Offline Text-Independent Chinese Writer Identification Using GLDM Features

Gloria Jennis Tan[1], Ghazali Sulong[1,2] and Mohd Shafry Mohd Rahim[1]
[1]*Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia.*
[2]*School of Informatics and Applied Mathematics,*
Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.
jtgloria@gmail.com

***Abstract*—This paper presents a method using retrieval mechanism along with Gray-Level Difference Method (GLDM) feature extraction, an approach based on the textural features which is firstly introduced for off-line, text-independent Chinese writer identification. A widely used performance evaluation database HIT-MW has been used for conducting the experiment. An improvement in the identification rates has been revealed in the experimental evaluations by decreasing the search space using a writer retrieval mechanism prior to identification.**

***Index Terms*—Chinese Handwriting Identification; Text-Independent Writer Identification; Writer Recognition; Writer Retrieval.**

I. Introduction

Writer identification is being studied over the past few years. Biometric recognition [1], personalized handwriting recognition systems [2], automatic forensic document examination [3], classification of ancient manuscripts [4], and smart meeting rooms [5] are various applications in which writer identification has applied. It is defined as a behavioral handwriting-based recognition modality which proceeds by matching unknown handwritings against a database of documents with known writers and is considered as a promising research area today.

Researchers in the field of off-line Text-Independent Writer Recognition have mainly focused on the Writer Identification without Retrieval Mechanism approach till now. Therefore, the aim of this paper is to initially introduce the writer retrieval mechanism to off-line text-independent Chinese writer identification, putting a focus on data heterogeneity and human interpretability of the results by forensic experts in order to reduce the time-consuming when dealing with large databases languages.

In this article, Section 2 gives a brief description of the current state of the art between writer identification without and with writer retrieval. Section 3 describes the methodology of this approach. Section 4 elaborates the experimental results and analysis including a brief description of the databases which has been used in our study. Finally, a short conclusion is given in Section 6.

II. Related Works

Determination of the writer of one document among a number of known writers where at least one sample is known is the objective of writer identification. There is need of building a database in advance with specific features for every writer which can used in identification of a new text features by comparing it to the ones already stored in the database. The writer of the new text is assigned to then one having maximum similarity of text in the database. On the other hand, in writer retrieval all documents are searched for one specific writer by generating a ranking of the similarity of handwriting in a dataset. The processes of writer identification and writer retrieval have been depicted in Figure 1.

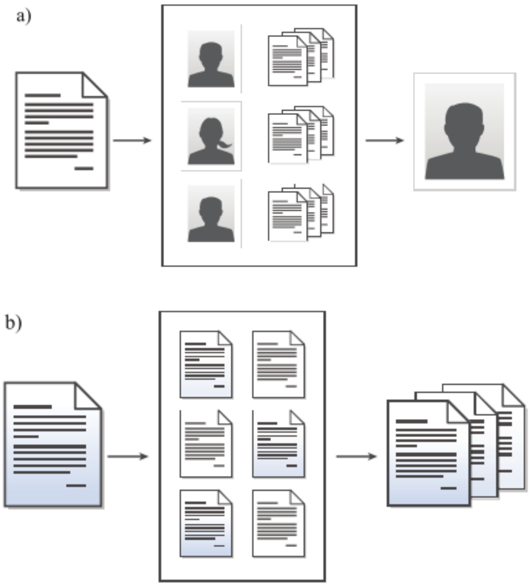

Figure 1: a) illustrates the task of writer identification. A new document is compared with all documents in the database and the result is the identity of the writer with the smallest difference. b) shows the task of writer retrieval. A new document is compared with all documents and a ranking of the similarity of the handwriting is generated [6]

If the document repositories of the forensic experts lack comprehensive writer classification metadata or if many handwriting datasets need to be merged or networked then, they need writer retrieval. Due to continuously increase in the handwritten documents in the digital form, the interest for the writer retrieval scenario has been scaled up.

In the upcoming section, we summarize selected related work on text-independent writer identification, which are broadly categorized into A) writer identification without retrieval and B) writer identification with retrieval. Our main focus is on writer identification with retrieval mechanism.

A. Writer Identification without Retrieval Mechanism

There has been a significant amount of research done on writer identification during last few decades. Most of the work is based on Latin script which began to evolve in 1990s. Plamondon et al. [7] has published a comprehensive survey on the research based on automatic writer identification dating before 1989. A survey of subsequent years 1989-1993 is compiled by Leclerc and Plamondonin in [8]. Further in chronological order the ref. [9] is composed of a research survey until the year 2000. Forensic based research on writer identification is compiled in [10] for the era until 2003. More approaches from before 2005 and 2011 are, respectively, reviewed by Schomakerin [11], and Sreeraj and Idiculain [12]. Several comprehensive surveys provide a broad overview of the efforts done in text-independent writer identification across 3 major world languages employed from year 2011 to 2016 has been published. The reader can refer to our recent publication [58, 59].

