OFFLINE TEXT-INDEPENDENT CHINESE WRITER IDENTIFICATION METHOD WITH TWO-TIER IMAGE RETRIEVAL

GLORIA JENNIS TAN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

APRIL 2019

In memory of my late grandfather, Kayam Ak Bawar

and

To my family with love and eternal appreciation

ACKNOWLEDGEMENT

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Dr. Mohd Shafry Mohd Rahim for his supervisory role throughout my PhD program. I am very thankful of his patience to understand and moral support throughout my study's journey. His guidance helped me in all the time along this research and the writing phase of this thesis.

Also, I would like to express my appreciation to Prof. Dr. Ghazali Bin Sulong for the continuous support and guidance along my research study, most importantly, of his patience, motivation, enthusiasm and powerful knowledge in this research scope and image processing. I am really grateful that he willing to take me as his student and study under his wing. Also, I have gained a wealth of experience and knowledge working under your supervision, which will always be my delight to share along my life's journey.

Last but not least, my heartfelt gratitude goes to my beloved mother and husband for their love, faith in me and encouragement all these years. To my adorable daughter Yvie Yu, you are the epitome of a perfect daughter. You created your own play-world just to afford me the time to bring out the best in me. There is nothing we cannot achieve. Special thanks to my friends especially Dr. Sean Tan, Dr. Wong Lih Fong, Mr. Chong Yuh Chwen, Ms. Aow Yong Li Yew and Mdm. Choo Yen Lee for their support along this challenging journey.

ABSTRACT

Writer identification is essential today to identify the authenticity of a document in forensic expert decision-making. However, handwriting in various languages specifically Chinese poses a different challenge in identifying the writer. The main challenge faced by current researchers is that they fail to adopt traditional methods over an offline text independent Chinese writer identification scheme due to the complexity of Chinese writing structure and style. Furthermore, the previous method relies heavily on the selection of window size, which causes an ambiguity and leads to inconsistent results if the previous method is applied on a large image repository while finding the best-matched document from the database. Thus, much uncertainty still exists about the insurmountable searching space and the method has failed to show the effectiveness in searching relevant documents from a large image repository. This research attempted to solve problems by developing a new identification scheme for offline text-independent Chinese writer identification with the enhancement of feature extraction method and two-tier image retrieval mechanism to reduce search space and increase identification rates. The technique involved three essential steps. Firstly, the first-tier phase used Slantlet Transform based Local Binary Pattern (SLT-LBP) to bring out fine details. Then, sixty matching handwriting images were short-listed for the second-tier phase using Hierarchical Centroid (HC) of image pixels method for feature extraction. Finally, thirty shortlisted images were used as the input in the identification phase using Gray-Level Difference Method (GLDM) features. Experiment results had remarkably improved as compared to the previous method and the increase was from 95.4% to 96.68% in terms of identification rate as reported in the HIT-MW dataset. The contribution of this study is that it highlights the importance of using a two-tier retrieval mechanism to reduce search space in a large database in order to achieve higher accuracy. Besides, the development of a size-independent writer identification mechanism is a novelty as it can corroborate real-world application.

ABSTRAK

Pengenalpastian penulis adalah sangat penting untuk mengenal pasti kesahihan penulis dokumen dalam proses membuat keputusan pakar forensik. Walau bagaimanapun, tulisan dengan pelbagai bahasa terutama tulisan Cina memberi cabaran berbeza kepada pengenalpastian penulis. Cabaran utama yang dihadapi oleh penyelidik semasa adalah mereka gagal menggunakan kaedah tradisional menerusi pengenalpastian penulis Cina kerana kesukaran memahami struktur dan gaya penulisan tulisan Cina. Selain itu, kaedah sebelumnya bergantung pada pemilihan saiz tetingkap, yang menyebabkan keputusan yang samar dan tidak konsisten jika kaedah terdahulu diterapkan pada repositori imej yang besar sambil mencari dokumen yang paling sesuai dari pangkalan data. Oleh itu, banyak ketidakpastian masih wujud tentang ruang pencarian yang tidak dapat diatasi dan kaedah tersebut gagal menunjukkan keberkesanan dalam mencari dokumen yang relevan dari repositori imej yang besar. Kajian ini cuba menyelesaikan masalah dengan membangunkan skim pengenalan baru untuk pengenalpastian penulis bebas teks di luar talian dengan peningkatan kaedah pengekstrakan ciri dan mekanisma pengambilan imej dua peringkat untuk mengurangkan ruang carian dan meningkatkan kadar pengenalan. Teknik ini melibatkan tiga langkah penting. Pertama, fasa peringkat-pertama menggunakan *Slantlet Transform* berdasarkan *Local Binary Pattern* (SLT-LBP) untuk membawa butiran halus. Kemudian, enam puluh tulisan tangan yang sepadan tersenarai pendek untuk fasa peringkat-kedua menggunakan *Hierarchical Centroid* (HC) kaedah pixel imej untuk pengekstrakan ciri. Akhirnya, tiga puluh gambar yang disenarai pendek digunakan sebagai input dalam fasa pengenalan menggunakan ciri *Gray-Level Difference Method* (GLDM). Keputusan eksperimen telah meningkat berbanding dengan kaedah sebelumnya dan peningkatan adalah dari 95.4% kepada 96.68% dari segi kadar pengenalan seperti yang dilaporkan dalam dataset HIT-MW. Sumbangan kajian ini adalah menekankan pentingnya menggunakan mekanisma pengambilan dua peringkat untuk mengurangkan ruang carian dalam pangkalan data yang besar untuk mencapai ketepatan yang lebih tinggi. Di samping itu, perkembangan mekanisma ukuran pendekatan penulis bebas adalah sesuatu yang baru kerana dapat menyokong aplikasi dunia nyata.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AWR	-	Automated writer recognition
AIST	-	Advanced Industrial Science and Technology
BoW	-	Bag-of-words
BoF	-	Bag-of-features
BSM	-	Blurred Shape Model
C	-	Contour Gradients
CVL	-	Computer Vision Lab dataset
CDIP	-	Complex Document Information Processing
CNN	-	Convoluted Neural Networks
CASIA	-	The Institute of Automation of Chinese Academy of Sciences dataset
CS-UMD	-	A method using An Alphabet of Contour Gradient Descriptors
CCF	-	Connected Component Feature
CVC	-	Computer Vision Center–Department of Computer Science, Universitat Autònoma de Barcelona
DWT	-	Discrete Wavelet Transforms
ETL	-	Electrotechnical Laboratory
EER	-	Equal Error Rate
ESC	-	Edge Structure Coding
FAR	-	False Alarm Rate
FRR	-	False Reject Rate
FFT	-	Fast Fourier Transform
GGD	-	Generalized Gaussian density
GLCM	-	Gray Level Co-occurrence Matrix
GMM	-	Gaussian Mixture Models

