A novel framework for writer identification based on pre-segmented Gurmukhi characters

MUNISH KUMAR^{1,*}, M K JINDAL², R K SHARMA³ and SIMPEL RANI JINDAL⁴

¹Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

²Department of Computer Science and Applications, Panjab University Regional Centre, Muktsar, India

³Department of Computer Science and Engineering, Thapar University, Patiala, India

⁴Computer Science and Engineering, Yadavindra College of Engineering, Talwandi Sabo, Bathinda, India
e-mail: munishcse@gmail.com; manishphd@rediffmail.com; rksharma@thapar.edu;
simpler_jindal@rediffmail.com

MS received 28 July 2017; revised 29 December 2017; accepted 28 July 2018; published online 30 October 2018

Abstract. Handwriting is an obtained apparatus utilized for correspondence of one's recognition or sentiments. Components that judge a person's handwriting is not merely subject to the individual's handwriting depends on the background, additionally considers like nervousness, inspiration and the reason for the handwriting. In spite of the high variation, in a man's handwriting, recent outcomes from various writers have demonstrated that it has adequate individual quality to be utilized as an identification strategy. In this paper, the authors are the pact with a novel approach to text dependent writer identification in view of pre-segmented Gurmukhi characters. The text dependent writer identification framework proposed in this paper includes distinctive stages like preprocessing, feature extraction, classification or identification. The feature extraction stage incorporates four schemes, zoning, diagonal, transitions and peak extent based features. To analyze the proposed framework execution, experiments are performed with two classifiers, namely, k-NN and SVM. SVM is also considered with linear-kernel in the present work. For experimental results, we have collected 31,500 samples from 90 different writers for 35 class problem. Maximum writer identification accuracy of 89.85% has been achieved by using a combination of zoning, transition and peak extent based features with Linear-SVM classifier when we have taken 70% data as the training set and remaining 30% data as the testing set. Using 10-fold cross validation, we have achieved an accuracy of 94.76% with a combination of zoning, transition and peak extent based features and Linear-SVM classifier.

Keywords. Feature extraction; writer identification; zoning; diagonal; transitions; peak extents; k-NN; SVM.

1. Introduction

Nowadays writer identification is a champion among the most captivating and testing, research regions in the field of pattern recognition. Notwithstanding the way that, sufficient reviews have been performed for writer identification framework in a few of the non-Indic scripts like Arabic, Chinese, Roman and Indic scripts, for example, Bangla, Devanagari and so forth, yet the limited work is accessible for writer identification based on the characters of the Gurmukhi script. In writer identification framework, databases with handwritten samples of known personalities are taken as training dataset, and afterward compare the questioned handwritten sample with that training dataset to recognize the identity of the writer. Writer identification can be offline writer identification or online writer

identification. Here, we have considered offline writer identification framework. In this framework, the scanned images are consequently changed into ASCII codes which are normally used as a part of information processing. In the wake of being put away, it is commonplace to perform further processing to allow superior recognition. In online writer identification framework, the extraordinary pen device is utilized, which involves the pen up and pen down exchanging developments that should be possible by using sensors on the exceptional digitizer. Writer identification frameworks are further classified as text dependent and text independent methods. In text dependent methodology, a writer needs to compose the same content to perform recognition, but in text independent methodology, any text might be utilized to build up the identity of the writer [1]. The recognition rates of writer identification framework rely on the number of writers, the type of handwriting, the span of training and testing dataset. The different uses of

*For correspondence

writer identification framework are; a forensic record examination, authentication of customers in the banking system, biometric recognition, etc. In this paper, a novel framework for text dependent writer identification has been presented based on isolated handwritten Gurmukhi characters. In this work, the author is exploring various features and classifiers for writer identification in light of the Gurmukhi text handwriting. The writer identification framework proposed in this work includes diverse stages like digitization, image preprocessing, feature extraction and classification. The framework first prepares as a skeleton of the character so that meaningful data about the handwriting of writers can be extracted. Various feature extraction techniques, namely, zoning, diagonal, transition and peak extent based features have been considered for extracting the properties of handwriting by individuals. Different combinations of these feature extraction techniques are also explored for improving the identification accuracy. In order to assess the prominence of these features, we have used two classification techniques, namely, Support Vector Machine (SVM) and k -Nearest Neighbours (k -NN). This paper is divided into 8 sections. The introduction has been presented in section 1. Section 2 illustrates the related work. Characteristics of Gurmukhi script and data collections are dealt with in section 3. Motivation and challenges of the proposed work are discussed in section 4. Section 5 presents the proposed framework for writer identification. Section 6 incorporates experimental results and discussions. State-of-the-art work with previous work is presented in section 7. Finally, conclusion and future work are presented in section 8.

