Image and Vision Computing 59 (2017) 17–30

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata ☆

Younes Akbari [a], Kazem Nouri [a], Javad Sadri [b,c,*], Chawki Djeddi [d], Imran Siddiqi [e]

[a] Dept. of Mathematics, Faculty of Mathematics, Statistics and Computer Sciences, Semnan University, Semnan, P.O. Box 35195-363, Iran
[b] Dept. of Computer Science & Software Engineering, Faculty of Engineering and Computer Science, Concordia University, Montreal, Quebec H3G 1M8, Canada
[c] Dept. of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, P.O. Box 97175-615, Iran
[d] Mathematics and Computer Science Department, Larbi Tebessi University, Tebessa, Algeria
[e] Bahria University, Islamabad, Pakistan

ARTICLE INFO

Article history:
Received 14 November 2015
Received in revised form 15 November 2016
Accepted 19 November 2016
Available online 7 December 2016

Keywords:
Off-line handwriting analysis
Gender detection
Texture analysis
Wavelet sub-band
Symbolic dynamics
Probabilistic finite state automata

ABSTRACT

Detection of gender from handwriting of an individual presents an interesting research problem with applications in forensic document examination, writer identification and psychological studies. This paper presents an effective technique to predict the gender of an individual from off-line images of handwriting. The proposed technique relies on a global approach that considers writing images as textures. Each handwritten image is converted into a textur\ed image which is decomposed into a series of wavelet sub-bands at a number of levels. The wavelet sub-bands are then extended into data sequences. Each data sequence is quantized to produce a probabilistic finite state automata (PFSA) that generates feature vectors. These features are used to train two classifiers, artificial neural network and support vector machine to discriminate between male and female writings. The performance of the proposed system was evaluated on two databases, QUWI and MSHD, within a number of challenging experimental scenarios and realized classification rates of up to 80%. The experimental results show the superiority of the proposed technique over existing techniques in terms of classification rates.

1. Introduction

Analysis of handwriting and hand-drawn shapes is an attractive area of research for psychologists, document examiners, palaeographers, graphologists, forensic analysts and computer science researchers. While the manual analysis of handwriting has been in practice for many decades, the automated analysis enjoys a renewed research interest thanks mainly to the relatively recent technological and algorithmic advancements in different areas of computer sciences. Although handwriting recognition remains the most significant application of automated handwriting analysis, a number of other interesting applications have also been investigated. These include classification of writing styles, keyword spotting in handwritten documents and, identification and verification of writers from their handwritten samples. A closely related problem to writer identification is the classification of user demographics from handwriting. Identification of demographic attributes including gender, handedness, age, race etc. has been investigated in the literature [1,2]. Also, a few psychological studies suggest the existence of correlation between handwriting and different attributes of the personality of the writer [3–8]. The claims of these graphological studies, however, remain subjective and are yet to be validated by experimental studies. As opposed to personal attributes which tend to be subjective, demographic attributes are objective and the classification of these attributes from handwriting can be experimentally validated through quantified results. Among various demographic attributes, a relatively strong correlation has been demonstrated between gender and features of handwriting in a number of studies [9–14]. Psychological studies suggest that the male writings, in general, tend to be more 'spiky', 'hurried' and 'untidy' while the female writings are likely to be more 'decorative', 'homogenous' and 'consistent' [13–15]. Automatic detection of gender based on these features, makes the subject of our research.

Automatic gender prediction from handwritten samples can be implied in a number of interesting applications not only for psychologists but also for forensic experts and document examiners.

☆ This paper has been recommended for acceptance by Cheng-Lin Liu.

* Corresponding author at: Dept. of Computer Science & Software Engineering, Faculty of Engineering and Computer Science, Concordia University, Montreal, Quebec H3G 1M8, Canada.

E-mail addresses: akbari_younes@semnan.ac.ir (Y. Akbari), knouri@semnan.ac.ir (K. Nouri), j_sadri@encs.concordia.ca (J. Sadri), c.djeddi@univ-tebessa.dz (C. Djeddi), imran.siddiqi@bahria.edu.pk (I. Siddiqi).

http://dx.doi.org/10.1016/j.imavis.2016.11.017
0262-8856/

It can serve as a filtering step allowing experts to reduce the search space and focus on a subset of writings coming from a specific gender group. Besides, gender classification can also lead to improved results in many writer identification and verification applications [16]. In addition to these applied aspects of gender detection, the study of relationship between gender and different writing manifestations presents an interesting problem of basic research. Gender classification is closely related to the writer identification problem. A critical analysis of the gender classification methods based on handwritings reveals that in many cases features used for writer identification have been adapted for gender classification [1,2,16]. The gender classification rates, however, are not as good as those realized for writer identification systems. Although a two-class problem, the high intra-class variation makes gender detection from handwriting a challenging task. Moreover, in some cases, male writings tend to have a 'feminist' visual appearance and vice versa. This inter-class similarity also explains the low classification rates on this problem. This is also illustrated in Fig. 1 which shows male writing samples and female writing samples sometimes are very similar.

All automated methods for analysis of handwriting can be categorized into one of global and local approaches from the perspective of feature extraction technique. While local approaches are preferred from the view point of forensic experts [3,17,18], a major challenge in such methods is the segmentation of handwriting into smaller units including lines, words and characters. This segmentation becomes even harder in unconstrained handwritings which are encountered in most of the real world scenarios. The existing research on computerized detection of gender from handwriting is mainly focused on local approaches and to the best of authors' knowledge, texture analysis to characterize gender is yet to be investigated. Moreover, in most cases, features primarily employed for problems like handwriting recognition and writer identification are applied to the gender prediction task and the average gender classification rates on standard databases vary from 65% to 75%.

This paper addresses automatic gender prediction from offline handwritten images. We propose a novel and effective technique that considers writings as textures and employs wavelets to characterize male and female writings. An overview of the steps involved in the proposed technique is presented in Fig. 2. The proposed feature extraction relies on wavelet transform using symbolic dynamic filtering (SDF) [19], a recently reported feature extraction algorithm. In a recent study, SDF based feature extraction from time series of robot motion behavior has been carried out by Mallapragada et al. [20]. Likewise, SDF-based feature extraction from time series data has been proposed by Jin et al. [21] for target detection and classification in border regions. To the best of our knowledge, this is the first attempt of exploiting symbolic dynamic filtering to characterize gender from handwriting and represents the main contribution of this work. In our feature extraction method, the wavelet sub-band is extended into a data sequence that is symbolized to construct probabilistic finite state automata (PFSA) [19] which generates the feature vectors. These feature vectors are classified by artificial neural network (ANN) and support vector machine (SVM). The proposed method is evaluated on the well-known QUWI and MSHD databases with writing samples in English, French and Arabic. In addition to the gender classification rates on the two database, a number of interesting experimental settings have been considered to study the impact of textual content and script of text on the classification