The methods in this category mainly differ in how the handwriting is segmented into graphemes and how the graphemes are clustered. On the other hand, feature-based approaches compare the handwriting samples according to geometrical [13], structural [14], or textural features [15, 16]. When a limited amount of handwriting data is available then feature-based approaches become efficient and are generally preferred. Current attempts in writer identification without retrieval mechanism push the performance by increasing complexity, either by combining already complex descriptors [17, 18], or introducing additional (pre)processing steps and heuristics [19, 20].

B. Writer Identification with Retrieval Mechanism

A new research interest of the document analysis and recognition community to improve writer identification methods has emerged due to the result of potential applications of writer identification. New research avenues have opened as a result of recent advancements in writer identification. Integration of writer identification and writer retrieval mechanism is one of its fruits and with the continuous progress and increasing security requirements for the developing the modern society it has become an important field of research.

There are numerous proposed approaches available on literature for document image retrieval and writer identification. To optimize document management and sustainability a system has been designed by Pirlo et al. [21] for layout based document image retrieval to retrieve waybills, commercial forms as invoices and receipts. A morphologic filtering technique and Radon Transform is used in it for layout description and Dynamic Time Warping of document image matching. Using Canberra distance, 100% precision accuracy has been achieved by Shirdhonkar et al. [22] by employing Counterlet Transform based features for Handwritten document image retrieval. an item content image retrieval system has been proposed by Daramola et al. [23]. In the form of texture and shape two image content attributes were extracted. By decomposition of images using Haar wavelet transform low level features were extracted. before the extraction of phase congruency feature, images at detailed bands were divided into non-overlapping blocks. Retrieval of images based on query images and fused feature were achieved using Support Vector Machine (SVM) and image shape and texture respectively. A set of 10 features extracted from the probability density function of the orientations of the writing contours for retrieving the documents belonging to the same writer was used by Atanasiu et al. [24]. IAM database was used for conducting the experiment. A method for writer identification, by recasting the traditional information retrieval (IR) problem of finding documents based on the frequency of occurrence of particular character shapes (allographs) was proposed by Ralph Niels et al. [25]. For small query documents containing only 10 characters, Top-1 performances of almost 60% were achieved alongwith perfect top-1 writer identification rates for bigger databases. Based on shape features extracted using gradient operators like Robert, Sobel, Prewitt and Canny, image retrieval methods were proposed by Ajinkya et al. [26] by using a database of 1000 variable sized images spread across different categories. Higher rate of precision and recall has been shown by Robert gradient operator based retrieval method. Based on local feature extraction using Speeded Up Robust Features (SURF), feature indexing and geometric verification of documents, a method for document image retrieval has been proposed by Rajiv Jain et al. [27] and it used Complex Document Information Processing (CDIP) Tobacco dataset. For writer identification Dhandra et al. [28, 29] used the properties of GLCM of input handwritten document images. In another article [28] they extracted a set of texture features based on correlation-homogeneity properties of gray level co-occurrence matrices of the input handwritten document images of Roman, Kannada and Devanagari writers. Accuracies above 80% for writer identification in documents of single, two and three scripts written by the same 100 writers have been obtained by them. Correlation has shown higher writer identification rate in another article [29] among four properties of GLCM namely Correlation, Contrast, Energy and Homogeneity, when single property is used. Hence it may be used as a potential property in writer identification problems.

Writer retrieval was introduced for the first time based on the selection of all documents authored by a writer [24] by extending the writer identification task in the specialty literature. The retrieval is based on a set of ten features correlated to perceived characteristics of the writing (orientation, regularity…). These features are derived from the probability density function of orientations of the writing contour. When best feature of a query is used then upper limit of performance is obtained according to an evaluation conducted on publicly available IAM database. Finding an automatic process to select the best feature for a given query will be considered in future work. Using pdf features for the whole writers' population may be used to perform this. Performance might also be improved by combining several features [18,30]. Particularly, a two-step sequential combination of global and local features to improve writer identification performance have been proposed by Siddiqi et al. [31]. It is a texture based global analysis which initially does a broad classification of writings then followed by the use of local features to identify the writer of the query document.

Incessantly, a writer identification system optimized with a retrieval mechanism has also been proposed [32]. There are two main steps in this system. Initially, a query document is given to the writer retrieval system which retrieves its Top-N nearest neighbours by comparing it with all the documents in the database. Following that, the query document is compared only with the Top-N documents returned by the retrieval system in the next step. Though, a pre-classification of writings is not performed. Instead, to improve the overall system performance, a retrieval mechanism as pre-processing stage to the writer identification system is intergraded by the authors. In order to characterize the writing style of the writer, it uses the probability distributions of run-lengths [1, 33] and edge-hinges [34] as global features. One of these features is used for writer retrieval and the other for identification in two different scenarios of evaluation. To combine both local and global features for producing more reliable classification accuracy and carrying out some experiments with greater databases containing samples from different scripts, an integrated system will be considered. In order to reject any writer which is not a part of our databases, the proposed system can also be extended to include a rejection threshold.