GLDM	-	Gray-Level Difference Method
GLRL	-	Grey Level Run Length
H	-	High frequency band
HH	-	High-High frequency band
HL	-	High-Low frequency band
HMT	-	Hidden Markov Tree
HC	-	Hierarchical Centroids
HV	-	Hard Voting
HCSD	-	Hierarchical Centroid Shape Descriptor
HanjaDB1	-	A database is collected by the Korea Advanced Institute of Science and Technology
HCL2000	-	Handwritten Character Library 2000
HIT-MW	-	HIT is the abbreviation of Harbin Institute of Technology, and MW means Multiple Writers dataset
HIT-CG	-	Extracts SIFT features (SDS and SOH) for word to characterize individuality of the writer
IAM	-	Institut fur Informatik und angewandte Mathematik (Department of Computer Science and Applied Mathematics), University of Bern, Bern, Switzerland
ICA	-	Independent Component Analysis
ICT	-	Information & Communication Technology
IR	-	Information Retrieval
IFK	-	Fisher Kernels
IFK	-	Improved Fisher kernel
ICDAR	-	International Conference on Document Analysis and Recognition
ICFHR	-	International Conference on Frontiers in Handwriting Recognition
IFN/ENIT	-	Institute for Communications Technology (IFN) Ecole Nationale d’Ingénieurs de Tunis (ENIT)
K	-	K-Adjacent Segments
KAS	-	K-Adjacent Segment
KNN	-	K-nearest-neighbour

KHATT	-	KFUPM Handwritten Arabic TexT
KDA	-	Kernel Discriminant Analysis
KAIST	-	Korea Advanced Institute of Science & Technology
L	-	Low frequency band
LBP	-	Local Binary Pattern
LH	-	Low-High frequency band
LL	-	Low-Low frequency band
LLC	-	Locally-constrained Linear Coding
LPQ	-	Local Phase Quantization
MADCAT	-	Multilingual Automatic Document Classification Analysis and Translation
MSD	-	Modified SD
MSDH	-	MSD Histogram
MIAM	-	Modified IAM dataset
MSHDArabic	-	Multi-script Handwritten Data set - Arabic
NN	-	Neural Networks
oBIF	-	Oriented Basic Image Feature
QUWI	-	Qatar University Writer Identification dataset
ROI	-	Region of Interest
S	-	Speeded Up Robust Feature SURF
SVM	-	Support Vector machine
SLT	-	Slantlet Transform
SURF	-	Speeded Up Robust Features
SIFT	-	Scale Invariant Feature Transform
SLM	-	Statistical Language Models
SDS	-	SIFT Descriptor Signature
SOH	-	Scale and Orientation Histogram
SD	-	SIFT Descriptor
SR-KDA	-	Kernel Discriminant Analysis using Spectral Regression
SOM	-	Self-Organizing Map
SYSU	-	Sun Yat-sen University
TBLR	-	Special Quadrilateral

TD	-	Triangular Descriptor
TDH	-	TD Histogram
TN	-	True Negative
TP	-	True Positive
VQ	-	Vector Quantization
WED	-	Weighted Euclidean Distance

LIST OF SYMBOLS

$g_i(n)$, $f_i(n)$, and $h_i(n)$	-	The coefficients of the filters
a_0, a_1, b_0, b_1, c_0, c_1, d_0, and d_1	-	Filters-bank parameters of image block
SLTfilter	-	SLT filter matrix
SLTfilterT	-	SLT transposing filter matrix
SLTimage	-	SLT image matrix
m x m	-	Size of block image
P	-	Number of pixels in the neighborhood
R	-	Radius
P_c	-	Center pixel value
Q	-	Feature vector of query image
G_k	-	Feature vector of the kth image of the dataset
x1 and x2	-	Two feature vectors of both images
d1, d2, …, dN	-	Distances for all images of the dataset
g1	-	Mean
g2	-	Standard deviation
g3	-	Maximum
H_k	-	Feature vector of the shortlisted images
F	-	Feature vector of the query images
d	-	Depth of the features extraction process
I	-	Input binary image
$m \times n$	-	Size of binary image
Ifg	-	Binary image with foreground
Ibg	-	Binary image with background
$C(x_c, y_c)$	-	Input centroid at the root level
$(p+q)$	-	Moment of order (raw moment)
n_{ij}	-	Number of pixels