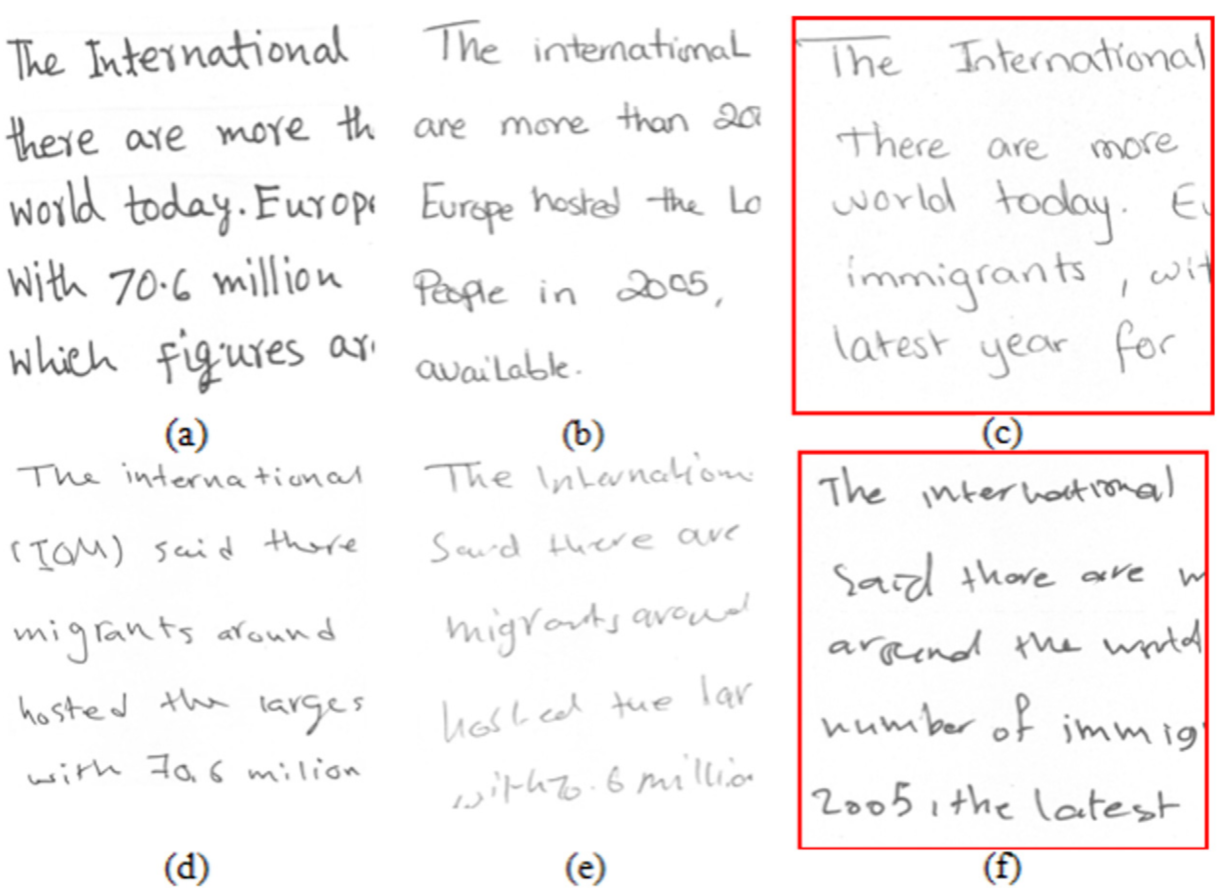

Fig. 1. Similarity and misclassification between male and female handwritings. (a–b) True female writings. (c) Similar male writing with 'decorative', 'homogenous' and 'consistent' attributes. (d–e) True male writings. (f) Similar female writing with 'spiky', 'hurried' and 'untidy' attributes.

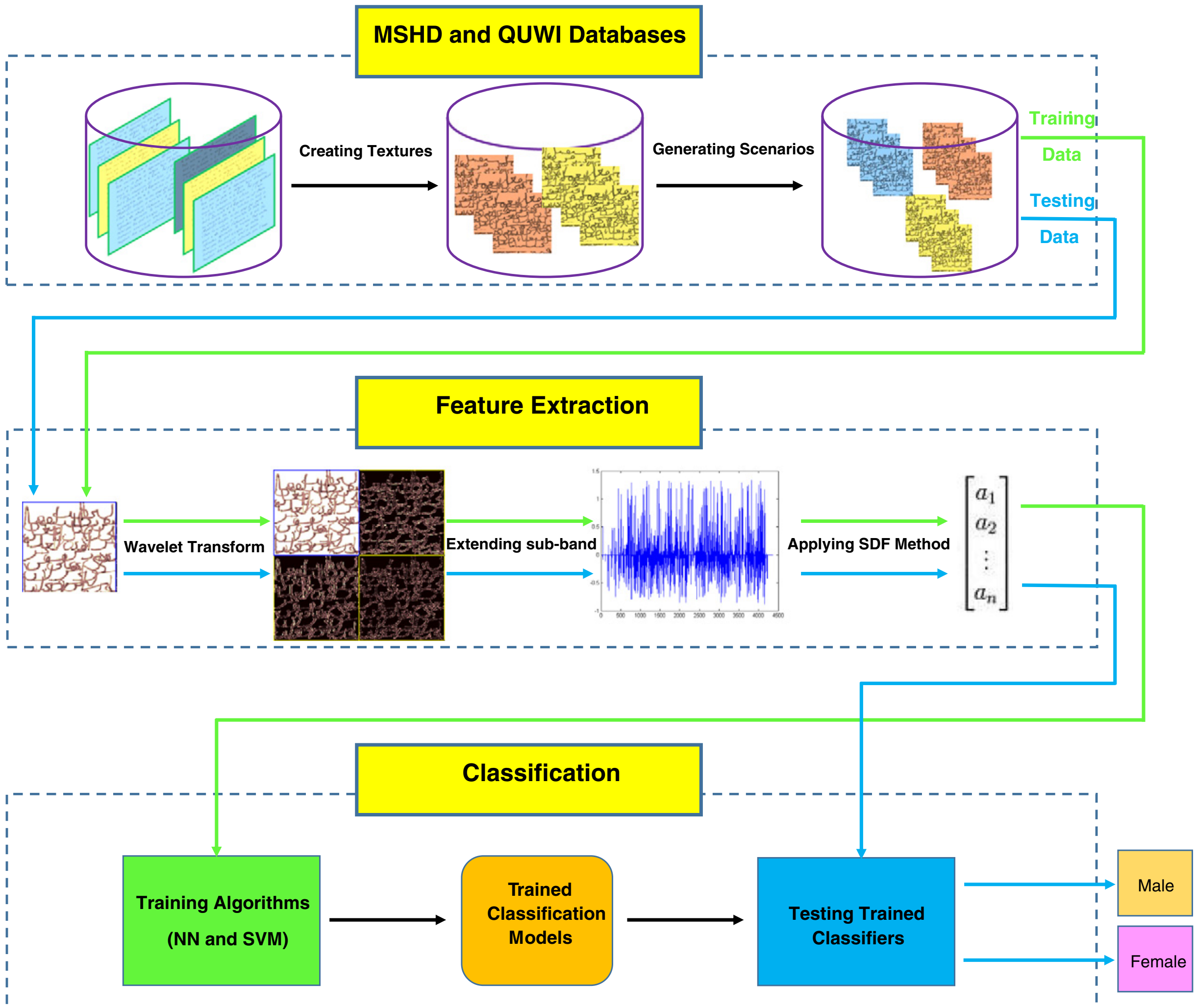

Fig. 2. Overview of the proposed gender prediction system.

performance. Cross database evaluations have also been conducted to study the impact of geographical distribution of writers on the classification task. An objective comparison with existing handwriting based gender classification techniques reveals the superiority of the proposed technique in terms of classification rates.

As discussed earlier, the task of gender prediction from handwriting is related to the writer identification problem. Writer identification aims at extracting writer-specific characteristics from a given set of writing samples. Gender classification, on the other hand, aims at exploiting the specific writing attributes that are common across individuals belonging to the same gender group (male or female). State-of-the-art features employed for writer recognition have been exploited for gender classification in a number of studies. For instance, the writer identification features presented in [22], [23] and [24] are exploited for gender classification in [1], [2] and [25], respectively. In our work, for comparison purposes, we also study the performance of well-known writer identification features for gender classification. These include the orientation and curvature features [22], fractal dimension [26], auto-regressive (AR) coefficients [27] and local binary patterns (LBP) [28]. These features are known to be effective for identification of writers and we analyze their performance in characterizing gender from handwriting.

The organization of the rest of this paper is as follows. As a background, brief discussions on relationship between gender and handwriting and the computerized detection of gender from handwriting are presented in Section 2. Section 3 describes the databases employed in our study while Section 4 details our proposed methodology. Experiments, results and the accompanying analysis is presented in Section 5. The last section concludes the paper with a discussion on interesting research directions and future works.

2. Related works

In this section, we first discuss the relationship between handwriting and gender as identified in the manual analysis of handwriting. Later, we present an overview of the notable contributions to automated analysis of handwriting for gender classification.

2.1. Gender and handwriting

A number of studies [12–14] have concluded that gender can be predicted from handwriting though the accuracy of predication may vary. It has also been observed that individuals who frequently deal with handwritten images become skilled at discriminating the writings of the two gender classes [1]. These differences are generally attributed to the differences in motor coordination among the two genders. The grapho-analytical investigation of Lester et al. [29] considered the influence of gender on handwriting. Comparing male and female handwritings is an objective re-examination of Lester's graphologic findings that there are differences between the writings coming from different genders [9]. As discussed earlier, in general, female writings tend to be neat, even, well-organized, rounded, small and symmetrical. Male writings, on the other hand, are generally characterize by attributes like rushed, uneven, messier, slanted and sloping [13]. A set of detailed discriminating factors between male and female writings can be found in [30].

Eames and Loewenthal [31] carried out a study using samples from 12 undergraduate students to determine whether handwriting has any correlation with the grades of the students and whether the gender of a student could be recognized through his or her handwriting. The results of the study showed that grades of students were not influenced by the quality of handwriting. However, in many cases, it was possible to determine the gender of the student. In another study, Hayes [14] performed several experiments with male and female students and reported a 75% correct identification of gender even with very limited writing samples. The author concluded that in most cases, one letter or symbol is often enough for identification of gender. In a separate study on manual analysis of handwriting by Hamid and Loewenthal [12], the authors analyzed English and Urdu handwriting samples collected from 30 different individuals. Each contributor copied a 50-word passage (both in Urdu and English) and the textual content of the passage was same for all the individuals. These samples were then analyzed by 25 examiners for identification of gender. It was concluded that gender identification is equally reliable for text in both the languages with an average accuracy of 68%.