2. Previous work

Awaida and Mahmoud [2] have introduced an overview on writer identification work for Arabic handwritten documents. They have talked about different feature extraction techniques and different classifiers like minimum distance, statistical classifiers which are used by different people and distinctive research teams in their work. Akbari *et al* [1] have proposed another approach for writer identification framework for Persian script. The features used are wavelet transformation and co-occurrence matrix and k -NN classifier for classification. They have accomplished a recognition rate of 93.3%. Maadeed [3] has proposed a writer identification framework based on Arabic handwritten text. He has compared edge direction distribution features with other features and used k -NN for recognition. Brink *et al* [4] have used the ink width, direction feature, brush feature, white runs feature, hinges and franglet features for writer identification. They have used the nearest neighbor as a classifier and acquired exactness in the range of 63–95% with hinge and franglets and 70–97% with quill hinge features. Ghiasi and Safabakhsh [5] have proposed an

effective technique for text independent writer recognition using a code book. They have utilized Farsi database which incorporates short, medium and extensive messages and results demonstrate that the productivity of short messages is more effective. Leng and Shamsuddin [6] have done research on handwriting identification on the Chinese text. They examined the literature about the quality and shortcoming of all feature elements which are used as a part of writer identification in various scripts. Pratama *et al* [7] have presented various filters and wrapper feature selection for writer identification by using an IAM standard database. The classifiers Naive Bayesian, Random Forest and Negative Selection algorithms are used with important feature selection technique. These strategies are analyzed and result acquired is only up to 50%. Bulacu *et al* [8] have utilized the combination of textural features with allo-graphic features to distinguish and verify the writer in view of offline Arabic script. Gazzah and Amara [9] have proposed an approach for writer identification framework by using offline handwritten Arabic script documents. Their proposed technique is depending on combining the global and neighbourhood feature sets by utilizing the genetic algorithm so as to take out the redundant and irrelevant features. They have utilized two classifiers SVM and MLP for recognition. The outcome acquired shows that MLP performs preferable outcomes over SVM and got precision of around 94%. Sreeraj and Idicula [10] introduce a review on different writer identification schemes. Sreeraj and Idicula [11] introduced a writer identification framework for Malayalam reports. They have utilized 8 features like loop roundness, slant, direction, angle, relative height and width/height ratio of loops, direction, angle, slant, curvature of the character and distance from the centroid. They have accomplished 95.62% precision by using k -NN classifier. Additionally, they utilized WD-LBP feature and acquired accuracy of 70.35%. Jain and Doermann [12] have presented a new method for writer identification, which emulates the approach taken by forensic document examiners. They combine a novel feature, which uses contour gradients to capture local shape and curvature, with character segmentation to create a pseudo-alphabet for a given handwriting sample. Fiel and Sablatnig [13] have presented writer identification and retrieval system using a Convolutional Neural Network. The method presented is using Convolutional Neural Networks (CNN) to generate a feature vector for each writer, which is then compared with the pre-calculated feature vectors stored in the database. He and Schomaker [14] have proposed two novel and curvature-free features: run-lengths of local binary pattern (LBPruns) and cloud of line distribution (COLD) features for writer identification. The LBPruns is the joint distribution of run-length and local binary pattern and the COLD is the joint distribution of the relation between orientations and lengths of line segments. The combination of the LBPruns and COLD features provides a promising results on the curvature-less CERUG data set. Christlein *et al* [15]

have faced a problem in the case of writer identification, where the writers of the training set are different from those in the test set of the typical used benchmark datasets. On top of that, current datasets have only one to five images per writer. While a form of writer adaptation with exemplar Support Vector Machines (E-SVM) is possible. Adak and Chaudhuri [16] have presented a framework for writer identification based on isolated Bangla characters and numerals. Initially, they detected key points by structural analysis and SIFT based detector. After that, they calculated a set of features within a certain neighbourhood of the key point and employ fusion rule on multiple probabilistic SVM classifier output for writer identification. We have seen that the vast majority of work has been done for Arabic, Chinese and so forth scripts, however, no review is accessible for writer identification of Gurmukhi script. In the literature of the Gurmukhi script, few systems are available for character recognition. For example, Lehal and Singh [17] have presented a hybrid classification scheme for printed Gurmukhi script recognition. Kumar *et al* [18] have presented a novel feature extraction technique for offline handwritten Gurmukhi character recognition. They have also presented efficient feature extraction techniques based on curvature features for offline handwritten Gurmukhi character recognition [19]. Kumar *et al* [20] have also presented a character recognition using principal component analysis. In this paper, we have exhibited a novel methodology for text dependent writer identification framework in light of segmented handwritten Gurmukhi characters.

