IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 25, 2019, accepted July 3, 2019, date of publication July 8, 2019, date of current version July 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927286

Length Independent Writer Identification Based on the Fusion of Deep and Hand-Crafted Descriptors

ALAA SULAIMAN, KHAIRUDDIN OMAR, MOHAMMAD F. NASRUDIN, AND ANAS ARRAM
Center of Artificial Intelligence Technology, Pattern Recognition Research Group, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia

Corresponding author: Alaa Sulaiman (alaasol@gmail.com)

This work was supported by the Ministry of Higher Education in Malaysia under Grant FRGS/1/2016/ICT02/UKM/02/1.

ABSTRACT Writer's identification from a handwritten text is one of the most challenging machines learning problems because of the variable handwritten sources, various languages, the similarity between writer's pattern, context variation, and implicit characteristics of handwriting styles. In this paper, a combination of the deep and hand-crafted descriptor is utilized to learn patterns from the handwritten images. First, to do so, the local patches are extracted from the handwritten images. Then, these patches are simultaneously fed to deep and hand-crafted descriptors to generate the local descriptions. The extracted local features are then assembled to make the whole description matrix. Finally, by applying the vector of locally aggregated descriptors (VLAD) encoding on the description matrix, a 1-D feature vector is extracted to represent the writer's pattern. It is worthwhile to mention that the generated description does not rely on any language model or context information. Thus, the proposed approach is language and content independent. In addition, the proposed method does not have any restriction on the input length, hence, the writer's sample can be a passage, paragraph, line, sentence, or even a word. The obtained results on three public benchmark datasets of IAM, CVL, and Khatt indicate that the proposed method has a high-accuracy rate in writing identification task. Furthermore, the performance of the proposed method on CVL dataset using both German and English samples demonstrates that the proposed approach has a high capability in learning a writer's pattern from both languages at the same time.

INDEX TERMS Writer identification, deep descriptor, hand-crafted feature, feature fusion, feature length independent.

I. INTRODUCTION

During recent years, writer identification from handwriting text images became an interesting application and a hot research topic in the areas of computer vision and machine learning. Writer identification is to assign a handwriting text image to a certain writer from a set of pre-defined characteristics, i.e., recognizing a person based on his/her handwriting text images. It is usually employed for confirming the authentication of the handwriting text images; such as forensic documents (suspected criminals), a financial district, and historical documents. Moreover, after digitalizing and recognizing a historical text image, an interesting research topic would be to find the authorship of the historical document.

The associate editor coordinating the review of this manuscript and approving it for publication was Xi Peng.

In spite of the large number of approaches which have been proposed for the writer identification, this field of research is still challenging in computer vision and machine learning. This is due to the large intra-class variability and large inter-class similarities in the shape of handwriting text images. This task even becomes more challenging when we are dealing with different languages with completely different structural and statistical properties. There are many languages with different characteristics and properties in the word. Hence, it is obvious that writer identification is challenging because of language variation.

The general pipeline of the writer identification systems involves three steps of preprocessing, feature extraction and classification. The purpose of the preprocessing step is to clean the handwriting (i.e. remove noise), normalize the size of the pieces, and do some operations which contribute to appropriate feature representation [1]. In the classification step, the classifier is trained over the features of training data. Then, in the test phase, it assigns the unknown query pattern to one of the known patterns, i.e., classify the written sample to one of the writers. The main crucial step of writer identification is the feature extraction stage. The feature extraction step is employed to capture useful, efficient, and discriminative information from the handwriting text image. It is aimed at preserving relevant information to distinguish different writers from each other [2]. Feature extraction approaches can be categorized into two groups of handcrafted methods [3], [4] and deep learning based one [5], [6]. Handcrafted features include shape-based features, features extracted from the spatial domain and features extracted from the frequency domain. Handcrafted features are easy to compute and the time complexity of them is feasible. However, most of these approaches are unsupervised. Therefore, their performance may be low in the case of large inter-class similarities and large intra-class variabilities.

During recent years, deep learning networks provided an analysis and a learning of massive amounts of data with state-of-the-art performance in different research fields. The network learns features through learning many non-linear functions, and, simultaneously, the network classifies the handwriting text images. Utilizing the labels for extracting features (i.e., supervised feature extraction method) significantly improves the quality of the features. Literature work has also applied both convolutional neural network (CNN) and recurrent neural network (RNN) with long short-term memory (LSTM) in writer identification systems, which significantly improved the model performance. However, the performance of the deep models largely depends on the availability of the large labeled samples.

In this paper, we propose a method based on the combination of both deep learning-based and handcrafted features for writer identification. To do that, we divided the inputted handwriting text image into some local patches. Then, from all of the local patches, handcrafted features are extracted. In this part, the local binary pattern (LBP) is employed as the feature extractor from the patches. Along with these features, a CNN structure is also utilized to extract deep learning-based features. These features are then assembled to make the whole description matrix. The final feature representation of the writer pattern is computed by encoding the matrix features using VLAD encoding approach. The main properties of this pipeline are that it has no restriction on the input text length (e.g., passage, paragraph, line, or even word), and, more importantly, it is independent of the input language and text content. To sum it up, the proposed method has the following characteristics:

1- Language independent (The proposed approach does not rely on any language model)

2- Content independent (A handwritten text can contain any content and unseen words)

3- Length independent (The train samples can be paragraphs, lines, or words)

4- Variable input sample on the test time (Input for the proposed model on test time can be a paragraph, a sentence or even a word)

Accordingly, the rest of the paper is organized as follows: in section II, we will review the related Work. The proposed method is subsequently explained in Section III. Experimental results are then evaluated in section IV, and, finally, in section V, conclusion and future work will be summarized.

II. RELATED WORK

Writer identification approaches can be both online and offline. In the online writer identification, the input data is captured by some special equipment such as a tablet, stylus, or digital pen for writing. The offline writer identification includes the scanned texts which are written by conventional pen and paper and captured by cameras or scanners. The classification rates of the online approaches are better than that of the offline ones due to the available dynamic information in online data. However, having dynamic information of data (i.e., temporal order) for some applications, like historical documents, is impossible or very difficult to capture. Therefore, we focus on the offline data for writer identification task.

Based on the feature extraction algorithms, writer identification methods can be categorized into two groups: handcrafted features and deep learning based features. Handcrafted features for writer identification is mainly divided into two groups of texture features and shape features. In texture features, the inputted handwriting text is described as a series of texture properties. In shape-based features, handwriting text is indicated as a group of segmented shapes [1]. Texture features include two groups of features: features extracted in the frequency domain and features extracted in the spatial domain. In the first group, the global traits of handwriting are described in the frequency domain while the second group concerns local spatial structures of handwriting [1].

Some methods utilize frequency transform approaches for feature extraction from the whole handwriting text image. A handwriting text image is a grayscale image with particular textures that can be considered as repeated patterns. Thus, texture analysis methods, like frequency transform ones, can be employed to extract features from these images [1]. In [7], multichannel 2D Gabor filters are utilized for writer identification. This filter is a series of Gabor filters with different orientations and spatial frequencies. The authors' extract means values and standard deviations of the filtered images as the features in the frequency domain. Gabor filters are also employed as features for writer identification of Persian [8] and Chinese [9] languages. The extended Gabor filter (XGabor) is also utilized for writer recognition in Persian [10]. He *et al.* [11] proposed a method based on wavelet features and used hidden Markov tree model (HMTM) for classification. Fourier transform is another feature extractor method for writer identification [12].