The interesting findings of researchers on the relationship between handwriting and gender inspired the computer scientists to apply image analysis and pattern classification techniques to automate this study. The following subsection presents an overview of computerized classification of gender from handwriting.

2.2. Automated gender prediction from handwriting

Limited literature is available when it comes to automated detection of gender and other demographic attributes of individuals form their handwriting samples [16]. Among one of the preliminary studies on this problem, Cha and Srihari [32] proposed a system for classification of user demographics from handwritten images in text-dependent mode and reported a gender classification accuracy of 70.2%. In a later study [25], the authors employed a set of macro-features [24,33] to classify writers on the basis of gender, handedness and age. Artificial neural networks are used for classification and the results are improved by classifier combination using bagging and boosting. The authors report classification accuracies of 77.5%, 86.6%, and 74.4% for gender, age, and handedness, respectively.

Liwicki et al. [2] proposed a gender and handedness classification technique from online handwriting that, in addition to the offline representation of character shapes, allows to exploit the temporal information of writing as well. The authors extract a set of 29 features from the online-samples and its offline representation and employ support vector machine and Gaussian mixture models for classification. Accuracies of up to 67.06% and 84.66% are reported for gender and handedness classification respectively. A separate analysis of offline and online writings revealed that the classification rates on online handwritings are better than those on offline images of writing. In [34], the authors exploit features like Fourier descriptors and curvature information to identify gender from handwriting. The authors, however, do not present any quantified results and discuss the values of these parameters for male and female writings.

Two competitions on gender classification using the QUWI database have been held in conjunction with ICDAR 2013 and ICDAR 2015. The winning system in the ICDAR 2013 competition [35] used the Gradient Boosted Decision Trees (GBDT) for feature selection and classification and reported classification rates of 74% and 79% on Arabic and English writings, respectively. Later, a number of studies [1,16,30] investigated the prediction of gender from handwriting using the competition database (QUWI). In [1], Siddiqi et al. exploit writing attributes like orientation, legibility and curvature to distinguish between male and female writings. For classification, the authors employ artificial neural networks and support vector machine and report classification rates of around 68% on the QUWI database. In another study [30], the wavelet domain local binary patterns (WDLBP) are extracted from handwritten images to characterize the gender of the writer. SVM based classifiers reported correct classification rates of 68.6% and 85.7% for Arabic and English writings of the QUWI database, respectively. A variety of features aimed at charaterizing the gender of writer are combined using random forests (RF) and kernel discriminant analysis (KDA) in [16]. Gender classification rates of 74.05% (text-dependent mode) and 73.59% (text-independent mode) are realized on the QUWI database.

In the next section, details of the handwriting databases used in this study are presented.

3. Databases

The experimental study of the proposed system is carried out on two well-known handwriting databases, the Qatar University Writer Identification (QUWI) database [36] and Multi-script Handwritten Database (MSHD) [37]. Although primarily developed for evaluation of writer identification systems, the gender information of the handwriting images also allows using these databases for gender classification experiments. The images in the two databases were divided into several groups to analyze the performance of the proposed methodology in a number of experimental scenarios including script-dependent and script-independent, text-dependent and text-independent and cross database evaluations. As mentioned earlier, we extract features from texture images of writing created by concatenating the writing blocks of an individual. This resulted in a collection of textured images (256×256) to support further experiments. More details on generating the texture images are presented in the next section.

The details of the two databases along with the distribution for different experimental scenarios are presented in the following sections. The experimental protocol is similar to the one presented in [1].

3.1. QUWI database

The QUWI database [36] comprises writing samples contributed by 1017 writers. Each writer provided 4 samples, two in English and two in Arabic. Page 1 and Page 3 of all writers contain a text from writer's own imagination in Arabic and English respectively. Likewise, page 2 (Arabic) and page 4 (English) of each writer contains the same textual content. Having four pages per writer in two scripts with same as well as different textual content allows this database to be employed in a number of interesting experimental scenarios. To allow meaningful comparison of the performance of our system with other studies, we carried out the experiments on writing samples of 475 writers, same as in the ICDAR 2013 competition database. The training set includes writing samples of 282 writers while the test set comprises 193 writers. The distribution of writers is kept same in different experiments whereas the number of samples per writer varies from scenario to scenario as described in following:

- Text-dependent and text-independent experiments:
  The text-dependent evaluations are carried out using page 2 (Arabic) and page 4 (English) of each individual having the same textual content. Similarly, page 1 and page 3 of all writers which contain, respectively, an arbitrary text in Arabic and English are used in text-independent evaluations.

- Script-dependent and script-independent experiments:
The script dependent experiments are carried out separately for Arabic (pages 1 and 2) and English (pages 3 and 4) and involve writing samples in the same script in both training and test sets. The more challenging script-independent evaluations, on the other hand, require the training and test samples to be in different scripts. In the first series of script-independent evaluations, the training set comprises English samples of 282 writers while the Arabic samples of 193 writers are used as the test set. In a similar fashion, another series of experiments is carried out by reversing the scenario and using Arabic samples in the training set and English samples in the test set.

3.2. MSHD database

The MSHD database [37] contains handwritten samples collected from a total of 84 writers. Each writer was required to write 12 pages, 6each in French and Arabic and each page had the same textual content for all writers. In most of the evaluations the training and test sets comprise writing samples of 42 writers each. The distribution of samples in different experimental scenarios is described in the following:

- Text-dependent and text-independent experiments:
For text-dependent experiments, 6 samples of each writer containing the same text in French (Arabic) are used. In text-independent evaluations, the idea is to compare writing samples with different textual content. Consequently, we employ the first 3 French (Arabic) samples of the 42 writers as the training set and the last 3 French (Arabic) samples of the remaining 42 writers as the test. The training and test sets are then reversed and the classification rate is computed by taking the average of the two experiments.
- Script-dependent and script-independent experiments:
Similar to the QUWI database, the script-dependent evaluations employ the text in the same script (Arabic or French) in both training and test sets with 42 writers each in the training and test sets. In script-independent evaluations, the learning base comprises French writings of the first 42 writers while the evaluation base contains Arabic writings of the other 42 writers. Likewise, the last experiment involves reversing the training and test sets similar to the QUWI database.

It should be noted that in all experimental settings, the training and test sets do not contain any writing sample of the same writer so that the problem corresponds to gender classification and not writer identification. The division of writers in training and test sets along with other information is presented in Table 1.

4. Methodology

This section presents the details of the proposed methodology including pre-processing, feature extraction and classification. Prior to these steps, for completeness, we present the basic theoretical background of the techniques used in our system.

4.1. Background

The proposed characterization of gender from images of handwriting mainly relies on wavelet transform and symbolic dynamic filtering (SDF). We present a brief overview of these concepts in the following sections before discussing their application to our particular problem.

4.1.1. Wavelet transform

Wavelets are generally termed as numerical microscope in signal and image processing. Characterizing time-dependent data using wavelet basis results in a powerful representation of information that is simultaneously localized in time and frequency domains. This is in contrast to the Fourier representation where it is not possible to associate specific frequencies to specific intervals of time. Wavelets are specifically powerful when it comes to analysis of signals and images with discontinuities and sharp spikes and find applications in a wide variety of problems in Physics, Mathematics and Electrical Engineering, replacing the conventional Fourier transform in many cases. Typical applications of wavelets in signal processing include image compression, image denoising, speech recognition, EEG, EMG and ECG analyses etc. An integrable function $\psi \subseteq L^2(R)$ is considered a wavelet function if it satisfies the zero-moment condition [38,39]

$$\int_{-\infty}^{\infty} \psi(t)dt = 0, \quad t \in R. \tag{1}$$

The zero-moment condition is valid when it satisfies the following admissibility condition that is needed to obtain the inverse of the wavelet transform,

$$C_\psi = 2\pi \int_{-\infty}^{\infty} \frac{\mid \psi(\hat{\xi}))\mid^2}{\mid \xi \mid} d\xi < \infty, \quad t \in R. \tag{2}$$

The continuous wavelet transform (CWT) of a function $f \in L^2(R)$ is defined as

$$W_\psi f(\mu, s) = \langle f, \psi_{\mu,s} \rangle = \int_{-\infty}^{\infty} f(t)\psi^*_{\mu,s} dt, \quad t \in R, \tag{3}$$

where $\psi_{(\mu,s)}(t) \ = |s|^{-1/2}\psi((t-\mu)/s)$, with $\mu, s \in R, s \neq 0$. μ, s vary continuously over R and are termed as dilation parameter and translation parameter, respectively. ψ is called the mother wavelet and the $\psi_{\mu,s}$ are called wavelets [38]. The wavelets $\psi_{\mu,s}$ cover different frequency ranges when s changes.

