 2013 competition using CLBP

μ	CLBP 4,1	CLBP 4,2	CLBP 4,4	CLBP 8,1	CLBP 8,2	CLBP 8,4	CLBP 8,8	CLBP 16,1	CLBP 16,2	CLBP 16,4	CLBP 16,8	CLBP 16,16
60	9,20	7,80	4,00	8,00	61,90	33,40	29,60	9,10	67,60	88,60	89,70	76,90
62	9,40	8,10	4,00	8,00	63,00	34,30	30,80	9,20	68,00	88,80	90,00	77,10
64	9,40	8,40	4,00	8,00	63,40	35,50	31,80	9,10	68,70	89,00	90,10	77,00
66	9,40	8,40	4,10	7,80	64,00	37,00	32,40	8,80	69,40	89,50	90,10	77,50
68	9,60	8,70	4,40	8,20	64,20	38,10	33,40	9,10	69,80	89,50	90,20	78,10
70	9,70	9,30	4,40	8,20	64,40	39,00	33,70	9,10	70,10	89,50	90,20	78,50
72	9,90	9,80	4,40	8,30	64,70	40,20	34,20	9,20	70,20	89,50	90,30	78,80
74	10,00	10,20	4,40	8,30	65,10	41,40	35,10	9,10	70,60	89,40	90,40	78,90
76	10,00	10,60	4,50	8,30	65,50	42,60	35,80	9,10	70,70	89,70	90,70	79,20
78	10,10	10,70	4,70	8,70	65,90	43,70	36,40	9,40	71,40	90,10	90,80	79,60
80	10,30	11,00	4,70	8,70	66,90	44,30	37,70	9,50	71,20	90,40	91,00	79,90
82	10,40	10,80	4,80	8,90	67,10	44,50	38,50	9,70	71,50	90,40	91,20	80,20
84	10,50	11,10	4,80	8,90	67,10	45,00	39,70	9,70	71,70	90,40	91,30	80,20

Table 2: Top1 rates on ICDAR 2013 competition using LBPV.

μ	LBPV 4,1	LBPV 4,2	LBPV 4,4	LBPV 8,1	LBPV 8,2	LBPV 8,4	LBPV 8,8	LBPV 16,1	LBPV 16,2	LBPV 16,4	LBPV 16,8	LBPV 16,16
60	6,80	9,50	6,90	21,10	61,70	52,30	67,20	21,00	56,60	81,90	90,10	79,70
62	6,90	9,40	7,20	20,90	61,80	52,70	67,90	21,00	56,30	82,00	90,10	79,80
64	7,00	9,40	7,60	20,90	61,60	53,00	68,30	20,80	56,20	82,00	90,10	80,00
66	6,90	9,80	8,00	20,90	62,10	53,40	68,90	20,90	56,20	81,40	90,30	79,90
68	6,90	9,80	8,10	20,90	61,70	54,50	69,30	20,80	55,70	81,20	90,30	80,20
70	6,90	10,10	8,40	20,90	61,90	55,20	69,90	20,70	55,80	81,30	90,10	80,20
72	7,00	10,20	8,50	20,40	61,80	55,50	70,30	20,80	55,80	81,50	90,10	80,20
74	7,50	10,10	9,00	20,50	61,80	56,30	70,70	20,90	55,70	81,20	90,20	80,40
76	7,80	10,20	9,10	20,30	61,80	57,20	70,70	21,10	55,40	81,20	90,30	80,50
78	7,70	10,20	9,20	20,50	61,70	57,70	70,60	21,00	55,00	80,80	90,30	80,60
80	8,00	10,20	9,30	20,70	61,60	58,20	70,90	21,10	54,60	80,50	90,20	80,80
82	8,10	10,50	9,80	20,90	61,60	58,70	71,30	21,20	54,50	80,40	90,20	80,80
84	8,10	10,60	10,00	21,10	61,50	59,40	71,40	21,20	54,50	80,30	90,30	81,00

To increase the classification rates, we applied several distance metrics such as correlation distance, cosine distance, Spearman distance, and city block distance in addition to the Euclidean distance.

Table 4 summarizes the results of the several distance metrics used to calculate the performance of these features. Using correlation distance, LBPV $_{16,8}$, CLBP $_{16,8}$, and CLBP $_{16,4}$ achieved the highest precision Top1 of 90.40 percent, 91.30 percent, and 90.30 percent, respectively.

Table 3: Writer identification rates with different features using the different distance metrics.

	Feature Histogram Description		
	CLBP $_{16,4}$	**CLBP $_{16,8}$**	**LBPV $_{16,8}$**
DIM	486	486	243
EUCL	87.70	88.90	82.10
CORR	**90.40**	**91.30**	**90.30**
COSINE	90.00	90.90	85.10
SPEARMAN	89.50	93.70	90.60
CITYBLOCK	87.20	88.50	84.00

Also, to increase the classification rate, we calculated the precision Top1 for several combinations of best distance metrics. The proposed method's performance was investigated by calculating the minimum of the product (Prod) of several distance metrics with matching characteristics (features).

Table 5 summarizes the results obtained using these combination schemes. In general, the classification rates of combination scheme based on the minimum of product of the cosine distance of CLBP$_{16,4}$ and Spearman distance of LBPV$_{16,8}$ are relatively high as compared to other combinations distance metrics as well achieving the precision Top1 of 95.70%.

Table 4: writer identification rates for various combination schemes.

Combination schemes			**Top 1 (%)**	
D1: **CLBP $_{16,4}$**	**D2:** **CLBP $_{16,8}$**	**D3:** **LBPV$_{16,8}$**	**Prod (D1, D3)**	**Prod (D2, D3)**
CORR	CORR	CORR	95.40	93.40
		COSINE	94.10	92.70
		SPEARMAN	95.50	93.40
		EUCL	94.50	93.00
		CITYBLOCK	94.40	93.00
COSINE	**COSINE**	CORR	95.30	93.00
		COSINE	93.60	91.70
		SPEARMAN	**95.70**	93.30
		EUCL	94.10	92.00
		CITYBLOCK	94.30	92.20
SPEARMAN	SPEARMAN	CORR	93.20	94.20
		COSINE	91.80	93.50
		SPEARMAN	92.40	93.60
		EUCL	91.00	**94.30**
		CITYBLOCK	91.60	94.20
EUCL	EUCL	CORR	94.40	91.70
		COSINE	91.80	89.10
		SPEARMAN	95.50	92.90
		EUCL	92.40	89.00
		CITYBLOCK	92.40	89.10
CITYBLOCK	CITYBLOCK	CORR	93.30	91.40
		COSINE	91.00	88.80
		SPEARMAN	95.30	92.50
		EUCL	90.70	88.50
		CITYBLOCK	91.50	88.90

According the previous results, we are evaluated the proposed method using the optimal decision strategy compared with the four (4) best systems submitted to ICDAR 2013 competition on Writer Identification. Table 6 reports the comparison of the proposed method with the same as that of the ICDAR 2013 competition (1000 document images (Greek and English)).

Table 5 : Comparison of proposed method with ICDAR 2013 methods

Rank	Method	Top1
1	**Proposed Method**	**95.70**
2	CS-UMD-a [8]	95,10
3	CS-UMD-b [8]	95,00
4	HIT-ICG [8]	94,80
5	TEBESSA-c [8]	93,40

Table 6 shows clearly that our approach outperforms the other methods. This fact, proves the validity of the texture features with optimal combination schemes for Writer Identification.

The second phase of experimentation was conducted using only the Greek part of the benchmarking dataset (500 images) and only the English part of the benchmarking dataset (500 images). The evaluation results of proposed system with optimal combination shames for each language independently described in the Table 7.

Table 6 : writer identification rates using only the Greek part and the English part of the benchmarking dataset.

Rank	Method	Script	Top1	Average
1	**Proposed Method**	Greek	**97.20**	**95.20**
		English	93.20	
2	CS-UMD-a [8]	Greek	95.60	95.10
		English	**94.60**	
3	CS-UMD-b [8]	Greek	95.20	94.80
		English	94.40	
4	HIT-ICG [8]	Greek	93.80	93.00
		English	92.20	
5	TEBESSA-c [8]	Greek	92.60	91.90
		English	91.20	

It can be seen from Table 7 that the proposed method outperforms other methods using for each language independently (script-dependent) Greek and English. It should however be noted that the proposed system does not require any preprocessing and the features are directly extracted from document images.

7. Conclusion

In this chapter, we proposed an original approach to identify individuals based on their handwritten documents. we proposed to characterize the handwriting a couple of features: CLBP and LBPV histograms. These features offer the benefit of capturing both the curvature and the textual information. To measure the distance between two features vectors, we used different metric distances. The experimentations were carried out according to the same experimental protocol of the ICDAR 2013 . The obtained results have showed clearly that our approach outperformed the methods reported in the ICDAR competition.

In the next chapter we investigated the effectiveness of texture feature column scheme to characterize writer from his handwriting documents written in different languages.

CHAPTER 03

OFFLINE WRITER IDENTIFICATION BASED ON MULTI SCRIPT

1. Introduction

Computerized analysis of handwriting and handwritten documents is one of the most active types of research are as for many decades that have attracted considerable research interest of document examiners, forensic experts, psychologists, neurologists and paleographers. Contrary to machine printed text, handwritten text and Hand-drawn figures carry valuable and varied information about the person who produces these documents. The unique writing characteristics of an individual (depicted through writing style) make it possible to employ handwriting as a behavioral biometric modality hence allowing identification and verification of writers from handwritten documents. Formally, writer identification task includes finding a writer's identity based on a query of a handwritten document comparing it with a set of writing samples with known writers. Writer verification, on the other hand, includes deciding whether two writing samples have been produced by the same individual or not.

Physiological biometrics (Iris, fingerprint, hand geometry, retinal blood vessels, DNA…etc.) are strong modalities for person identification due to the high complexity of the biometric templates used. On the contrary, behavioral biometrics such as handwriting [46], [51],[53],[54] are less invasive, but the achievable identification accuracy is less impressive due to the large variability of the behavior-derived biometric templates.

Analysis of handwriting system can be integrated an expert system to identify the writer of handwritten text. Design of a practical application of writer identification can perform as well as a human expert and exhibit characteristics of a traditional expert system [52], [55], [56]. It is Important to mention that such solutions are aimed at facilitating and not replacing the human experts. Hence, the automatic systems can be employed to reduce the search space so that human experts can focus on a manageable sized list of potential candidates. A number of works have been carried out to integrate the information from forensic experts into expert system where a subset of features employed by the experts is algorithmically computed by computer programs. As discussed earlier, such computerized systems can be employed to reduce the search space and the hit-list returned by such systems can be subjected to examination by human experts to come to a conclusion.

Many researchers were attracted towards handwritten writer identification which is the most promising area of image analysis and machine learning. Due to advancements in technology pattern recognition and image processing lot of improvement was observed in Writer Identification from offline handwriting images. The latest developments on this problem can be found in [57], [58], [55],[56]. These systems aim to capture the visual differences in the individual's handwriting, including allograph variations, slope of lines and slant of characters, line spacing, inter and intra-word spacing, legibility, cursivity, readability, line tilt and tilt characters, etc. in Figure 11.

A number of recent studies on writer identification rely on extracting the writing features to characterize the writer. In our study, we aim to use texture feature based on LBP Column histogram and oBIFs columns histogram to build a feature extraction. The feature vector is then employed to classify writer using Support Vector Machine (SVM). An overview of the proposed method is presented in Figure 12. The key theme of this study is to combine oBIF or LBP at two different scales or parameters as we aim to characterize the writing style based on 'how' the text is written rather than 'what' is written. The proposed approach is validated on single-script and multi-script writer identification approach and high identification rates are reported in a number of experimental scenarios.

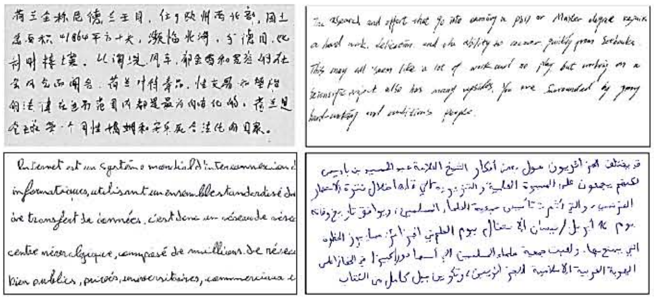


Figure 11.Samples of two different writers

This chapter is organized as follows. In the next section, we discuss the relevant literature on this problem. We introduce the texture features employed in our study in Section III. Section IV presents the experimental results, the classification scheme and the accompanying analysis.

Finally, the last section concludes the chapter with some useful insights for further research on this problem.

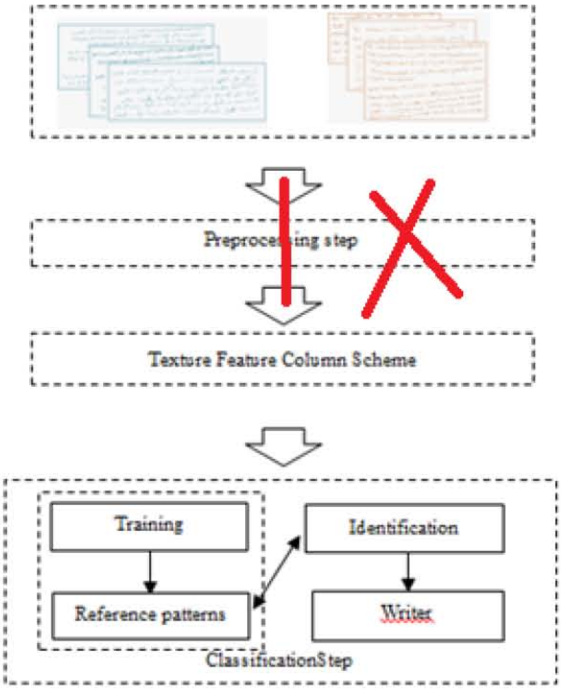


Figure 12. Overview of the Proposed System.

2. Related work

Writer identification techniques reported in the literature are traditionally categorized into single-script and multi-script approaches. Single-script methods for writer identification have been a popular choice of researchers, they are used for to identify a writer based on a particular script employed by forensic experts [53],[61], [62],[63]. Multi-script approaches are also a very practical situation, where aggregation of a variety of handwritings script coming from same writer is used to recognize his identity. Performance of multi-script approaches has hardly been studied in the literature [62], [28], [6].

Among one of the contributions, based on textural features to identification of writers from handwriting, [61] proposed to use textural measures, including Local Phase Quantization (LPQ) and Local Binary Patterns (LBP), that were extracted from normalized blocks of handwriting to capture the writing style. Experimental study on two different datasets, the IAM dataset and the Brazilian forensic letter dataset, realized high identification rates. In another study, [63] proposed to divide handwriting into small windows and consider each window as a unique textural pattern. Textural descriptors including Local Ternary Patterns (LTP), Local Phase Quantization (LPQ), and Local Binary Patterns (LBP) are then computed from these fragments. This technique was evaluated on the complete set of samples in IFN/ENIT and IAM datasets, and among the employed descriptors LPQ reported the best identification rates on both the datasets. The oriented Basic Image Feature Columns (oBIF Columns) proposed by [40] is the probability distribution of the bank of six Derivative-of-Gaussian filters on two scales. The features have been evaluated using the IAM dataset and by making entries to two top international competitions for assessing the state-of-the-art in writer identification, these features provide a significant improvement for writer identification.