LIST OF ALGORITHMS

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

Research in writer identification has received significant interest in recent years due to its forensic applicability. The problem of writer identification arises frequently in the court of justice where one must come to a conclusion about the authenticity of a document. Forensic analysis of handwriting requires to query large databases of handwritten samples of known writers due to the large number of individuals to be considered. The task is to establish the identity of the writer of a questioned handwritten document, by comparing the questioned handwriting to handwritten samples with known identities which are stored in a database. Hence, the research has history of decades and recently draws more and more attention because of its significance in criminal justice proceedings owing to its significance in forensic, security, financial transaction and archaeological investigations, both academic and industrial researchers are now more interested in it than ever before (H.E.S. Said, Tan and Baker, 2000; Louloudis *et al.*, 2013; Kore and Apte, 2012; Saranya and Vijaya, 2013). This performance however, remains far from being achieved for the time being tend to be computationally over-expensive, especially searching for relevant document from large complex document image repositories is a crucial problem in document image analysis and retrieval. It is as yet unclear such methods can be seamlessly integrated in current forensic handwriting expertise.

Furthermore, there are numerous languages throughout the world. Each language with different structure characters and writing style poses a different challenge to the writer identification problem depending on its characteristics. Although, Western handwriting identification technology has been experimented on large handwriting databases and has shown practical effectiveness, but many identification methods proposed for Western handwriting are not suitable for Eastern handwriting, such as Chinese and Japanese (J. Tan *et al.*, 2011; Bulacu and Schomaker, 2007; Li, X., Ding, X., and Wang, 2008). Chinese handwriting identification is a rather challenging task because different writing styles with unique stroke shapes and structures of Chinese characteristics are embedded which is more difficult from those of other languages.

1.2 Problem Background

The problem of writer identification arises frequently in the court of justice where one must come to a conclusion about the authenticity of a document. Hence, it has received significant interest in recent years due to its forensic applicability and most of the studies in this field share the same goal of identifying authorship of a script by acquiring individual features of the handwriting. Current traditional method involves a process that to generate and find all documents features, then by comparing the feature vector distance between query and entire library database. Extensive research has been carried out on traditional method which is without retrieval is being done in real-world scenario as of now across three major languages in Chinese, English and Arabic around the world. It is observed that during that period, there are significant progresses achieved on English and Arabic; however, the growth on Chinese is rather slow and far from satisfactory in comparison to its wide usage. The systems developed for Latin scripts have been tried on Arabic scripts in some studies that achieved variable degree of success but Chinese handwritten text is comparably rare and not proportional to its widely usage (Bulacu, Schomaker and Brink, 2007). This is due to each language poses a new challenge to the writer identification because of unique characteristics each language with its complex writing structure poses a different set of difficulties to the writer identification problem and requires a unique identification approach specific for that language.

The complications arising because of complicated Chinese and unique stroke Arabic script has motivated only a few studies on off-line text independent writer identification (J. T. J. Tan *et al.*, 2011; Tan, Lai and Zheng, 2013; J. Tan *et al.*, 2011; Sargur N Srihari, Yang and Ball, 2007; Cheng, 1998; Ahmed and Sulong, 2014; M. A. Abdullah *et al.*, 2012; Chawki and Labiba, 2010; Al-Maadeed, 2012; Hashemi, Fatemi and Safavi, 1995; Saeed, 2001; Assayony and Mahmoud, 2016). The best performance of text-independent Chinese writer identification attained to date is 95.4% matching rate with Top-1 (Wen *et al.*, 2012). They proposed a method based on edge structure code (ESC) distribution feature, which is extracted by window that scans the edge detected binary image of handwritten text. It requires pre-extraction of fragmented edge structure and code-based structural probability distribution of all writers in the database. Such a method segments texts into small square windows has been proven to effectively perform although in this context, window-based extraction calls for a tedious, challenging from a size-adjustable sliding window and the selection of window size directly affect identification performance. However, the performance of ESC is limited by window size and its accuracy decreases when operated on large size databases.

Few of the recent techniques, stated for all languages, performed ambiguously when tried on different languages. In addition to the challenges presented by characteristics of different language scripts, data size negatively affects the identification rate. Thus, it is difficult to find multiple researches on different languages using same dataset for benchmarking which makes the comparison, on common grounds, impossible. It would be valuable to prepare comparison mechanism based on a common benchmark to compare various identification schemes introduced by different researchers on common grounds and to avoid ambiguous results which needs to be summarized and compared among other researchers on a standard public database (Sreeraj and Idicula, 2011). In another word, a method toward language invariant instead being specific to a particular language.

Undoubtedly, many achievements have been made and only focused on identification performance on this very subject but a major problem with this kind of traditional method is to search for the relevant document from large complex document image repositories. This performance however, remains far from being achieved for the time being tend to be computationally over-expensive. Time complexity is however beyond the scope of this study. Searching and retrieval of a document from large image repositories is currently a big issue. Issues on datasets used by previous studies are also highlighted because the size matter – accuracy of the writer identification deteriorates as database size increases. There is a problem of data heterogeneity apart from problems of scale involve database size. Alternatively, it has become an important field of research to integrate writer identification with retrieval mechanism for improving the writer identification performance in real life applications. Extension of writer identification with retrieval was first introduced by Atanasiu *et al.* (2011) in English language to enhance the writer identification performance using single retrieval mechanism. So far, very little attention has been paid to the role of retrieval mechanism. Seamless integration of such methods in current forensic handwriting expertise is as yet unclear. Until now, in the research on off-line text-independent writer recognition researchers have been mainly interested in writer identification without retrieval mechanism approach.