Large values of the dilation parameter $|s|$ correspond to small frequencies whereas the small values correspond to high-frequencies. Changing the translation parameter μ moves the time localization center. Each $\psi_{\mu,s}(t)$ is localized around $t = \mu$. Hence, a perfect time frequency description of function f can be obtained using the wavelet transform. In our case, the handwritten images are decomposed into a series of wavelet sub-bands using the Mallat algorithm [40]. This decomposition is similar to the one presented in [39] for writer identification.

Table 1
The overall distribution of samples used in two databases.

Database	QUWI database	MSHD database
Text	Arabic and English	Arabic and French
Number of writers (training+test)	475 (282+193)	84 (42+42)
Distribution of writers	221 male & 254 female	37 male & 47 female
Samples per writer	4 (2 Arabic, 2 English)	12 (6 Arabic, 6 French)
Total samples	1900 (4 × 475)	1008 (12 × 84)
Size of texture block	256 × 256	256 × 256
Texture blocks from n pages ($p_1 + p_2 + \ldots + p_n$)	$n = 4$, 52 (11+15+11+15)	$n = 12$, 180 (15+15+…+15)
Total number of texture blocks	24,700 (52 × 475)	15,120 (180 × 84)

4.1.2. Symbolic dynamic filtering (SDF)

This section briefly presents the underlying concepts of symbolic dynamic filtering (SDF) introduced by Ray [19] for extraction of features. In our implementation, we have applied the SDF algorithm presented in [41] to extract gender specific features from the writing images. The key steps involved in SDF based feature extraction are discussed in the following:

A) Symbolization/quantization of data

The first step in SDF is the quantization of data (wavelet transformed images in our case). For symbolization, the given data sequence is divided into a finite number of cells. Each cell is the assigned a unique label or symbol and the total number of unique symbols is the same as the total number of cells. This can be elaborated by considering the data sequence in Fig. 3. Assuming the symbol alphabet to be A = {a1,a2,a3,a4,a5} with card(A) = 5, the data sequence has been divided into five partitions on the y-axis. These regions are mutually exclusive as well as exhaustive (the complete data profile is covered). Each of the five regions is assigned a label from the alphabet *A*. For a data sequence under consideration, the signal value at a given point in time is assigned the symbol corresponding to the cell in which it is located. This allows representing the data sequence by a finite string of symbols often termed as the symbol block. Further details on the symbolization of data can be found in [42].

For partitioning of data, strategies like uniform partitioning (UP) maximum entropy partitioning (MEP) have been investigated [42]. Uniform partitioning, as the name suggests, generates cells of equal size. Maximum entropy partitioning, on the contrary, attempts to maximize the entropy of the generated symbols by ensuring that each cell contains approximately the same number of data points. Consequently, the cell size is small in information rich regions while it is large in sparse regions. In both cases, the size of alphabet is chosen as a function of the data under study and the target system performance.

B) Generation of probabilistic finite state automata (PFSA)

The construction of probabilistic finite state automata (PFSA) is based on the assumption that the symbol generation process can be modeled as a Markov chain, termed as the *D*-Markov machine [19]. A *D* order Markov chain is a stochastic process where the probability of occurrence of a symbol is a function of previous *D* symbols,

$$P(s_i|s_{i-1}, s_{i-2}, \ldots, s_{i-D}, \ldots) = P(s_i|s_{i-1}, \ldots, s_{i-D}) \tag{4}$$

The states in PFSA represent combinations of words in the sequence of symbols while the edges correspond to transition between different states. In our work, we take $D = 1$, hence, the number of states is the same as the number of

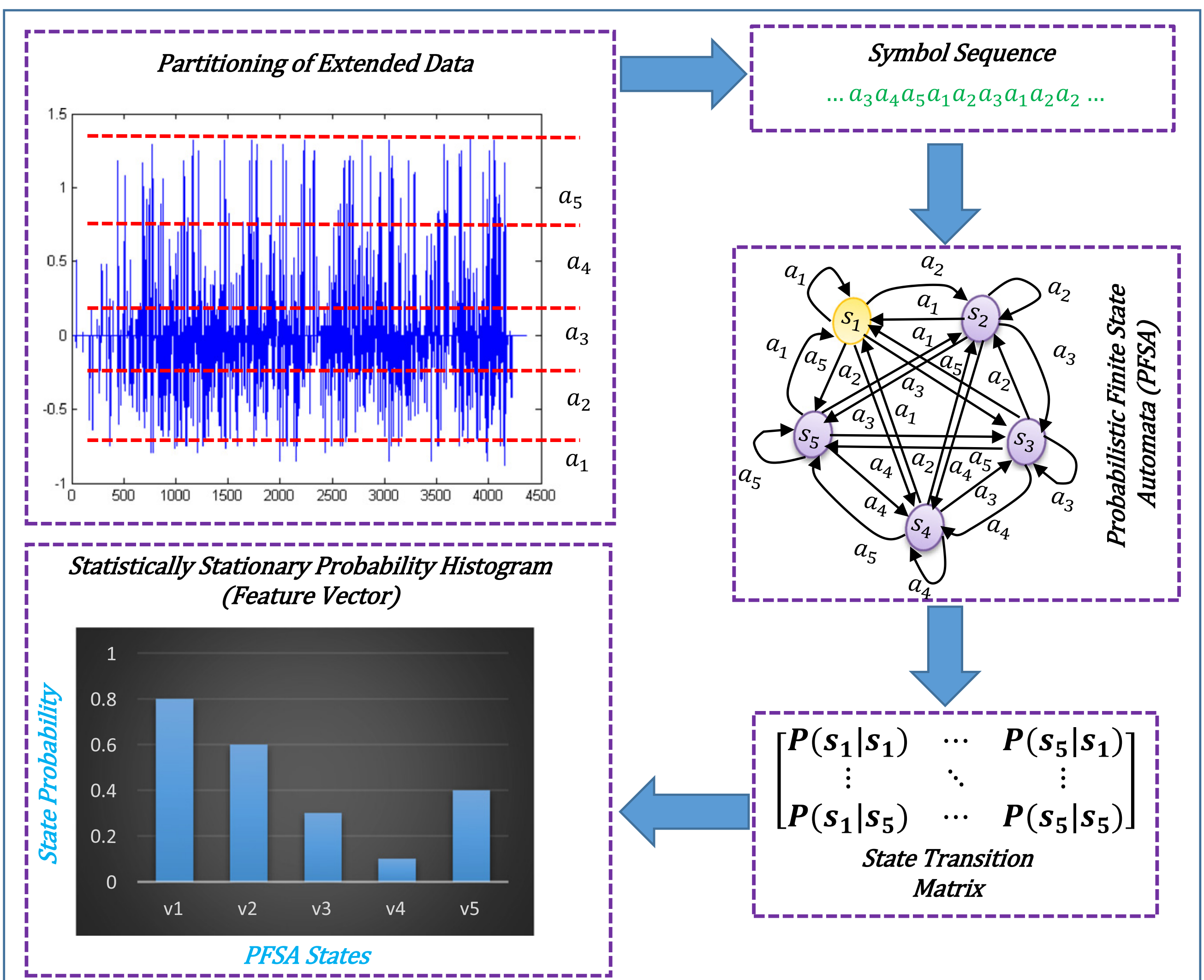

Fig. 3. Overview of symbolic dynamic filtering (SDF) for feature extraction (based on [41]).

symbols in the alphabet. The set of possible states is given by $S = \{s_1, s_2, \ldots, s_N\}$ with N being the total number of states. The probability of transition from state s_i to state s_j can then be defined as follows:

$$P(s_j|s_i) = \frac{N(s_i, s_j)}{\sum_{k=1,2,\ldots,N} N(s_i, s_k)}, \forall s_i, s_j \in S, \quad (5)$$

where $N(s_i, s_j)$ is the total number of events when s_j appears adjacent to s_i. Once all the transition probabilities $P(s_j|s_i), \forall s_j, s_i \in S$ have been estimated, they are collected the $N \times N$ state transition matrix.

Once the matrix of transition probabilities is computed for a given data, its left eigenvector $V = \{v_1, v_2, \ldots, v_N\}$ with respect to the unity eigenvalue is calculated which serves as the feature vector and can be used in recognition tasks.

After having presented an overview of the wavelet transform, the SDF and PFSA, we now discuss its application to handwritten images in the sections to follow.