An approach based on textural features using local features for writer retrieval has been proposed by Fiel and Sablatnig [6], which can also be used for writer identification and therefore making the proposed method independent of binarization step. Local features of the image are calculated initially and an occurrence histogram can be created with the help of a predefined codebook. To determine the identity of the writer or the similarity of other handwritten documents this histogram is compared. The writer identification is achieved by analysing the style of the characters in forensics. It is compulsory for this analysis to separate the foreground from the background in the images, making results of the writer identification dependent on the binarization algorithm. Additionally, these algorithms do not function properly with faded out or blurred ink. While no such separation is required using textural features for writer identification. Thus, a method without this binarization step is proposed as wrong binarization leads to wrong features at character level. The need of more text for the identification is its drawback. The IAM dataset containing 650 writers and the TrigraphSlant dataset containing 47 writers are the two datasets on which the proposed method has been evaluated. The experiments have shown that the proposed method can keep up with previous approaches for writer identification and outperforms previous work for writer retrieval.

Based on their previously work [6] using local textural features and on the writer identification from Gordo et al. [35] using a Bag of Words (BoW) approach, a writer identification and writer retrieval method has been proposed by Them [36]. Not only to identify the writer but to find all documents written by the same writer as a reference document is the main goal of this approach. By using the Scale Invariant Feature Transform (SIFT) from Lowe [37], local features are computed on the normalized image. To calculate a feature vector of each document image the Fisher kernels, which were presented by Perronnin and Dance [38] and improved by Perronnin et al. [39] are applied to the visual vocabulary. To determine the similarity between two handwritings and for writer identification and writer retrieval this vector is then used. Contrary to their previous method proposed in [6], this approach is different from the feature to the center of one cluster into account. ICDAR 2011 Writer Identification Contest dataset consisting of 208 documents from 26 writers, and the CVL Database containing 1539 documents from 309 writers are the two databases on which the proposed method is evaluated. Experiments show that the proposed method performs has shown to perform slightly better than previous writer identification methods. Since, it is important to get all documents in writer retrieval which is written from the same writer as a reference document, so, a new criterion is introduced to evaluate writer retrieval. The correct documents percentage in the top N from the ranking is calculated. The proposed scheme has been evaluated on two different datasets and it can keep up with the state of the art techniques and its performance for writer retrieval tasks is slightly better. The use of different pens, which changes the style of the writing for each person are the challenges for writer identification and writer retrieval. It is based on the fact if the writer was in hurry or not while writing the text, and also that one word is infrequently written twice exactly the same way. The handwriting of a person may change over a period of time making the identification a harder task is another challenge, which is not covered by any database. Independent of binarization algorithm and segmentation is the main advantage of this approach but still need power normalization.

A new approach is proposed for texture based handwritten document image retrieval based on writer using directional multiresolution property of DWT and features of correlation of GLCM along four directions and five distances of image.

A new global approach is proposed [40] for texture based handwritten document image retrieval based on writer using unique directional multiresolution property of conventional two-dimensional (2-D) discrete wavelet transforms (DWTs) and features of correlation of GLCM is used to extract spatial features along four directions and five distances of image. Handwritten documents from 100 writers each in English, Kannada and Hindi scripts are collected. 2000 image blocks of each script writers are used separately for validation of the proposed method after the segmentation of these handwritten documents into image blocks. Using City Block measure for English and Kannada documents, Euclidean measure for Hindi documents, the document matching using Euclidean and City block distance measures is achieved with higher retrieval results. The retrieval results are high for writer document retrieval in each English, Kannada and Hindi script document writers, although the content of text in each image block is very less. In the process of document image retrieval, the features based on discrete wavelet transform and correlation property of GLCM of input handwritten document image have given encouraging results through exhaustive experimentation and hence, in the proposed writer document image retrieval method these features of text are extracted.

To generate a feature vector for each writer, Convolutional Neural Networks (CNN) is used in the latest method [41]. Then comparison is done with the pre-calculated feature vectors already stored in the database. CNN is trained on a database with known writers to generate this vector and after that classification layer is cut off and second last fully connected layer's output is used as feature vector. To identify a writer or the retrieval of similar writers these vectors are used implying a nearest neighbour approach. However, an image with fixed size is required by CNNs as input, thus, it is necessary to pre-process the document images. Binarization, sliding windows and text line segmentation are included in pre-processing. ICDAR2013 Competition on Writer Identification, ICDAR 2011 Writer Identification Contest, and the CVL-Database datasets are three datasets on which the proposed method has been evaluated. Slightly better results on two of three datasets, but worse results on the remaining dataset originating mainly from the pre-processing steps have been demonstrated by the evaluation of proposed method. Designing a new CNN customized to the input data and capable of attaining better performance on several datasets is included in future work. Furthermore, a better normalization of the image patches will be used to improve pre-processing step and a voting strategy on complete page and the rejection of insignificant image patches for improvement of post-processing step will be introduced. Also, if some image patches do not show any relevant information they may be skipped during pre-processing step for successful writer identification and writer retrieval.

Based on the same idea, we propose a writer retrieval approach based on textural features, which can also be used for writer identification.

