Can Handwriting Style Help Strengthen the Person Identity?

Abdallah Meraoumia
LAMIS Laboratory
University of Larbi Tebessi
Tebessa, Algeria
ameraoumia@gmail.com

Maarouf Korichi
Lab. de Génie électrique
university of KASDI Merbah
Ouargla, Algeria
korichi.maarouf@gmail.com

Chlaoua Rachid
Lab. de Génie électrique
university of KASDI Merbah
Ouargla, Algeria
r.chlaoua@gmail.com

Hakim Bendjenna
LAMIS Laboratory
University of Larbi Tebessi
Tebessa, Algeria
hbendjenna@gmail.com

***Abstract*—Recently, user identification is an essential foundation for protecting information in several applications. However, the need for heightened this identification has expanded the research to focus on the biometric traits of the users. Obviously, one of the biometrics based behavioral characteristics advantage is their simplicity of use as well as their acceptability by the users. For that, among several behavioral characteristics, person handwriting has been systematically used to make identification for last years. Handwriting identification aims to determine which person or writer has written a given text sample. In this paper, we aim to design a Latin script identification system based on Local Phase Quantization (LPQ) and then the obtained results are used to constrict a biometric based person identification system. The proposed method is validated for their efficacy on the available CVL-Database of 310 writers. Our experimental results show the effectiveness and reliability of the proposed method, which brings both high recognition and accuracy rate.**

***Index Terms*—Handwriting recognition, Biometrics identification, Security, Local Phase Quantization, CVL database.**

I. Introduction

GENERALLY, at any language, great information about writer's personality and identity can be provided from their Handwriting text. So, this information can be used for construct an effectiveness system for identifying any writer. Recently, writer identification has become an important and active research area in document analysis and recognition community. In addition, the person handwriting represents a useful biometric technique, which used in a variety of applications covering such as forensics, digital rights management, crime and terrorism. Generally, Writer Identification (WI) based on their handwritten text is a behavioral biometrics due to the fact that personal style of writing is a habit that is learnt and refined from an early age. Thus, one of their advantages is the simplicity of use as well as the acceptability by the users. Handwriting identification aims to determine which person or writer has written a given text sample [1], [2].

The analysis of handwriting and hand-drawn shapes has been an interesting research area for many centuries that has attracted a significant number of psychologists, graphologists, palaeographers and forensic experts to solve a wide variety of problems. Formally, given a set of handwritten documents with known authorship, the writer identification task involves finding the writer of a query document comparing it with the samples in the database [3].

The challenges for (WI) and writer retrieval (WR) include the use of different pens, which changes a person's writing style, the physical condition of the writer, distractions like multitasking and noise, and also that the writing style changes with age. The changing of the style with increasing age is not covered by any available dataset and cannot be examined, but makes the identification or retrieval harder for real life data [4].

Literally, a handwritten based identification or retrieval system can operate into two different way: on-line or off-line handwriting. At this context, this research paper aims to design an off-line handwriting identification/retrieval system. For the design and evaluation of handwriting recognition algorithms and systems, the availability of large-scale, unconstrained handwriting dataset is very important. Among the off-line sample datasets are IAM [5], RIMES [6]. Although most of these databases have been developed using text in languages based on the Latin alphabet, development of databases in Chinese [7], Korean [8], Arabic [9], Farsi [10] and Indian scripts [11] is also on the rise. The trend of multi-script handwritten databases [12] has also been observed in the last few years.

Feature extraction is an essential tasks in any pattern recognition application since a more accurate classification results are directly depend on the choice of the feature extraction techniques. However, the distinctiveness and unevenness of the extracted features used to differentiate between different handwritten images [13]. Thus, the objective of this work is twofold: first, we aim to design a Latin script identification system based on Local Phase Quantization (LPQ). second, the obtained results are used to construct a biometric based person identification system. The proposed method is validated for their efficacy on the available CVL-Database of 310 writers. Our experimental results show the effectiveness and reliability of the proposed method, which brings both high recognition and accuracy rate.