Compared to the frequency-based features, in the spatial-based features, statistical information of spatial structures is extracted as features; such as edges, key-points, lines, corners, etc. Among all spatial features, gradient-based ones might be the most popular ones. Since these features contain both magnitude and direction information, they can describe the properties of the texture very well. The direction information of the gradient features is employed for writer identification [13]. Gradient-based features along with the chain code based direction features are utilized for Bengali [14] and Arabic [15] writer identification. In [16], the authors utilized edge direction features and edge hinge features which are extracted from a binary image of edge information.

In [17], scale invariant feature transform (SIFT) is used for writer identification for the Arabic language. In [18], each writer is encoded by the root SIFT-based GMM supervectors. From text lines, different kinds of spatial features like the width and the height of text line, slopes of the second and the third line segments, and slant information have been used for writer identification in different languages [19], [20]. Shape-based features utilize the local closed regions of characters as the representation of handwriting text. In [21], the authors proposed the connected component contours for describing allograph. Khalifa *et al.* [22] created an ensemble of graphemes codebooks in feature extraction method. Ghiasi and Safabakhsh [23] utilized the fragments of connected components based on the idea that the fragmented parts of different people's handwriting are quite different. The authors divided the connected components into shorter fragments. Jain and Doermann [24] exploited the contour gradient descriptor (CGD) for shape representation.

Deep learning based approaches have become popular in recent years thanks to their high influence and superb performance in different fields of computer vision approaches. Fiel and Sablatnig [25] proposed the first deep learning based method for writer identification. They utilized Caffe net. The last layer of this network is extracted as the feature vector for writer identification. To compute the similarity of handwriting text images, this vector is then used for distance measurement (Chi-square distance in this method). DCNN [26] is utilized for writer identification from path signature images which consist of six convolutional layers, five pooling layers and two fully connected layers at the top. Inspired by the Dropout idea, the authors introduce the Drop Stroke which randomly omits some strokes in the image of characters. The set of created images along with the original images are used as the input data to the network. Hosoe *et al.* [27] utilized an Auto-Encoder network with the condition of character class to separate the character class information and to extract personal writing style. The encoder part is based on the LeNet, and the decoder part is the layer configuration which is roughly an opposite to the encoder network.

In [28], the authors employed a CNN network for learning local activation features. The extracted features from CNN are then encoded by means of GMM super vector encoding. Multi-Branch Encoding net (Mul-BEnc) [29] is proposed for letter level writer identification problem which takes advantages of both RNN and CNN networks. In this method, the authors used CNN-RNN (CRNN) network to automatically learn an encoding feature. Zhang *et al.* [30] proposed an end-to-end architecture for writer identification by utilizing RNN. The image of a special character is represented by a set of random hybrid strokes (RHSs). Based on imaginary information (pen up and pen down), a series of RHS samples are randomly selected and an RNN model is utilized to encode the RHS strokes.

Many deep learning-based approaches have been proposed to extract local features from images. For instance, Hu *et al.* [31] proposed a method for extracting local features for multi-view discriminative analysis. To do that, the authors divided the input image into overlapping patches. Local feature descriptors (LFD) are then extracted from those patches. To make the descriptor more compact and to preserve the discriminative information, the linear coefficients of the LFDs for different views are calculated. LFDs of different views are then projected into a common space using the Fisher criterion. In the end, the relationship between different views is learned by a view-similarity constraint. Peng *et al.* [32] proposed Structured Auto Encoder (StructAE) to learn a set of explicit transformations to progressively map input data points into nonlinear latent spaces. In that method, subspace clustering is obtained by calculating the structured reconstruction relation from raw data. StructAE preserves the local information by minimizing reconstruction error w.r.t. itself and the global information is preserved by specified reconstruction patterns over the entire dataset.

III. PROPOSED METHOD

In this section, the proposed writer identification method will be presented in more detail. Generally speaking, in writer identification, the scanned handwritten image could be a paragraph, line or word. Thus, it is much desired that the algorithm has the capability to recognize a writer through the variable sample length. In addition, for being a general solver approach, the working algorithm should not be restricted by the passage content or language model. Hence, the purpose of our writer identification algorithm, besides being content and language independent, is to recognize a writer's pattern by using a different length of samples. We consider handwritten line images as an input for our model. In fact, we choose handwritten line images because the number of written words inside the line images may vary from one word to several words; which results in variable input length. The proposed method for writer identification consists of two phases: visual word learning and writer recognition. In the next sub-section, we will discuss the learning and recognition phases in more detail.

A. VISUAL WORD LEARNING PHASE

The general overview of the learning phase is depicted in Figure 1 below. The purpose of the learning phase is to extract visual words from the training set. To do so, the following steps are applied on the input handwritten line images.

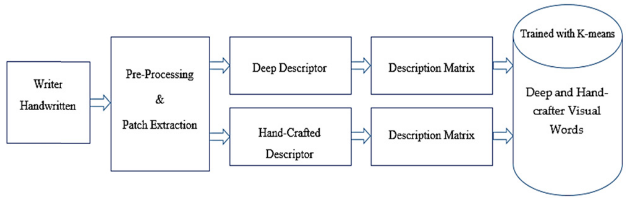

FIGURE 1. General overview of the proposed visual word learning phase.

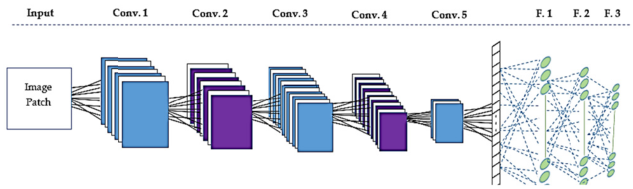

FIGURE 2. Structure of the deep descriptor model, the output of the fifth layer used for feature description.

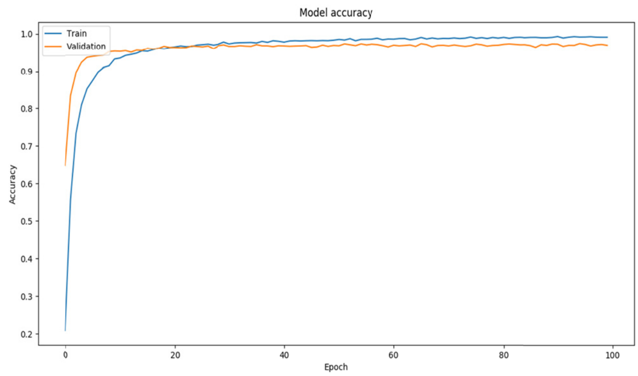

FIGURE 3. Deep model convergence through the learning process.

1) PRE-PROCESSING

The purpose of the preprocessing step is to enhance image quality by eliminating unwanted data; such as: noise, variations and impractical details from image content, and to improve image readability information for generating suitable image [33]. The preprocessing step contains several tasks; such as: background elimination, noise removing and image resizing. In the pre-processing step, we apply binarization and image resizing. It is worthwhile to mention that we only apply row based image resizing, and, we do not change the column size since the input image is a line image with the variable number of words.

2) PATCH EXTRACTION

In many approaches, the input text images are divided into small patches and the features are extracted from those patches for writer identification. For instance, Rehman *et al.* [5] used a sliding window strategy to extract patches from handwriting text images. AlexNet is fine-tuned over the extracted patches. Christlein and Maier [34] randomly sampled four million patches for training. They exploited LeNet for extracting features for writer identification. In this step, we extract overlapped patches from the input handwritten images. To do this, we divide the entire input image into several overlapping patches [35] and, then, for each patch, we apply the feature description step.