III. Methodology

Idea of the search space reduction before an identification task is the inspiration for our research. With an increase in size of the database, there in a decrease in accuracy of the writer identification systems [34]. Hence, for large-scale identification systems this deterioration can be very significant. In such cases, to reduce the search space for a writer identification system, a writer retrieval mechanism whose objective is to retrieve all the documents written by a specific writer from the database may be used.

The technique (see Figure 2) consists of sequentially executed two main phases: (1) First phase – At this stage, the method proposes a first matching to separate out dissimilar images and select only the images that closely resemble the query image. The shortlist is then used as the input to the second phase to select the closest matching handwriting image; (2) Second phase – At this stage, the second matching phase determines the closest match from the shortlist rather than the entire dataset that nearly resembles the query image. In simple words, the writer retrieval task serves as a filtering step before the identification task, when dealing with large databases.

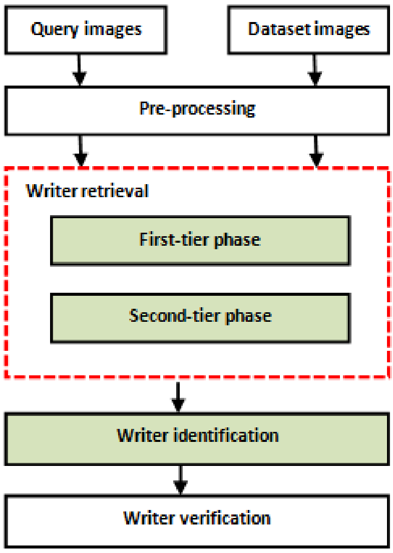

Figure 2: A block diagram of the methodology

The method involves three components: (1) Feature vector formation from query image; (2) Feature vector formations from dataset images; (3) Computation of Euclidian Distances and shortlisting. The different steps for both methods are now described in more detail.

A. *Writer Retrieval*

All documents feature in the dataset need to be generated for the writer retrieval in order to compare the query document with every document in the dataset. In feature space, an ordered list of relevant documents (regards as the query content) is obtained by the similarity measure between query and each document. The retrieval phase is concerned with the calculation of relevance score of each document for a certain query.

At this stage, SLT transformed image was inversed white background of page image becomes black and black ink of handwriting becomes white but not as an operation simply making the black pixels white and making white pixels black. Inverse image is giving good result. Some feature extraction techniques give better results on image inverse that original image. Next, Sober edge detection is computed to inverse images to get edge information for feature extraction. Feature from edge is extracted using Local Binary Pattern (LBP). [47] [48] used texture features based on LBP, where the features are directly extracted from each overlapping block. The image is partitioned into several overlapping circular blocks to extract the feature vectors using LBP. To describe the local structure of images, a dense local texture descriptor LBP can be used [49]. Usually, in LBP feature extraction a comparison of each central pixel with its neighbors is done for creating an image of integer valued code, then these codes are pooled into histograms. Instead of whole text LBP feature is extracted only from the edge points because Contour point (center point) contains much individuality information of the writer. But, it will take longer time for computation if LBP feature is extracted from the whole text lines, as it will contain many stoke points [52]. So, rapid identification algorithm can be achieved if only contour points are used to extract these features without losing identification accuracy. There is no need of extracting local or global features from the whole image as majority characteristics of writing style lies in contour points.

These feature vectors are then arranged into a matrix to determine the similar blocks and sorted the matrix lexicographically to reduce the computation complexity. Finally, a set of features is extracted and compared with that of the shortlist to find the closest match. For each dataset image, a distance is computed using Euclidean Distance to differentiate between image and query image.

Following that, the second phase determines the closest match from the shortlist rather than the entire dataset that nearly resembles the query image. At this stage, the Hierarchical Centroids (HC) is applied to extract a set of features and compared with that of the shortlist to find the closest match. The method is based on recursive subdivisions of input binary image by measuring centroids at each division and outputs a fixed length features vector. Further, feature vector is normalized according the size of the input image. Therefore, the size of images does not affect the final feature set. Illumination invariant and insensitive to scaling are some interesting properties of these Hierarchical Centroids. Figure 3 illustrates Hierarchical Centroids (HC) of image pixels method for feature extraction which does not require fixed size.

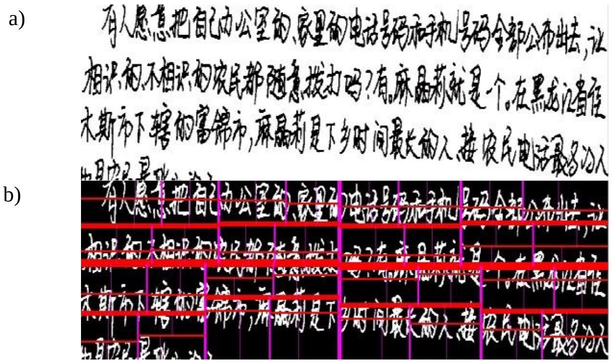

Figure 3: a) input image. b) Hierarchical Centroids (HC) of image pixels

B. *Writer Identification*

In this phase, the texture coarseness or fineness of an image can be interpreted as the distribution of the elements in the matrix in Gray Level Difference Method (GLDM). GLDM proposed by [55] is similar to the co-occurrence matrices. Grey level difference method probability density functions for the pre-processed grey image are calculated by GLDM process. Extraction of statistical texture features of a digital image is done by using this method. Entropy, contrast, angular second moment, mean and inverse difference moment are the five texture aspects outlined from each density functions. Each represents a different probability density function and each is a vector with 256 elements. In our experiments, we used the first probability density function because it was giving best results.