The remaining of this paper is organized as fellow: at the first section, the proposed handwritten identification/retrieval system is presented, nest section deals with the description of used features extraction technique (Multi level Local Phase Quantization ML-LPQ). the criterion's that used to evaluate the proposed handwritten recognition system are discussed in

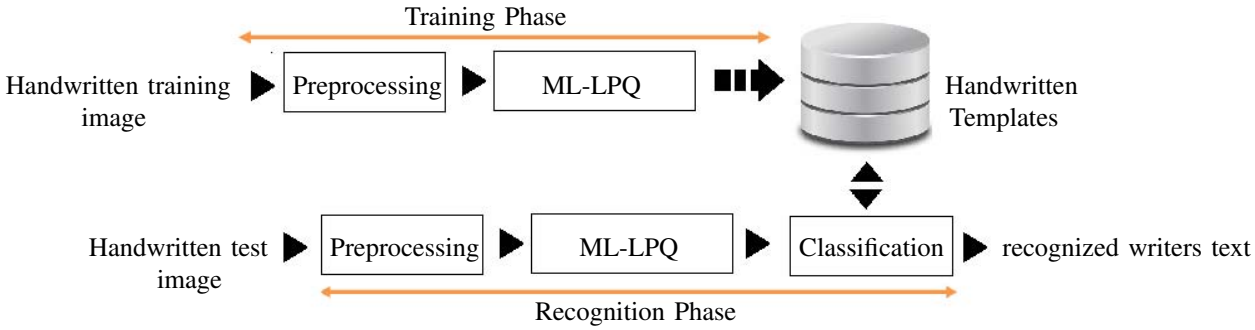


Fig. 1. Block diagram of the writers identification/retrieval system using Latin Handwritten images based on ML-LPQ Features.

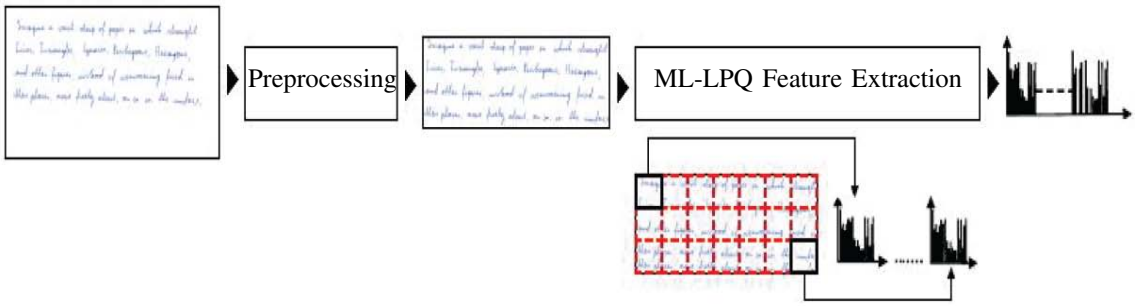


Fig. 2. Block diagram of the handwriting text feature extraction.

section 3 followed by a discussion on the different obtained experimental results. Finally, Section 5 present the paper conclusion with future works.

II. Proposed Methodology

In this section, the description of the proposed Latin handwritten identification/retrieval system is presented where the aim is to develop a system which is capable of identifying/retrieving an unknown handwritten image by providing a likely list of candidates from a known database of writers. The block diagram of the proposed system based on the Latin handwritten script images is shown in Fig.1. For the training phase, the handwritten training images are preprocessed and normalized. After that, the feature vector for each input images is represented by LPQ features.Next, all feature vectors (extracted from all writers handwritten) are used to construct the Handwritten templates database. For the recognition phase, the same feature vector is extracted from the handwritten test image and then it uses as an input vector in order to recognize or to identify the writers of the handwritten document image.

III. Feature Extraction

After a preprocessing phase applied on the text images, a feature extraction phase is necessary to obtain some effective features. The feature extraction module processes the acquired Handwritten data and extracts only the salient information to form a new representation of the data. Ideally, this new representation should be unique for each person. In our work, we propose to use the Multi-Level-LPQ (ML-LPQ) algorithm.