3) DESCRIPTOR

A significant step in writer identification task is to extract robust local and discriminative information. To do this, we apply both hand-crafted and deep descriptors on the extracted patches. The LBP algorithm is used as a hand-crafted feature descriptor. The LBP is a local description algorithm that has been utilized in many computer vision applications. The computational speed of this descriptor is very high, and it also has enough discriminative power. It is a binary code for all image pixels which describes the local texture pattern of an image. For calculating this binary code, a predefined circle with constant radius r; ($r > 0$), is centered at every pixel q of the image. Then at each center with neighbors $\mathbb{N}(q.r)$, the LBP is extracted as [35]:

$$LBP\,(q) = \sum_{i=0}^{m-1} f\left(q - q_i^{'}\right) 2^i$$
$$\text{s.t. } q_i^{'} \in \mathrm{N}\,(q.r)$$
$$f\,(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{1}$$

where $\mathbb{N}(q.r)$ is the set of m neighbors in the circle with radius r with center q, and $q_i^{'}$ is its i^{th} neighbors, and $f(x)$ is a function that map $x > 0$ to one and otherwise to zero.

As a second descriptor, we use deep description. We modify the Alex-net [36] structure by changing the input size to be equal to the patch size and reducing the number of filters on the fifth convolutional layer. Therefore, our model input is a grayscale image with a size equal to the patch size. For extracting deep description, we feed each patch to the trained model. Then, we use the output of the target layer as a description vector. Figure 2 below shows the structure of the deep model. We use the output of the Conv. 5 as a target layer for describing the input query.

Since we do not have labels for the patches, we learned the deep model by using auxiliary dataset. To do this, we used Alexuw dataset. Alexuw is a publicly available Arabic handwritten dataset that contains 25114 samples of 109 unique Arabic words [37]. All the samples in this dataset are scanned with a resolution of 600 dpi and saved with tif format. Training, validation and test sets in this dataset are separated and their labels are provided. In our implementation, we trained the model for 100 epochs on the training and validation sets. Figure 3 below shows the model convergence on the training set through the learning process.

4) VISUAL WORDS

The extracted local descriptions are used for learning visual words. To do this, first, both deep and hand-crafted features are extracted from each image patches. These descriptions are then put together to create the description matrix for the input image. The description matrix is extracted for the entire training samples. These description matrices are then

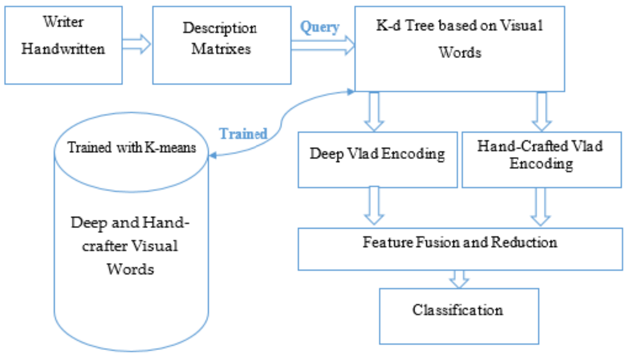

FIGURE 4. **General overview of the proposed writer recognition phase.**

assembled to make the whole description matrix. The purpose of the whole description matrix is to use it for generating visual words. To do this, we apply the k-means clustering algorithm on the whole description matrix, and we cluster each row of the whole description matrix to one of the N clusters as visual words. These visual words are extracted for deep and hand-crafted features separately. Algorithm 1 below shows the steps of generating visual words. The algorithm takes the handwritten line images input along with the desired number of visual words (NVW), and, generates the visual words for both hand-crafted and deep descriptors.

Algorithm 1 Generating Visual Word

```
Input: handwritten line images L, 'NVW
Output: LBP_vw, Deep_vw
L' = resize (L, fixed row)
For s = 1 to N do
    For r = 1 to row samples do
        For c = 1 to column samples do
            patch = window (L', r, c)
    HC_(s,r,c) = LBP(patch)
            Deep_(s,r,c) = CNN(patch)
        End for
    End for
End for
LBP_vw = kmeans(HC_(s,r,c), NVW)
Deep_vw = kmeans(Deep_(s,r,c), NVW)
```

B. WRITER RECOGNITION

The general overview of the writer recognition phase is illustrated in Figure 4. The recognition phase consists of VLAD encoding, feature reduction, and classification.

1) VLAD-ENCODING

VLAD [38] is employed for encoding the extracted deep and hand-crafted features from local patches. VLAD encoding can be considered as a feature mapping method which is mostly employed to transform the features from local patches into a fixed-size vector representation. It is a kind of super vector encoding which encodes samples to the learned visual words, and, then locally aggregates its encoding to generate the feature vector. The visual words are extracted in the first phase for both deep and hand-crafted features. For fast assigning, KD-tree structures are built for both kinds of visual words as shown in Figure 4.

The output of the VLAD descriptor is $v_{i,j}$ where indices $i = 1 \ldots k$ and $j = 1 \ldots d$ respectively index the visual word and the local descriptor component. Then one component of v is calculated as [36]:

$$v_{i,j} = \sum_{x \text{ such that } NN(x)=c_i} x_j - c_{i,j} \tag{2}$$

where x_j and $c_{i,j}$ are the j^{th} component of the descriptor x, and its corresponding visual word c_i, respectively. The vector v is subsequently L2-normalized.

2) FEATURE REDUCTION AND CLASSIFICATION

The extracted VLAD-encoding for both deep and hand-crafted features are concatenated to form the 1-D feature vector. Finally, the representation vector is obtained by applying PCA algorithm. Table 1 below shows the output shape in each step of the recognition algorithm. For classification, a single hidden layer feed-forward neural network is utilized. The ELM [39] method used for learning model weights. The ELM algorithm uses a random value for parameters initialization. Then, it performs an analytical approach for finding the best value for each parameters using training data [39].

IV. EXPERIMENTAL RESULTS

The experimental result of the proposed writer identification method is evaluated on three public benchmarks of IAM, CVL and Khatt datasets. Brief description of the datasets will be provided in sub-section A. The effect of different parameters, such as patch size, number of visual words and PCA components, on the performance of the proposed method will be evaluated in sub-section B. In sub-section C comparison results between the proposed method and state-of-the-art approaches will be discussed in more details. In sub-sections D and E, the results of different fusion level on model generalization performance and path-based learning vs. image-based learning will be elaborated. Finally, in the last sub-section, effect of using key point descriptor as a third feature vector will be evaluated.

A. DATASET

Three public datasets CVL, IAM and Khatt are utilized in the experimental results.

1) CVL DATASET

The CVL dataset contains 7 different passages, which 6 of these passages are in the English language and the last one in the German language [40]. There are 311 writers in CVL datasets and each writer wrote 5 or 7 of the predefined passages. All handwritten texts are scanned and saved in RGB form. Separate line images for this data along with their

TABLE 1. Output shape of the input sample in each step of the describing algorithm.

Step	Input Shape	Output Shape
Input Handwritten Image	$r \times c$	-
Pre-processing	$r \times c$	$r' \times c$
Patch Extracting	$r' \times c$	Total of N patches with constant size $p \times p$. $N = \left(\frac{r'}{p} \times \frac{c}{p}\right) * \text{overlapratio}$
Local Patches	N patches with size $p \times p$	-
Deep Descriptor	N patches with size $p \times p$	N (1-D vector with length d_1)
Hand-crafted Descriptor	N patches with size $p \times p$	N (1-D vector with length d_2)
Description Matrix for deep	N 1-D vector with length d_1	2D Matrix with shape $N \times d_1$
Description Matrix for Hand-crafted	N 1-D vector with length d_2	2D Matrix with shape $N \times d_2$
Deep VLAD Encoding	$N \times d_1$	1D vector length ($d_1 \times NVW$)
Hand-Crafted VLAD Encoding	$N \times d_2$	1D vector length ($d_2 \times NVW$)
Feature Fusion and Reduction	2 1D vector	1-D vector with constant length

FIGURE 5. A sample of handwritten line images in CVL dataset [40].