By computing the distance between the query image Q and all the images in the training data set, writer identification is performed, and the writer of Q is identified as the writer of the document with the smallest distance. This relates to the nearest neighbour classification (KNN with k=1). For a query document, not only the nearest neighbour (Top-1) is found, but a longer list up to a given rank (Top-K) which simultaneously increases the chance of finding the correct writer in the retrieved list.

IV. Experimental Evaluations

Experiment performed to confirm the effectiveness of the proposed features for writer identification and retrieval are presented in coming section. In start of this section, a brief description of the databases used in our study has been provided. Metrics used for retrieval and identification are discussed in section B while the performance without retrieval task is presented in section C. Section D presents the identification result with retrieval mechanism; next a short conclusion and contribution for further future work opportunity.

A. *Database*

The database used in the experiments is HIT-MW Chinese database [56, 57]. This database is the first collection of Chinese handwritten texts in hand writing recognition domain with 300 dpi resolution scanned handwritings. Although the basic purpose of this database is the facilitation of the fundamental study on off-line Chinese handwriting recognition, but, other research directions such as writer identification and real text-line segmentation also use this database. There are 853 handwriting Chinese samples, in which 254 images from 241 writers are labelled by ID make the dataset of HIT-MW database. Only one page of handwritten text belongs to most of the writers. In our experiments, the HIT-MW dataset has been modified to form a new dataset. Only one page for each writer is used for the 241 writers' labeled images and there are two commensurate parts of each page. Figure 4 illustrates this division which form a new dataset, called L-HIT-MW. In this way, two samples handwriting from one writer can be obtained, first image is used for enrolment (verification) or training (identification) and the second on is used for verification or testing (identification). For identification and verification of the proposed method, the segmentation of these handwritten documents into image blocks is done to use them separately.

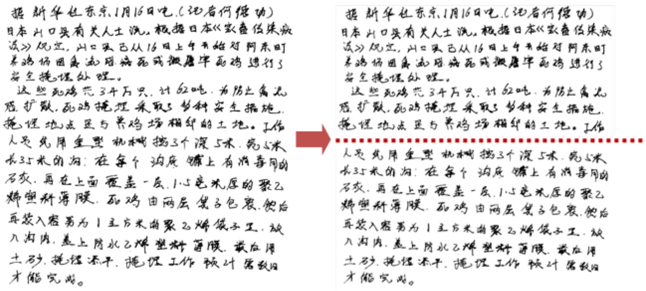

Figure 4: One page for each writer is used and each page is segmented into two commensurate parts.

B. *Evaluation Metrics*

Soft TOP-N and the hard TOP-N criterion are the two criteria defined to evaluate the accuracy of the submitted methodologies. Correct hits are counted for all document images of the benchmarking dataset. A correct hit for the Soft TOP-N criterion is considered when at least one document image of the same writer is included in the N most similar document images. Concerning the hard TOP-N criterion, consider a correct hit when all N most similar document images. Whereas, a correct hit for the hard TOP-N criterion is considered when all N most similar document images are written by the same writer. But in our study soft TOP-N is used to evaluate the performance of all participating algorithms.

The performance of writer identification, the TOP-N accuracy is the quotient of the total number of correct hits to the total number of the document images in the benchmarking dataset. The report of Top 1, Top 5 and Top 10 identification rates is involved in every evaluation. For every document in the database, Euclidean distance metric is used to calculate the distance among all documents in this database. A ranked list starting from the most similar to the least similar document image will be generated and the evaluation will measure the accuracy in terms of Top 1, Top 5 with the Top 5 writers in the list, and Top 10 with the query document within the top 10 writers in the list.

The identification results are presented in the following section. Two different experiments (scenarios) are considered; in the first one, identification without writer retrieval in Table I, while in the second the inverse is employed which is writer identification with writer retrieval in table II. Since the number of pages per writer are not equally distributed in the HIT-MW dataset, so, checking of total 241 documents in the rankings is carried out. Only one page for each writer is used and there are two commensurate parts of each page after segmentation. First image is used for enrolment (verification) or training (identification) and the second on is used for verification or testing (identification).

C. *Writer Identification without Retrieval Mechanism*

The performance of SLT and Hierarchical Centroids features independently on writer identification without retrieval mechanism is presented in this section. Chinese database is used for conducting the experiments. Table I shows the summary of the Top1 identification rates, the results corresponding to the highest accuracy are marked bold. It is very clear from Table I, that identification rates vary: from 92% to 66.4% for the SLT features independently, from 68% to 36.9% for the Hierarchical Centroids features independently and from 92% to 65.1% for GLDM features independently. We can conclude that the accuracy of the writer identification is found to decrease for larger databases with writers increasingly.