A. Multi Level Local Phase Quantization (ML-LPQ)

The local phase quantization (LPQ) method is based on the blur invariance property of the Fourier phase spectrum. It uses the local phase information extracted using the 2-D DFT or, more precisely, a short-term Fourier transform (STFT) computed over a rectangular M-by-M neighborhood $\aleph_x$ at each pixel position x of the image $f(x)$ defined by [14]:

$$F(u,x) = \sum_{y \in \aleph_x} f(x-y) e^{-j2\pi u^T y} = w_u^T \qquad (1)$$

where w_u u is the basis vector of the 2-D DFT at frequency u, and f_x is another vector containing all M^2 image samples from $\aleph_x$.

the handwritten document image is divided into n sub-blocks where $n = 1; 2; 3; 4; ....$,then the LPQ feature extraction methods are applied to each sub-block, this method is called multi block LPQ (MB-LPQ), in our work, based on the LPQ method, instead of using single MB-LPQ division we use Multi Level Local Phase Quantization (ML-LPQ). The main idea of ML-LPQ is to extract features from different MB-LPQ divisions and then combine them. In other words, extracting features from the whole image, then dividing the image into 2×2 sub-blocks and extracting the features from each sub-block and so on until we reach the intended level. The final result of ML-LPQ is $1^2+2^2+...+n^2$ histograms. We combine these histograms to get the feature vector. Figure. 2 resume the ML-LPQ approach.

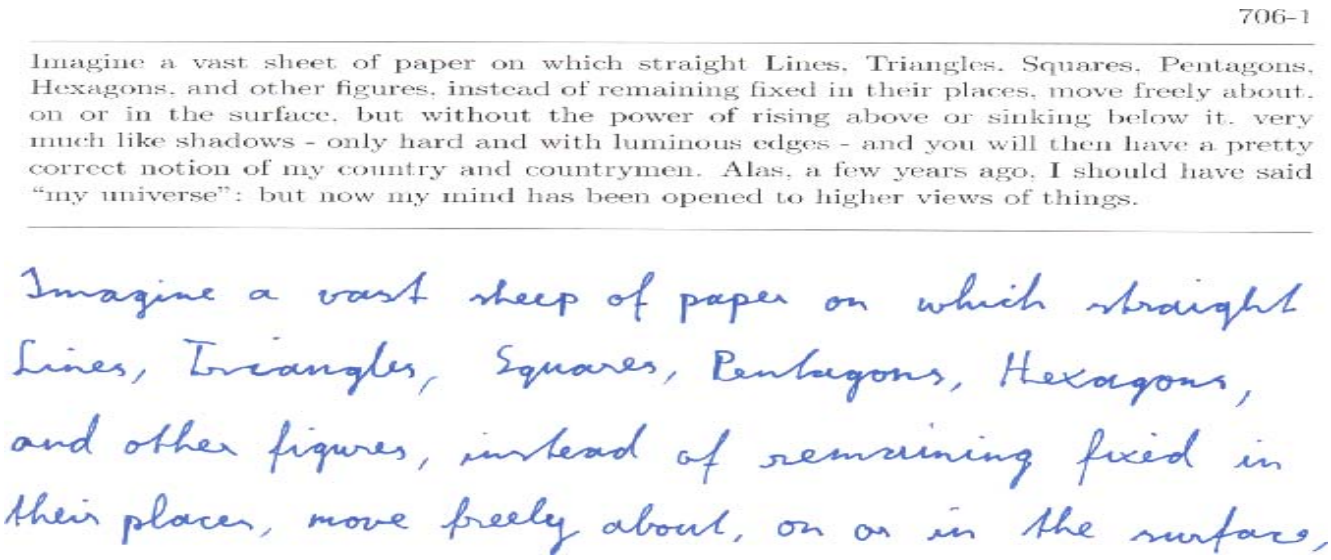


Fig. 3. Sample handwritten text from CVL Database (Writer ID 706,Text # 1).

IV. Performance Evaluation

In this section, the description of the criterion that is used in the evaluation of the performance of the proposed handwritten writer's identification/retrieval system is presented. Thus, For each document, a ranking of the other documents according to the similarity is generated. There the top N documents are examined whether they are from the same writer or not. so, In order to measure the accuracy of the proposed methodologies we use the soft *TOP-N* and the hard *TOP-N* criterion. The results are sorted from the most similar to the least similar document image.