Current progress in writer identification is forging new areas of research and applications has inspired this study to introduce retrieval mechanism in off-line text-independent writer identification approach for Chinese language while emphasizing on data heterogeneity aiming for reducing big search space, which requires a large amount of memory, computation power and time consumption in interpretability of results when large databases are involved. This research is enlivened by the similar idea of reducing the feature dimension before to next identification task. Up to now, there are still remains research opportunity and gap in existing method in terms of offline text- independent Chinese writer identification.

1.3 Problem Statements

Based on the problem background, unimpressive identification accuracy (Wen *et* *al.*, 2012) lead ambiguous and inconsistent result strictly depends upon the selection window size involve manipulation constant window size. Thus, the manipulation process is not reliable and efficient involve insurmountable search space especially finding best matched document in large and complex databases across large number of individuals with complex writing structure by comparing the feature vector distance between query and entire library database directly impacts the writer identification performance deteriorates as database size increases is a crucial problem in this field.

1.4 Research Goal

This study aims to develop a new scheme of off-line text-independent Chinese writer identification with retrieval mechanism for better and remarkable accuracy.

1.5 Research Objectives

In order to achieve the above mentioned goal, the following three major objectives must be fulfilled:

1. To develop feature extraction method for image retrieval process which realizes better identification rates than the existing state-of-the-art writer identification methods.
2. To enhance writer identification method that integrates with two-tier image retrieval for reducing search space and improving identification rates.
3. To propose and implement the identification enhancement scheme of Chinese writer for off-line text-independent.

1.6 Research Scope

The research objectives are achieved by identifying the problem scope which focuses the following aspects:

i. Evaluations on public standard HIT-MW Chinese database (Tonghua Su, 2006) comprising three major writing styles in regular script, fluent script and cursive script.
ii. Type of features: texture feature is the concern of this study, other features such as shape and color are beyond the scope of study.
iii. Use of offline or scanned Chinese handwriting samples.
iv. Performance evaluation: Accuracy is used in line with previous works. Time complexity, precision and recall are however beyond the scope of this study.

1.7 Significance of the Study

Handwriting based personal identification is great interest in image processing to correctly recognise an individual through handwriting and important research area of forensic interest which has been lately focused on automatic writer identification (Zhu, Tan and Wang, 2000). The need for an efficient, fully automated writer recognition (AWR) and a powerful handwritten document retrieval system as an emerging technology, tackles the challenges and issues that were reported by the researchers regarding the current writer identification, and that it will be considered as a consistently reliable biometric system with retrieval mechanism. After examining the issues involved in managing writer retrieval in some depth, the participants concluded that handwritten documents retrieval were indeed likely to play an increasingly important role in many fields. Several writer retrieval methods have been described and large number of the new writer retrieval techniques invented but most of them are not adequate of effective writer identification representation.

It is also hoped and expected that AWR may minimize the size of the search area of writer identification as the size of the forensics databases and standard datasets are immense. Such issue directly impacts the system performance of writer identification and it consumes time. AWR also extracts and employs the most important information from the handwriting to accurately emphasize the writer’s characteristics.

Considering the above issues, the research outcomes are expected to contribute to the currently information regarding writer identification systems with retrieval mechanism. Despite the above described significance of this study, it must understand that it will not confine only to the enrichment of knowledge. It is self-evident of potentially capable of practical applications.

1.8 Thesis Outline

In this chapter along presenting an overview of the research problem and a brief background, it cites recent work in the area particularly those dedicated to window-based methods to stress on the issues that exist in the current methods. The objectives of the research are also described in this chapter.

In the first part of Chapter 2, an overview of writing identification; the significant contributions in the writer identification field is presented, and different methods are synthesized in the text independent methods for writer identification including the selection of database, features extraction, similarity measurement, and performance evaluation method are described. Whereas, in the second part of this chapter, several writer identifications with retrieval mechanisms is described.

In Chapter 3, presents a clear roadmap that describe research framework in details is presented. It provides explanation of every step involved in systematic approach used for carrying out this research project. A discussion on phases, techniques and performance measures of the proposed technique is appropriately

REFERENCES

Ahmed, A. and Sulong, G. (2014). Arabic Writer Identification: A Review of Literature. *Journal of Theoretical and Applied Information Technology*. 69(3), 474–484.

Al-Ma'adeed, S., Mohammed, E., Al Kassis, D. and Al-Muslih, F. (2008). Writer identification using edge-based directional probability distribution features for Arabic words. *AICCSA 08 - 6th IEEE/ACS International Conference on Computer Systems and Applications*., 582–590.

Al-Maadeed, S. (2012). Text-dependent writer identification for arabic handwriting. *Journal of Electrical and Computer Engineering*. 2012.

Al-Maadeed, S., Hassaine, A., Bouridane, A. and Tahir, M.A. (2016). Novel geometric features for off-line writer identification. *Pattern Analysis and Applications*. 19(3), 699–708.

Armon, S. (2011). *Handwriting recognition and fast retrieval for Hebrew historical manuscripts*.

Assayony, M.O. and Mahmoud, S.A. (2016). An Enhanced Bag-of-Features Framework for Arabic Handwritten Sub-words and Digits Recognition. *Journal of Pattern Recognition and Intelligent Systems*. 4(1), 27–38.

Atanasiu, V. (2003). Allographic Biometrics and Behavior Synthesis. *EuroTeX 2003 Proceedings*. 24(3), 328–333.

Atanasiu, V., Likforman-Sulem, L. and Vincent, N. (2011). Writer retrieval - Exploration of a novel biometric scenario using perceptual features derived from script orientation. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*., 628–632.

Awaida, S.M. and Mahmoud, S. a. (2013). Writer Identification of Arabic Text Using Statistical and Structural Features. *Cybernetics and Systems*. 44(1), 57–76.