On the other hand, other works are based on writing segmentation into graphemes and codebook to characterize the writer [61],[60]. Among well-known codebook-based writer characterization techniques, [17] propose to employ a codebook of fragmented connected-component contours (FCO3) extracted from character fragments (fraglets). This codebook is generated by clustering these fragments using Self Organizing Maps (SOMs), the probability distribution of fraglet contours was computed for an independent test set. The study describes new algorithms for forensic or historical writer identification. Likewise, [18] extended the same idea and used small writing fragments to extract writer-specific frequent patterns. Distribution of these patterns was combined with curvature and orientation features extracted from writing contours. In another study, [66] proposed to use a beta-elliptic model in order to generate a synthetic codebook. In this method, a feature selection was proposed to reduce the codebook's size where the feature extraction is performed using a template matching approach. The proposed technique was evaluated on the 411 writers in the IFN/ENIT dataset. In another notable work [38], the authors propose a junction detection method using crossings of strokes in the handwriting. The detected junctions are then employed to generate a codebook of 'Junclets' to characterize the writer. the study of [38] proposed a writer identification method using codebook extension model with an ensemble of codebooks (of graphemes) in which a kernel discriminant analysis using spectral regression (SR-KDA) is deployed as a dimensionality reduction technique to avoid over-fitting problem. However, [40] employ the bagged discrete cosine transform (BDCT) descriptors for offline text independent writer identification. Multiple models are generated using universal codebooks and a final decision is then obtained by using the majority voting rule from these predictor models.

In another recent work, [69], writer characterization is investigated using an implicit shape codebook technique. The proposed approach relies on identifying the key points in handwriting and clustering the patches around these key points to generate an implicit shape codebook. The authors in [70] employ RootSIFT descriptors computed densely at contour edges. GMM super vectors are created by adapting a background model to the distribution of local feature descriptors. The local descriptors are used as input to an Exemplar-SVM which ranks the documents as a function of similarity with the query document.

Recently, the study specifically targets deep learning based automatic feature extraction techniques have also been investigated for the text-independent approach to writer identification. In [71] for example, authors proposed an end-to-end deep-learning method using convolutional neural networks to automatically extract local features from writing samples. Experiments on multiple datasets demonstrated the effectiveness of machine learned features over hand-crafted features. A summary of well-known contributions to writer identification reported in the literature is presented in Table 8.

Table 8: Performance comparison of well-known writer identification systems

Study		Database	Language	Number of writers	Classification rates (%)
Features	Methods				
Textural features	Bertolini et al. (2013)	IAM	English	650	96.70
		BFL	Portuguese	315	99.20
	Hannad et al. (2016)	IFN/ENIT	Arabic	411	94.89
		IAM	English	657	89.54
		CVL	English	310	96.20
	Siddiqi and Vincent (2010)	IAM	English	650	91.00
	He et al. (2015)	IAM	English	650	91.10
		Firemaker	Dutch	250	89.80
		CERUG-MIXED	Chinese and English	105	96.20
	Abdi and Khemakhem(2015)	IFN/ENIT	Arabic	411	90.00
Codebook-based	Khalifa et al. (2015)	IAM	English	650	92.00
	Khan et al. (2017)	IAM	English	650	97.20
		CVL	English	310	99.60
		IFN/ENIT	Arabic	411	76.00
	Bennour et al. (2019)	BFL	Portuguese	300	98.33
		CVL	English	300	94.00
		KHAT	Arabic	1000	62.81
Local descriptors	Christlein et al. (2017)	IAM	English	650	88.00
		CVL	English	310	99.20
		KHATT	Arabic	1000	98.80
CNN	Nguyen et al. (2019)	JEITA-HP	Japanese	400	93.82
		Firemaker	English	250	92.38
		IAM	English	650	90.12

The writer identification systems on single-script approach have been generally employed for writer identification evaluation while IAM and CVL datasets used for English script, the IFN/ENIT and KHATT dataset have been considered in most of the studies for Arabic script.

It is important to mention that many of the datasets, such as: IAM, and IFN/ENIT, were primarily developed for segmentation and recognition evaluation tasks and not for writer identification task. However, these datasets were employed in writer identification tasks include the CEDAR letter [46], the BFL dataset [72], ICDAR 2015 writer identification dataset [62], ICFHR 2016 writer identification dataset [28] and ICFHR 2018 writer identification dataset [6].

An analysis of the writer identification methods based on feature extraction techniques reported in the literature reveals the textural features remain a popular choice of researchers. Since, we are investigating mainly the multi-script identification task; the feature characterizing a writer cannot be local features, and can relies only on the texture. In the present study, we investigate a new method based on LBP Column histogram and the oBIF Colum histograms to identify writers independently from the script used.

3. Proposed Approach

The proposed writer identification approach depends on two main components: the feature extraction component and classification component (Figure 12). A preprocessing step, where the handwriting image is binarized using global thresholding is performed before the feature extraction and the SVM classifier training. Each of these modules is discussed in detail in the following.

3.1. Texture Feature Column Scheme

The most important component in any writer identification system is the features extraction module. Indeed, the performance of the identification system relies mainly on the choice of the features set used to characterize the writer. These features have to embed a strong discriminant power to distinguish efficiently one writer from another.

Our approach is based on the combination of two sets of texture features: the LBP and oBIF column histograms described next.

3.1.1. LBP Column Histogram Scheme

As a continuity of the Texture Unit (TU) proposed by [76], a new efficient texture tool analysis: LBP (Local Binary Pattern) has been proposed in [74], where each pixel is encoded with a decimal number based on its local surrounding structure. This encoding is operated first by subtracting the intensity value of a pixel from its eight neighbors. If the resulted value is negative, the neighbor is assigned 1, and 0 otherwise. Finally, a binary sequence of the eight neighbors is obtained by a clockwise concatenation; this sequence is then converted into a decimal value which represents it LBP code.

The basic LBP calculated in a local 3 x 3 neighborhood suffers from capturing large scale structure that may be the dominant features captured using neighborhoods of different sizes [75]. Local neighborhood can be defined using circular neighborhood, which is centered at the pixel to be labeled, and the sampling points which are not part of the pixels are interpolated by bilinear interpolation, thus making it possible to define a circle of radius R and any number of sampling points P in the neighborhood denoted (P, R). Figure 13 illustrates two neighbor-sets for different values of P and R. Also, it is referred to Extended LBP.

.

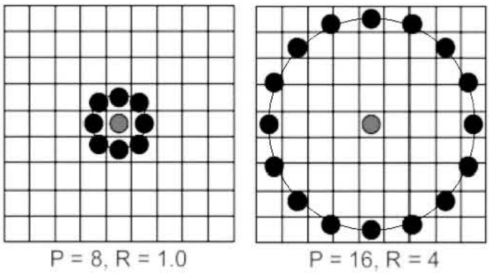


Figure 13. Circularly neighbor sets for two different values of P and R

The coordinates of P neighbors (x_p, y_p) from the coordinates of the center pixel (x_c, y_c) on the edge of the circle with radius R can be calculated with the sinus and cosines, as follows:

$$x_p = x_c + R\cos\left(\frac{2\pi p}{P}\right)$$

$$y_p = y_c + R\sin\left(\frac{2\pi p}{P}\right)$$

If the intensity value of the center pixel is g_c and the intensity values of its neighbors are g_p, with p= 0... P−1, then the texture T are considered only the signs of the differences in the local neighborhood of pixel (x_c, y_c), and is defined as:

$$T \approx \left(s(g_0 - g_c), \ldots, s(g_{P-1} - g_c)\right)$$

Thus, the local texture is represented as a joint distribution of the value of the center pixel and the differences. However, the extended LBP operator [28] is computed to each sign $s(g_P - g_c)$ at each pixel location by taking into account the values of a small circular neighborhood with radius R pixels around the value of a central pixel (x_c, y_c) a binomial weight 2^p, as follows:

$$\mathrm{LBP}_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} s(g_P - g_c)2^P$$

where P is the number of pixels in the neighborhood, R is the radius, and s($g_P - g_c$) = 1, if $(g_P - g_c) \geq 0$, and 0 otherwise. The histogram of these binary numbers is then used to designate the texture of the image.

In our case, a uniformity measure of a pattern is used when it includes at most two bitwise transitions from 0 to 1, or from 1 to 0, when the corresponding bit string is measured circular the number of bitwise transitions from 0 to 1, or the opposites, the bit pattern is measured circular. Therefore, P(P−1) +3 is the number of different output labels for mapping for patterns of P bits. For illustration, the uniform mapping of neighborhoods of 8 sampling points produces 59 output labels, and 243 labels is produced in the case of neighborhoods of 16 sampling points.

For the problem of writer identification, the LBP calculation takes two parameters, the sampling points P on a circle of radius R; the encoding process is illustrated in Figure 14, The LBP column features are then calculated across the image to generate a histogram. The histograms are then normalized, by dividing by the total number of locations which are at both LBP images. Finally, to equalize the variances of the bins of the histogram, the normalized histograms are square rooted. Figure 14 shows the different steps in the coding scheme. The encoded image can be used to train a classifier for recognition.

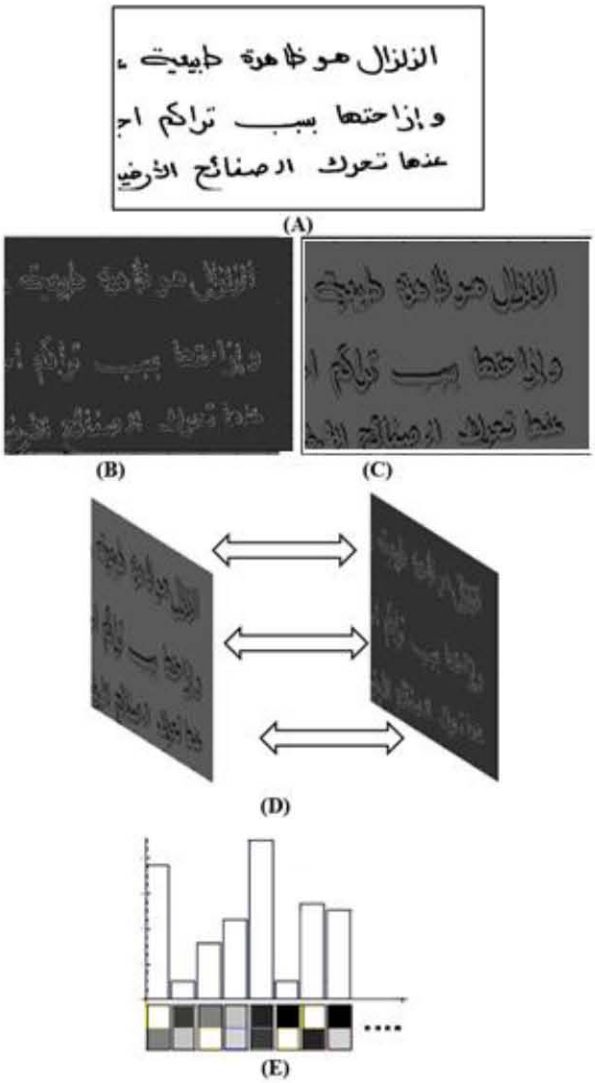


Figure 14. Different steps of the LBP Column scheme (A) Original image (B) image of LBP (P=8, R=2) (C) image of LBP (P=4, R=16) (D) Both LBPs images at different parameters are crossed to form columns at each location (E) the histogram is computed with columns.

3.1.2. oBIF column histogram scheme

The second set of features used in our writer identification and retrieval task the oriented Basic Image Features (oBIF) that capture the textural information in a handwriting for a discriminatory representation. These descriptors have been successfully useful to problems, including: texture classification [77], gender classification [85], modern writer identification document [16], digit recognition [76], and Historical handwritten document [87]. The texture-based descriptor oriented Basic Image Features (oBIFs) is an extension to the Basic Image Features (BIFs) [80],[81] which can be slope, flat, dark rotational, light rotational, saddle, light line on dark or dark line on light. Each location in the image is categorized into one of seven local symmetry classes depending on local symmetry type, which can be slope, flat, dark rotational, light rotational, saddle, light line on dark or dark line on light.

This feature calculated using a set of six derivatives of Gaussian filters (up to second order) determinate by a scale parameter σ, is very efficient in detecting local symmetry classes. The parameter ε determines if a location is to be classified as flat. The dimension of the oBIF feature vectors 5n + 3. In our study, the orientations calculated into n = 4 levels resulting in 23 entries in the oBIFs dictionary.

We combine oBIF at two distinct scales to provide the oBIF column features [60] by ignoring any symmetry type flat to improve the performance of the oBIF descriptor. As a result, the oBIF column histogram scheme produces 484 entries up to 5 (n + 2)2. The features of the oBIF column histogram are determined by two parameters: the parameter ε, which is set to one of three small values of ε∈{0.1,0.01,0.001}, and the scale parameter where σ∈{1,2,4,8,16}.

Finally, the generated feature vector is normalized. Figure 15 shows the various steps in the oBIF column histogram scheme.

The final features vector of our writer identification system is formed by concatenating the LBP or oBIFs column histograms extracted from each handwritings image. This vector has been normalized so that the mean and unit variance are zero. Different configurations are formed by changing the two parameters of the oBIFs column histograms σ and ε, and the parameters of the LBP column histograms, the sample points P on a circle of radius R, these configurations are described in detail in Section 4.

4. Classification

We used the Support Vector Machine (SVM) as a classifier for the classification task [82], [83]. The following are the features that were extracted: The column histograms of LBP and oBIFs are used to train our (SVM) and initiate the learning of the defined classes classifier. The Radial Basis Function (RBF) kernel was used, with a kernel parameter set to [0, 100] and the soft margin parameter C is fixed to 10.

To improve the reliability of system performance, the final decision is made by selecting the maximum value from both normalized decision functions (LBP column histogram, oBIFs column histogram) for each feature.

More precisely, each feature is allied to an SVM classifier providing a normalized response termed $(fn)_j(x)$, j=1,...n such that x is the handwriting feature vector (LBP column histograms or oBIFs column histograms).