Table 1
Identification Performance without Retrieval Mechanism

# of writers	SLT features			Hierarchical Centroids features			GLDM features		
	Top1	Top 5	Top10	Top1	Top5	Top10	Top 1	Top 5	Top 10
50	92	100	100	68	100	100	92	100	100
100	84	100	100	60	100	100	83	100	100
150	73.3	100	100	45.3	93.3	98.7	74	100	100
200	70	99	100	41	89	98.5	69	100	100
241	66.4	98.8	100	36.9	84.2	97.5	65.1	99.1	100

D. *Writer Identification with Retrieval Mechanism*

The impact of integrating a retrieval mechanism in a writer identification system is aimed to study in the final series of experiments. Table II summarizes these results. The primary step of writer retrieval is based on texture features (SLT and Hierarchical Centroids features) as described before. Euclidean distance metric is used to compare the features. In this stage, a small subset of N documents (N chosen to be 60 in our case) with maximum similarity to the query is selected and all others are rejected. Therefore, this step serves as a filter excluding more than 87% (from 241 to 181 documents) of handwritten documents in the database and considering only less than 25% of the documents (60 documents) for the next step.

In the second step, GLDM is used to compare every returned document by the retrieval step with the query. The documents with increasing distance to the query are stored and those having minimum distances are supposed to be written by the same writer as that of the query image.

An improvement in the identification rates by integrating the retrieval mechanism in the writer identification scheme can be seen after comparing with the performance of individual features (Table I). It is clear from achieved identification rates that integration has resulted in a rise in the identification rates from 66.4% (the best score) to 88.8% for Chinese database.

The proposed idea in this paper i.e., a writer identification system can be optimized when the query document is compared with a top few retrieved documents (returned by a writer retrieval system) rather than the entire database is clearly supported by the experimental results reported in this section.

Related to the previous study, an author [40] proposed feature extraction based on spatial domain and combine it with DWT to enhance the outcomes of image handwriting retrieval and obtain high retrieval performance. However, this study proposed the use of SLT instead of DWT to bring out fine details prior to feature extractions. To the best of the author's knowledge, this is the first time such transform is used in handwriting images to extract features. Secondly, to our best knowledge, writer retrieval mechanism is started in research to English language but so far not in Chinese & Arabic language. This is the first attempt to bring writer retrieval mechanism to the field of Chinese writer identification.

Table 2
Identification Performance with Retrieval Mechanism

# of writers	Proposed method (with retrieval mechanism)		
	Top1	Top5	Top10
50	84	100	100
100	88	100	100
150	85.3	100	100
200	85	100	100
241	88.8	100	100

V. CONCLUSION

A writer identification scheme based on a retrieval mechanism has been proposed in this article along with reduction of search space of the identification process. We used new approach using SLT instead of DWT to bring out fine details prior to feature extractions. However, SLT requires that image is partitioned into a number of block sizes. We introduced Hierarchical centroid of image pixels method for feature extraction which does not require fixed size as compared to existing work. The main advantage of using this method is that it gives size invariant feature vector. We also introduced GLDM features require that image as a whole with no divisions.

A contribution, this is the first time such features extraction methods (SLT, Hierarchical Centroid and GLDM) are used in handwriting images to extract features in writer identification scheme. Secondly, to our best knowledge, writer retrieval mechanism is started in research to English language but so far not in Chinese & Arabic language. This is the first attempt to bring writer retrieval mechanism to the field of Chinese writer identification.

Currently, the extraction of global features is the main focus of our work, further focus will be on the use of local features as future work. In order to produce more reliable classification accuracy an integrated system i.e. combination of both local and global features will be considered. Some experiments with larger databases containing samples from various scripts are being conducted. Also, the proposed system can be extended to a system which will be able to incorporate variable sizes as compared to existing work.

REFERENCES

[1] D. Chawki and S. M. Labiba, "A texture based approach for arabic writer identification and verification," in *2010 International Conference on Machine and Web Intelligence, ICMWI 2010 - Proceedings*, 2010, pp. 115–120.

[2] A. Nosary, L. Heutte, and T. Paquet, "Unsupervised writer adaptation applied to handwritten text recognition," *Pattern Recognit.*, vol. 37, no. 2, pp. 385–388, 2004.

[3] L. S. M. Van Erp, L. Vuurpijl, K. Franke, "The WANDA measurement tool for forensic document examination," *J. Forensic Doc. Exam.*, no. 16, pp. 103–118, 2005.

[4] I. Siddiqi, F. Cloppet, and N. Vincent, "Contour based features for the classification of ancient manuscripts," in *Conference of the International Graphonomics Society*, 2009, pp. 226–229.

[5] M. Liwicki, A. Schlapbach, H. Bunke, S. Bengio, J. Mariéthoz, and J. Richiardi, "Writer identification for smart meeting room systems," in *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2006, vol. 3872 LNCS, pp. 186–195.