B.V.Dhandra, V.M.. Text Independent Writer Identification for Tamil Script.

National conference on Advances in Modern Computing and Application Trends, AIT, Bangalore.

Bensefia, A., Nosary, A., Paquet, T. and Heutte, L. (2002). Writer identification by writer's invariants. *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*., 274–279.

Bensefia, A., Paquet, T. and Heutte, L. (2004). Handwriting analysis for writer verification. *Frontiers in Handwriting …*., 196–201.

Bensefia, A., Paquet, T. and Heutte, L. (2005). Handwritten document analysis for automatic writer recognition. *... vision and image analysis*. 5(2), 72–86.

Bentley, J.L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*. 18(9), 509–517.

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

Brink, A.A., Smit, J., Bulacu, M.L. and Schomaker, L.R.B. (2012). Writer identification using directional ink-trace width measurements. *Pattern Recognition*. 45(1), 162–171.

Buckland, M. and Gey, F. (1994). The relationship between Recall and Precision. *Journal of the American Society for Information Science*. 45(1), 12–19.

Bui, Q.A., Visani, M., Prum, S. and Ogier, J.-M. (2011). Writer Identification Using TF-IDF for Cursive Handwritten Word Recognition. *2011 International Conference on Document Analysis and Recognition*. 844(1), 844–848.

Bulacu, M. (2007). *Statistical pattern recognition for automatic writer identification and verification,*

Bulacu, M. and Schomaker, L. (2007). Text-independent writer identification and verification using textural and allographic features. *IEEE transactions on pattern analysis and machine intelligence*. 29(4), 701–17.

Bulacu, M., Schomaker, L. and Brink, A. (2007). Text-independent writer identification and verification on offline arabic handwriting. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. 2, 769–773.

Bulacu, M., Schomaker, L. and Vuurpijl, L. (2003). Writer identification using edge-based directional features. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. 2003–Janua, 937–941.

Bulacu, M., Schomaker, L. and Vuurpijl, L. (2003). Writer Identification Using Edge-Based Directional Features. In *ICDAR'2003: International Conference on Document Analysis and Recognition*. IEEE Computer Society, p.937.

Cao, H., Prasad, R. and Natarajan, P. (2010). Improvements in HMM adaptation for handwriting recognition using writer identification and duration adaptation. *Proceedings - 12th International Conference on Frontiers in Handwriting Recognition, ICFHR 2010*., 154–159.

Chaabouni, A., Boubaker, H., Kherallah, M., Alimi, A.M. and Abed, H. El (2011)(a). Combining of Off-line and On-line Feature Extraction Approaches for Writer Identification. *2011 International Conference on Document Analysis and Recognition*., 1299–1303.

Chaabouni, A., Boubaker, H., Kherallah, M., Alimi, A.M. and Abed, H. El (2011)(b). Multi-fractal modeling for on-line text-independent writer identification. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*., 623–627.

Chawki, D. and Labiba, S.M. (2010). A texture based approach for Arabic Writer Identification and Verification. In *2010 International Conference on Machine and Web Intelligence, ICMWI 2010 - Proceedings*. pp.115–120.

Chen, J., Lopresti, D. and Kavallieratou, E. (2010). The impact of ruling lines on writer identification. *Proceedings - 12th International Conference on Frontiers in Handwriting Recognition, ICFHR 2010*., 439–444.

Chen, Q., Yan, Y., Deng, W. and Yuan, F. (2008). Handwriting Identification Based on Constructing Texture. *2008 First International Conference on Intelligent Networks and Intelligent Systems*., 523–526.

Cheng, F. (1998). Multi-stroke relaxation matching method for handwritten Chinese character recognition. *Pattern recognition*. 31(4).

Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. 13(1), 21–27.

Csurka, G., Dance, C.R., Fan, L., Willamowski, J. and Bray, C. (2004). Visual categorization with bags of keypoints. *Proceedings of the ECCV International Workshop on Statistical Learning in Computer Vision*., 59–74.

Dai, R., Liu, C. and Xiao, B. (2007). Chinese character recognition: history, status and prospects. *Frontiers of Computer Science in China*. 1(2), 126–136.

Daramola, S.A., Abdulkareem, A. and Adinfona, K.J. (2014). Efficient Item Image

Retrieval System. (2), 109–113.

Dhandra, B. V. and Vijayalaxmi, M.B. (2014). Text and script independent writer identification. *Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014.*, 586–590.

Dimauro, G. and Impedovo, S. (2002). A new database for research on bank-check processing. *IEEE Computer Society.*, 524–528.

Djeddi, C., Siddiqi, I., Souici-Meslati, L. and Ennaji, A. (2012). Multi-script Writer Identification Optimized with Retrieval Mechanism. *2012 International Conference on Frontiers in Handwriting Recognition.*, 509–514.

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

Du, L.D.L., You, X.Y.X., Xu, H.X.H., Gao, Z.G.Z. and Tang, Y.T.Y. (2010). Wavelet Domain Local Binary Pattern Features For Writer Identification. *Pattern Recognition (ICPR), 2010 20th International Conference on.*, 3695–3698.

Duda, R.O., Hart, P.E. and Stork, D.G. (1995). Pattern Classification and Scene Analysis. *New York: John Wiley, Section.*, 69.

E. N. Zois and V. Anastassopoulos (2000). Morphological waveform coding for writer identification. *Pattern Recognition*. 33(3), 385–398.

Fernandez-de-Sevilla, R., Alonso-Fernandez, F., Fierrez, J. and Ortega-Garcia, J. (2010). Forensic writer identification using allographic features. *Proceedings - 12th International Conference on Frontiers in Handwriting Recognition, ICFHR 2010*. 2010, 308–313.