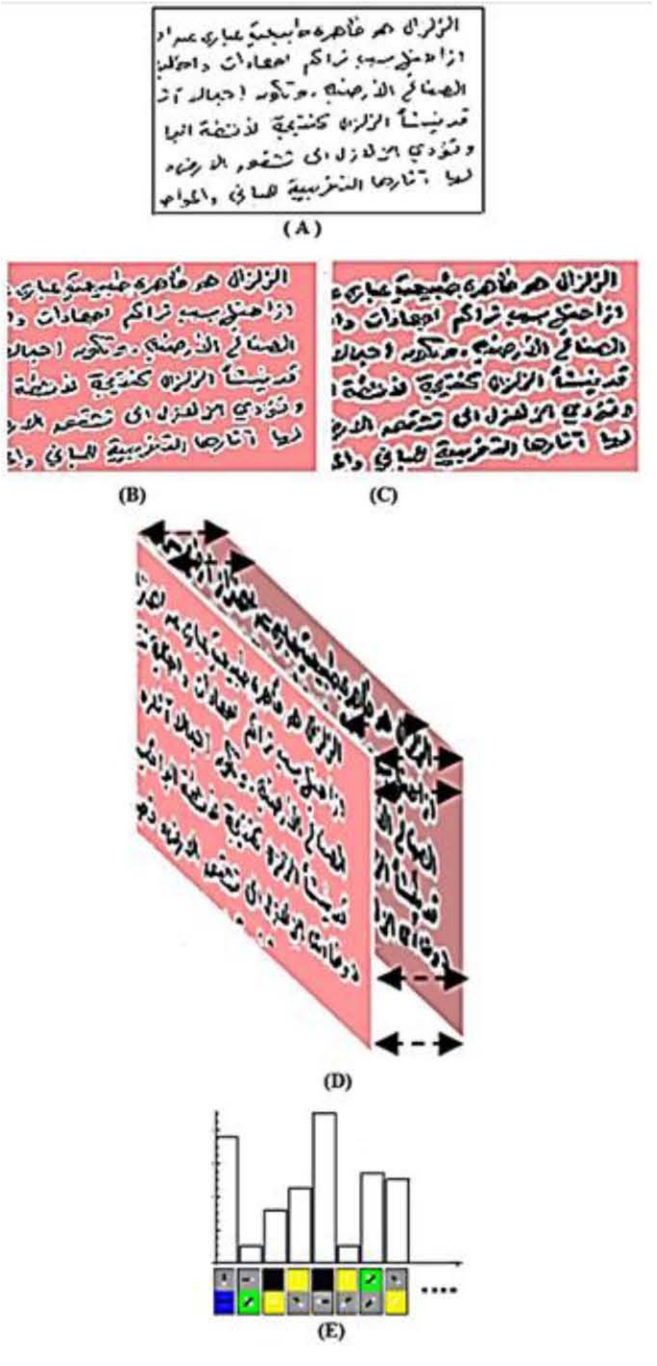


Figure 15. The steps of the oBIF column histogram (A) Original handwriting image (B) oBIF computation for scale parameter σ =8 and σ=4 (c) the oBIFs at two scales are crossed to form columns at each location (D) the histogram is computed with non-flat columns

Therefore, n is number of writers according to the following decision rule:

$$\mathrm{f}_{\max}(x) = \max\{\mathrm{fn}_{\mathrm{j}}(x); \mathrm{j} = 0, .., \mathrm{n}\}$$

$\mathrm{f}_{\max}(\mathrm{x})$ is the maximum value selected from the n responses provided by the SVM classifier, Where, $\mathrm{fn}(\mathrm{x})$ is the nth mapping to the interval [0 1] of the original SVM output values $\mathrm{f}(\mathrm{x})$ as following:

$$\mathrm{fn}(x) = \frac{1}{1 + e^{-f(x)}}$$

Where f(x) represents the output of the SVM obtained for a handwriting *x* to be classified.

5. System evaluation

Two main series of experiments validate the efficacy of the proposed approach. We first try to determine the optimal configuration of the LBP and oBIFs column histograms in traditional writer identification systems on single-script approach (Latin/Arabic), where the training and the test samples are within the same script. We used two standard datasets for this purpose: the Brazilian Forensic Letter (BFL) dataset [72] and the KHATT dataset [84], [85].

The second phase of our experiments is focuses on multi-script writer identification, the experiments are performed on three datasets, CERUG [46], WDAD [67] and LAMIS-MSHD [86] using the same experimental protocol as that of the ICFHR 2018 competition [6].

5.1. Dataset and Experimental Protocol

The experiments are carried out on single-script approach and multi-script writer identification. however, the datasets were developed primarily for the evaluation of writer identification systems. In the single script datasets, each writer provided samples in Latin and in Arabic. therefore, each writer in the multi-script datasets provided handwriting samples in English, Arabic, French, Farsi and Chinese. The details of the respective experimental protocols are shown below.

5.2. Single-script protocol

Two distinct datasets were employed in the single-script writer identification protocol: the Brazilian Forensic Letter (BFL) and KHATT datasets.

The BFL dataset consists of 945 handwriting samples representing 315 writers (3 samples per writer). The textual content on every sample covers punctuations, words, special symbols, character combinations and diacritics in Portuguese making this dataset applicable for evaluation of text-dependent writer identification systems. In our experiments, we perform triple cross validation. In each of the experiments, 630 samples from 315 writers are used as a training set, and the remaining 315 samples from these writers formed the test set.

KHATT is the second dataset employed in our experiments, it comprises of 4000 different handwritings samples (Arabic Texts) representing 1000 different writers. However, the experiments are carried out five folds cross validation using 2000 samples as test set and 2000 samples as a training set for each run.

5.3. Multi-script protocol

As explained above, during the second phase of our experimentations, we used three different datasets: CERUG, LAMIS-MSHD and WDAD by performing six different tasks as described in the framework presented in the ICFHR 2018 competition [64]

The most exciting aspect of the ICFHR to evaluate the performance of any writer identification approach to learn the writing style of an individual, independently from of the writing script; the writing samples proposed are written in Chinese, Arabic, English, Farsi and French. ICFHR competition proposes to consider the six tasks below:

- Task 1: is created on English and Chinese samples of the CERUG dataset. The test dataset comprises 80 unlabeled handwritings images written in English by 40 different writers while the training dataset includes 80 handwritings samples written in Chinese by 40 different writers.
- Task 2: is created on Chinese and English samples of the CERUG dataset. The training dataset contains 80 samples produced in English by 40 different writers while the test dataset comprises 80 unlabeled handwritings images written in Chinese by 40 different writers.

- Task 3:is targeted writing samples in French and Arabic produced by the LAMIS-MSHD dataset. The training test dataset comprises 240 unlabeled handwritten images written in French by 40 different writers while the training dataset contains 240 samples in Arabic produced by 40 different writers.

- Task 4: is targeted writing samples in Arabic and French produced by of the LAMIS-MSHD dataset. The test dataset comprises 240 unlabeled handwritten images produced in Arabic by 40 different writers while the training dataset comprises 240 samples written in French by 40 different writers.

- Tasks 5: is carried out on handwritings samples in the WDAD dataset. 160 samples written in Farsi by 80 different writers were provided as training dataset while the test dataset contains 160 unlabeled handwritings images written in English.

- Tasks 6: is carried out on handwritings samples in the WDAD dataset. 160 samples produced in English by 80 different writers were used as a training dataset while the test dataset contains 160 unlabeled handwritings images written in Farsi by 80 different writers.

All evaluations are performed using the mentioned experimental protocols. Classification rates are analyzed, according to the parameter σ and ε in the oBIFs column histograms and according to the parameter P on a circle of radius R for the LBP columns histograms. The combined decisions from the best features of oBIFs and LBP column histograms are used to improve the classification rates. The results achieved are also compared to many of well-known current approaches. The next sections present the details of these experiments.

Table 9: Distribution of Writers and Samples in the five datasets.

	KHATT	**BFL**	**CERUG**		**LAMIS-MSHD**		**WDAD**	
Evaluation type	**Single Script**		**Multi-Script**					
	-	-	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6
Language	**Arabic**	**Portuguese**	English/ Chinese		Arabic/ French		Farsi/ English	
samples #	4000	945	160		480		320	
Writers #	1000	315	40		40		80	
Samples in # Test set	2000	315	80(English)	80(Chinese)	240(French)	240(Arabic)	(English) 160	160(Farsi)
Samples # Validation set	2000	630	80(Chinese)	80(English)	240(Arabic)	240(French)	160(Farsi)	(English) 160
Samples # per writer	4	3	4(Chinese,2 English 2)		12(Arabic, 6 French 6)		4(Farsi, 2 English 2)	

5.4. Evaluations on Single-script protocol

The evaluation protocol on Single script means that the languages of the samples used in the test dataset and the training dataset is the same. The experiments on Single script aim to study the effect of the parameters (P, R) in computing the LBP columns histogram as well as scale parameter σ and the parameter ε in computing the oBIFs columns histogram. The realized classification rates shown in Figures 16 and 17 correspond to the BFL and KHATT datasets, respectively.

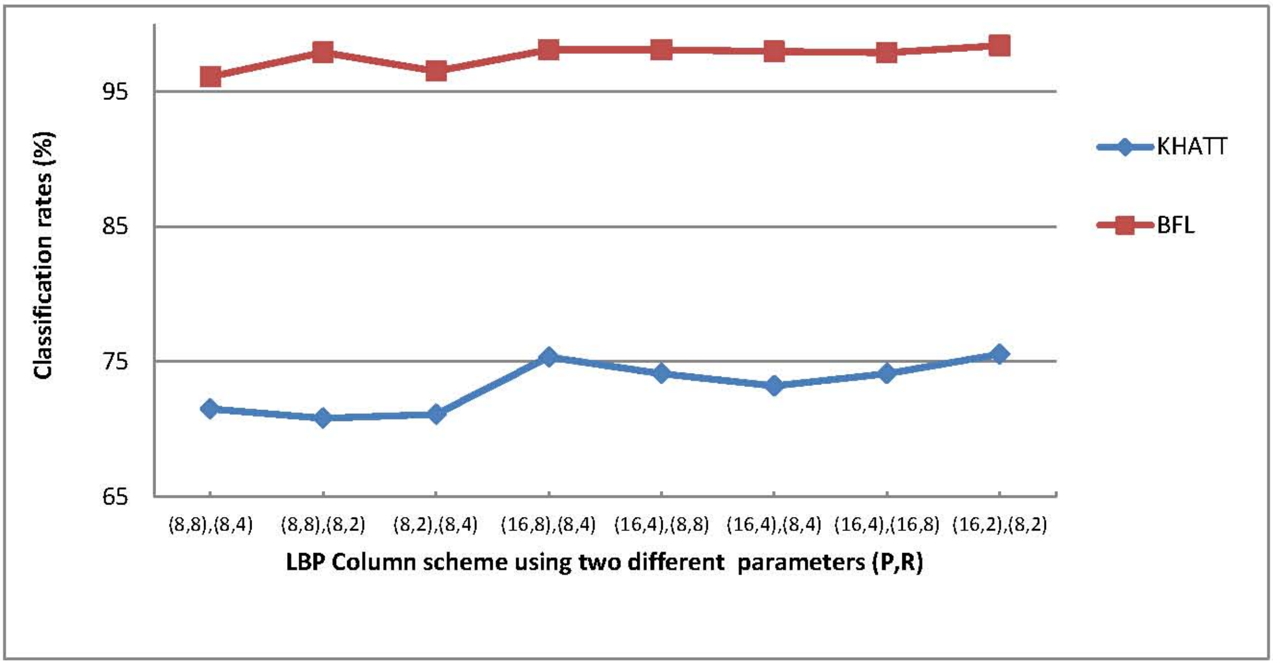


Figure. 16. Classification rates on BFL and KHATT dataset using the LBP column histograms

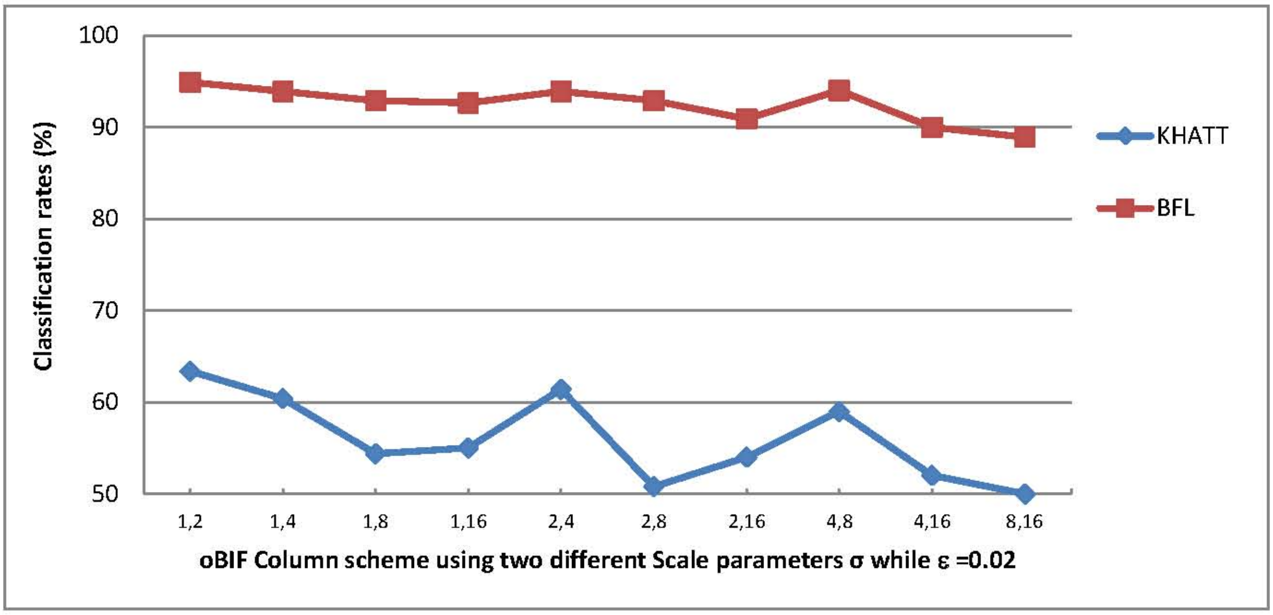


Figure. 17. Classification rates on BFL and KHATT dataset using the OBIF column histograms

From Figure 17 and Figure 18, we can see that the histogram for the LBP column is better than the histogram for the oBIF column in both datasets. This is because of using the cross of best both LBPs images at two parameters (P, R). Table 10 summarizes the best performing configuration of the LBP column histogram, the oBIF column histogram, and their combined decisions. It can be observed that the combined decision at different configurations of LBP column histogram and oBIFs columns histogram, produce up to 98.63 % and 77.10 % classification rates (average) for the BFL and KHATT dataset respectively.

The reason for these rates is the fact that the BFL dataset is text-dependent (all writers wrote the same text), while the KHATT dataset is text-independent. In addition, KHATT dataset contains more writers (1000 writers) and every writer is represented by few words per sample only. However, the rates obtained with our approach, in the KHATT dataset, are higher than those obtained in [72], using the same evaluation protocol. The authors in [39] who have used 15% only of KHATT dataset in their identification task have obtained better results comparing to our approach where we used 100% of this same dataset.

Table 10: Classification rates of writer on BFL and KHATT datasets.

Datasets		Features	Parameters	Dim.	SVM Parameters	Average Identification rates (%)			
						Top 1	Top 2	Top 5	Top 10
BFL	*f1*	LBP column histogram	LBP at $\mathbf{P} = \mathbf{16}\&\mathbf{R} = \mathbf{2}$ LBP at $\mathbf{P} = \mathbf{8}\ \&\mathbf{R} = \mathbf{2}$	14337	C=10, $\boldsymbol{\sigma}$= 50	98.41	99.37	99.37	100.00
	f2	oBIF column histogram	oBIF at $\boldsymbol{\sigma} = \mathbf{2}\ \&\ \boldsymbol{\varepsilon} = \mathbf{0.01}$ oBIF at $\boldsymbol{\sigma} = \mathbf{4}\ \&\ \boldsymbol{\varepsilon} = \mathbf{0.01}$	529	C=10, $\boldsymbol{\sigma}$= 40	94.92	98.01	98.01	98.01
		Combined decision *(f1, f2)*		-	C=10, $\boldsymbol{\sigma}$= 60	98.63	99.05	99.05	99.65
KHATT	*f1*	LBP column histogram	LBP at $\mathbf{P} = \mathbf{16}\&\mathbf{R} = \mathbf{2}$ LBP at $\mathbf{P} = \mathbf{8}\ \&\mathbf{R} = \mathbf{2}$	14337	C=10, $\boldsymbol{\sigma}$= 65	75.55	86.20	92.70	94.90
	f2	oBIF column histogram	oBIF at $\boldsymbol{\sigma} = \mathbf{1}\&\boldsymbol{\varepsilon} = \mathbf{0.02}$ oBIF at $\boldsymbol{\sigma} = \mathbf{2}\ \&\boldsymbol{\varepsilon} = \mathbf{0.02}$	529	C=10, $\boldsymbol{\sigma}$= 41	63.40	64.30	76.70	83.00
		Combined decision *(f1, f2)*		-	C=10, $\boldsymbol{\sigma}$= 20	77.10	81.10	88.10	92.90

5.5. *Evaluations on Multi-script protocol*

The Multi-script protocol evaluations is a more challenging protocol because the training and the test samples belong from distinct scripts. These experiments aim to validate that writing patterns that are common across different scripts may be utilized to identify the writer. Writers in multiple scripts share common behaviors and patterns which are persevering across different scripts where our LBP and oBIFs column histograms are very relevant to capture these patterns. The experiments are first performed using the Chinese samples in the training dataset and English samples in the test dataset (Task 1). The task is then reversed by using the English samples as the training dataset and the Chinese writings as the test dataset (Task 2). This is also the same for Arabic and French script (Task 3, Task 4) and the same procedure for English and Farsi (Task 5, Task 6).