4.2. Preprocessing

Prior to feature extraction, we need to generate texture blocks from the handwritten images. Methods for automatic generation of texture blocks [39,43], in general, work well for handwritten documents having a uniform layout but face a number of challenges when it comes to documents with irregular layout of handwriting. The main problems arise from the difficulties in segmenting the writing into lines, words, and characters. The databases considered in this study comprise unconstrained handwritings with irregular layout. We, therefore, employ the texture generation algorithm presented in [28,44] which is known to handle such writings. Irrespective of the way handwritten text is arranged on the page, the connected components in the writing are rearranged into a new space keeping the original slant but reducing the spaces between lines of text and components. This generates texture images which preserves the over all look and feel of the writing and allows using a global approach avoiding the complexity of segmentation. The major steps in this process are listed in the following:

- The image is binarized using global thresholding and the connected components are extracted using 8-connectivity.
- Small components which are likely to correspond to punctuation marks and noise are removed using area based filtering.
- The bounding boxes of the remaining components are then used to extract the corresponding components from the gray scale image.
- Components in the first line of the original image are aligned in a new image using the center of mass of the bounding box. The slant of the text lines, therefore, is normalized.
- After filling the first line, the average height of the components is calculated and this value is used to define the y-coordinate of the next line, which is given as follows [28]:

$$y_{new} = y_{prev} + \frac{h}{2}, \quad (6)$$

where y_{prev} is the y-coordinate used to fill the previous line and h is the average height of all connected components used to fill the previous line. The term $h/2$ allows generating a more representative texture eliminating white spaces and reducing the gap between the lines.
- In the final step, if the end of page is reached, but the texture window is not completely filled, the algorithm is repeated starting from the beginning of window as illustrated in Fig. 4.

As mentioned previously, a block size of 256 × 256 is employed to segment the handwriting images. For the QUWI database, we extract eleven blocks of texture each from the first and the third page and fifteen blocks of texture each from the second and the fourth pages. Likewise, for the MSHD database, we extract fifteen blocks of texture from each of the pages. Fig. 5 illustrates an example each of texture blocks created from the gray scale images from the two databases. It is important to note that in generating the textured images, the original line spacing and slant of writing (skew of text lines) are not preserved. However, a number of classical as well as recent techniques on identification of writers [28,44,45] employed similar packing of text lines to effectively characterize writer of a document. In our study, to investigate whether the line packing has any impact on the classification performance, we carry out feature extraction with and without packing the lines. More details on this are presented in Section 5 of the paper.

Once the texture blocks are generated, we proceed to extraction of features from each block as described in the following section.

4.3. Feature extraction

The key steps in feature extraction involve application of wavelet transform to the images and SDF based feature extraction explained as follows. The major steps involved in the extraction of features are summarized in Fig. 6.

1. **Step 1**: The handwriting image I is decomposed into sub-bands using the discrete wavelet transform (DWT). DWT produces a series of sub-bands including one low-frequency and three high-frequency bands. Experiments reveal that a three level decomposition suffices and a decomposition beyond three

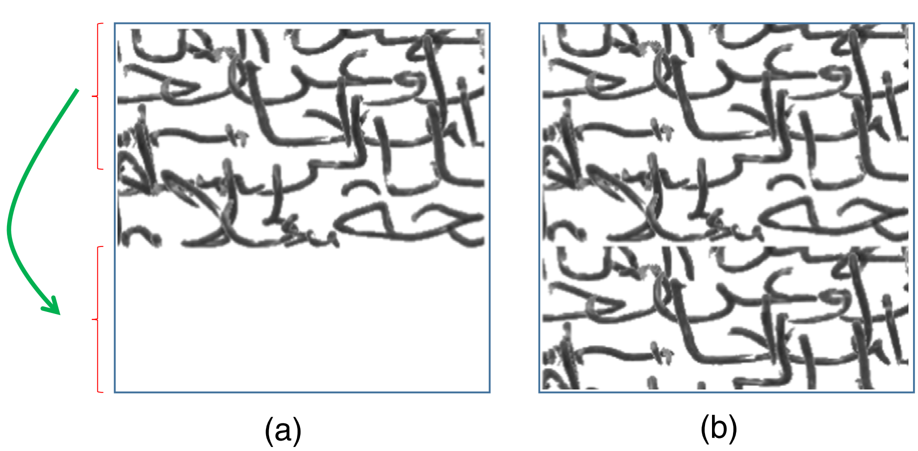

Fig. 4. Generation of texture blocks. (a) Unfilled window. (b) Filled window.

Fig. 5. Original page and texture blocks. (a) Samples from the QUWI database (Arabic and English) and (b) samples from the MSHD database (Arabic and French).

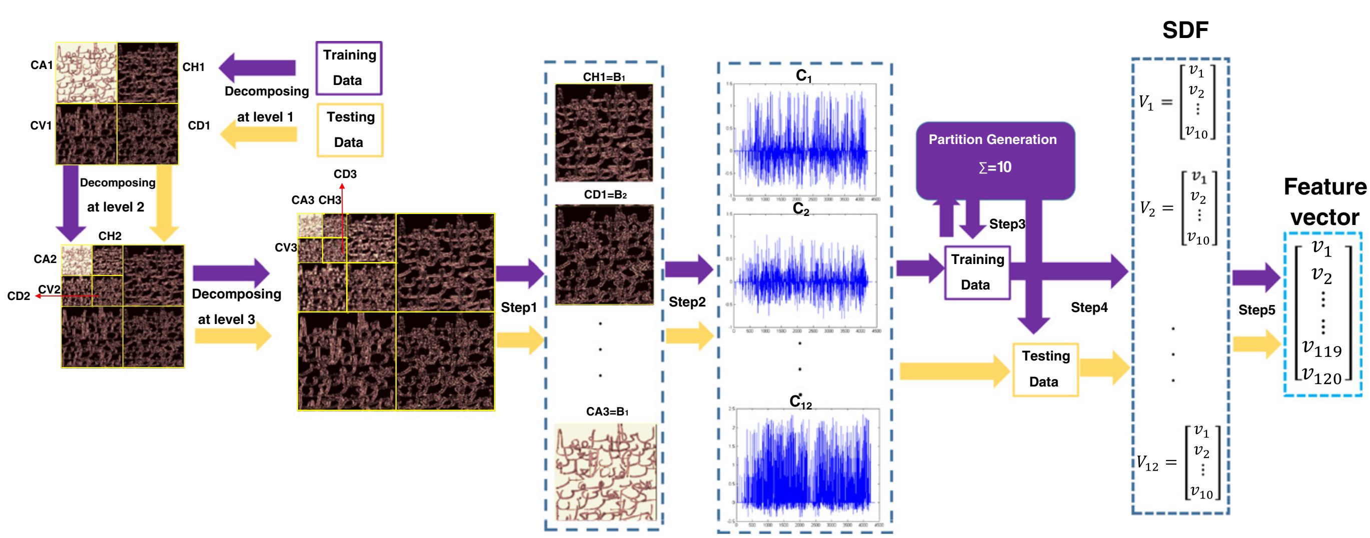

Fig. 6. Block diagram of the proposed process of feature extraction.

Table 2
Classification rates on QUWI and MSHD databases.

Method	Database	Script	Classification method	Classification rate (%)
Proposed method (with line packing)	QUWI	(Arabic+English)	(SVM, NN)	(77.80, 79.30)
	MSHD	(Arabic+French)		(79.90, 79.00)
Proposed method (without line packing)	QUWI	(Arabic+English)	(SVM, NN)	(59.00, 64.00)
	MSHD	(Arabic+French)		(68.50, 69.20)
Orientation and curvature [22]	QUWI	(Arabic+English)	(SVM, NN)	(68.75, 67.00)
	MSHD	(Arabic+French)		(72.82, 69.25)
Fractal dimension [26]	QUWI	(Arabic+English)	(SVM, NN)	(61.50, 62.50)
	MSHD	(Arabic+French)		(62.30, 61.90)
Combination of LBP [28] and AR [27])	QUWI	(Arabic+English)	(SVM, NN)	(59.75, 61.50)
	MSHD	(Arabic+French)		(68.65, 64.88)
Siddiqi et al. [1]	QUWI	(Arabic+English)		(68.75, 67.50)
	MSHD	(Arabic+French)		(73.02, 69.44)
Winner of ICDAR2013 [35]	QUWI	(Arabic+English)	GBDT	76.00
ICDAR2013 features [35]	QUWI	(Arabic+English)	SVM	75.20
Youssef et al. [30]	QUWI	(Arabic+English)	SVM	74.30

levels does not cause any significant improvement in the classification rates. Hence, for the subsequent discussion: $I \Rightarrow \{B_1, \ldots, B_{12}\}$ with B being the set of sub-bands.

2. **Step 2**: The second step comprises extending the sub-bands B_i into a data sequence C_i [43]. $B_i \Rightarrow C_i : C_i(y \times width + x) = b_i(x, y)$. with *width* being the width of the sub-bands B_i.
3. **Step 3**: The maximum entropy partitioning (MEP) is applied on the generated data sequence. For the purpose of pattern classification, the training data set is partitioned with a given alphabet size $card(A)$ that is selected by trade-off between information loss and computational complexity [42] and the partitioning is subsequently kept constant for the test data.
4. **Step 4**: SDF based features (using an alphabet size of 10) are extracted as discussed earlier in the paper. The algorithmic details of these steps can be found in [41].
5. **Step 5**: The feature vector is extracted by concatenating vectors from step 4.