Fiel, S. and Sablatnig, R. (2013). Writer identification and writer retrieval using the fisher vector on visual vocabularies. In *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. pp.545–549.

Fiel, S. and Sablatnig, R. (2012). Writer Retrieval and Writer Identification Using Local Features. *2012 10th IAPR International Workshop on Document Analysis Systems.*, 145–149.

Fornés, A., Dutta, A., Gordo, A. and Lladós, J. (2011). The ICDAR 2011 music scores competition: Staff removal and writer identification. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR.*, 1511–1515.

Fornes, A. and Llados, J. (2010). A symbol-dependent writer identification approach in old handwritten music scores. *Proceedings - 12th International Conference on Frontiers in Handwriting Recognition, ICFHR 2010*., 634–639.

Franke, K., Biinnemeyer, O. and Sy, T. (2002). Ink Texture Analysis for Writer Identification. 00, 268–273.

Garz, A., W, M., Fischer, A. and Ingold, R. (2016). Simple and Fast Geometrical Descriptors for Writer Identification. *Document Recognition and Retrieval*., 1–12.

Ghiasi, G. and Safabakhsh, R. (2013)(a). Offline text-independent writer identification using codebook and efficient code extraction methods. *Image and Vision Computing*. 31(5), 379–391.

Ghiasi, G. and Safabakhsh, R. (2013)(b). Offline text-independent writer identification using codebook and efficient code extraction methods. *Image and Vision Computing*. 31(5), 379–391.

Gordo, A., Fornés, A. and Valveny, E. (2013). Writer identification in handwritten musical scores with bags of notes. *Pattern Recognition*. 46(5), 1337–1345.

Gordo, A., Fornés, A., Valveny, E. and Lladós, J. (2010). A bag of notes approach to writer identification in old handwritten musical scores. *Proceedings of the 8th IAPR International Workshop on Document Analysis Systems - DAS '10*., 247–254.

Gupta, S. (2008). Automatic Person Identification and Verification using Online Handwriting. *Online*.

H Ya-pin, L Si-wei, C.E. (2003). Writer recognition based on independent component analysis. *Journal of Chinese Information Processing*. 4(17), 52–58.

Hashemi, M.R., Fatemi, O. and Safavi, R. (1995). Persian cursive script recognition. *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*. 2, 869–873.

Hassaine, A., Al-Maadeed, S., Alja'am, J.M., Jaoua, A. and Bouridane, A. (2011). The ICDAR2011 Arabic writer identification contest. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*., 1470–1474.

Hassaïne, A. and Al Maadeed, S. (2012). ICFHR2012 competition on writer identification - Challenge 2: Arabic scripts. *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*., 835–840.

He, Z., Fang, B. and Du, J. (2005). A novel method for offline handwriting-based writer identification. *Document Analysis and Recognition, Proceedings. Eighth International Conference*. 1, 242–246.

He, Z. and Tang, Y. (2004). Chinese handwriting-based writer identification by texture analysis. *Proceedings of the Third International Conference on Machine Learning and Cybernetics, 2004.* 6, 3488–3491.

He, Z., You, X. and Tang, Y.Y. (2008). Writer identification of Chinese handwriting documents using hidden Markov tree model. *Pattern Recognition*. 41(4), 1295–1307.

Hiremath, P.S. and Shivashankar, S. (2008). Wavelet based co-occurrence histogram features for texture classification with an application to script identification in a document image. *Pattern Recognition Letters*. 29(9), 1182–1189.

Hu, Yongjie, Wenming Yang, Y.C. (2014). Bag of features approach for offline text-independent Chinese writer identification. *Image Processing (ICIP), 2014 IEEE International Conference*., 2609–2613.

Jain, R. and Doermann, D. (2014). Combining Local Features for Offline Writer Identification. *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*. 2014–Decem, 583–588.

Jain, R. and Doermann, D. (2011). Offline Writer Identification Using K-Adjacent Segments. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*., 769–773.

Jain, R. and Doermann, D. (2013). Writer identification using an alphabet of contour gradient descriptors. In *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. pp.550–554.

Jain, R., Oard, D.W. and Doermann, D. (2013). Scalable ranked retrieval using document images. 9021, 90210K.

Khalifa, E., Al-Maadeed, S., Tahir, M.A., Bouridane, A. and Jamshed, A. (2015). Off-line writer identification using an ensemble of grapheme codebook features. *Pattern Recognition Letters*. 59, 18–25.

Kırlı, Ö. and Gülmezoğlu, M.B. (2011). Automatic writer identification from text line images. *International Journal on Document Analysis and Recognition (IJDAR)*. 15(2), 85–99.

Kore, S. and Apte, S. (2012). The current state of art: handwriting a behavioral biometric for person identification and verification. *Proceedings of the*

International Conference on Advances in Computing, Communications and Informatics., 925–930.

Kumar, R., Chanda, B. and Sharma, J.D. (2014). A novel sparse model based forensic writer identification. *Pattern Recognition Letters*. 35(1), 105–112.

L Hong (2003). Writer Identification Using Support Vector Machines and Texture Feature. *JOURNAL OF COMPUTER AIDED DESIGN*.

Leclerc, F. and Plamondon, R. (1994). Automatic Signature Verification: The State of the Art --1989-1993. *International Journal of Pattern Recognition and Artificial Intelligence*. 08(03), 643–660.

Lee, S., Member, S. and Society, I.C. (1998). Offline Recognition of Chinese handwriting by Multifeature and Multilevel Classification. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 20(5), 556–561.