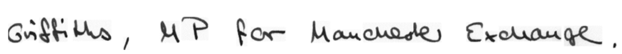

FIGURE 6. A sample of handwritten line images in IAM dataset [41].

writer's ID is provided. A sample of line images in CVL dataset is shown in Figure 5.

2) IAM DATASET

IAM dataset is a publicly available handwritten dataset that consist of 657 writers, each contributed with, at least, one English page. This dataset contains 115320 isolated and labeled words in the English language. All the samples are scanned at 300 DPI and saved in grey scale [41]. Separated line images along with their writer's ID are also provided for this dataset. A sample of line images is shown in Figure 6.

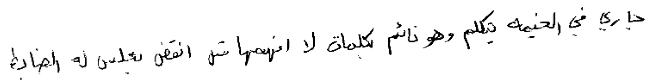

FIGURE 7. A sample of handwritten line images in Khattdataset [42].

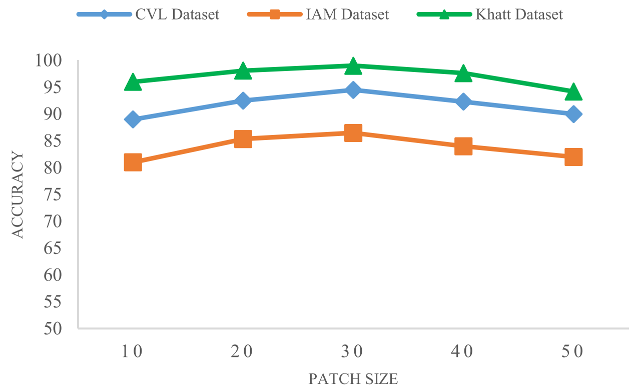

FIGURE 8. Effect of choosing different patch size against accuracy rate.

3) KHATT DATASET

Khatt is public and large Arabic handwritten dataset. This dataset includes samples from 1000 writers. These writers are from different origins and they participated in collecting this large dataset. For each writer, there are six handwritten forms, in which two of these forms are written in an arbitrary topic chosen by the writer [42]. The Khatt dataset designed in a way to cover all the Arabic characters with different style, shape, and variations. Sample of the handwritten line in Khatt dataset is shown in figure 7.

B. PARAMETERS EFFECT

In the proposed method, there are some parameters that contribute to the final performance of the proposed algorithm. Among these parameters: patch size, number of visual words and PCA components which are needed to be selected effectively. To do this, in each dataset, we randomly selected 15% of the training data as a validation set to analyze the effect of choosing different parameters. In the next subsections, we will briefly discuss the effect of using these parameters on the final accuracy rate of the proposed method.

1) PATCH SIZE

One of the important parameters in the proposed method is the window size (patch size) for extracting local information. In fact, the window size should be chosen in a way that best describes the local information, and, in the meantime, preserves the global information. In other words, a too small window size causes the loss of the global information. Whereas a too large window size will result in the loss of the local information. Therefore, there should be a trade-off between local and global information for selecting a window size. In our experiment, we use a square window with a fixed size (30×30), which showed promising results for all CVL, IAM and Khatt datasets. Figure 8 shows the effect of choosing different values for a window size on the final accuracy rate.

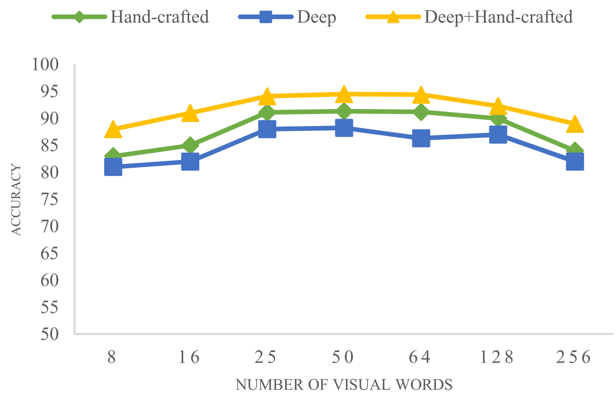

FIGURE 9. Effect of choosing number of visual words on the accuracy rate of CVL dataset.

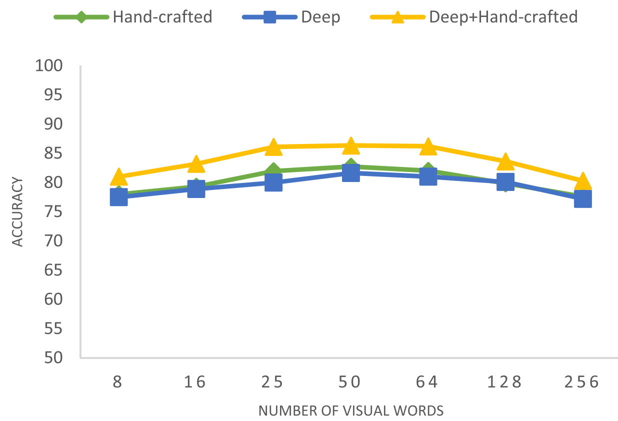

FIGURE 10. Effect of choosing number of visual words on the accuracy rate of IAMdataset.

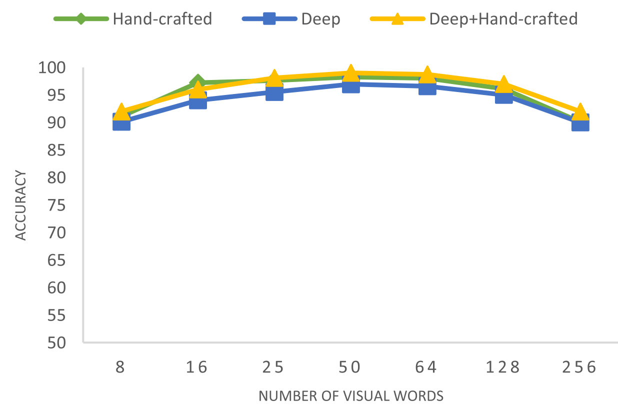

FIGURE 11. Effect of choosing number of visual words on the accuracy rate of Khattdataset.

2) THE NUMBER OF VISUAL WORDS

The second important parameter in the proposed method is related to choosing an appropriate number of visual words. As stated earlier, we utilize two different descriptors for extracting both hand-crafted and deep descriptions from each image patches. These descriptors are needed to be quantized and assigned to the limited number or feature vectors (visual words) to best describe the content information. In fact, the purpose of creating visual words is to assign the extracted features vectors into a limited set of clusters. Each of these clusters (visual words) is designed to describe the different local information. Thus, the number of visual words needs to be chosen wisely. For example, using small amount of visual words does not have the capability in describing different local information. Hence, it is less effective in distinguishing different patterns. From other point of view, a large number of visual words causes noisy information and is less effective in grouping related local features into the same category. Consequently it has less discriminative capability for recognizing different writer's cues. Therefore, the number of visual words should be chosen in a way that contains sufficient local information for discriminating different classes. To do this, in our experimental results we evaluated the effect of choosing the number of visual words on each descriptor and their combination which range from 8 to 256 visual words. Figures 9, 10 and 11 show the effect of choosing different numbers of visual words on the final accuracy rate for the CVL, IAM and Khatt validation sets, respectively.