For the soft *TOP-N* criterion, one can defined as the accuracy of at least one of the same writer is included in the N most similar document images. Concerning the hard TOP-N criterion, it is defined as the accuracy of all the N most similar document images are written by the same writer.

The values of N used for the soft criterion are 1, 2, until 10 while for the hard criterion are 1,2,3 and 4. Since we have 5 document images per writer, 4 is the maximum value of N for the hard criterion.

Furthermore, the CVL dataset has a retrieval criterion, which is defined as the percentage of the documents of the corresponding writer in the first N documents. For this criterion the values of N are the same as for the soft criterion.

As it is mentioned previously, this research study has twofold, the second part of this study is the conception of a behavioral biometric system based on a Latin handwritten scripts. So, considering the evaluation of the biometric system the Equal Error Rate (EER) is used.

V. Experimental Results And Discussion

A. Experimental database: CVL Database

The proposed method is tested and validated on CVL database from Institute of Visual Computing & Human-Centered Technology, Computer Vision Laboratory [15]. The CVL Database is a public database for writer identification/retrieval and word spotting. The database consists of 7 different handwritten texts (1 German and 6 English Texts). In total 310 writers participated in the dataset. 27 of which wrote 7 texts and 283 writers had to write 5 texts. For each text a RGB color image (300 dpi) comprising the handwritten text and the printed text sample is available as well as a cropped version (only handwritten). An unique id identifies the writer, whereas the Bounding Boxes for each single word are stored in an XML file.

hereafter,samples of the following texts have been used:

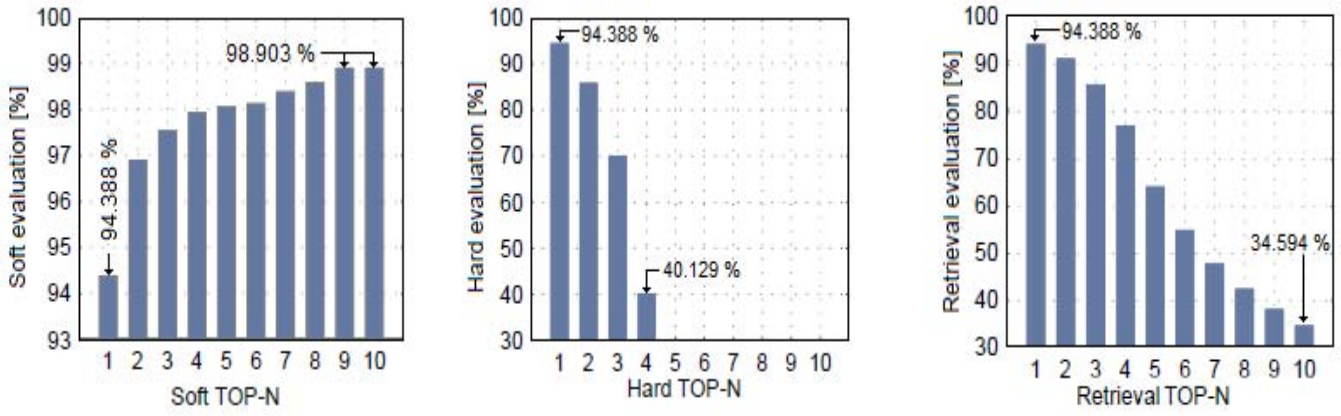
- Edwin A. Abbot - Flatland: A Romance of Many Dimension (92 words).
- William Shakespeare - Mac Beth (49 words).
- Wikipedia - Mailüfterl (73 words, under CC Attribution-ShareALike License).
- Charles Darwin - Origin of Species (52 words).
- Johann Wolfgang von Goethe - Faust. Eine Tragödie (50 words).
- Oscar Wilde - The Picture of Dorian Gray (66 words).
- Edgar Allan Poe - The Fall of the House of Usher (78 words).

The last update of the database is done on 12/09/2013 since one writer ID (265/266) was wrong. The version number was changed to 1.1.

Figure 3 shows an example of a filled-out form. All pages have a unique writer id and the text number (separated by a dash) at the upper right corner, followed by the printed sample text.