Leng, W.Y. and Shamsuddin, S.M. (2010). Writer Identification for Chinese Handwriting. *Int. J. Advance. Soft Comput. Appl*. 2(2), 143–173.

Li, X., Ding, X., and Wang, X. (2008). Semi-text-independent Writer Verification of Chinese Handwriting. *Proc. of the 11th ICFHR*., 100–105.

Li, X. and Ding, X. (2009). Writer identification of chinese handwriting using grid microstructure feature. *Advances in Biometrics*., 1230–1239.

Liu, C.-L., Yin, F., Wang, Q.-F. and Wang, D.-H. (2011). ICDAR 2011 Chinese Handwriting Recognition Competition. *2011 International Conference on Document Analysis and Recognition*., 1464–1469.

Liu, C., Dai, R. and Liu, Y. (1995). Extracting individual features from moments for Chinese writer identification. *Proceedings of 3rd International Conference on Document Analysis and Recognition*. 1, 438–441.

Louloudis, G., Gatos, B. and Stamatopoulos, N. (2012). ICFHR 2012 Competition on Writer Identification Challenge 1: Latin/Greek Documents. *2012 International Conference on Frontiers in Handwriting Recognition*., 829–834.

Louloudis, G., Gatos, B., Stamatopoulos, N. and Papandreou, a. (2013). ICDAR 2013 Competition on Writer Identification. *2013 12th International Conference on Document Analysis and Recognition*., 1397–1401.

Louloudis, G., Stamatopoulos, N. and Gatos, B. (2011). ICDAR 2011 Writer Identification Contest. In *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. pp.1475–1479.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints.

International Journal of Computer Vision. 60(2), 91–110.

M. A. Abdullah, L. M. Al-Harigy, H. H. Al-Fraidi, Abdullah, M. a, Al-harigy, L.M. and Al-fraidi, H.H. (2012). Off-Line Arabic Handwriting Character Recognition Using Word Segmentation. *Journal of Computing*. 4(3), 40–44.

M. Awaida, S. (2012). State of the art in off-line writer identification of handwritten text and survey of writer identification of Arabic text. *Educational Research and Reviews*. 7(20), 445–463.

Maadeed, S. Al, Ayouby, W., Hassaïne, A. and Aljaam, J.M. (2012). QUWI: An Arabic and English handwriting dataset for offline writer identification. *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*., 746–751.

Maaten, L. van der (2009). *A new benchmark dataset for handwritten character recognition*,

Mahmoud, S.A., Ahmad, I., Al-Khatib, W.G., Alshayeb, M., Tanvir Parvez, M., Märgner, V. and Fink, G.A. (2014). KHATT: An open Arabic offline handwritten text database. *Pattern Recognition*. 47(3), 1096–1112.

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

Marti, U. and Bunke, H. (1999). A full English sentence database for off-line handwriting recognition. *Document Analysis and Recognition, 1999. ICDAR '99. Proceedings of the Fifth International Conference*., 705–708.

Marti, U. and Bunke, H. (2002). The IAM-database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*. 5(1), 39–46.

Mezghani, A., Kanoun, S., Khemakhem, M. and Abed, H. El (2012). A Database for Arabic Handwritten Text Image Recognition and Writer Identification. *2012 International Conference on Frontiers in Handwriting Recognition*., 399–402.

Newell, A.J. and Griffin, L.D. (2014). Writer identification using oriented Basic Image Features and the Delta encoding. *Pattern Recognition*. 47(6), 2255–2265.

Niels, R., Gootjen, F. and Vuurpijl, L. (2008). Writer Identification through Information Retrieval: The Allograph Weight Vector. *International Conference on the Frontiers of Handwriting Recognition*., 481–486.

Nigam, A., Kumar, B. and Gupta, P. (2015). Writer Identification and Retrieval Using a Convolutional Neural Network. 1, 702–714.

Nilawar, A.P. and Prof, A. (2014). Image Retrieval Using Gradient operators. (January), 2–5.

Ojala, T., Pietikäinen, M. and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*. 29(1), 51–59.

Ojala, T., Pietikainen, M. and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 24(7), 971–987.

Perronnin, F. and Dance, C. (2006). Fisher Kernels on Visual Vocabularies for Image Categorization. *Proc. {CVPR}*.

Perronnin, F., Sánchez, J. and Mensink, T. (2010). Improving the Fisher kernel for large-scale image classification. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp.143–156.

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

Pirlo, G., Chimienti, M., Dassisti, M., Impedovo, D. and Galiano, A. (2014). A layout-analysis based system for document image retrieval. *Mondo Digitale*. 13(49).

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

Rivard, D., Granger, E. and Sabourin, R. (2013). Multi-feature extraction and selection in writer-independent off-line signature verification. *International Journal on Document Analysis and Recognition*. 16(1), 83–103.

Saeed, K. (2001). New Approaches for Cursive Languages Recognition : Machine and Hand Written Scripts and Texts. *Invited Paper in Proceedings of World Scientific and Engineering Society WSES-NNA*.

Said, H.E.S., Tan, T.N. and Baker, K.D. (2000). Personal identification based on handwriting. *Pattern Recognition*. 33, 149–160.

Said, H.E.S., Tan, T.N. and Baker, K.D. (2000). Personal identification based on handwriting. *Pattern Recognition*. 33, 149–160.

Said, Y., Saidani, T., Smach, F., Atri, M. and Snoussi, H. (2012). Writer recognition on arabic handwritten documents. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 7340(JUNE), 85–92.

Saranya, K. and Vijaya, M. (2013). An interactive Tool for Writer Identification based on Offline Text Dependent Approach. *International Journal*. 2(1), 33–40.