For all CVL, IAM and Khatt datasets, we achieve the best performance rate on the validation set by using 50 visual words. Hence, we consider 50 visual words in the final implementation. According to the diagrams(figure 9, figure 10 and figure 11), an amount between 8-16 as the number of visual words is less effective in describing different local information and, consequently, has low accuracy rate. Similarly, a too large number of visual words (256) has low accuracy rate since it has a noisy description and contains less discriminative power for separating patterns. For all datasets, the best accuracy rate is achieved with 50 visual words, which shows that, in these datasets, 50 visual words have sufficient discriminative power for distinguishing different writers' patterns.

3) THE NUMBER OF PRINCIPAL COMPONENTS

Another parameter utilized in the proposed method is to select the appropriate number of principal components. In the proposed method, after extracting the VLAD-encoding for both hand-crafted and deep features, we reduced the feature vector by applying PCA algorithm. In fact, the description extracted by VLAD-encoding contains a considerable amount of less discriminative and noisy components. Thus, to enhance the final feature vector and provide more compact representation, it is mandatory to apply a feature reduction algorithm like PCA. In our experimental result, we applied the PCA algorithm on the validation sets of all three datasets to evaluate the feature reduction effect. Figure 12 shows the effect of using a percentage of ordered PCA components on the final accuracy rate.

C. COMPARISION RESULTS

In this section, the performance of the proposed method against the state-of-the-art approaches will be evaluated. As stated earlier, for evaluating the proposed method we used three public datasets: CVL, IAM and Khatt. For all datasets, we use handwritten line images for writer

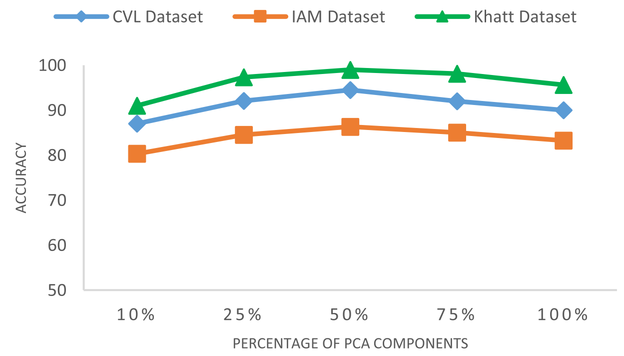

FIGURE 12. Percentage of PCA components vs accuracy rate on both CVL and IAM datasets.

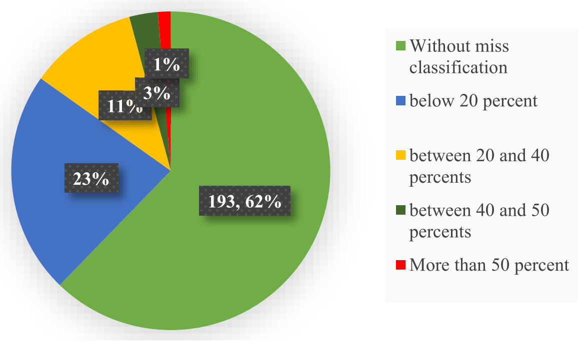

FIGURE 13. Statistical error information of the proposed method on CVL dataset using both English and German languages.

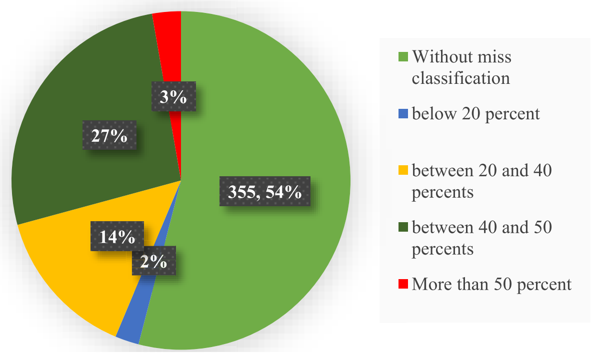

FIGURE 14. Statistical error information of the proposed method on IAM dataset.

identification. We choose line images because each line in the written passage can contain variable numbers of words. In this way, we can evaluate the performance of the proposed method on the input with various length. Since the number of writers in all CVL, IAM and Khatt datasets is so large, providing a confusion matrix is almost impossible for error analyzing. As an alternative approach, we provide statistical information for evaluating the error rate. Also, for comparing the results, we utilize an accuracy matrix:

$$\text{Accuracy} = \frac{Pc}{N} \tag{3}$$

where Pc shows the number of correct classification and N shows the total number of predictions.

As stated earlier, the CVL dataset contains both English and German passages. Hence, some related works such as [43], [44] considered only English passages for evaluating the method, whereas some research work [45], [46] evaluated their method on both English and German passages. Since the purpose of our proposed pipeline is for language-independent writer identification, we consider an evaluation on both languages. Figure 13 shows the statistical information of the proposed method on CVL dataset (both languages). According to figure 13, the whole handwritten samples of the 63% (193) writers are recognized without even a single miss-classification. For the rest of writers, the ratio in the diagram shows the percentages of writers whose samples are incorrectly classified. It is worth to mention that just for 1% percent of the writers, the miss-classification rate was higher than half. Also, the algorithm effectively recognizes samples for most of the writers with zero or low miss-classification rate. The miss-classification rate on this dataset has a mean 0.68 sample per writer, which demonstrates that the proposed algorithm has a high capability in recognizing patterns with low misclassification rate.

Table 2, shows the accuracy rate of the proposed method versus the state-of-the-art approaches on CVL dataset using both English and German passages. The best accuracy rate is achieved by using a combination of deep and hand-crafted features.

According to table 2, it is clear that the hand-crafted features, (LBP) compared to deep features, has a higher accuracy. In fact, the LBP encodes the local information by using a binary pattern of the neighbor information, and has more capability in describing local and content information. On the other hand, the deep descriptor is designed to extract different local information from the image content to be complementary for the hand-crafted feature. The significance of this description is more obvious when we combine deep description with LBP features for improving the performance of the model. Furthermore, the proposed method effectively recognizes a writer's sample by using patterns from two different languages. It is worthwhile to mention that learning a writer's pattern from two different languages at the same time is quite a more difficult task than from single language. Our accuracy rate increased by 3% when we trained the model on only English samples. Our method outperforms SVM combination approach [44] with the same setting.

The statistical error information of the IAM dataset is shown in Figure 14. For more than half of the writers, all of their samples are recognized correctly. The method has approximately 0.49 misclassification sample per writer. This demonstrates a high overall performance rate. Moreover, for 3% of the writers, most of their samples are incorrectly recognized, which shows the high inter-class similarity and intra-class variation in this dataset.

The comparison result of the proposed with state-of-the-art approaches on IAM dataset are shown in table 3. Like CVL dataset, the proposed method obtains the best

TABLE 2. Performance of the proposed method for writer identification on CVL dataset.