B. Assessment Protocol

As it is mentioned, in our test, we have used the CVL Latin handwritten database, Since, each writer's have at least five document images. For the training phase, one image of each writer's is randomly selected. the other four images of



(a) (b) (c)

Fig. 4. writer's identification/retrieval results: (a) Soft Evaluation, (b) Hard evaluation, (c) Retrieval Evaluation

TABLE I : SOFT, HARD ET RETRIEVAL EVALUATION USING THE ENTIRE DATASET (%)

	Top 1	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10
Soft	94.388	96.903	97.548	97.935	98.065	98.129	98.387	98.581	98.903	98.903
Hard	94.388	85.742	70.000	40.129	/	/	/	/	/	/
Retrv	94.388	91.323	85.828	77.032	64.194	54.925	47.834	42.435	38.108	34.594

each writer's were selected for the test phase of the proposed system.

Furthermore, the handwritten identification/retrieval tests results are divided into two parts. First part presents the performance of the ML-LPQ-based Latin handwritten recognition algorithm. For this, experiment was conducted using the CVL database to identify/retrieve the writer's of the text image(based on the *SOFT TOP-N*,*HARD TOP-N* criteria). In the second part, we present the performance of proposed ML-LPQ method when we use the handwritten images as a behavioral biometric modality. this part comport a discussion of the performance of handwritten based biometric system.

C. Experiment 1: writer's identification/retrieval results

In this part of experiments, the evaluation was done using the soft *TOP-N* and the *hard TOP-N*. So, for every document image of the database the distance to all other document images is calculated and this distance is sorted from the most similar to the least similar document image. Figure 4 give, the soft, hard retrieval evaluation results.

The first evaluation is carried out on the Soft criterion, The results for this criterion is presented in **Table I**. Starting from this table, It can be seen that for the soft criterion, our proposed method have a good performance. It is very effectively for represent the modality features (Soft TOP 1 = 94.388%) since the experimental results was performed on the cropped CVL database. So, it will be contain less written text in the image,the proposed method is very efficiency compared with state-of-arts according to the obtained SOFT *TOP-1* accuracy which is the exact identification of the writer. Also, a high Soft TOP 10 = 98.903%.

The evaluation of the hard criterion are shown also in the same Table,For the hard criterion similar results can be seen like for the soft criterion. the ML-LPQ algorithm show a good performance and high robustness, by giving a HARD TOP-1 equal to 94.388%.

Finally, based on the Soft and the Hard evaluation results, These results can also be seen in the retrieval criterion, where the proposed method achieves higher results compared to state of the art of Handwritten based recognition methods(see **Table I**).

D. Experiment 2: Handwritten based Biometric identification System

In this experiment, the biometric identification system performance, reported in Fig. 5, is aimed to shows the advantage of using Handwriting text as biometric modality, in this figure we report the experimental results as a Receiver Operating Characteristic (ROC) curve. Therefore, an experimental result at the EER points show that this modality get the best performance with a minimum Equal Error Rate (EER) equal to 2.308% at the threshold To = 0.0313. The results from the investigation in these results suggest that ML-LPQ method can be effectively used to achieve superior performance than the conventional features extraction and classification methods for the Handwriting based biometric identification system.

VI. Conclusion and future works

In this paper, a robust system for off-line Handwritten Latin text writer identification has been proposed based on a new feature descriptor called Multi Level Local Phase Quantization(ML-LPQ). the experimental results was divided into two subpart, the first subpart discuss the performance

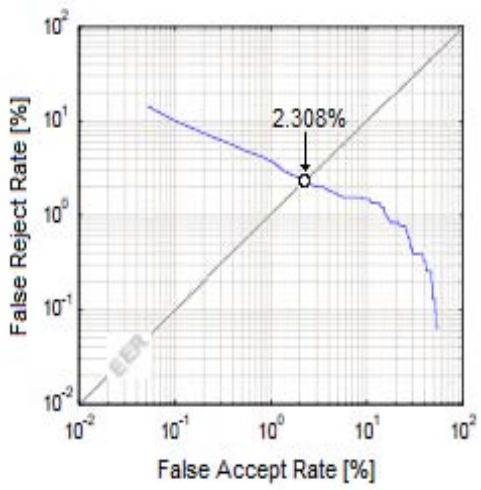


Fig. 5. ROC curve: FRR vs FAR.