It can be observed from Figure 18 that the classification rates are more or less coherent for different configurations of LBP column histogram as well as oBIFs column histogram. The highest reported classification rate is achieved with the SVM classifier by LBP column histogram. It can also be seen that the optimal configuration from the LBP column histogram outperforms other configurations of the oBIF column histogram. Knowing that, the system based on oBIFs column histogram in the ICFHR competition was placed in the second position. in general, it can be seen that the classification rates of Task1, Task2 and Task4 are relatively higher to those realized in of Task3, Task5 and Task6.

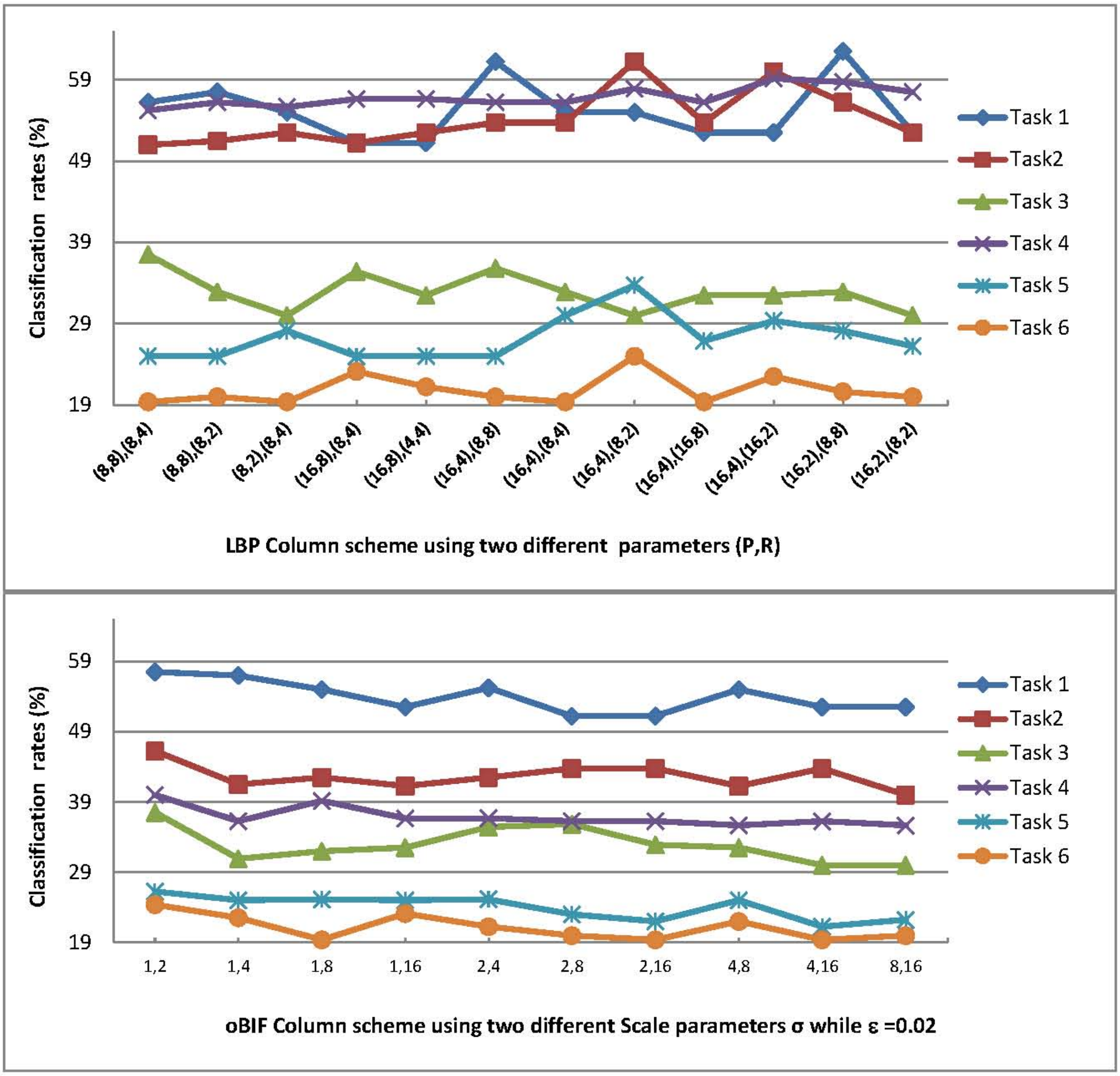


Figure.18. Classification rates on three datasets - ICFHR 2018 Experimental settings

On Task1, Task2, Task3, Task4, Task5 and Task6, the highest reported classification rate using the combined decision is 63.75%, 68.75%, 50.42%, 63.75%, 41.25% and 31.88% respectively. Similarly, the Top1, Top2, Top5 and Top10 using the ICFHR 2018 Experimental settings are summarized in Table 11. In general, it can be observed that, the Top1 using combined decision are relatively high compared to those achieved by other features.

Table 11: Writer classification rates of different tasks- ICFHR 2018 Experimental settings.

Tasks	Column Histogram Features		Parameters	Dim.	SVM Parameters	Classification rates (%)			
						Top 1	Top 2	Top 5	Top 10
Task1	*f1*	LBP	LBP at $\mathbf{P = 16 \& R = 4}$ LBP at $\mathbf{P = 8 \& R = 8}$	14337	C=10, $\boldsymbol{\sigma}$= 77	62.50	71.25	83.75	92.50
	f2	oBIF	oBIF at $\boldsymbol{\sigma} = \mathbf{1 \& \varepsilon = 0.02}$ oBIF at $\boldsymbol{\sigma} = \mathbf{2 \& \varepsilon = 0.02}$	529	C=10, $\boldsymbol{\sigma}$= 40	57.50	67.50	80.00	86.25
		Combined decision (*f1, f2)*		-	C=10, $\boldsymbol{\sigma}$= 60	**63.75**	**75.00**	**86.25**	**95**
Task2	*f1*	LBP	LBP at $\mathbf{P = 16 \& R = 4}$ LBP at $\mathbf{P = 8 \& R = 2}$	14337	C=10, $\boldsymbol{\sigma}$= 85	61.25	70.00	81.25	88.75
	f2	oBIF	oBIF at $\boldsymbol{\sigma} = \mathbf{1 \& \varepsilon = 0.02}$ oBIF at $\boldsymbol{\sigma} = \mathbf{2 \& \varepsilon = 0.02}$	529	C=10, $\boldsymbol{\sigma}$= 41	46.25	55.00	71.25	80.00
		Combined decision (*f1, f2)*		-	C=10, $\boldsymbol{\sigma}$= 20	**68.75**	**78.75**	**87.50**	**91.25**
Task3	*f1*	LBP	LBP at $\mathbf{P = 8 \& R = 8}$ LBP at $\mathbf{P = 8 \& R = 4}$	3481	C=10, $\boldsymbol{\sigma}$= 98	37.50	51.25	72.5	81.67
	f2	oBIF	oBIF at $\boldsymbol{\sigma} = \mathbf{1 \& \varepsilon = 0.02}$ oBIF at $\boldsymbol{\sigma} = \mathbf{2 \& \varepsilon = 0.02}$	529	C=10, $\boldsymbol{\sigma}$= 40	37.50	50.00	68.75	80.83
		Combined decision (*f1, f2)*		-	C=10, $\boldsymbol{\sigma}$= 55	**50.42**	**65.83**	**79.17**	**86.25**
Task4	*f1*	LBP	LBP at $\mathbf{P = 16 \& R = 4}$ LBP at $\mathbf{P = 16 \& R = 2}$	59049	C=10, $\boldsymbol{\sigma}$= 100	63.33	77.92	85.83	94.16
	f2	oBIF	oBIF at $\boldsymbol{\sigma} = \mathbf{1 \& \varepsilon = 0.02}$ oBIF at $\boldsymbol{\sigma} = \mathbf{2 \& \varepsilon = 0.02}$	529	C=10, $\boldsymbol{\sigma}$= 55	40.00	47.50	68.75	79.58
		Combined decision (*f1, f2)*		-	C=10, $\boldsymbol{\sigma}$= 51	**63.75**	**75.42**	**87.5**	**96.67**
Task5	*f1*	LBP	LBP at $\mathbf{P = 16 \& R = 4}$ LBP at $\mathbf{P = 8 \& R = 2}$	14337	C=10, $\boldsymbol{\sigma}$= 59	36.25	45.00	61.25	72.50
	f2	oBIF	oBIF at $\boldsymbol{\sigma} = \mathbf{1 \& \varepsilon = 0.02}$ oBIF at $\boldsymbol{\sigma} = \mathbf{2 \& \varepsilon = 0.02}$	529	C=10, $\boldsymbol{\sigma}$= 40	26.25	36.87	49.37	64.37
		Combined decision (*f1, f2)*		-	C=10, $\boldsymbol{\sigma}$= 45	**41.25**	**47.50**	**66.25**	**76.88**
Task6	*f1*	LBP	LBP at $\mathbf{P = 16 \& R = 4}$ LBP at $\mathbf{P = 8 \& R = 2}$	14337	C=10, $\boldsymbol{\sigma}$= 33	25.00	36.88	53.13	68.75
	f2	oBIF	oBIF at $\boldsymbol{\sigma} = \mathbf{1 \& \varepsilon = 0.02}$ oBIF at $\boldsymbol{\sigma} = \mathbf{2 \& \varepsilon = 0.02}$	529	C=10, $\boldsymbol{\sigma}$= 40	24.37	33.75	47.50	64.37
		Combined decision (*f1, f2)*		-	C=10, $\boldsymbol{\sigma}$= 30	**31.88**	**45.00**	**61.25**	**75.63**

As a long way as we know, no current research has stated the results on multi-script protocol at the ICFHR 2018 datasets which makes a component our comparative study. We compared the proposed approach with the other four (4) other approaches that have been submitted to the competition. Table 12 summarizes the classification rates of different approaches in multi-script based on the experimental setup of the competitions.

Table 12: Comparison of classification rates in Multi-script protocol -ICFHR 2018 Experimental settings.

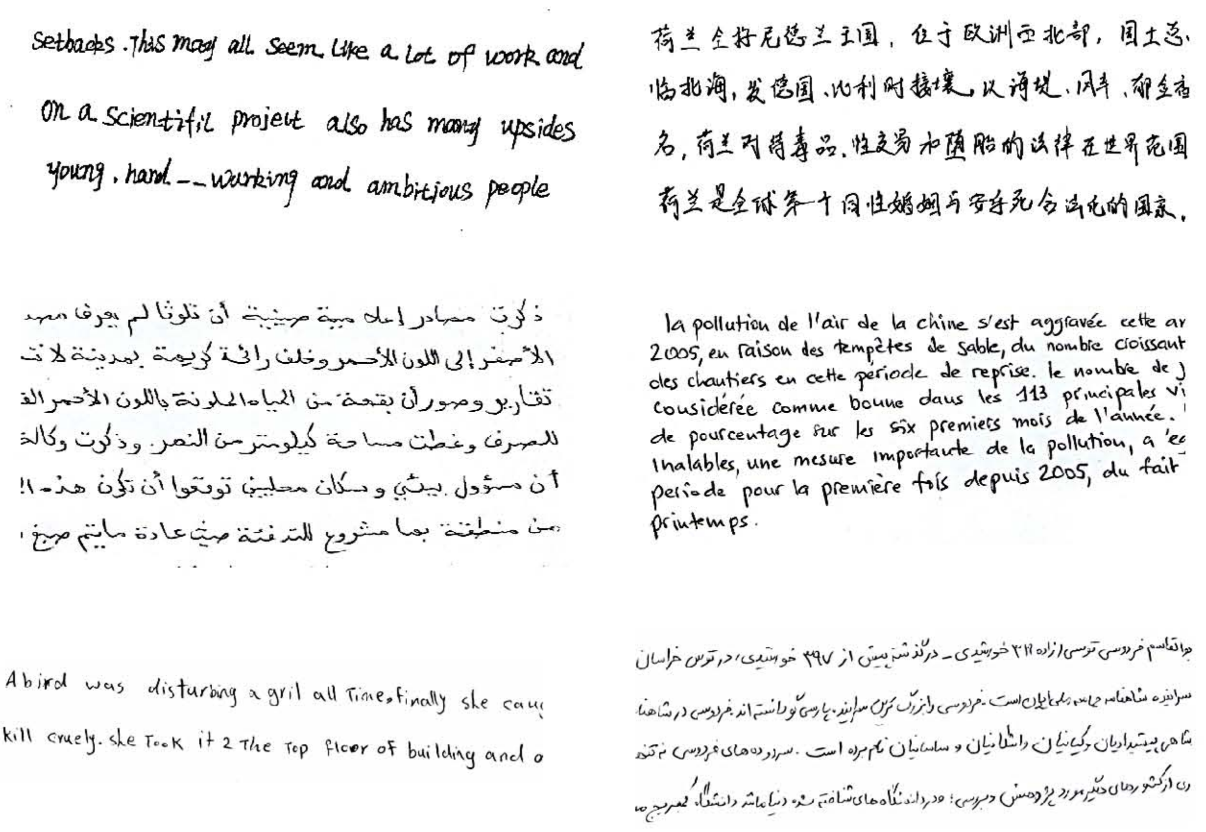
Tasks	Features Parameters	Classifier	Classification rates (%)			
			Top 1	Top 2	Top 5	Top 10
Task1	**Proposed Approach**	SVM	**63.75**	**75.00**	**86.25**	**95.00**
	LIMPAF-1	SVM	42.50	53.75	72.50	83.75
	LIMPAF-2	SVM	57.50	67.50	80.00	86.25
	Tokyo	KNN	23.75	42.50	60.00	68.75
	Nuremberg	Cosine distance	32.50	46.25	66.25	82.50
Task2	**Proposed Approach**	SVM	**68.75**	**78.75**	**87.50**	**91.25**
	LIMPAF-1	SVM	56.25	70.00	81.25	90.00
	LIMPAF-2	SVM	46.25	55.00	71.25	80.00
	Tokyo	KNN	16.25	28.75	46.25	57.50
	Nuremberg	Cosine distance	27.50	40.00	61.25	80.00
Task3	**Proposed Approach**	SVM	**50.42**	**65.83**	**79.17**	**86.25**
	LIMPAF-1	SVM	40.83	52.92	67.92	83.33
	LIMPAF-2	SVM	37.50	50.00	68.75	80.83
	Tokyo	KNN	30.00	40.42	56.67	71.25
	Nuremberg	Cosine distance	19.58	24.17	36.67	55.42
Task4	**Proposed Approach**	SVM	**63.75**	**75.42**	**87.5**	**96.67**
	LIMPAF-1	SVM	42.08	51.67	73.83	85.00
	LIMPAF-2	SVM	40.00	47.50	68.75	79.58
	Tokyo	KNN	17.08	29.17	51.25	60.83
	Nuremberg	Cosine distance	31.25	36.67	46.67	63.75
Task5	**Proposed Approach**	SVM	**41.25**	**47.50**	**66.25**	**76.88**
	LIMPAF-1	SVM	29.37	38.75	58.12	70.62
	LIMPAF-2	SVM	26.25	36.87	49.37	64.37
	Tokyo	KNN	09.37	16.87	31.25	50.62
	Nuremberg	Cosine distance	20.62	28.12	45.00	59.37
Task6	**Proposed Approach**	SVM	**31.88**	**45.00**	**61.25**	**75.63**
	LIMPAF-1	SVM	28.75	38.12	59.37	68.75
	LIMPAF-2	SVM	24.37	33.75	47.50	64.37
	Tokyo	KNN	06.87	17.50	33.12	47.50
	Nuremberg	Cosine distance	17.50	21.87	38.75	51.87

From Table 12, we can see that our approach outperforms the other ones in the Multi-script protocol and achieving the highest classification rates for all tasks across the three datasets. In general, the classification rates for WDAD dataset (Task5 and Task6) are relatively low compared to those of the script-dependent evaluations. The possible reason is that the number of words and the amount of text per page in this dataset is relatively limited compared to the CERUG (Task1 and Task2) and LAMIS-MSHD (Task3 and Task4) datasets. This observation is consistent not only for our proposed approach but for all those mentioned in Table 12.