Method	Passages	Top 1 Accuracy	Top 5 accuracy
curvature-free features [4]	English and German	12.8	29.6
textural and allographic features [47]	English and German	25.8	48.0
contour-based orientation [45]	English and German	28.8	51.4
Co-occurrence features [3]	English and German	30	52.4
directional ink-trace [46]	English and German	29.4	52.6
Deep adaptive learning [6]	English and German	79.1	93.7
AR coefficient [43]	English	93.87	-
SVM combinations [44]	English	94.83	-
Proposed method with only deep descriptor	English and German	88.24	93.57
Proposed method with only hand-crafted descriptor	English and German	91.33	94.03
Proposed method with both descriptors	English and German	94.5	97.34
Proposed method with both descriptors	English	97.55	99.69

accuracy rate using both deep and hand-crafted features. The top 5 accuracy rates of the proposed method demonstrate that the proposed pipeline has a high capability in finding the most related writers among 657 writers with an acceptable accuracy rate. It is worth mentioning that the deep adaptive learning method [6] was unable to achieve very good results. In that method, the authors utilized an auxiliary part for the network which extracts features for word recognition. The purpose of handwriting word recognition approaches is to recognize an underlying text by extracting the style invariant features to eliminate variations added due to different handwriting styles, while the purpose of writer identification approaches is to identify the style of writers' text. As a result, extracted features from the auxiliary part of the network might not be useful for the writer identification.

TABLE 3. Performance of the proposed method for writer identification on IAM dataset.

Method	Top 1 Accuracy	Top 5 accuracy
curvature-free features [4]	15.7	32.1
textural and allographic features [47]	26.7	45.4
contour-based orientation [45]	21.6	39.7
Co-occurrence features [3]	37.2	57.8
directional ink-trace [46]	35.9	57.8
Deep adaptive learning [6]	69.5	86.1
Edge based features [48]	82.5	-
Proposed method with only deep descriptor	80.65	86.72
Proposed method with only hand-crafted descriptor	80.92	87.32
Proposed method with both descriptors	86.33	96.1

The statistical error information of the Khatt dataset is shown in Figure 15. For 98% of the writers, all of their samples are recognized correctly. The method has approximately 0.01 misclassification sample per writer; which demonstrates a high overall performance rate. Moreover, only 2% of the writers had approximately half of their samples incorrectly recognized. This shows that these writers have similar handwriting style.

The comparison results of the proposed with state-of-the-art approaches on Khatt dataset is shown in table 4. Like CVL and IAM datasets, the proposed method achieves the best performance by using a fusion of deep and handcrafted features. It is manifested that the proposed method outperformed most of the state of the art approaches. The best top 1 accuracy rate is achieved by a combination of GMM super vector and exemplar SVM [52]. However, this method has the same accuracy rate for the top 5 predictions, which demonstrates that this method lacks in estimating the top writers in comparison with the proposed method. In other words, the advantage of

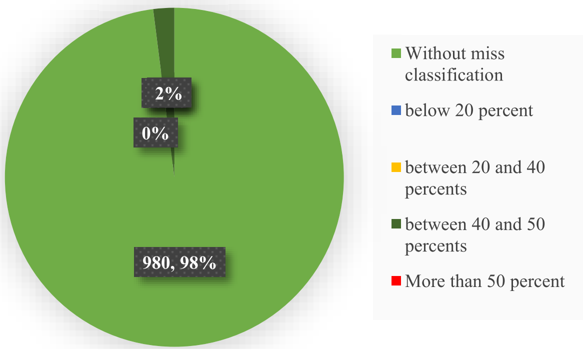

FIGURE 15. Statistical error information of the proposed method on Khattdataset.

TABLE 4. Performance of the proposed method for writer identification on Khatt dataset.

Method	Top 1 Accuracy	Top 5 accuracy
GMM [49]	73.4	84.3
Edge Hings [50]	84.1	91.8
HMM [51]	94.82	-
R-Sift [18]	97.8	99.3
GMM Super vector [52]	96	99.5
GMM Super vector+ESVM [52]	99.5	99.5
Proposed method with only deep descriptor	96.85	99.0
Proposed method with only hand-crafted descriptor	98.25	99.60
Proposed method with both descriptors	99.0	99.70

our proposed method compared to [52] is to retrieve the top writers with a high probability of being true.

D. FUSION LEVEL

The proposed method recognizes writer's pattern with a fusion of deep and hand-crafted features. Generally speaking, feature fusion can perform in 3 different levels as follows:

1- Early fusion
2- Middle fusion
3- Late fusion

In early fusion, first the extracted deep and hand-crafted features are concatenated and then followed by VLAD-encoding and classification steps to perform the recognition. In middle fusion, first of all, VLAD-encoding algorithm is applied on each of the deep and hand-crafted features separately and then the encoded features are concatenated and followed by the classification step. In the late fusion, encoding and classification are done separately for each feature set, and, then, a score of these two classifications is aggregated to perform the final recognition. Experimental results reveal that in both CVL and IAM dataset the middle fusion performs better than early and late fusion. In fact, the domains of the deep and hand-crafted feature sets are quite different from each other, and having an early concatenation without a domain adaptation decreases the recognition performance. Furthermore, extracted deep and hand-crafted features are somehow complementary for each other. Thus, performing late fusion with separate classification will miss this hidden information. On the other hand, in middle fusion, first, the VLAD-encoding performs on both deep and hand-crafted features separately to generate the encoded features. These encoded features are in the same domain. Thus, concatenation in this level no longer has a domain fusion problem. Moreover, as stated earlier these features are complimentary for each other. Therefore, applying the classification step on the concatenated features can capture the hidden information for increasing model generalization performance. Figure 16 shows the accuracy rate achieved by different fusion levels on all datasets.

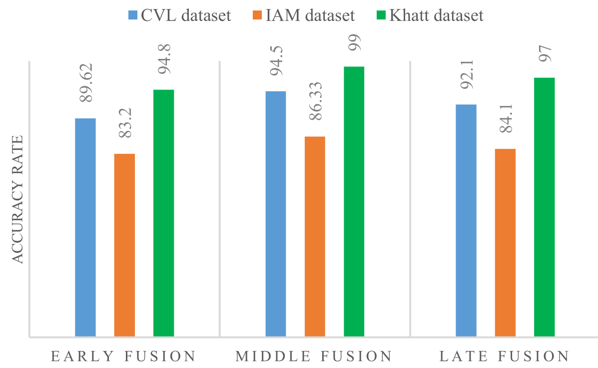

FIGURE 16. Fusion level vs accuracy rate for writer identification.

E. PATCH-BASED VS IMAGE-BASED LEARNING

In the proposed pipeline, we applied patch-based descriptor to extract local features from the input samples. Given the fact that the deep structure has the capability in extracting local features, the main purpose of our patch-based deep descriptor is to cope with a variable length of input samples. Generally speaking, in the handwritten line images the number of written words can vary from one to several words. Thus, the purpose of our algorithm is to recognize the writer's sample with a variable length. Figure 17 shows three different line images of the CVL dataset. In figure 17, the first handwritten line image contains 8 words, while the second and third line images contain only two and one words, respectively. Therefore, it is mandatory to use the patch-based descriptor to cope with a variable length of input samples.

In order to compare the performance of patch-based deep descriptor with the image-based deep descriptor, we only considered complete line images of the English passages for

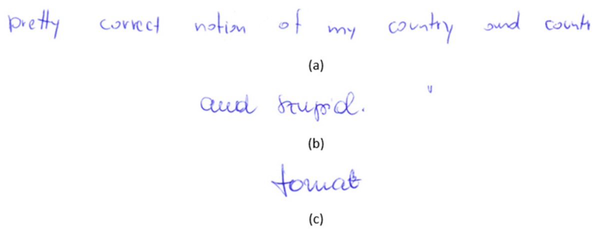

FIGURE 17. **Sample of three different length line images in CVL dataset.**

TABLE 5. **Accuracy rate comparison between patch-based and image-based deep descriptors on full-line images.**

Dataset	Patch-based descriptor	Image-based descriptor
IAM	80.81	78.36
CVL	89.32	86.1
Khatt	97.10	94.21

CVL and IAM and Arabic line images for Khatt datasets. Then, we resized all the samples to the same sizes and applied both patch-based and image-based descriptors. For patch-based descriptor, we used the proposed pipeline, and, for the image-based descriptor, we modified the input size of Alex-net structure and trained the model for 100 epochs. Table 5 shows the experimental results.