4.4. Classification

For classification of handwritten documents into gender classes (male and female), two classifiers have been employed, the artificial neural networks (ANN) and the support vector machine (SVM).

1. **Neural networks**: Neural network classification is a classical machine learning algorithm that, in addition to other classification tasks, has been widely employed for document and handwriting recognition purposes [46]. A neural network aims to simulate the human nervous system through a large number of processing units (neurons) connected with other units in several layers. In this study, the multi-layer perceptron (MLP) network with three layers (input, hidden and output) has been employed. The number of neurons in the input layer is the same as the dimensionality of the feature vectors while the

Table 3
Classification rates of text-dependent and text-independent evaluations on the QUWI and MSHD databases.

Method	Database	Script	Classification method	Scenario	
				Text-dependent rate (%)	Text-independent rate (%)
Proposed method	QUWI	Arabic	(SVM, NN)	(76.20, 68.20)	(74.20, 74.60)
		English		(75.20, 73.90)	(76.10, 69.20)
	MSHD	Arabic		(78.10, 76.00)	(75.20, 74.30)
		French		(80.10, 76.20)	(75.80, 72.20)
Siddiqi et al. [1]	QUWI	Arabic		(69.00, 71.00)	(63.00, 65.00)
		English		(68.00, 70.00)	(70.00, 66.00)
	MSHD	Arabic		(74.20, 72.62)	(72.22, 71.43)
		French		(68.25, 67.06)	(67.46, 66.27)
Al-Maadeed and Hassaine [16]	QUWI	Arabic	(RF, KDA)	(71.10, 70.00)	(69.00, 71.60)
		English		(71.10, 69.30)	(74.70, 73.70)

Table 4
Classification rates of script-dependent and script-independent evaluations on the QUWI and MSHD databases.

Method	Database	Scenario	Script		Classification	
			Train	Test	SVM (%)	NN (%)
Proposed method	QUWI	Script-dependent	Arabic	Arabic	77.70	71.90
			English	English	75.50	69.00
	MSHD		Arabic	Arabic	75.80	76.40
			French	French	77.60	75.10
	QUWI	Script-independent	Arabic	English	69.40	67.80
			English	Arabic	68.60	68.70
	MSHD		Arabic	French	77.90	74.20
			French	Arabic	75.70	75.20
Siddiqi et al. [1]	QUWI	Script-dependent	Arabic	Arabic	68.50	65.00
			English	English	68.50	66.50
	MSHD		Arabic	Arabic	76.98	73.41
			French	French	70.63	69.44
	QUWI	Script-independent	Arabic	English	62.50	65.00
			English	Arabic	67.00	65.00
	MSHD		Arabic	French	57.94	61.90
			French	Arabic	70.63	69.84
Youssef et al. [30]	QUWI	Script-dependent	Arabic	Arabic	68.60	–
			English	English	85.70	–
ICDAR2013 features [35]	QUWI	Script-dependent	Arabic	Arabic	62.30	–
			English	English	77.10	–

output layer has 2 neurons corresponding to the two gender classes. The number of neurons in the hidden layer was fixed to 6 using cross validation. The transfer functions used in the experiments include tangent sigmoid for the hidden layer and logistics sigmoid for the output layers.

2. **Support vector machine**: SVM, primarily a linear classifier, generally employs a kernel function to map data points from the original space to a higher dimensional space to find the classification decision boundary [47]. SVM classification relies on maximizing the margin between two classes and bounding the number of training errors to find the optimal hyper plane. We use the radial basis kernel based SVM with $\gamma = 0.3$ (RBF kernel parameter) and $c = 10$ (parameter to control the tolerance of the classification errors in learning) has been used.

We present the performance of these classifiers in the next section.

5. Experimental results

We now present the results of the experiments to validate the effectiveness of the proposed technique in characterizing gender from handwriting. First, we discuss the classification rates on the two databases (QUWI and MSHD) and later present the results of

Table 5
Classification rates of cross-database evaluations.

Method	Database		Script	Classification	
	Train	Test		SVM (%)	NN (%)
Proposed method	QUWI	MSHD	Arabic	74.30	67.70
	MSHD	QUWI		67.40	64.00
	QUWI	MSHD	English and French	67.90	66.90
	MSHD	QUWI		65.50	66.10
Siddiqi et al. [1]	QUWI	MSHD	Arabic	72.22	70.04
	MSHD	QUWI		58.13	58.88
	QUWI	MSHD	English and French	62.90	63.49
	MSHD	QUWI		57.87	58.13

a number of interesting experiments as discussed in Section 3. For comparison purposes, these experimental settings are the same as in [1]. It should also be noted that in all experiments no writers are common in the training and test sets so that the problem turly corresponds to gender classification and not writer identification [1].

In addition to the proposed features, we also evaluate state-of-the-art writer identification features for gender classification on the two databases. These include orientation and curvature features, fractal dimension, local binary patterns (LBP) and auto-regressive (AR) coefficients as discussed in Section 1. Likewise, to study the impact of line packing, we compute the classification rates without generation of textured images.

Table 2, summarizes the gender classification rates on the two databases along with the results of some recent studies on the same subject. Classification rates as high as 79.30% and 79.90% are realized on the QUWI and MSHD databases, respectively. Performing a comparative evaluation of the performance of ANN and SVM, Table 2 shows that the classification rates vary by less than 2%. Comparing the results of the proposed technique with and without line packing, it can be seen that extracting features directly from raw images realizes low classification rates. When compared with other gender classification systems reported in the literature [1,30,35], the classification rates of the proposed system outperform those reported in some of the well-known earlier studies. Likewise, the realized classification rates are also better than those achieved by using writer identification features (orientation and curvature, fractal dimension, local binary patterns and auto-regressive coefficients). Among these features, the highest classification rates are realized using orientation and curvature features [22] reading 68.75% and 72.82% for the QUWI and MSHD databases, respectively.

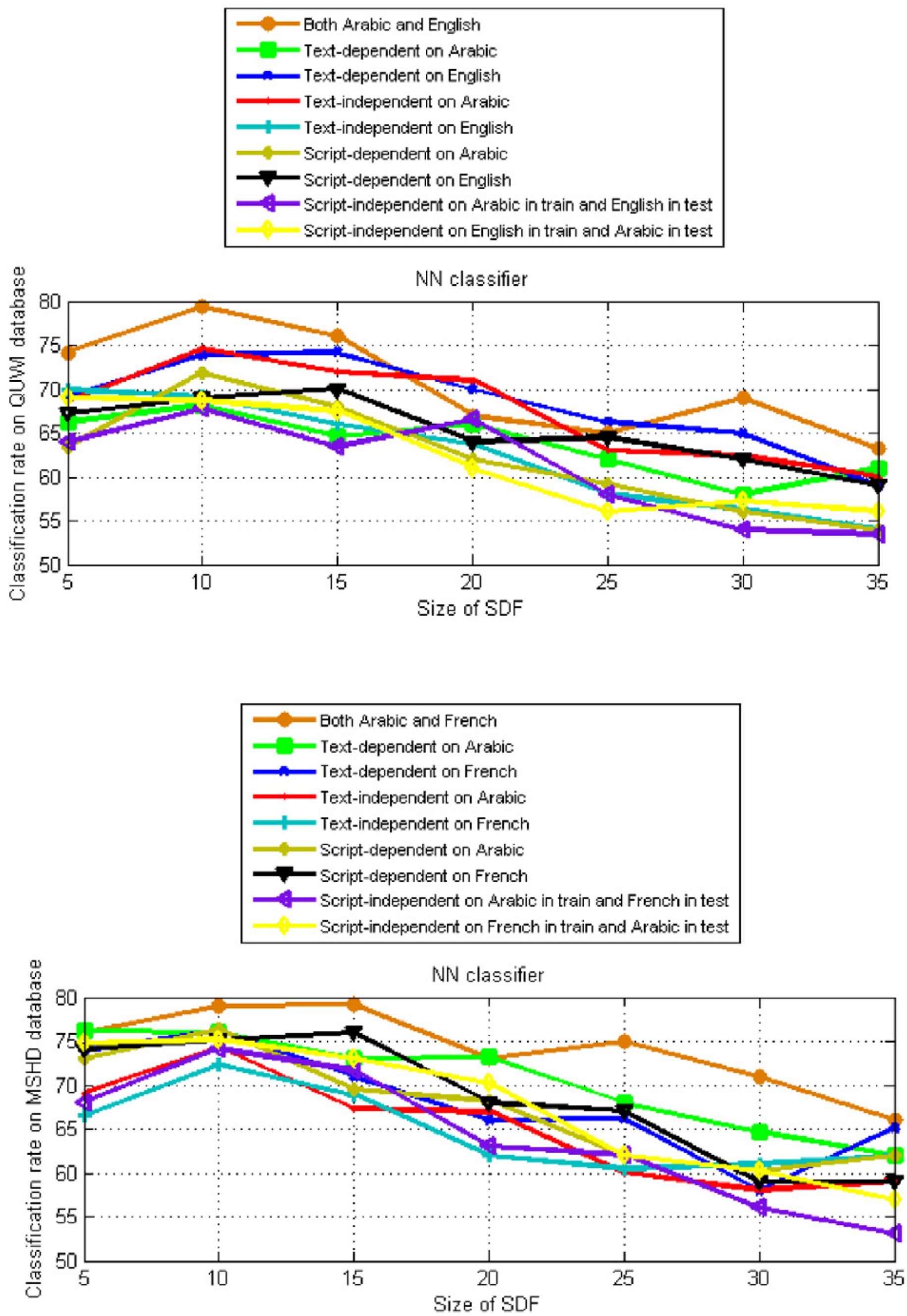

Fig. 7. Classification rates as a function of SDF alphabet size using NN classifier.