As previously discussed, multi-script protocol is more challenging as the training dataset and test dataset contain handwritings samples written in different scripts. Because the difficulty of this problem, the classification rates achieved are very promising. Indeed, as illustrated in Figure 18, in some cases, the writings shape to be visually similar making classification a very difficult task. Figure 19 shows some very similar handwritings samples provided by the proposed system.

Another interesting observation to be noted in our approach, is it consistency in the classification rates, whatever is the script: Arabic, English, Chinese, French and Farsi handwritings samples. Therefore, the proposed approach reports a more similar classification rates on different scripts and demonstrates that the LBP column histogram and oBIFs column histogram very well characterizes the writer based on his handwriting.

Throughout this study, we performed various configurations of texture feature column scheme (proposed LBP and oBIFs columns histogram) and a decision strategy that combined them, which allow to extract highly feature and informative elements of the handwriting. We have studied whether our proposed approach is able to find the relationship between the style of different handwriting that were simple and robust against handwriting dissimilarities.



Setbacks . This may all seem like a lot of work and
on a Scientific project also has many upsides
young . hard -- working and ambitious people

荷兰全称尼德兰王国，位于欧洲西北部，国土总、
临北海，与德国、比利时接壤，以海堤、风车、郁金香
名，荷兰对待毒品、性交易和堕胎的法律在世界范围
荷兰是全球第一个同性婚姻与安乐死合法化的国家。

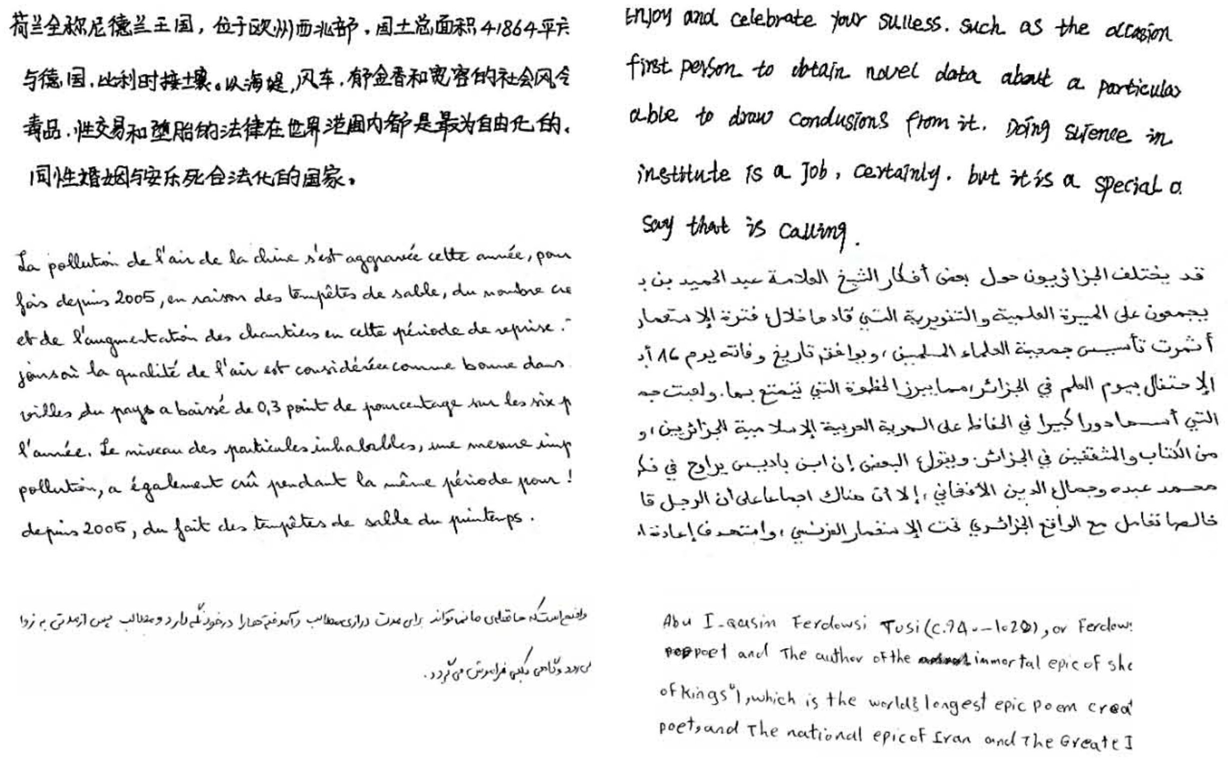
ذكرت مصادر إعلامية صينية أن تلوثا لم يعرف مصد
الأصفر إلى اللون الأحمر وخلف رائحة كريهة بمدينة لانت
تقارير وصور أن بقعة من المياه الملونة باللون الأحمر الق
للصرف وغطت مساحة كيلومتر من النهر. وذكرت وكالة
أن مسؤول بيئي وسكان محليين توقعوا أن تكون هذه ا!
من منطقة بها مشروع التدفئة حيث عادة ما يتم صبغ ،

la pollution de l'air de la chine s'est aggravée cette ar
2005, en raison des tempêtes de sable, du nombre croissant
des chantiers en cette période de reprise. le nombre de j
considérée comme bonne dans les 113 principales vi
de pourcentage sur les six premiers mois de l'année.
inalables, une mesure importante de la pollution, a 'ec
periode pour la première fois depuis 2005, du fait
printemps.

A bird was disturbing a gril all Time, finally she cau
kill cruely. she Took it 2 The Top floor of building and o

ابوالقاسم فردوسی توسی زاده ۳۱۲ خورشیدی ـ درگذشت پیش از ۳۹۷ خورشیدی، در توس خراسان
سراینده شاهنامه حماسه ملی ایران است. فردوسی را بزرگ ترین سرایندهٔ پارسی گو دانسته اند. فردوسی در شاهنا
شاهی پیشدادیان، کیانیان، اشکانیان و ساسانیان نام برده است. سروده های فردوسی ترجمه
ری از کشورهای دیگر مورد پژوهش و بررسی؛ در دانشگاه های شناخته شده دنیا مانند دانشگاه کمبریج و

Figure 19. Visually similar produced by the proposed method (misclassification) on three datasets.



荷兰全称尼德兰王国，位于欧洲西北部，国土总面积41864平方
与德国，比利时接壤。以海堤，风车，郁金香和宽容的社会风气
毒品，性交易和堕胎的法律在世界范围内都是最为自由化的。
同性婚姻与安乐死合法化的国家。

Enjoy and celebrate your sulless. such as the occasion
first person to obtain novel data about a particular
able to draw condusions from it. Doing sience in
institute is a Job, certainly. but it is a special o
say that is calling.

La pollution de l'air de la chine s'est aggravée cette année, pour
fois depuis 2005, en raison des tempêtes de sable, du nombre cr
et de l'augmentation des chantiers en cette période de reprise.
jours où la qualité de l'air est considérée comme bonne dans
villes du pays a baissé de 0,3 point de pourcentage sur les six p
l'année. Le niveau des particules inhalables, une mesure imp
pollution, a également crû pendant la même période pour !
depuis 2005, du fait des tempêtes de sable du printemps.

قد يختلف الجزائريون حول بعض أفكار الشيخ العلامة عبد الحميد بن ب
يجمعون على المسيرة العلمية والتنويرية التي قادها خلال فترة الاستعمار
أثمرت تأسيس جمعية العلماء المسلمين، وبلغت تاريخ وفاته يوم ١٦ أب
الاحتفال بيوم العلم في الجزائر ما يبرز الحظوة التي يتمتع بها. ولعبت جم
التي أسسها دورا كبيرا في الحفاظ على الهوية العربية الإسلامية الجزائريين، و
من الكتاب والمثقفين في الجزائر. ويقول البعض إن ابن باديس يراوح في فكر
محمد عبده وجمال الدين الأفغاني، إلا أن هناك اجماعا على أن الرجل قا
خالصا تفاعل مع الواقع الجزائري تحت الاستعمار الفرنسي، واستهدف إعادة ا

واضح است که ما قبلی ما نمی‌تواند برای مدت درازی مطالب را آموخته‌ها را در خود نگه دارد و مطالب پس از مدتی به زودا
می‌رود و گاهی بکلی فراموش می‌گردد.

Abu I-qasim Ferdowsi Tusi (c.94--1020), or Ferdow
~~pop~~poet and The author of the ~~[illegible]~~ immortal epic of sh
of Kings"1, which is the world's longest epic poem creat
poets, and The national epic of Iran and The Greate I

Figure 20. Examples of the correct classification produced by the proposed approach on three datasets.

6. Conclusion

This chapter explored the efficiency of texture feature column scheme to characterize writer based on his handwriting. Various configurations of the proposed LBP columns histogram and the oBIFs columns histogram are explored using an SVM classifier. The evaluations are performed using experiments with two publicly available datasets through a single-script protocol, and using three other datasets, according to the experimental settings of the international Competition ICFHR for Multi-script protocol. The results are compared with state-of-the-art existing approaches studies in the literature. The approach we proposed outperforms the existing methods in multi-script protocol. An important note that emerges from this type of research is to understand the relation between different script handwriting and characterizing writers from handwriting images. The performance of the proposed approach can improve its acceptance by the forensic experts. Such tools can be effectively employed to narrowing down the search in large repositories. Developing the expert system for this problem ensures that the knowledge provided by different writer classification systems is systematically and reliably applied. An important feature of the expert system is useful for both computer scientists and forensic experts. In our further research on this topic, we think of explore and compare the performance of proposed approach in characterizing other demographic attributes of writers including age and handedness. In addition, a study of other textural measures to characterize gender from handwriting and exploration of feature selection techniques to classify the most appropriate textural descriptors for this topic is also planned.

CHAPTER 04

OFFLINE WRITER IDENTIFICATION FOR HISTORICAL DOCUMENT

1 . Introduction

Automatic writer identification is one of common problem in document analysis, writer retrieval, Historical Document Writer Identification and many more. State-of-the-art methods is usually focused on the feature extraction processes using traditional or deep-learning-based techniques. However, the majority of research has been done on modern datasets. Generally, these documents do not contain noise or artifacts due to acquisition conditions or the quality of original document.

Many researchers were attracted towards handwritten writer identification and retrieval which is the most promising field of image analysis and machine learning. Due to progressions in technology pattern recognition and image processing have seen many improvements in Writer Identification of offline handwriting images. The latest researches on this problem are in [55],[57],[87],[88],[89],[8]. These systems aim to capture the visual differences in the individual's handwriting, including slope of lines and slant of characters, line spacing, inter and intra-word spacing, cursivity, allograph variations, legibility, readability, tilt characters and line tilt, etc.

This chapter explores how current state-of-the-art approaches in writer identification achieve on historical documents. In contrast to modern handwritten documents, historical dataset frequently contains artifacts include ripping, holes, tearing, folding, water stains that make reliable identification prone to errors. Many competitions intended to identify the handwriting have been held at well-known conferences such as the International Conference on Frontiers of Document Recognition handwriting (ICFHR) and the International Conference on Document Analysis and Recognition (ICDAR). therefore, these competitions were used to the evaluation of various methods [22],[40],[41],[69],[90],[91], on modern handwritten documents datasets.

Recently, both competitions on Historical handwritten Document Writer Identification have been organized in conjunction with ICDAR 2017 and ICDAR 2019 [92] [93]. In the ICDAR 2017 competition [92], Tébessa II system [92] ranked first place used oriented Basic Image Features (oBIFs) columns histograms [94],[95] to characterize writer from handwritings image. However, in the ICDAR 2019 competition [92], The winner system combined the SIFT features from the foreground and the pathlet feature. In addition, Euclidean distance of the normalized vectors was calculated. A detailed comparison and analysis of the results of these competitions and the results obtained by the proposed method is shown in Section 3.

In order to discover the best combination of different handcrafted features algorithms used in modern writer identification document, textural feature has been a finest solution of researchers for writer identification. Abbas et al. [95], for instance, exploit a novel descriptor by crossing the Local Binary Patterns (LBP) with different configurations that allows capturing the local textural information in handwriting image using a column histogram. In another study, Abbas et al. [96] proposed to introduce Complete Local Binary Patterns (CLBP) and Local Binary Pattern Variance (LBPV) for extracting the features from handwriting documents. These features are normalized using Probability Density Function (PDF). Hence, The Hinge feature [64],[98] is calculated different angle combinations evaluated along the handwritten contours of the ink traces in another similar work, The curvature-free COLD feature [64],[99], which is the joint distribution of the orientation and length of line fragments. Edge-hinge, run-length and likewise features [100] are combined to identify writers through samples in a multi-script writer identification document.

The goal of the current study is to prove how to achieve high accuracy rates on historical Latin manuscripts using a combination of different handcrafted features algorithms. therefore, we have developed a system based on handcrafted features algorithms to identify writers of historical handwritten manuscripts using a new distance based on the moment. Our main concern was to ensure that the system evaluates its performance against public dataset. Meanwhile, we evaluated the system using the leading and well-known competition ICDAR 2017 dataset. The proposed system uses a leave-one-out strategy for ranking according to the similarity between two handwritings. An overview of the proposed method is illustrated in Figure 21.