It is worthwhile to mention that experimental results demonstrate that using a patch-based descriptor, in comparison to image-based descriptor, is more effective in encoding the writer's pattern. In addition, in the content of writer identification with variable input samples, the image-based descriptor is not feasible and effective compared to the patch-based descriptor.

F. KEYPOINT DESCRIPTOR

In order to analyze the effect of adding the third descriptor to the proposed pipeline, we consider SIFT key-point detector [53]. SIFT algorithm performs feature extraction in three steps: first, it makes a Gaussian pyramid with different octaves and then performs convolution between the input image and the difference of Gaussian kernel in each octave with variable variance. In the second step, the stables key-points are extracted with an analytical approach. For each key-point, information such as location, ordination and scale are extracted. Finally, in the third step, based on the histogram of orientation gradient, a 128-dimensional feature vector is extracted for each key-point. Sample of the extracted key-points for the input handwritten line is shown in figure 18. In our experimental results, we included SIFT features as a third descriptor in the proposed pipeline and followed the same approach on this descriptor. The experimental results revealed that the SIFT key-point detector is less effective in encoding writer's pattern compared to LBP and deep descriptors. Thus, the model generalization performance does not increase when we add this descriptor to the pipeline.

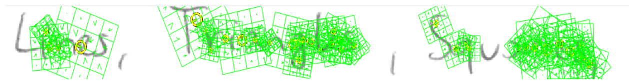

FIGURE 18. **SIFT-key point detector on the input handwritten sample.**

V. CONCLUSION

In this paper, we proposed a combination of hand-crafted and deep features for language independent writer identification. The proposed method extracted both LBP and CNN features from the overlapped patches, and, then, encoded the local information using VLAD algorithm. The proposed method does not rely on any language model, content information or restriction by the input length. Thus, it is independent of language, content and length, and has a high capability in learning complex patterns. The experimental results on two public datasets proved that the proposed method has a high performance in writer identification task. Furthermore, the method is evaluated through the statistical error analyzing approach which illustrated that the model has a high generalization performance. The effect of different parameters on the accuracy rate is evaluated using the validation set. This, accordingly, illustrated that the proposed approach does not have an over fitting with the parameter setting and can recognize writers in different datasets with the same parameter setting. We also evaluated the effect of adding SIFT key-point detector and fusion level on the model performance. In the future, we plan to extend this method for deep end-to-end recognition. More specifically, we will implement a deep model as an alternative for Vlad-encoding in order to combine it with our deep descriptor as an end-to-end network.

REFERENCES

[1] Y.-J. Xiong, Y. Lu, and P. S. P. Wang, "Off-line text-independent writer recognition: A survey," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31., no. 5, pp. 1–30, 2017.
[2] C. Adak, B. B. Chaudhuri, and M. Blumenstein, "An empirical study on writer identification and verification from intra-variable individual handwriting," *IEEE Access*, vol. 7, pp. 24738–24758, 2019.
[3] S. He and L. Schomaker, "Co-occurrence features for writer identification," in *Proc. 15th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Oct. 2016, pp. 78–83.
[4] S. He and L. Schomaker, "Writer identification using curvature-free features," *Pattern Recognit.*, vol. 63, no. 1, pp. 451–464, 2016.
[5] A. Rehman, S. Naz, M. I. Razak, and I. A. Hameed, "Automatic visual features for writer identification: A deep learning approach," *IEEE Access*, vol. 7, pp. 17149–17157, 2019.
[6] S. He and L. Schomaker, "Deep adaptive learning for writer identification based on single handwritten word images," *Pattern Recognit.*, vol. 88, no. 1, pp. 64–74, 2019.
[7] H. E. S. Said, T. N. Tan, and K. D. Baker, "Personal identification based on handwriting," *Pattern Recognit.*, vol. 33, no. 1, pp. 149–160, 2000.
[8] F. Shahabi and M. Rahmati, "A new method for writer identification of handwritten Farsi documents," in *Proc. 10th Int. Conf. Document Anal. Recognit.*, Jul. 2009, pp. 426–430.
[9] Y. Zhu, T. Tan, and Y. Wang, "Biometric personal identification based on handwriting," in *Proc. 15th Int. Conf. Pattern Recognit.*, Sep. 2000, pp. 797–800.
[10] B. Helli and M. E. Moghaddam, "A text-independent Persian writer identification system using LCS based classifier," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol.*, Dec. 2008, pp. 203–206.
[11] Z. He, X. You, and Y. Y. Tang, "Writer identification of Chinese handwriting documents using hidden Markov tree model," *Pattern Recognit.*, vol. 41, pp. 1295–1307, Apr. 2008.