Schomaker, L. (2007). Advances in writer identification and verification. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. 2, 1268–1273.

Schomaker, L. and Bulacu, M. (2004). Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 26(6), 787–798.

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

Selesnick, I.W. (1999). The Slantlet Transform. 47(5), 1304–1313.

Senior, a. W. and Robinson, A.J. (1998). An off-line cursive handwriting recognition system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 20(3), 309–321.

Sexton, A., Todman, A. and Woodward, K. (2000). Font Recognition Using Shape-Based Quad-tree and Kd-tree Decomposition. *Proceedings Of The Joint Conference On Information Sciences*. 5(2), 212–215.

Shirdhonkar, M.S. and Kokare, M.B. (2012). Handwritten Document Image Retrieval. *International Journal of Modeling and Optimization*. 2(6), 693–696.

Siddiqi, I. and Vincent, N. (2008). Combining global and local features for writer identification. *Proceedings of the 11. Int. Conference ...*.

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

Sreeraj, M. and Idicula, S.M. (2011). A Survey on Writer Identification Schemes. *International Journal of Computer Applications*. 26(2), 23–33.

Srihari, S.N., Cha, S.-H., Arora, H. and Lee, S. (2002). Individuality of handwriting.

Journal of forensic sciences. 47(4), 856–72.

Srihari, S.N., Yang, X. and Ball, G.R. (2007). Offline Chinese Handwriting Recognition : A Survey. *Frontiers of Computer Science in China*., 1–32.

Srihari, S.N., Yang, X. and Ball, G.R. (2007). Offline Chinese handwriting recognition: an assessment of current technology. *Frontiers of Computer Science in China*. 1(2), 137–155.

Su, T., Zhang, T. and Guan, D. (2007). Corpus-based HIT-MW database for offline recognition of general-purpose Chinese handwritten text. *International Journal of Document Analysis and Recognition (IJDAR)*. 10(1), 27–38.

Tan, G., Sulong, G. and Rahim, M. (2015). Off-Line Text-Independent Writer Recognition for Chinese Handwriting: A Review. *Jurnal Teknologi*. 2, 39–50.

Tan, J., Lai, J., Wang, C. and Feng, M. (2011). A Stroke Shape and Structure Based Approach for Off-line Chinese Handwriting Identification. *International Journal of Intelligent Systems and Application (IJISA)*., 1–8.

Tan, J., Lai, J. and Zheng, W. (2013). Chinese Handwritten Writer Identification based on Structure Features and Extreme Learning Machine. *Proceedings of the 2nd ICDAR International Workshop on Automated Forensic Handwriting Analysis, AFHA 2013*., 2–6.

Tan, J.T.J., Lai, J.-H.L.J.-H., Wang, C.-D.W.C.-D. and Feng, M.-S.F.M.-S. (2011). Off-Line Chinese Handwriting Identification Based on Stroke Shape and Structure. *2010 2nd International Conference on Information Engineering and Computer Science*. (March), 1–4.

Tang, X. (1998). Texture information in run-length matrices. *IEEE Transactions on Image Processing*. 7(11), 1602–1609.

Tang, Y., Bu, W. and Wu, X. (2014). Text-Independent Writer Identification Using Improved Structural Features. , 404–411.

Tang, Y., Wu, X. and Bu, W. (2013). Offline text-independent writer identification using stroke fragment and contour based features. In *Proceedings - 2013 International Conference on Biometrics, ICB 2013*.

Tonghua Su, T.Z. (2006). HIT-MW Dataset for Offline Chinese Handwritten Text Recognition. *Proceedings of the Tenth International Workshop on Frontiers in Handwriting Recognition*.

Vijayalaxmi, M.B. and Dhandra, B. V (2015). Writer based Handwritten Document Image Retrieval. , 30–34.

Wahyono, Kurnianggoro, L., Hariyono, J. and Jo, K.H. (2014). Similarity-based classification of 2-D shape using centroid-based tree-structured descriptor. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 8482 LNAI(PART 2), 279–288.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T. and Gong, Y. (2010). Locality-constrained linear coding for image classification. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp.3360–3367.

Wang, Y., Zhang, D. and Luo, W. (2012). Writer Identification Based on the Distribution of Character Skeleton. *Proceedings of international Conference on Teaching and Computational Science*. 117, 305–309.

Weiss, G.M. and Provost, F. (2003). Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*. 19, 315–354.

Wen, J., Fang, B., Chen, J., Tang, Y. and Chen, H. (2012). Fragmented edge structure coding for Chinese writer identification. *Neurocomputing*. 86, 45–51.

Weszka, J.S., Dyer, C.R. and Rosenfeld, A. (1976). A Comparative Study of Texture Measures for Terrain Classification. *IEEE Transactions on Systems, Man and Cybernetics*. SMC-6(4), 269–285.

Wilson, D.J.H., Irwin, G.W. and Lightbody, G. (1997). A step towards the use of writer's properties for text recognition. *IEEE Colloquium on Faults Diagnosis in Process Systems*. 1/5-5/5(MAY), 20080258.

Wu, X., Tang, Y. and 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.

Y Zhu, TN Tan, Y.W. (2001). Writer identification based on texture analysis. *Acta Automatica Sinica*. 27(2), 229–234.

Y Zi-hua, W Min, L.J. (2004). Handwriting Identification System Based on Texture Analysis. *Journal of Hunan Institute of Engineering*.

Zhu, Y., Tan, T. and Wang, Y. (2000). Biometric personal identification based on handwriting. *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*. 2, 797–800.

Zimmermann, M. and Bunke, H. (2002). Automatic Segmentation of the IAM Off-