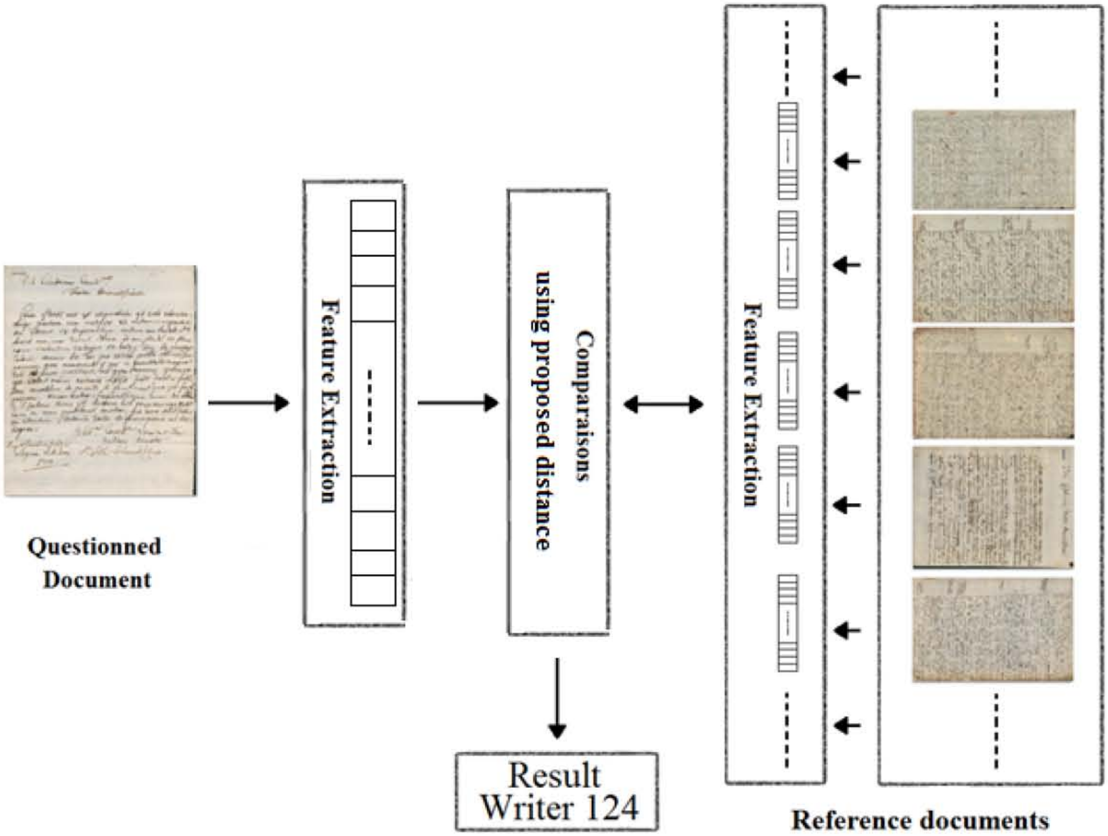


Figure 21.Overview of the Proposed System.

This chapter is organized as follows. We discuss in the next section the pertinent literature on this problem. We present the texture features employed in our study using the proposed moment distance in Section III and Section IV. Section V shows the experimental results, classification scheme and the analysis of the results. Finally, the last section concludes with some useful insights for further research on this problem.

2 . Related work

Offline historical document writer identification and retrieval methods can be grouped into codebook rely on deep learning methods (Deep Codebook) and handcrafted features methods. In codebook rely on deep learning methods, the codebooks based on Deep learning is used to compute statistics form the global descriptor. Conversely, the methods of the handcrafted features compute a global image descriptor directly from the handwriting.

Current methods based on codebook rely on deep learning and the most methods use Convolutional Neural Networks (CNN) to generate strong local descriptors by using the activation features [101],[102]. In the work of Christlein et al. [101] relies on both, traditional SIFT descriptors and the ResNet. The activations from the penultimate CNN layer serve as features that are afterwards encoded and classified.

In Jordan et al. [102], they used same features of [101] and experimented with two different re-ranking methods. First, the k-reciprocal Nearest Neighbours is integrated into the Jaccard distance. next, a proposed query expansion (QE) approach which extend the original ESVM and use Reciprocal Nearest Neighbours (rNN) to overcome the lack of spatial verification when using the features. As a result, their techniques has crossed the baseline of the ICDAR 2017 dataset.

In another study, Chammas et al. [103] proposed to use SIFT without any a pre-processing step where an unsupervised deep Convolutional Neural Network (CNN) trained with the extracted patches from SIFT descriptors. Then the results were encoded using multi-VLAD and used l2-norm to normalize the data. Finally, Exemplar Support Vector Machine (E-SVM) is used to calculate the distance between the VLAD vectors.

Another study of Lai et al. [104] proposed to encode SIFT features and pathlet for historical writer identification using bagged VLAD.

First, deep binarization based on U-Net and page-level rotation correction are used to remove the complex backgrounds and noise. To describe the handwriting styles, pathlet features and unidirectional SIFT features are extracted to capture structural information (junctions and corners) and rich shape (curvature and slant). To efficiently encode the proposed features, the encoding method based on bagged VLAD enhanced encoding performance using a much larger codebook. bVLAD feature vectors are whitened, normalized and classified using the cosine similarity.

Among the contributions to identification of writers from handwriting based on textural features, Gattal et al. [102] proposed to use textural measures by combining of oriented Basic Image Features (oBIFs) column Features at different scales extracted from whole binarized images of Historical handwritten document. Experimental studies on ICDAR 2017 datasets achieved high identification rates compared to the latest methods on this subject. Table1 summarizes the well-known contributions to historical document writer identification reported in the literature.

Table 1 Performance comparison of well-known historical writer identification systems

Study			Database	Top-1	mAP
	Methods	Features			
Handcrafted feature	Winner method in ICDAR 2017 competition [13]	oBIFs columns	Icdar17	76.40	55.60
	Gattal et al. (2018)	oBIFs columns	Icdar17	77.39	56.82
Deep codebook	Jordan et al. (2020)	SIFT descriptors and VLAD-encoded with CNN activation features	Icdar17	89.43	78.20
			CzByChron	98.04	80.10
			MusicDocs	98.62	78.64
	Christlein et al. (2017) [24]	SIFT descriptors and VLAD-encoded with CNN activation features	Icdar17	88.9	76.2
			CLaMM16	84.1	/
	Winner method in ICDAR 2019 competition [14]	SIFT features and the pathlet feature	Icdar19	97.4	92.5
	Chammas et al. (2020) [15]	SIFT descriptors, CNN and VLAD-encoded.	Icdar19	97.0	91.2
	Lai et al. (2020) [23]	SIFT features and the pathlet feature	Icdar17	90.1	77.2
			Icdar19	97.4	92.5

An analysis of the historical writer identification methods based on Handcrafted feature methods mentioned in the literature reveals the textural features continue to be the finest researchers choice. Then, we are investigating mainly the contributions based on textural features which characterizing a writer only on the texture. In the present study, we investigate a new method based on Hinge and the oBIF Column histograms using prposed New Moment Distance to identify writers independently from the script used.

3 . Feature extraction

Feature extraction is an significant step in image classification. It allows to extract the relevant features that represent the content of an images to form a feature vector. These features ara used by classifiers that to improve the accuracy rate of classification. In our study on Historical handwritten document, we will use the advantage of their strong discriminatory representation captured through the curvature information, textural information and contour information in Historical handwritten document. The sub-sections that follow a more detailed description of these features.

3.1 oBIF column histogram scheme

The descriptor of oriented Basic Image Features (oBIFs) is an extension of the Basic Image Features (BIFs) [44]. Each position in the image is categorized into one of seven local symmetry classes depending on the type of local symmetry, these classes could be sloping, light spinning, flat, dark rotational, light line on dark, dark line on light or saddle-like.

The feature is based on a bank of six (06) Derivative-of-Gaussian filters up to second order. These filters are used with various scale parameters σ to form a multiscale filter bank for capturing local symmetry classes. The parameter ε indicates whether the position is classified as flat.

The feature vector of twenty-three locations is attributed by twelve orientations, eight slopes and three no orientations. In order to improve the performance of the oBIF descriptor, the combination of oBIFs at two different scales yield the oBIF column features by ignoring all the symmetry type flat [102]. whoever, the histogram is fixed at 484 in this case. The oBIFs column features are generated using both values of the scale parameter $\sigma\in[\{2,8\}, \{2,4\}]$ with the parameter ε is set to a small value of ε=0.01 from the Historical handwritten document.

3.2 Hinge feature

The best contour-based features reported in the literature are the hinge [29] and Delta-n Hinge [101] which are developed to capture the ink-trace curvature of document images, which is very discriminatory between handwriting.

The Hinge feature [29] is calculated the joint *probability distribution* of *orientations* of the two legs of the obtained "contour/edge-*hinge*". Two parameters are used in the Hinge feature: the number of angle bins p and the leg length r. Therefore, it has been extended to the Delta-n Hinge [105] to reach the rotation-invariant property. There are four parameters, leg length r, the number of angle bins p, Manhattan distance Δl and the number of derivative *n*. In this case, we set n = 2 , p=40 as suggested in [30] and the Manhattan distance Δl= 7.

The generated feature vectors are standardized to have zero mean and unit variance. To map the normalized feature vector to the interval [0 1], the ensuing function is employed.

$$V(x) = \frac{1}{1+e^{\frac{-\pi}{2}(x)}} \qquad (1)$$

Where V(x) represents the new normalized feature vector of the feature vector x.

The two sets of features Hinge and oBIFs column histogram scheme are combined only at the decision level by submitting each set of features to a separate distance, as detailed in the following section.

4 . New moment matching method

Distance metrics are the key to getting similar writer of historical documents from different writers. Numerous distance metrics for measuring similarity between feature vectors have been proposed in the literature [106] such as the correlation distance, Euclidean distance, city block distance, Spearman distance and cosine distance. a new matching algorithm, called the New Moment Distance (NMD) algorithm is presented.

Moment has a wide range of application in image analysis, including object classification, pattern recognition, reconstruction and image coding [106]. If the feature vector is considered a discrete function of $f(x)$ with x = *0,1, ..., N* , the moment of order (k) is defined as:

$$\mu_k = \mathbb{E}[X^k] = \int_{-\infty}^{\infty} x^k \, f(x) dx, \qquad (2)$$

We understood that the first moment is a random variable’s mean, the second (central moment) is its variance, and so forth. However, we need the moments to identify the shape of the distribution known as skewness and kurtosis. The moments about mean are the deviations mean from the mean after raising them to integer powers. The kth population moment about mean is symbolized by μ_k is

$$\mu_k = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^k}{N} \qquad (3)$$

Note that the first moment is always zero if *k=1* as

$$\mu_1 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^1}{N} = 0 \qquad (3)$$

If *k=2* then the second moment is variance as

$$\mu_2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N} \qquad (4)$$

Similarly, the 3rd is used to define the skewness of the distribution and 4th moments is used to define the kurtosis of the distribution are

$$\mu_3 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^3}{N} \quad (5)$$

$$\mu_4 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^4}{N} \quad (6)$$

During the feature extraction step, the features is extracted for each image, and then the query image feature ($QueryFV$) used to compare with reference handwriting features ($ReleventFV$). In our case, we proposed a new moment matching method based on Moments about mean using Arbitrary Value. the kth sample moment about arbitrary source "a" denoted by m'_k is

$$m'_k = \frac{\sum_{i=1}^{N}(y_i - a\bar{y})^k}{N} \quad (7)$$

Where $y_i = QueryFV_i - ReleventFV_i$ and $\bar{y}$ is the Mean. Therefore,

$$a = \sqrt{|y_i|} \quad (8)$$

Using the eq.7 and eq.8, the kth *proposed moment* is typically defined as

$$m'_k = \frac{\sum_{i=1}^{N}\left(y_i - \bar{y}\sqrt{|y_i|}\right)^k}{N} \quad (9)$$

Therefore

$$m'_2 = \frac{\sum_{i=1}^{N}\left(y_i - \bar{y}\sqrt{|y_i|}\right)^2}{N}, \; m'_4 = \frac{\sum_{i=1}^{N}\left(y_i - \bar{y}\sqrt{|y_i|}\right)^4}{N}, m'_6 = \frac{\sum_{i=1}^{N}\left(y_i - \bar{y}\sqrt{|y_i|}\right)^6}{N}$$

We compute the new moment matching method from the following relations by calculating Moments about mean using Arbitrary Value

$$Dist = m'_2 + m'_4 + m'_6 \quad (10)$$

In the comparison step, the query image feature is compared to the feature vector of the reference image of documents, from which the final result of matching score reports the minimum distance is selected that are nearby matches to a query handwritten document. For the challenge of improving the reliability of the proposed system's accuracy rate, the decision module is designed to provide the final decision. This module is proposed with two different New Moment Distance. Each distance is calculated by a set of features: Hinge or oBIF column histogram scheme. The final decision is used to include the product (Prod), minimum (Min), sum (Sum) and average (Avr) are commonly applied to attain at final decision taking into account two different New Moment Distance. In our study, we adopted the minimum of product (Prod) distance from the best combination. The next section presents the experimental settings and the corresponding results.

5 . Experimental Results

The efficiency of the employed the proposed New Moment Distance using texture features in characterizing the writer of Historical Document is verified by a series of experiments. A series of experiments will be conducted to evaluate the effectiveness of the proposed system for competition Historical Document datasets. The test set contained 3600 historical document samples of 720 writers as each writer has 5 documents. The performance measurement used is the precision (Top1, Top5 and Top10) and the mean Average Precision (mAP). The first experiment aims to find the best features of different handcrafted features algorithms used in modern writer identification document. Hence, the correlation distance used to compute the distance between the handcrafted features vectors. The realized Top1, Top5, Top10 and mAP are showed in the table 2.

Table 2 Performance of well-known handcrafted features using correlation distance

Features	Parameters	Dim.	Average Identification rates (%)			
			Top 1	Top 5	Top 10	Map
f1: oBIF column histogram [15]	oBIF at $\boldsymbol{\sigma} = \{2\ ,4\}\&\ \boldsymbol{\varepsilon} = 0.1$	484	74.56	83.33	86.36	53.26
f2: oBIF column histogram [15]	oBIF at $\boldsymbol{\sigma} = \{2\ ,8\}\&\ \boldsymbol{\varepsilon} = 0.1$	484	72.69	80.92	84.14	52.34
f3: oBIF column histogram [15]	oBIF at $\boldsymbol{\sigma} = \{2\ ,4\}\&\ \boldsymbol{\varepsilon} = 0.1$ oBIF at $\boldsymbol{\sigma} = \{2\ ,8\}\&\ \boldsymbol{\varepsilon} = 0.1$	968	**76.17**	**84.25**	**86.94**	**55.27**
LBP column histogram [18]	LBP at $\mathbf{P} = 16\&\mathbf{R} = 2$ LBP at $\mathbf{P} = 8\ \&\mathbf{R} = 2$	14337	37.81	43.22	45.83	21.67
LBP	$\mathbf{P} = 12\ \&\ \mathbf{R} = 4$	529	49.31	55.72	58.14	29.40
CLBP [19]	$\mathbf{P} = 12\ \&\ \mathbf{R} = 4$	270	54.00	60.61	63.42	33.43
LBPV [19]	$\mathbf{P} = 12\ \&\ \mathbf{R} = 4$	529	56.06	63.75	65.89	34.00
Run-length distribution on white pixels [23]	/	200	39.92	47.56	51.17	25.18
Run-length distribution on black pixels [23]	/	400	27.14	32.67	35.03	15.96
Run-length distribution on white and black pixels [23]	/	600	42.11	49.36	52.39	26.08
Edge-direction distribution using 16 angles [23]	/	16	29.03	35.53	39.08	17.72
Edge-hinge with fragment of length equal to 7 pixels [23]	/	2304	61.11	68.64	72.03	40.40
Delta Hinge feature [20]	$\mathbf{r = 5\ , p = 40}$	5184	71.14	79.69	83.33	50.50
COLD feature [21]	$\mathbf{k = \{3,4,5,6,7\}}$	84*5	63.44	72.36	75.42	41.67

It can be noticed that the oBIFs column histogram at σ= {2,4} and σ= {2,8} while ε=0.1 outperforms the other features. Therefore, the Hinge feature with r = 10 while p = 40 does better than others remaining features. These features are achieved in the ICDAR2017 competition test dataset using correlation distance measures according to the experimental protocole of the competition. In addition to the correlation distance, we also evaluated the optimal features using the proposed New Moment Distance and with various distance metrics such as the correlation distance, Spearman distance, cosine distance and city block distance. Table 3 reports the optimal features using proposed distance and with different distance metrics.