[6] S. Fiel and R. Sablatnig, "Writer retrieval and writer identification using local features," in *2012 10th IAPR Int. Work. Doc. Anal. Syst.*, 2012, pp. 145–149.

[7] R. Plamondon and G. Lorette, "Automatic signature verification and writer identification—the state of the art," *Pattern Recognit.*, vol. 22, no. 2, pp. 107–131, 1989.

[8] F. Leclerc and R. Plamondon, "Automatic signature verification: the state of the art --1989-1993," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 8, no. 3, pp. 643–660, 1994.

[9] R. Plamondon and S. N. Srihari, "On-line and off-line handwriting recognition : a comprehensive survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 63–84, 2000.

[10] C. F. Romero, C. M. Travieso, J. B. Alonso, and M. A. Ferrer, "Writer identification by handwritten text analysis," in *Proc. 5th WSEAS Int. Conf. Syst. Sci. Simul. Eng.*, 2006, pp. 204–208.

[11] L. Schomaker, "Advances in writer identification and verification," in *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, vol. 2, pp. 1268–1273, 2007.

[12] M. Sreeraj and S. M. Idicula, "A survey on writer identification schemes," *Int. J. Comput. Appl.*, vol. 26, no. 2, pp. 23–33, 2011.

[13] S. Al-Ma'adeed, E. Mohammed, D. Al Kassis, and F. Al-Muslih, "Writer identification using edge-based directional probability distribution features for Arabic words," in *AICCSA 08 - 6th IEEE/ACS Int. Conf. Comput. Syst. Appl.*, 2008, pp. 582–590.

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

[15] K. Franke, O. Biinnemeyer, and T. Sy, "Ink texture analysis for writer identification," in *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, 2002, pp. 268–273.

[16] H. E. S. Said, T. N. Tan, and K. D. Baker, "Personal identification based on handwriting," *Pattern Recognition*, vol. 33, pp. 149–160, 2000.

[17] R. Jain and D. Doermann, "Combining local features for offline writer identification," in *Proc. Int. Conf. Front. Handwrit. Recognition, ICFHR*, 2014, pp. 583–588.

[18] M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and allographic features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 701–17, Apr. 2007.

[19] X. Wu, Y. Tang, and W. Bu, "Offline text-independent writer identification based on scale invariant feature transform," *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 3, pp. 526–536, 2014.

[20] R. Kumar, B. Chanda, and J. D. Sharma, "A novel sparse model based forensic writer identification," *Pattern Recognition Lett.*, vol. 35, no. 1, pp. 105–112, 2014.

[21] G. Pirlo, M. Chimienti, M. Dassisti, D. Impedovo, and A. Galiano, "A layout-analysis based system for document image retrieval," *Mondo Digit.*, vol. 13, no. 49, 2014.

[22] M. S. Shirdhonkar and M. B. Kokare, "Handwritten document image retrieval," *Int. J. Model. Optim.*, vol. 2, no. 6, pp. 693–696, 2012.

[23] S. A. Daramola, A. Abdulkareem, and K. J. Adinfona, "Efficient Item Image Retrieval System," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 4, no. 2, pp. 109–113, 2014.

[24] V. Atanasiu, L. Likforman-Sulem, and N. Vincent, "Writer retrieval - Exploration of a novel biometric scenario using perceptual features derived from script orientation," in *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, 2011, pp. 628–632.

[25] R. Niels, F. Gootjen, and L. Vuurpijl, "Writer identification through information retrieval: The Allograph Weight Vector," in *Int. Conf. Front. Handwrit. Recognition*, 2008, pp. 481–486.

[26] A. P. Nilawar and A. Prof, "Image retrieval using gradient operators," *International Journal on Recent and Innovation Trends in Computing and Communication.*, vol. 2., no. 1, pp. 2–5, 2014.

[27] R. Jain, D. W. Oard, and D. Doermann, "Scalable ranked retrieval using document images," in *Proc. of SPIE-IS&T Electronic Imaging*, 2013, pp. 90210K-1-90210K-15.

[28] B. V. Dhandra and M. B. Vijayalaxmi, "Text and script independent writer identification," in *Proc. 2014 Int. Conf. Contemp. Comput. Informatics, IC3I 2014*, 2014, pp. 586–590.

[29] V. M., B. V. Dhandra, "Text independent writer identification for tamil script," in *Natl. Conf. Adv. Mod. Comput. Appl. Trends, AIT, Bangalore,* 5-6 Dec 2014, pp. 8-15

[30] V. Atanasiu, "Allographic biometrics and behavior synthesis," in *EuroTeX 2003 Proc.*, vol. 24, no. 3, 2003, pp. 328–333.

[31] I. Siddiqi and N. Vincent, "Combining global and local features for writer identification," in *Proc. of the 11th International Conference on Frontiers in Handwriting Recognition*, 2008, pp. 48 – 53.

[32] C. Djeddi, I. Siddiqi, L. Souici-Meslati, and A. Ennaji, "Multi-script writer identification optimized with retrieval mechanism," in *2012 Int. Conf. Front. Handwrit. Recognition*, Sep. 2012, pp. 509–514.