[12] 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.
[13] S. S. Ram and M. E. Moghaddam, "A Persian writer identification method based on gradient features and neural networks," in *Proc. 2nd Int. Congr. Image Signal Process.*, Oct. 2009, pp. 1–4.
[14] S. Chanda, K. Franke, U. Pal, and T. Wakabayashi, "Text independent writer identification for bengali script," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2005–2008.
[15] S. M. Awaida and S. A. Mahmoud, "Writer identification of arabic text using statistical and structural features," *Cybernetic Syst. Int. J.*, vol. 44, no. 1, pp. 57–76, 2003.
[16] M. Bulacu, L. Schomaker, and L. Vuurpijl, "Writer identification using edge-based directional features," in *Proc. Int. Conf. Document Anal. Recognit.*, 2003, pp. 937–941.
[17] D. Fecker, A. Asit, V. Märgner, J. El-Sana, and T. Fingscheidt, "Writer identification for historical arabic documents," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 3050–3055.
[18] V. Christlein, D. Bernecker, F. Hönig, and E. Angelopoulou, "Writer identification and verification using GMM supervectors," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2014, pp. 998–1005.
[19] A. Rafiee and H. Motavalli, "Off-line writer recognition for farsi text," in *Proc. 6th Mex. Int. Conf. Artif. Intell., Special Session*, Nov. 2007, pp. 193–197.
[20] A. J. Newell and L. D. Griffin, "Writer identification using oriented basic image features and the delta encoding," *Pattern Recognit.*, vol. 47, no. 6, pp. 2255–2265, 2014.
[21] A. Bensefia, A. Nosary, T. Paquet, and L. Heutte, "Writer identification by writer's invariants," in *Proc. 8th Int. Workshop Frontiers Handwriting Recognit.*, Aug. 2002, pp. 274–279.
[22] E. Khalifa, S. Al-maadeed, M. A. Tahir, A. Bouridane, and A. Jamshed, "Off-line writer identification using an ensemble of grapheme codebook features," *Pattern Recognit. Lett.*, vol. 59, no. 1, pp. 18–25, 2015.
[23] G. Ghiasi and R. Safabakhsh, "An efficient method for offline text independent writer identification," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 1245–1248.
[24] R. Jain and D. Doermann, "Writer identification using an alphabet of contour gradient descriptors," in *Proc. 12th Int. Conf. Document Anal. Recognit.*, Aug. 2013, pp. 550–554.
[25] S. Fiel and R. Sablatnig, "Writer identification and retrieval using a convolutional neural network," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2015, pp. 26–37.
[26] W. Yang, L. Jin, and M. Liu, "Chinese character-level writer identification using path signature feature, DropStroke and deep CNN," in *Proc. 13th Int. Conf. Document Anal. Recognit.*, Aug. 2015, pp. 546–550.
[27] M. Hosoe, T. Yamada, K. Kato, and K. Yamamoto, "Offline text-independent writer identification based on writer-independent model using conditional AutoEncoder," in *Proc. 16th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Aug. 2018, pp. 441–446.
[28] V. Christlein, D. Bernecker, A. Maier, and E. Angelopoulou, "Offline writer identification using convolutional neural network activation features," in *Proc. German Conf. Pattern Recognit.*, 2015, pp. 540–552.
[29] C. Zelin, Y. Hong-Xing, W. Ancong, and Z. Wei-Shi, "Letter-level writer identification," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 381–388.
[30] X.-Y. Zhang, G.-S. Xie, C.-L. Liu, and Y. Bengio, "End-to-end online writer identification with recurrent neural network," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 2, pp. 285–292, Apr. 2016.
[31] P. Hu, D. Peng, J. Guo, and L. Zhen, "Local feature based multi-view discriminant analysis," *Knowl.-Based Syst.*, vol. 149, pp. 34–46, Jun. 2018.
[32] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. T. Zhou, and S. Yang, "Structured autoencoders for subspace clustering," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5076–5086, Oct. 2018.
[33] A. Sulaiman, K. Omar, and M. F. Nasrudin, "Degraded historical document binarization: A review on issues, challenges, techniques, and future directions," *J. Imag.*, vol. 5, no. 4, p. 48, 2019.
[34] V. Christlein and A. Maier, "Encoding CNN activations for writer recognition," in *Proc. 13th IAPR Int. Workshop Document Anal. Syst. (DAS)*, Apr. 2018, pp. 169–174.
[35] R. Azad, M. Asadi-Aghbolaghi, S. Kasaei, and S. Escalera, "Dynamic 3D hand gesture recognition by learning weighted depth motion maps," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 6, pp. 1729–1740, Jun. 2019.
[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. NIPS Conf.*, 2012, pp. 1097–1105.
[37] M. E. Hussein, M. Torki, A. Elsallamy, and M. Fayyaz, "AlexU-word: A new dataset for isolated-word closed-vocabulary offline arabic handwriting recognition," Nov. 2014, *arXiv:1411.4670*. [Online]. Available: https://arxiv.org/abs/1411.4670
[38] H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *Proc. IEEE Computer Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 3304–3311.
[39] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.
[40] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig, "CVL-DataBase: An off-line database for writer retrieval, writer identification and word spotting," in *Proc. 12th Int. Conf. Document Anal. Recognit.*, Aug. 2013, pp. 560–564.
[41] U.-V. Marti and H. Bunke, "The IAM-database: An English sentence database for offline handwriting recognition," *Int. J. Document Anal. Recognit.*, vol. 5, no. 1, pp. 39–46, 2002.
[42] S. A. Mahmoud, I. Ahmad, W. G. Al-Khatib, M. Alshayeb, M. T. Parvez, V. Märgner, and G. A. Fink, "KHATT: An open Arabic offline handwritten text database," *Pattern Recognit.*, vol. 47, no. 3, pp. 1096–1112, 2014.
[43] U. Garain and T. Paquet, "Off-line multi-script writer identification using AR coefficients," in *Proc. 10th Int. Conf. Document Anal. Recognit. (ICDAR)*, Jul. 2009, pp. 991–995.
[44] Y. Kessentini, S. BenAbderrahim, and C. Djeddi, "Evidential combination of svm classifiers for writer recognition," *Neurocomputing*, vol. 313, pp. 1–40, Nov. 2019.
[45] I. Siddiqi and N. Vincent, "Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features," *Pattern Recognit.*, vol. 43, no. 11, pp. 3853–3865, 2010.
[46] A. A. Brink, J. Smit, M. L. Bulacu, and L. R. B. Schomaker, "Writer identification using directional ink-trace width measurements," *Pattern Recognit.*, vol. 45, no. 1, pp. 162–171, 2012.
[47] 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. 701–717, Apr. 2007.
[48] L. Schomaker and M. Bulacu, "Automatic writer identification using connected-component contours and edge-based features of uppercase Western script," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 787–798, Jun. 2004.
[49] F. Slimane, S. Awaida, A. Mezghani, M. T. Parvez, S. Kanoun, S. A. Mahmoud, and V. Märgner, "ICFHR2014 competition on arabic writer identification using AHTID/MW AND KHATT databases," in *Proc. 14th Int. Conf. Frontiers Handwriting Recognit.*, Sep. 2014, pp. 797–802.
[50] C. Djeddi, L.-S. Meslati, I. Siddiqi, A. Ennaji, H. El Abed, and A. Gattal, "Evaluation of texture features for offline arabic writer identification," in *Proc. 11th IAPR Int. Workshop Document Anal. Syst.*, Apr. 2014, pp. 106–110.
[51] A. Cheikh, R. Email, Z. Abdelhedi, and Y. Kessentini, "A HMM-based arabic/latin handwritten/printed identification system," in *Proc. 16th Int. Conf. Hybrid Intell. Syst.*, 2016, pp. 298–307.
[52] V. Christlein, D. Bernecker, F. Hönig, A. Maier, and E. Angelopoulou, "Writer identification using GMM supervectors and exemplar-SVMs," *Pattern Recognit.*, vol. 63, pp. 258–267, Mar. 2016.
[53] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

ALAA SULAIMAN received the B.A. degree in electrical and computer engineering from An-Najah National University, Nablus, Palestine, in 2003, and the M.Sc. degree in scientific computing minor in computer science and engineering from Bierzit University, Ramallah, Palestine, in 2010. He is currently pursuing the Ph.D. degree from the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Malaysia. His areas of interest includes deep learning, machine learning, image processing, and pattern recognition.

KHAIRUDDIN OMAR received the B.Sc. and M.Sc. degrees in computer science from Universiti Kebangsaan Malaysia (UKM), in 1986 and 1989, respectively, and the Ph.D. degree from Universiti Putra Malaysia, in 2000. He is currently a Professor with the Center for Artificial Intelligence Technology (CAIT), Pattern Recognition Research Group, Faculty of Technology and Information Science, UKM. He has authored over 220 articles and over 1500 citations. His research interests include artificial intelligence, pattern recognition, image processing, and feature extraction.

MOHAMMAD F. NASRUDIN received the B.B.A. degree *(cum laude)* in computer information system (CIS) minor in business and the master's degree in information technology (computer science) from Universiti Kebangsaan Malaysia (UKM), in 1999 and 2001, respectively, and the Ph.D. degree from the Faculty of Information Science and Technology, UKM, in 2010, where he is currently an Associate Professor with the Center for Artificial Intelligence Technology (CAIT), Pattern Recognition Research Group. His current research interests include document analysis and recognition, and robotics.

ANAS ARRAM received the Ph.D. degree in computer science from the National University of Malaysia, in 2017. He was a member of the Centre for Artificial Intelligent (CAIT), Data Mining and Optimization Research Group (DMO), National University of Malaysia. He is currently a lecturer with different Palestinian universities. He is currently a Computer Vision Engineer with PrevPain. His main research areas include meta-heuristics, hyper-heuristics, scheduling and timetabling, machine learning, and deep learning. He has published over six papers at international journals peer-reviewed international conferences. He has received the Outstanding Researcher Award upon completing the Ph.D. degree.