Table 3. Performance of optimal features using different distance metrics

Features	Distance metrics		Average Identification rates (%)			
			Top1	Top5	Top10	mAP
f1: oBIF column histogram [15]	**NMD**	**New Moment Distance**	**74.61**	**82.83**	**86.08**	**53.44**
	Eu	Euclidean Distance	73.53	81.81	85.14	52.33
	CB	City-Block Distance	74.33	82.33	85.00	52.46
	Corr	Correlation Distance	74.56	83.33	86.36	53.26
	Cos	Cosine Distance	74.17	82.78	85.42	53.18
	Sp	Spearman Distance	73.89	82.72	86.19	52.87
f2: oBIF column histogram [15]	**NMD**	**New Moment Distance**	**74.17**	**83.56**	**86.58**	**53.20**
	Eu	Euclidean Distance	72.83	81.42	84.36	51.30
	CB	City-Block Distance	72.64	82.03	84.72	51.08
	Corr	Correlation Distance	72.69	80.92	84.14	52.34
	Cos	Cosine Distance	73.22	82.67	86.03	52.67
	Sp	Spearman Distance	72.19	80.25	83.17	51.68
f3: oBIF column histogram [15]	**NMD**	**New Moment Distance**	**77.36**	**86.42**	**89.39**	**55.88**
	Eu	Euclidean Distance	76.03	84.17	87.11	54.39
	CB	City-Block Distance	75.72	84.53	86.97	54.32
	Corr	Correlation Distance	76.17	84.25	86.94	55.27
	Cos	Cosine Distance	75.97	84.39	87.28	55.14
	Sp	Spearman Distance	75.42	83.33	86.08	54.72
f4: Delta Hinge feature [20]	**NMD**	**New Moment Distance**	**75.00**	**84.00**	**86.47**	**54.06**
	Eu	Euclidean Distance	72.11	80.72	83.69	50.51
	CB	City-Block Distance	72.94	82.19	85.11	51.30
	Corr	Correlation Distance	71.14	79.69	83.33	50.50
	Cos	Cosine Distance	72.25	80.33	83.47	51.08
	Sp	Spearman Distance	69.94	78.19	81.22	50.49

It can be noticed by using the proposed distance, the Top1 as high as 77.36% and 75.00% are realized with oBIF column histogram (f3) and Delta Hinge feature (f4) respectively. following the previous results, we evaluated the proposed method using a decision module with the both optimal features. We also calculated the precision Top1, Top5, Top10, mAP for combination of proposed distance from the oBIF column histogram and Delta Hinge feature to increase the classification rate. The robustness of the proposed method using decision combination was studies including minimum of the product distances from features (Prod-Min), minimum of the sum (Sum-Min), minimum of the average (Avg-Min), and minimum of the minimum (Min-Min) of the corresponding the proposed distance. The effects of these combination decisions are summarized in Table 4.

Table 4. performance rate for various decision combination.

Combination decision	Average Identification rates (%)			
	Top1	**Top5**	**Top10**	**mAP**
Sum-Min(*f1*, *f2*)	77.25	86.08	88.39	56.07
Sum-Min(*f3*, *f4*)	**78.72**	**88.33**	**91.00**	**58.45**
Sum-Min(*f1*, *f2*, *f4*)	78.69	87.06	89.89	58.37
Sum-Min(*f1*, *f2*, *f3*, *f4*)	78.42	86.75	89.67	57.99
Prod-Min(*f1*, *f2*)	77.22	86.58	89.28	56.11
Prod-Min(*f3*, *f4*)	**78.75**	**88.31**	**91.08**	**58.62**
Prod-Min(*f1*, *f2*, *f4*)	78.64	86.86	89.92	58.48
Prod-Min(*f1*, *f2*, *f3*, *f4*)	78.53	86.97	90.08	58.03
Avg-Min(*f1*, *f2*)	77.25	86.08	88.39	56.07
Avg-Min(*f3*, *f4*)	**78.72**	**88.33**	**91.00**	**58.45**
Avg-Min(*f1*, *f2*, *f4*)	78.69	87.06	89.89	58.37
Avg-Min(*f1*, *f2*, *f3*, *f4*)	78.42	86.75	89.67	57.99
Min-Min(*f1*, *f2*)	75.47	84.14	87.50	54.70
Min-Min(*f3*, *f4*)	**77.94**	**87.25**	**89.53**	**57.29**
Min-Min(*f1*, *f2*, *f4*)	76.61	85.86	89.33	56.08
Min-Min(*f1*, *f2*, *f3*, *f4*)	76.61	85.86	89.33	56.08

Generally, the classification rates of combination scheme based on the minimum of the product (Prod-Min) corresponding proposed distance are comparatively high as compared to others combinations likewise realizing the precision Top1 and mAP of 78.75 % and 58.62 % respectively. The high-performance rates are an indicative of the efficiency of the proposed system. Though, it should be noted that through the curvature information, contour information and local textural information can validate the efficacy of the proposed method.

We also compare the performance of the proposed system with handcrafted feature methods of state-of-the-art according to the competition protocol ICDAR 2017 in Historical handwritten Document. As summarized in Table 5, The evaluation protocol measured in our experiments is the same as the competition evaluation protocol to allow a significant.

Table 5. Comparison of proposed method with handcrafted feature methods of state-of-the-art.

Rank	Method	Average Identification rates (%)	
		Top1	mAP
1	**Proposed Method**	**78.75**	**58.62**
2	Gattal et al. (2018)	77.39	56.82
3	Winner method in ICDAR 2017 competition [13]	76.40	55.60

It turns out that the proposed methodology achieves higher Identification rates in Historical handwritten Document evaluations than other methods based on handcrafted feature in the literature. However, it should be noted that the proposed system does not need any preprocessing and the features are extracted directly from the binarized Historical Document images. These results confirm the effectiveness of the proposed New Moment Distance using Delta Hinge and oBIF column histogram Features for identify the writer of Historical Documents.

Considering the difficulty of the Historical Document Writer Identification problem, the Top1 and mAP as 78.75% and 58.62 % are indeed very promising, also, in some cases, the handwritings writer found has a homogeneous vision comparing another writer and vice versa making classification of writer a very challenging task.

6 . Conclusions

This chapter presented an effective technique for characterizing writer from historical handwriting by exploiting oBIFs columns histograms and Delta Hinge Features using the proposed New Moment Distance. Different handcrafted features are examined with various decision combination. The system performed and evaluated on the same experimental protocol as that of the Historical Document Writer Identification overcome the existing methods based on handcrafted features.

CONCLUSION

Identifying the writer from handwritings image is an interesting research topic in the field of handwriting analysis and recognition. In particular, the writer Identification and retrieval in handwriting can be used by intelligent systems and expert who of forensic survey. The intelligent systems involve the design and development of an expert system to help forensic experts in the diagnosis of handwriting document and the retrieval of pages which have been written by the same writer, index archives and in financial issues to know Who is the authorized person to perform the operation.

Identifying individuals based on their handwriting is still a defying task due the variability between writers, where the same writer may write differently the same text. To overcome these difficulties, we proposed in this thesis to use the textural features that characterizes the handwritten document along with their writing language and document status , to go deeply the search we exploit the icfhr2018 dataset of multi-script by employs combinations of different configurations of both Texture Features such as Oriented Basic Image Features (oBIFs) column scheme and Local Binary Patterns (LBP) column scheme to identify writer from handwriting , A series of evaluations trained with a support vector machine classifier (SVM) using different combinations and configurations realized high recognition rates which are compared with the state-of-the-art methods on this topic.

In pursuit of further evaluation of Texture Features that characterize writer from historical handwriting, it is presented that exploits Delta Hinge and oBIFs columns histograms Features using the proposed New Moment Distance. a set of evaluations using different configurations and combinations employed to improve the performance of the systems for Identifying writer of Historical Documents which realized high recognition rates compared with to the latest methods on this subject.

Generally, the proposed methods allows managing all the datasets of handwriting single /multi script and historical / modern datasets, so the first results obtained are encouraging since This combination uses few rules and has the advantage of providing the correct presentation of handwriting without using the contextual knowledge.

REFERENCES

[1]. Bulacu, M., & Schomaker, L. (2007). Text-independent writer identification and verification using textural and allographic features. IEEE transactions on pattern analysis and machine intelligence, 29(4), 701-717.

[2]. Plamondon, R., & Lorette, G. (1989). Automatic signature verification and writer identification—the state of the art. Pattern recognition, 22(2), 107-131.

[3]. Bar-Yosef, I., Beckman, I., Kedem, K., & Dinstein, I. (2007). Binarization, character extraction, and writer identification of historical Hebrew calligraphy documents. *International Journal of Document Analysis and Recognition (IJDAR)*, *9*(2), 89-99..

[4]. Siddiqi, I., & Vincent, N. (2007, September). Writer identification in handwritten documents. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) (Vol. 1, pp. 108-112). IEEE.

[5]. Pervouchine, V., & Leedham, G. (2007). Extraction and analysis of forensic document examiner features used for writer identification. *Pattern Recognition*, *40*(3), 1004-1013..

[6]. Bensefia, A., Paquet, T., & Heutte, L. (2005). A writer identification and verification system. Pattern Recognition Letters, 26(13), 2080-2092.

[7]. He, Z. Y., & Tang, Y. Y. (2004, August). Chinese handwriting-based writer identification by texture analysis. In Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826) (Vol. 6, pp. 3488-3491). IEEE.

[8]. Louloudis, G., Gatos, B., Stamatopoulos, N., & Papandreou, A. (2013, August). ICDAR 2013 competition on writer identification. In 2013 12th International Conference on Document Analysis and Recognition (pp. 1397-1401). IEEE.

[9]. Djeddi, C., Gattal, A., Souici-Meslati, L., Siddiqi, I., Chibani, Y., & El Abed, H. (2014, September). LAMIS-MSHD: a multi-script offline handwriting database. In 2014 14th International Conference on Frontiers in Handwriting Recognition (pp. 93-97). IEEE.

[10]. He, S., Wiering, M., & Schomaker, L. (2015). Junction detection in handwritten documents and its application to writer identification. Pattern Recognition, 48(12), 4036-4048.

[11]. Masomi, A., Ghafari, H. R., Nouri, K., Akbari, Y., Bouamra, W., & Djeddi, C. (2016, November). A new database for writer demographics attributes detection based on off-line persian and english handwriting. In Proceedings of the Mediterranean Conference on Pattern Recognition and Artificial Intelligence (pp. 125-130).

[12]. Fiel, S., Kleber, F., Diem, M., Christlein, V., Louloudis, G., Nikos, S., & Gatos, B. (2017, November). Icdar2017 competition on historical document writer identification (historical-wi). In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 1377-1382). IEEE.

[13]. Freitas, C., Oliveira, L. S., Sabourin, R., & Bortolozzi, F. (2008). Brazilian forensic letter database. In 11th International workshop on frontiers on handwriting recognition, Montreal, Canada.

[14]. Mahmoud, S. A., Ahmad, I., Alshayeb, M., Al-Khatib, W. G., Parvez, M. T., Fink, G. A., ... & El Abed, H. (2012, September). Khatt: Arabic offline handwritten text database. In *2012 International conference on frontiers in handwriting recognition* (pp. 449-454). IEEE.

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

[16]. Saba, T., Rehman, A., Al-Dhelaan, A., & Al-Rodhaan, M. (2014). Evaluation of current documents image denoising techniques: a comparative study. *Applied Artificial Intelligence*, *28*(9), 879-887.

[17]. Otsu, N. (1979). A threshold selection method form gray-level histograms. Proceedings of the 1986 IEEE Transactions Systems, 9(1), 62-66

[18]. Bernsen, J. (1986). Dynamic thresholding of gray-level images, Proceedings 8th International Conference on Pattern Recognition, Paris, 1251-1255.

[19]. Sauvola, J., & Pietikainen, M. (2000). Adaptive document image binarization. Pattern Recognition, 33(2), 225-236.

[20]. Rukhsar Firdousi, Shaheen Parveen," Local Thresholding Techniques in Image Binarization",International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 3 Issue 3 March, 2014 Page No. 4062-4065.

[21]. Fiel, S., Sablatnig, R., 2015. Writer Identification and Retrieval Using a Convolutional Neural Network. In: International Conference on Computer Analysis of Images and Patterns, Valetta, 2-4 September, pp. 26–37.

[22]. Fiel, S., Kleber, F., Diem, M., Christlein, V., Louloudis, G., Niko, S., Gatos, B., 2017. ICDAR 2017 Competition on Historical Document Writer Identification (Historical-WI). In: 14th IAPR International Conference on Document Analysis and Recognition, Kyoto, 13-15 November, pp. 1377–1382.

[23]. Liwicki, M., Schlapbach, A., Bunke, H., 2011. Automatic gender detection using on-line and off-line information. Pattern Analysis and Applications 14 (1), 87–92.

[24]. Bouadjenek, N., Nemmour, H., Chibani, Y., 2016. Robust soft-biometrics prediction from off-line handwriting analysis. Applied Soft Computing 46, 980–990.

[25]. Jain, R., Doermann, D., 2014. Combining local features for offline writer identification. In: 14th International Conference on Frontiers in Handwriting Recognition, Crete, 1-4 September, pp. 583–588.

[26]. Wu, X., Tang, Y., & Bu, W. (2014). Offline text-independent writer identification based on scale invariant feature transform. IEEE Transactions on Information Forensics and Security, 9(3), 526-536.

[27]. Christlein, V., Bernecker, D., Hönig, F., & Angelopoulou, E. (2014, March). Writer identification and verification using GMM supervectors. In IEEE Winter Conference on Applications of Computer Vision (pp. 998-1005). IEEE.

[28]. Bertolini, D., Oliveira, L. S., Justino, E., & Sabourin, R. (2013). Texture-based descriptors for writer identification and verification. *Expert Systems with Applications*, *40*(6), 2069-2080.

[29]. Brink, J. Smit, M. L. Bulacu, and L. R. B. Schomaker, "Writer identification using directional ink trace width measurements," Pattern Recognit., vol. 45, no. 1, pp. 162–171, 2012,doi: 10.1016/j.patcog.2011.07.005.

[30]. Marti, U. V., Messerli, R., & Bunke, H. (2001, September). Writer identification using text line based features. In Proceedings of Sixth International Conference on Document Analysis and Recognition (pp. 101-105). IEEE.

[31]. Djeddi, C., Meslati, L. S., Siddiqi, I., Ennaji, A., El Abed, H., & Gattal, A. (2014, April). Evaluation of texture features for offline arabic writer identification. In 2014 11th IAPR international workshop on document analysis systems (pp. 106-110). IEEE.

[32]. Vapnik, V. N. (1995). The nature of statistical learning. Theory.. Springer-Verlag, London, UK, 1995.