We next present the analysis of system performance in text-dependent and text-independent, script-dependent and script-independent and, cross-database evaluations.

5.1. Text-dependent vs. text-independent evaluations

These experiments aim to study the variation of system performance as a function of the textual content of the writing samples. The results of these experiments are presented in Table 3 along with the results of same evaluations reported in [1,16]. In most cases, the two modes (text-dependent and text-independent) report similar results which are better than those reported in [1,16].

5.2. Script-dependent vs. script-independent evaluations

In these experiments, we study how the classification rates vary if the same/different scripts are used as training and test sets. The results of script-dependent and script-independent experiments are summarized in Table 4. Comparing the classification rates across Table 4, the classification rates of the script-dependent evaluations are higher than those of the script-independent experiments. This observation is very much natural and is consistent with previous studies [1]. Having the textual content in the same script in the training and test databases, the realized classification rates are higher than the ones achieved with different scripts in for learning and evaluation. Except one experiment reported in [30] on English samples of QUWI database, the proposed system outperforms other systems in these experiments as well.

5.3. Cross-database evaluations

In the final series of experiments, similar to the experimental protocol presented in [1], samples of one database are used for training

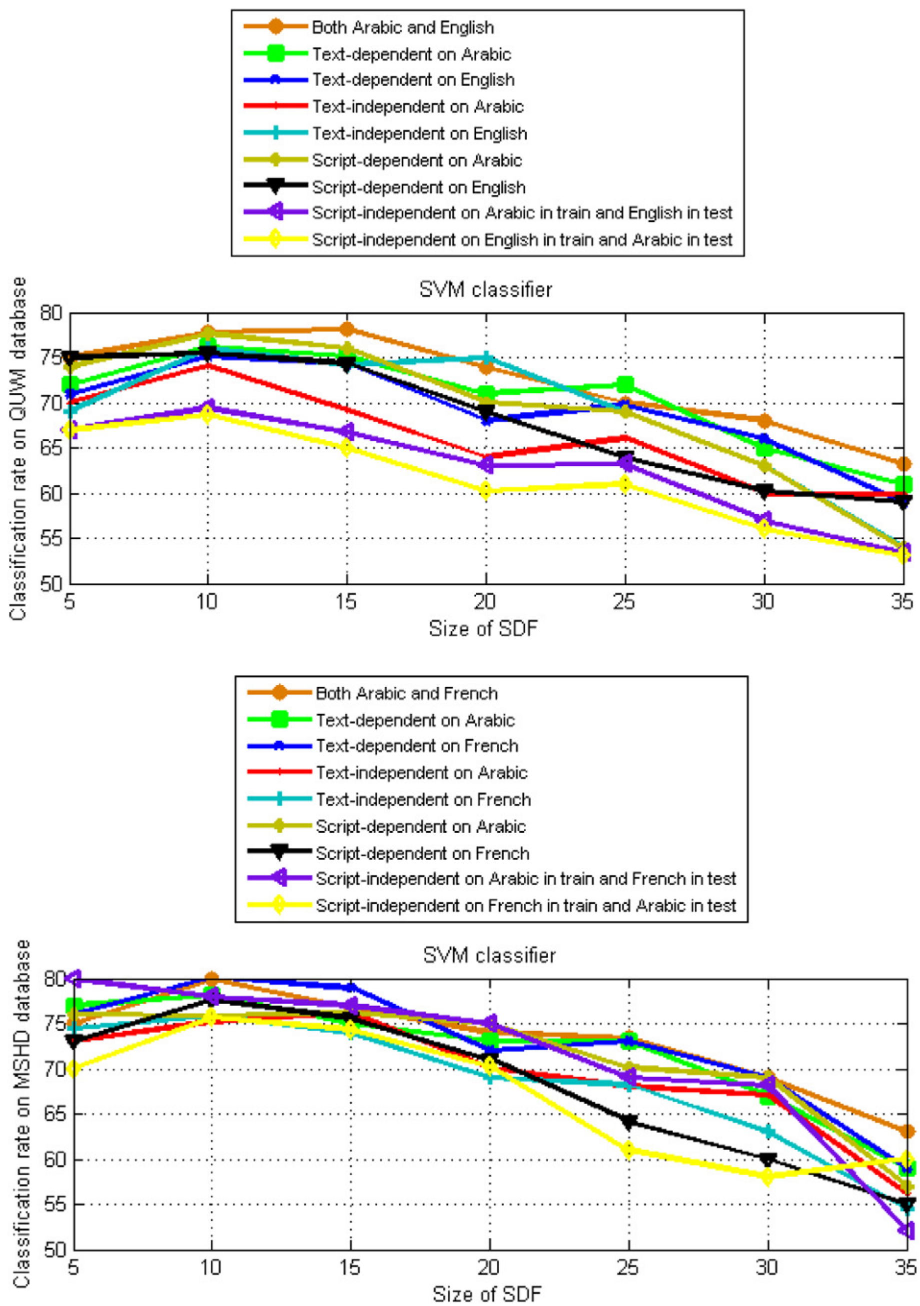

Fig. 8. Classification rates as a function of SDF alphabet size using SVM classifier.

and those of the other database for evaluation. The two databases have been collected in different parts of the world and these experiments would allow studying if individuals belonging to the same gender share common characteristics irrespective of their geographical locations. First, we use the set of Arabic writings in the QUWI database for training and the set of Arabic writings in the MSHD database for evaluation and vice versa. In the next experiments, the complete set of English writings in the QUWI database is used as training base while the complete set of French writings in the MSHD database is employed as the test base. In a similar fashion, the training and test sets are reversed for completeness. These experimental settings are exactly the same as in [1]. The results of these experiments are detailed in Table 5.

The classification rates in Table 5 indicate that acceptably good classification rates are realized when the training and test sets come from two different databases. When QUWI database is used in training, the classification rates are relatively higher as opposed to when the MSHD database is used as training set. This observation is consistent with the findings reported in [1] where the authors attribute this to the fact to the significant difference in the sizes of the two databases. The high classification rates (74% on Arabic and 68% on English–French) validate the idea that individuals of a particular gender share common attributes which are manifested in their handwritings irrespective of the cultural and social circumstances.

5.4. Performance sensitivity with respect to size of SDF alphabet

One of the most important parameters in feature extraction that influences the gender classification performance is the size of SDF alphabet. To study the impact of alphabet size on the classification rates, we carry out a series of experiments on both QUWI and MSHD databases by varying the alphabet size from 5 to 35 with a step size of 5. The classification rates are computed for both classifiers and are illustrated in Figs. 7 and 8. It can be seen that for both databases, alphabet sizes of 5 to 15 yields high classification rates and this observation is consistent for both our classifiers (ANN and SVM).

6. Conclusion and future works

We presented a novel approach for gender detection of writers from images of handwritings. The technique relies on a global texture based approach by considering each writing as a texture. The handwritten images are used to generate texture blocks which are decomposed into a series of wavelet sub-bands. Each of the subbands is then extended into a data sequence that is symbolized to construct a probabilist finite state automata (PFSA). The PFSA is then used to generate the feature vector characterizing gender of the writer of a given handwriting sample. For classification, artificial neural networks and support vector machines are employed. The technique was evaluated on two databases comprising handwritten samples in English, French and Arabic. A series of experiments in a number of challenging scenarios showed encouraging classification rates which outperform results reported in the literature on these databases.