of writer identification/retrieval system when using the CVL Latin Handwritten script. the performance evaluation was carried out by using two criteria : the *SOFT TOP-N* and the *HARD TOP-N*, the obtained results show the efficiency of the use of the proposed algorithm. The second subpart describe the performance of a Behavioral Handwritten based biometric recognition system. The performance and the evaluation of the biometric system are tested throw the CVL database. As a Future work, we will focus our research work on the development of a new feature descriptor which its aims is to outperforms the existed state of the art of Latin handwritten recognition method, also we will focus on the development of other new raised research area such as Arabic handwritten recognition.

REFERENCES

[1] Hannad, Y., Siddiqi, I., and El Kettani, M. E. Y. . "Writer identification using texture descriptors of handwritten fragments", *Expert Systems with Applications*, 47:14-22,2016.

[2] Hanusiak, R. K., Oliveira, L. S., Justino, E., and Sabourin, R. " Writer verification using texture-based features," *International Journal on Document Analysis and Recognition (IJDAR)*, 15(3):213-226, 2012.

[3] Fiel, S. and Sablatnig, R. "Writer identification and retrieval using a convolutional neural network, "*In Computer Analysis of Images and Patterns*, pages 26-37. Springer,2015.

[4] Rejean Plamondon, and Sargur N. Srihari,"On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey,"*1EEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*,. VOL. 22, NO. 1. JANUARY 2000.

[5] U-V Marti, H Bunke," A full english sentence database for off-line handwriting recognition,"*in Proceedings of the Fifth International Conference on Document Analysis and Recognition*, pp. 705-708,1999.

[6] E. Augustin, M. Carré, G. E., J. M. Brodin, E. Geoffrois, and F. Preteux,"Rimes evaluation campaign for handwritten mail processing,? *In Proceedings of the Workshop on Frontiers in Handwriting Recognition*, pp. 231-235 , 2006.

[7] Cheng-Lin Liu, Fei Yin, Da-Han Wang, Qiu-Feng Wang,"CASIA Online and Offline Chinese Handwriting Databases," *In International Conference on Document Analysis and Recognition*,2011.

[8] D Kim, Y Hwang, S Park, E Kim, S Paek, S Bang,"Handwritten korean character image database pe92," *in Proceedings of the 2nd International Conference on Document Analysis and Recognition*, pp. 470-473,1993.

[9] M. Pechwitz, S. S. Maddouri, V.Märgner, N. Ellouze, and H. Amiri,"IFN/ENIT - Database of Handwritten Arabic Words," *In 7th Colloque International Francophone sur l'Ecrit et le Document* ,pp.129-136 , CIFED 2002.

[10] HP Jifroodian, N Nicola, CL He, CY Suen,"multi-purpose handwritten farsi database," *in Image Analysis and Recognition Lecture Notes in Computer Science. A new large-scale* vol. 5627, pp. 278-286,2009.

[11] U. Pal, B.B. Chaudhuri,"Indian script character recognition: a survey," *In Pattern Recognition*, Vol 37, pp. 1887-1899, 2004.

[12] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig, "CVL-Database: An Off-line Database for Writer Retrieval, Writer Identification and Word Spotting,?*In Proceedigns of thhe 12th International Conference on Document Analysis and Recogniition (ICDAR 2013)*, Washington, USA,pp. 560-564 ,2013.

[13] YOU, Jane, LI, Wenxin, et ZHANG, David. "Hierarchical palmprint identification via multiple feature extraction". *In Pattern recognition*, vol. 35, no 4, p. 847-859.2002.

[14] V. Ojansivu and J.Heikkil, "Blur insensitive texture classification using local phase quantization," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5099 LNCS, pp. 236-243, 2008.

[15] https://cvl.tuwien.ac.at/research/cvl-databases/an-off-line-database-for-writer-retrieval-writer-identification-and-word-spotting/.