[33]. Gazzah, S., & Amara, N. B. (2015). Neural networks and support vector machines classifiers for writer identification using Arabic script. 10.13140/RG.2.1.1327.7287.

[34]. Dargan, S., Kumar, M., Garg, A. et al. Writer identification system for pre-segmented offline handwritten Devanagari characters using k-NN and SVM. Soft Comput 24, 10111–10122 (2020). https://doi.org/10.1007/s00500-019-04525-y.

[35]. Thendral, T., Vijaya, M. S., & Karpagavalli, S. (2013). Supervised learning approach for Tamil writer identity prediction using global and local features. Int J Res Eng Technol (IJRET), 2(5), 204-208.

[36]. Feng, J., & Zhu, Y. (2006, October). Text independent writer identification based on Gabor filter and SVM classifier. In Sixth International Symposium on Instrumentation and Control Technology: Signal Analysis, Measurement Theory, Photo-Electronic Technology, and Artificial Intelligence (Vol. 6357, pp. 207-213). SPIE.https://doi.org/10.1117/12.716914.

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

[38]. Siddiqi, I., Vincent, N.: 'Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features', Pattern Recognition, 2010, 43, (11), pp. 3853–3865

[39]. Christlein, V., Bernecker, D., Hönig, F., Maier, A., & Angelopoulou, E. (2017). Writer identification using GMM supervectors and exemplar-SVMs. Pattern Recognition, 63, 258-267.

[40]. Djeddi, C., Al-Maadeed, S., Gattal, A., Siddiqi, I., Souici-Meslati, L., & El Abed, H. (2015, August). ICDAR2015 competition on multi-script writer identification and gender classification using 'QUWI'database. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 1191-1195). IEEE.

[41]. Djeddi, C., Al-Maadeed, S., Gattal, A., Siddiqi, I., Ennaji, A., & El Abed, H. (2016, October). ICFHR2016 Competition on multi-script writer demographics classification using" QUWI" database. In 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR) (pp. 602-606). IEEE.

[42]. He, S., & Schomaker, L. (2017). Writer identification using curvature-free features. Pattern Recognition, 63, 451-464.

[43]. He, S., & Schomaker, L. (2014, August). Delta-n hinge: rotation-invariant features for writer identification. In 2014 22nd International Conference on Pattern Recognition (pp. 2023-2028). IEEE.

[44]. Bulacu, M., & Schomaker, L. (2007). Text-independent writer identification and verification using textural and allographic features. IEEE transactions on pattern analysis and machine intelligence, 29(4), 701-717.

[45]. Tan, G. J., Sulong, G., & Rahim, M. S. M. (2017). Writer identification: a comparative study across three world major languages. Forensic science international, 279, 41-52.

[46]. Srihari, S. N., Cha, S. H., Arora, H., & Lee, S. (2002). Individuality of handwriting. Journal of forensic sciences, 47(4), 856-872.

[47]. Bensefia, A., Paquet, T., & Heutte, L. (2003, November). Grapheme based writer verification. In 11th Conference of the International Graphonomics Society, IGS.

[48]. Brink, A. A., Niels, R. M. J., van Batenburg, R. A., van den Heuvel, C. E., & Schomaker, L. R. B. (2011). Towards robust writer verification by correcting unnatural slant. Pattern Recognition Letters, 32(3), 449-457.

[49]. Schomaker, L., Franke, K., & Bulacu, M. (2007). Using codebooks of fragmented connected-component contours in forensic and historic writer identification. Pattern Recognition Letters, 28(6), 719-727.

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

[51]. Guo, Z., Zhang, L., & Zhang, D. (2010). Rotation invariant texture classification using LBP variance (LBPV) with global matching. Pattern recognition, 43(3), 706-719.

[52]. Srihari, S.N., Meng, L., Hanson, L.: 'Development of individuality in children's handwriting', Journal of forensic sciences, 2016, 61, (5), pp. 1292–1300.

[53]. Saunders, C.P., Davis, L.J., Buscaglia, J.: 'Using automated comparisons to quantify handwriting individuality', Journal of forensic sciences, 2011, 56, (3),pp. 683–689.

[54]. Jain, A.K., Flynn, P., Ross, A.A.: 'Handbook of biometrics. (Springer Science & Business Media, 2007)

[55]. Kumar, P., Sharma, A.: 'Dcwi: Distribution descriptive curve and cellular automata based writer identification', Expert Systems with Applications, 2019, 128, pp. 187– 200.

[56]. Chahi, A., Ruichek, Y., Touahni, R., et al.: 'An effective and conceptually simple feature representation for off-line text-independent writer identification', Expert Systems with Applications, 2019, 123, pp. 357–376.

[57]. Tan, G.J., Sulong, G., Rahim, M.S.M.: 'Writer identification: A comparative study across three world major languages', Forensic science international, 2017, 279, pp. 41–52.

[58]. Rehman, A., Naz, S., Razzak, M.I.: 'Writer identification using machine learning approaches: a comprehensive review', Multimedia Tools and Applications, 2019, 78, (8), pp. 10889–10931.

[59]. Schomaker, L. 'Advances in writer identification and verification'. In: Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). vol. 2. (IEEE, 2007. pp. 1268–1273.

[60]. Garain, U., Paquet, T. 'Off-line multi-script writer identification using ar coef-ficients'. In: 2009 10th International Conference on Document Analysis and Recognition. (IEEE, 2009. pp. 991–995.

[61]. Bertolini, D., Oliveira, L.S., Sabourin, R. 'Multi-script writer identification using dissimilarity'. In: 2016 23rd International Conference on Pattern Recognition (ICPR). (IEEE, 2016. pp. 3025–3030.

[62]. Djeddi, C., Siddiqi, I., Souici.Meslati, L., Ennaji, A. 'Multi-script writer identification optimized with retrieval mechanism'. In: 2012 International Conference on Frontiers in Handwriting Recognition. (IEEE, 2012. pp. 509–514.

[63]. Brink, A., Niels, R., Van.Batenburg, R., Van den Heuvel, C., Schomaker, L.: 'Towards robust writer verification by correcting unnatural slant', Pattern Recog-nition Letters, 2011, 32, (3), pp. 449–457.

[64]. Djeddi, C., Al.Maadeed, S., Siddiqi, I., Abdeljalil, G., He, S., Akbari, Y. 'Icfhr 2018 competition on multi-script writer identification'. In: 2018 16th Interna-tional Conference on Frontiers in Handwriting Recognition (ICFHR). (IEEE, 2018. pp. 506–510.

[65]. Hannad, Y., Siddiqi, I., El.Kettani, M.E.Y.: 'Writer identification using texture descriptors of handwritten fragments', Expert Systems with Applications, 2016, 47, pp. 14–22.

[66]. Newell, A.J., Griffin, L.D.: 'Writer identification using oriented basic image features and the delta encoding', Pattern Recognition, 2014, 47, (6), pp. 2255–2265.

[67]. Schomaker, L., Franke, K., Bulacu, M.: 'Using codebooks of fragmented connected-component contours in forensic and historic writer identification', Pattern Recognition Letters, 2007, 28, (6), pp. 719–727.

[68]. Abdi, M.N., Khemakhem, M.: 'A model-based approach to offline text-independent arabic writer identification and verification', Pattern Recognition, 2015, 48, (5), pp. 1890–1903.

[69]. He, S., Wiering, M., Schomaker, L.: 'Junction detection in handwritten documents and its application to writer identification', Pattern Recognition, 2015, 48, (12), pp. 4036–4048.

[70]. Khalifa, E., Al.Maadeed, S., Tahir, M.A., Bouridane, A., Jamshed, A.: 'Off-line writer identification using an ensemble of grapheme codebook features', Pattern Recognition Letters, 2015, 59, pp. 18–25.

[71]. Khan, F.A., Tahir, M.A., Khelifi, F., Bouridane, A., Almotaeryi, R.: 'Robust off-line text independent writer identification using bagged discrete cosine transform features', Expert Systems with Applications, 2017, 71, pp. 404–415.

[72]. Bennour, A., Djeddi, C., Gattal, A., Siddiqi, I., Mekhaznia, T.: 'Handwriting based writer recognition using implicit shape codebook', Forensic science international, 2019, 301, pp. 91–100.

[73]. Nguyen, H.T., Nguyen, C.T., Ino, T., Indurkhya, B., Nakagawa, M.: 'Text-independent writer identification using convolutional neural network', Pattern Recognition Letters, 2019, 121, pp. 104–112.

[74]. Xing, L., Qiao, Y. 'Deepwriter: A multi-stream deep cnn for text-independent writer identification'. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). (IEEE, 2016. pp. 584–589.

[75]. Yang, W., Jin, L., Liu, M.: 'Deepwriterid: An end-to-end online text-independent writer identification system', IEEE Intelligent Systems, 2016, 31, (2), pp. 45–53.

[76]. Christlein, V., Bernecker, D., Maier, A., Angelopoulou, E. 'Offline writer iden-tification using convolutional neural network activation features'. In: German Conference on Pattern Recognition. (Springer, 2015. pp. 540–552.

[77]. Zhang, X.Y., Xie, G.S., Liu, C.L., Bengio, Y.: 'End-to-end online writer identi-fication with recurrent neural network', IEEE Transactions on Human-Machine Systems, 2016, 47, (2), pp. 285–292

[78]. Djeddi, C., Siddiqi, I., Souici.Meslati, L., Ennaji, A.: 'Text-independent writer recognition using multi-script handwritten texts', Pattern Recognition Letters, 2013, 34, (10), pp. 1196–1202.

[79]. Freitas, C., Oliveira, L. S., Sabourin, R., & Bortolozzi, F. (2008). Brazilian forensic letter database. In 11th International workshop on frontiers on handwriting recognition, Montreal, Canada.

[80]. Wang, L., He, D.C.: 'Texture classification using texture spectrum', Pattern Recognition, 1990, 23, (8), pp. 905–910.

[81]. Ojala, T., Pietikainen, M., Harwood, D. 'Performance evaluation of texture mea-sures with classification based on kullback discrimination of distributions'. In: Proceedings of 12th International Conference on Pattern Recognition. vol. 1. (IEEE, 1994. pp. 582–585.

[82]. Ojala, T., Pietikäinen, M., Harwood, D.: 'A comparative study of texture measures with classification based on featured distributions', Pattern recognition, 1996, 29, (1), pp. 51–59.

[83]. Ojala, T., Pietikainen, M., Maenpaa, T.: 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', IEEE Transactions on pattern analysis and machine intelligence, 2002, 24, (7), pp. 971–987

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

[85]. Gattal, A., Djeddi, C., Chibani, Y., Siddiqi, I. 'Isolated handwritten digit recognition using obifs and background features'. In: 2016 12th IAPR Workshop on Document Analysis Systems (DAS). (IEEE, 2016. pp. 305–310

[86]. Gattal, A., Djeddi, C., Siddiqi, I., Chibani, Y.: 'Gender classification from offline multi-script handwriting images using oriented basic image features (obifs)', Expert Systems with Applications, 2018, 99, pp. 155–167.

[87]. Chahi, A., Ruichek, Y., & Touahni, R. (2019). An effective and conceptually simple feature representation for off-line text-independent writer identification. Expert Systems with Applications, 123, 357-376.

[88]. Tan, Gloria Jennis, Ghazali Sulong, and Mohd Shafry Mohd Rahim. 2017. "Writer Identification: A Comparative Study across Three World Major Languages." Forensic Science International 279:41–52.

[89]. Rehman, Arshia, Saeeda Naz, and Muhammad Imran Razzak. 2019. "Writer Identification Using Machine Learning Approaches: A Comprehensive Review." Multimedia Tools and Applications 78(8):10889–931

[90]. Malik, M. I., Ahmed, S., Marcelli, A., Pal, U., Blumenstein, M., Alewijns, L., & Liwicki, M. (2015, August). ICDAR2015 competition on signature verification and writer identification for on-and off-line skilled forgeries (SigWIcomp2015). In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 1186-1190). IEEE.

[91]. Slimane, F., Awaida, S., Mezghani, A., Parvez, M. T., Kanoun, S., Mahmoud, S. A., & Märgner, V. (2014, September). Icfhr2014 competition on arabic writer identification using ahtid/mw and khatt databases. In 2014 14th international conference on frontiers in handwriting recognition (pp. 797-802). IEEE.

[92]. Christlein, V., Nicolaou, A., Seuret, M., Stutzmann, D., & Maier, A. (2019, September). ICDAR 2019 competition on image retrieval for historical handwritten documents. In 2019 International Conference on Document Analysis and Recognition (ICDAR) (pp. 1505-1509). IEEE.

[93]. Abdeljalil, G., Djeddi, C., Siddiqi, I., & Al-Maadeed, S. (2018, August). Writer identification on historical documents using oriented basic image features. In 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR) (pp. 369-373). IEEE.

[94]. Gahmousse, A., Gattal, A., Djeddi, C., & Siddiqi, I. (2020, October). Handwriting based personality identification using textural features. In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI) (pp. 1-6). IEEE.

[95]. Abbas, F., Gattal, A., Djeddi, C., Siddiqi, I., Bensefia, A., & Saoudi, K. (2021). Texture feature column scheme for single-and multi-script writer identification. IET Biometrics, 10(2), 179-193.

[96]. Abbas, F., Gattal, A., Djeddi, C., Bensefia, A., Jamil, A., & Saoudi, K. (2020, December). Offline writer identification based on clbp and vlbp. In Mediterranean Conference on Pattern Recognition and Artificial Intelligence (pp. 188-199). Springer, Cham.

[97]. He, S., Schomaker, L. 'Delta-n hinge: rotation-invariant features for writer identification'.In: 2014 22nd International Conference on Pattern Recognition. (IEEE,2014. pp. 2023–2028

[98]. He, S., Schomaker, L.: 'Writer identification using curvature-free features', Pattern Recognition, 2017, 63, pp. 451–464

[99]. Gattal, A., Djeddi, C., Bensefia, A., & Ennaji, A. (2020, June). Handwriting based gender classification using cold and hinge features. In International Conference on Image and Signal Processing (pp. 233-242). Springer, Cham.

[100]. Christlein, V., Gropp, M., Fiel, S., & Maier, A. (2017, November). Unsupervised feature learning for writer identification and writer retrieval. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 991-997). IEEE.

[101]. Jordan, S., Seuret, M., Král, P., Lenc, L., Martínek, J., Wiermann, B., ... & Christlein, V. (2020, July). Re-ranking for Writer Identification and Writer Retrieval. In International Workshop on Document Analysis Systems (pp. 572-586). Springer, Cham.

[102]. Chammas, M., Makhoul, A., & Demerjian, J. (2020, December). Writer identification for historical handwritten documents using a single feature extraction method. In 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1-6). IEEE.

[103]. Lai, S., Zhu, Y., & Jin, L. (2020). Encoding pathlet and sift features with bagged vlad for historical writer identification. IEEE Transactions on Information Forensics and Security, 15, 3553-3566.

[104]. Lai, Songxuan, Yecheng Zhu, and Lianwen Jin. "Encoding pathlet and sift features with bagged vlad for historical writer identification." IEEE Transactions on Information Forensics and Security 15 (2020): 3553-3566.

[105]. Hafner, J., Sawhney, H. S., Equitz, W., Flickner, M., & Niblack, W. (1995). Efficient color histogram indexing for quadratic form distance functions. IEEE transactions on pattern analysis and machine intelligence, 17(7), 729-736.

[106]. Mukundan, R., & Ramakrishnan, K. R. (1998). Moment functions in image analysis: theory and applications. World scientific.