For our further studies on this problem, We aim to enhance the feature set and apply a feature selection technique to find the optimal set of features for this problem. We also plan to include the detection of other demographic attributes of writers from handwritten images. These may include handedness (left or right), age group and race etc. Correlation between handwriting and personal and intellectual attributes of writers can also be studied.

Acknowledgments

The authors are grateful to the referees for their careful reading, insightful comments and helpful suggestions which have led to the improvement of the paper.

References

[1] I. Siddiqi, C. Djeddi, A. Raza, L. Souici-Meslati, Automatic analysis of handwriting for gender classification, Pattern. Anal. Applic. 18 (2014) 887–899.
[2] M. Liwicki, A. Schlapbach, H. Bunke, Automatic gender detection using on-line and off-line information, Pattern. Anal. Applic. 14 (2011) 87–92.
[3] R.A. Huber, Handwriting Identification: Facts and Fundamentals, CRC Press, Boca Raton, 1999.
[4] A. Furnham, B. Gunter, Graphology and personality: another failure to validate graphological analysis, Personal. Individ. Differ. 8 (1987) 433–435.
[5] R.J. Klimoski, A. Rafaeli, Inferring personal qualities through handwriting analysis, J. Occup. Organ. Psychol. 56 (1983) 191–202.
[6] E. Neter, G. Ben-Shakhar, The predictive validity of graphological inferences: a meta-analytic approach, Personal. Individ. Differ. 10 (1989) 735–745.
[7] R.P. Tett, C.A. Palmer, The validity of handwriting elements in relation to self-report personality trait measures, Personal. Individ. Differ. 22 (1997) 11–18.
[8] B. Gawda, A graphical analysis of handwriting of prisoners diagnosed with antisocial personality, Percept. Mot. Skills 107 (2008) 862–872.
[9] R. Mergl, P. Tigges, A. Schröter, H.J. Möller, U. Hegerl, Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives, J. Neurosci. Methods 90 (1999) 157–169.
[10] J. Beech, I. Mackintosh, Do differences in sex hormones affect handwriting style? Evidence from digit ratio and sex role identity as determinants of the sex of handwriting, Personal. Individ. Differ. 39 (2005) 459–468.
[11] F.L. Goodenough, Sex differences in judging the sex of handwriting, J. Soc. Psychol. 22 (1945) 61–68.
[12] S. Hamid, K.M. Loewenthal, Inferring gender from handwriting in Urdu and English, J. Soc. Psychol. 136 (1996) 778–782.
[13] J. Hartley, Sex differences in handwriting: a comment on spear, Br. Educ. Res. J. 17 (1991) 141–145.
[14] W.N. Hayes, Identifying sex from handwriting, Percept. Mot. Skills 83 (1996) 791–800.
[15] V. Burr, Judging gender from samples of adult handwriting: accuracy and use of cues, J. Soc. Psychol. 142 (2002) 691–700.
[16] S. Al-Maadeed, A. Hassaine, Automatic prediction of age, gender, and nationality in offline handwriting, EURASIP J. Image Video Process. (2014) http://dx.doi.org/10.1186/1687-5281-2014-10.
[17] G. Ghiasi, R. Safabakhsh, Offline text-independent writer identification using codebook and efficient code extraction methods, Image Vis. Comput. 31 (2013) 379–391.
[18] N. Morris, Forensic Handwriting Identification Fundamental Concepts and Principles, Academic Press, London, 2000.
[19] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Process. 84 (2004) 1115–1130.
[20] G. Mallapragada, A. Ray, X. Jin, Symbolic dynamic filtering and language measure for behavior identification of mobile robots, IEEE Trans. Syst. Man Cybern. B 42 (2012) 647–659.
[21] X. Jin, S. Sarkar, A. Ray, S. Gupta, T. Damarla, Target detection and classification using seismic and PIR sensors, IEEE Sensors J. 12 (2012) 1709–1718.
[22] I. Siddiqi, N. Vincent, Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features, Pattern Recogn. 43 (2010) 3853–3865.
[23] M. Liwicki, A. Schlapbach, H. Bunke, S. Bengio, J. Mariethoz, J. Richiardi, Writer identification for smart meeting room systems, Proceedings of 7th IAPR Workshop on Document Analysis Systems, 2006. pp. 186–195.
[24] S.N. Srihari, S.H. Cha, H. Arora, S. Lee, Individuality of handwriting, J. Forensic Sci. 47 (2002) 856–872.
[25] K.R. Bandi, S.N. Srihari, Writer demographic classification using bagging and boosting, Proceedings 12th International Graphonomics Society Conference, 2005. pp. 133–137.
[26] V. Bouletreau, N. Vincent, R. Sabourin, H. Emptoz, Handwriting and signature: one or two personality identifiers? ICPR '98: Proceedings of the 14th International Conference on Pattern Recognition, 1998. pp. 1758–1760.
[27] U. Garain, T. Paquet, Off-line multi-script writer identification using ar coefficients, 10th International Conference on Document Analysis and Recognition, 2009. pp. 991–995.
[28] D. Bertolini, L.S. Oliveira, E. Justino, R. Sabourin, Texture-based descriptors for writer identification and verification, Expert Syst. Appl. 40 (2013) 2069–2080.
[29] D. Lester, N. Werling, N.H. Heinle, Graphoanalytic differences by sex and handedness, Percept. Mot. Skills 55 (1982) (1190).
[30] A.E. Youssef, A.S. Ibrahim, A.L. Abbott, Automated gender identification for Arabic and English handwriting, 5th International Conference on Imaging for Crime Detection and Prevention, The Institution of Engineering & Technology, London, UK, 2013.
[31] K. Eames, K. Loewenthal, Effects of handwriting and examiner's expertise on assessment of essays, J. Soc. Psychol. 130 (1990) 831–833.
[32] S.H. Cha, S.N. Srihari, A priori algorithm for sub-category classification analysis of handwriting, Proceedings of the 6th International Conference on Document Analysis and Recognition, 2001. pp. 1022–1025.
[33] S.N. Srihari, S.H. Cha, S. Lee, Establishing handwriting individuality using pattern recognition techniques, International Conference on Document Analysis and Recognition (ICDAR), 2001. pp. 1195–1204.
[34] E. Sokic, A. Salihbegovic, M. Ahic-Djokic, Analysis of offline handwritten text samples of different gender using shape descriptors, Proceedings of IX International Symposium on Telecommunications (BIHTEL), 2012. pp. 1–6.

[35] A. Hassaine, S. Al-Maadeed, J. Aljaam, A. Jaoua, ICDAR 2013 competition on gender prediction from handwriting, Proceedings of IEEE 12th International Conference on Document Analysis and Recognition, 2013. pp. 1417–1421.
[36] S. Al-Maadeed, W. Ayouby, A. Hassaine, J.M. Aljaam, QUWI: an Arabic and English handwriting dataset for offline writer identification, Proceedings of 13th International Conference on Frontiers in Handwriting Recognition, 2012. pp. 746–751.
[37] C. Djeddi, I. Siddiqi, A. Gattal, Y. Chibani, L. Souici-Meslati, H. El Abed, LAMIS-MSHD: a multi-script offline handwriting database, Proc. of the 14th International Conference on Frontiers in Handwriting Recognition, 2014. pp. 93–97.
[38] I. Daubechies, Ten lectures on wavelets, Soc. Ind. Appl. Math. (1992)
[39] Z. He, X. You, Y.Y. Tang, Writer identification using global wavelet-based features, Neurocomputing 71 (2008) 1832–1841.
[40] S. Mallat, S. Zhong, Characterization of signals from multiscale edges, IEEE Trans. Pattern Anal. Mach. Intell. 14 (1992) 710–732.
[41] S. Bahrampour, A. Ray, S. Sarkar, T. Damarla, N.M. Nasrabadi, Performance comparison of feature extraction algorithms for target detection and classification, Pattern Recogn. Lett. 34 (2013) 2126–2134.
[42] V. Rajagopalan, A. Ray, Symbolic time series analysis via wavelet-based partitioning, Signal Process. 86 (2006) 3309–3320.
[43] Z. He, X. You, L. Zhou, Y. Cheung, J. Du, Writer identification using fractal dimension of wavelet subbands in gabor domain, Integr. Comp.-Aided Eng. 17 (2010) 157–165.
[44] R. Hanusiak, L.S. Oliveira, E. Justino, R. Sabourin, Writer verification using texture-based features, Int. J. Doc. Anal. Recognit. 15 (2011) 213–226.
[45] H.E.S. Said, G.S. Peake, T.N. Tan, K.D. Baker, Writer identification from non-uniformly skewed handwriting images, Proc. of British Machine Vision Conference (BMVC), 1998.
[46] P.Y. Simard, D. Steinkraus, J.C. Platt, Best practices for convolutional neural networks applied to visual document analysis, Proceedings of International Conference on Document Analysis and Recognition, 2003. pp. 958–963.
[47] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.