[33] X. Tang, "Texture information in run-length matrices," *IEEE Trans. Image Process.*, vol. 7, no. 11, pp. 1602–1609, 1998.

[34] M. Bulacu, *Statistical Pattern Recognition for Automatic Writer Identification and Verification*. s.n., 2007. 140 p.

[35] A. Gordo, A. Fornés, E. Valveny, and J. Lladós, "A bag of notes approach to writer identification in old handwritten musical scores," in *Proc. 8th IAPR Int. Work. Doc. Anal. Syst. - DAS '10*, 2010, pp. 247–254.

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

[37] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[38] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[39] F. Perronnin, J. Sánchez, and T. Mensink, "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)*, 2010, vol. 6314 LNCS, no. PART 4, pp. 143–156.

[40] M. B. Vijayalaxmi and B. V Dhandra, "Writer based Handwritten Document Image Retrieval," in *Advances in Digital Image Processing and Information Technology*, D. Nagamalai, E. Renault, and M. Dhanuskodi, Eds. Springer, 2015, pp. 30–34.

[41] A. Nigam, B. Kumar, and P. Gupta, "Writer identification and retrieval using a convolutional neural network," in *Computer Analysis of Images and Patterns. CAIP 2015*, G. Azzopardi, and N. Petkov, Eds. Springer, 2015, pp. 702–714.

[42] S. Bhagavathy and K. Chhabra, "A wavelet-based image retrieval system," *Computer. Engineering*, pp. 1–7, 2007.

[43] N. Suematsu, Y. Ishida, A. Hayashi, and T. Kanbara, "Region-based image retrieval using wavelet transform," in *10th International Workshop on Database and Expert Systems Applications*, 2002, pp. 167173.

[44] X. Xiang and B. Shi, "Evolving generation and fast algorithms of slantlet transform and slantlet-Walsh transform," *Appl. Math. Comput.*, vol. 269, pp. 731–743, 2015.

[45] L. Yang, W. Gao, and Z. Liu, "An improved sobel algorithm based on median filter," in *2nd International Conference on Mechanical and Electronics Engineering (ICMEE 2010)*, vol. 1, 2010, pp. 88–92.

[46] R. Maini, "Study and comparison of various image edge detection techniques," *International Journal of Image Processing (IJIP)*, vol. 3, no. 1, 2009, pp. 1–12.

[47] D. Bertolini, L. S. Oliveira, E. Justino, and R. Sabourin, "Texture-based descriptors for writer identification and verification," *Expert Syst. Appl.*, vol. 40, no. 6, pp. 2069–2080, 2013.

[48] A. Nicolaou, A. D. Bagdanov, M. Liwickit, and D. Karatzas, "Sparse radial sampling LBP for writer identification," in *13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015, pp. 716–720.

[49] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[50] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu, "Boosting local binary pattern (LBP)-based face recognition," in *Advances in Biometric Person Authentication,* S. Z. Li, J. Lai, T. Tan, G. Feng, and Y. Wang, Eds. Springer, 2004, pp. 179–186.

[51] M. B. Yilmaz, B. Yanikoglu, C. Tirkaz, and A. Kholmatov, "Offline signature verification using classifier combination of HOG and LBP features," in *2011 Int. Jt. Conf. Biometrics, IJCB 2011*, 2011.

[52] Z. Fan, Z. Guo, and Y. Chen, "Writer identification using edge based features," in *3rd IAPR Asian Conference on Pattern Recognition (ACPR),* 2015, pp. 416–420.

[53] G. Vamvakas, B. Gatos, and S. J. Perantonis, "Handwritten character recognition through two-stage foreground sub-sampling," *Pattern Recognit.*, vol. 43, no. 8, pp. 2807–2816, 2010.

[54] S. Armon, *Handwriting Recognition and Fast Retrieval for Hebrew Historical Manuscripts*. Hebrew University, 2011.

[55] L. Van Gool, P. Dewaele, and A. Oosterlinck, “Texture analysis anno 1983,” *Comput. Vision, Graph. Image Processing*, vol. 29, pp. 336--357, 1985.

[56] T. Su, T. Zhang, and D. Guan, “Corpus-based HIT-MW database for offline recognition of general-purpose Chinese handwritten text,” *Int. J. Doc. Anal. Recognition*, vol. 10, no. 1, pp. 27–38, Mar. 2007.

[57] T. Z. Tonghua Su, “HIT-MW dataset for offline chinese handwritten text recognition,” in *The 10th International Workshop on Frontiers in Handwriting Recognition.,* 2006.

[58] G. J. Tan, G. Sulong, and M. S. M. Rahim, “Writer Identification: A comparative study across three world major language,” *Forensic Science International*, 2017, in press.

[59] G. Tan, G. Sulong, and M. Rahim, “Off-Line Text-Independent Writer Recognition for Chinese Handwriting: A Review,” *J. Teknol.,* vol. 75, no. 2, pp. 39-50, 2015.