3. Gurmukhi script and data collection

In Gurmukhi script, there are 35 basic character constants out of which the initial three are vowel bearers. There are six consonants which are made by setting a speck (*Bindi*) at the foot (match) of the consonant which are called Punjabi constants with subscript dots. Ten vowels are utilized as a part of the Gurmukhi script. In this work, we have considered 35 fundamental Gurmukhi characters. Samples of these characters are collected from schools, colleges, public places, etc. written by 90 different writers. 10 samples of each character written by each writer are taken. In this manner, we have collected 31,500 samples of 35-class problem.

4. Motivation and challenges

The main motivations of the present work of the writer identification system is a reading help for forensic record examination, authentication of customers in the banking system, biometric recognition, etc. Various challenges are identified which may provide interest to the researchers for

writer identification based on their handwriting styles. These challenges are: difficulty to identify the various styles of human writing, different angles of letters, different shapes and size of letter, pure input quality, low accuracy rate in recognition, constrained and unconstrained handwriting, etc. The quality of the input document and varying font styles is a challenging task for Indic scripts. As compared to non-Indic scripts, Indic scripts have many additional challenges like larger character set due to modifiers, lack of standard test databases, etc. Poor quality document consisting of holes, noise, spots, broken strokes, etc. is also a major challenge in writer identification task.

5. Proposed writer identification system

Proposed writer identification framework includes different stages, like, digitization, preprocessing, feature extraction and classification. The block diagram of the proposed framework is depicted in figure 1.

5.1 Digitization

It is the way toward changing over the paper based handwritten character in a digital image. This procedure is accomplished by using a scanner with 300 dpi. Digitization process produced the digitized image which is fed to the preprocessing phase for further processing.

5.2 Preprocessing

In preprocessing stage, it regularly incorporates numerous operations which will be applied to a character image. In this stage, initially the character image is normalized into 100×100 pixels by using the nearest neighborhood interpolation scheme and after that change over that image into a bitmap image as [0, 1]. Zhang and Suen [21] algorithm has been considered in this phase for reducing the width of text to only single pixel width by using parallel thinning technique.

5.3 Feature extraction

Feature extraction stage is the most imperative stage in writer recognition framework as it is used to create the list of features for writer identification based on pre-segmented character image. It gives the pertinent data about that character and perceives and classifies that writer with the least measure of time and calculation. In this paper, we have considered zoning features, diagonal features, transition features and peak extent based feature elements for writer identification in view of Gurmukhi characters. These

Figure 1. Block diagram of writer identification system.

feature extraction procedures are briefly discussed in subsections 5.3a–5.3d.

5.3a Zoning features (F_1): In this methodology, we divide the thinned image of a character into n ($=100$) number of equivalent estimated zones. After that, number of foreground pixels in each zone are calculated. These numbers $p_1, p_2, \dots, p_n$, acquired for all n zones, are standardized to $[0, 1]$ coming about into a list of capabilities of n components [22].

Various steps for zoning based feature extraction are:

- Step I Input thinned bitmapped image of pre-segmented Gurmukhi character
- Step II The thinned character image is divided into n ($=100$) numbers of equal sized zones

- Step III For each zone, pixel density is calculated using the number of object pixels in each region divided by the total number of pixels in each zone is given by

$$\text{Density} = \frac{\text{Number of Object Pixels in Every region}}{\text{Total Number of Pixels}}$$

- Step IV Normalize the values of the feature vector

Using these steps, n number of features extracted for each image of the character.

5.3b Diagonal features (F_2): In this technique, we divide the original thinned image of a character into n ($=100$)

number of proportionate evaluated zones. These features are taken out by moving along diagonals of the pixels of each zone. Each zone has $2n - 1$ diagonals and ON (foreground) pixels activated along each diagonal is computed up in order to acquire a single sub-feature. These $2n - 1$ sub-features values are averaged to form a single value and put into comparing zone as its feature. Here, we will get n features relating to each sample.

The steps that have been used to extract these features are:

- Step I Divide the thinned image into n ($=100$) number of zones, each of size 10×10 pixels
- Step II Each zone has 19 diagonals; foreground pixels present along each diagonal are summed up in order to get a single sub-feature

Table 1. Feature wise writer identification accuracy using k -NN and linear-SVM classifier (70% data as Training Set and 30% data as Testing set).

Feature extraction technique	Classification technique	
	k -NN ($k = 1$)	Linear-SVM
Zoning features (F_1)	86.62%	88.57%
Diagonal features (F_2)	86.58%	88.50%
Transition features (F_3)	86.04%	88.57%
Peak extent based features (F_4)	82.55%	85.54%
$F_1 + F_2$	86.63%	88.78%
$F_1 + F_3$	86.93%	89.42%
$F_1 + F_4$	85.42%	88.38%
$F_2 + F_3$	86.93%	89.38%
$F_2 + F_4$	85.32%	88.16%
$F_3 + F_4$	86.15%	89.08%
$F_1 + F_2 + F_3$	86.81%	89.52%
$F_1 + F_2 + F_4$	86.07%	88.92%
$F_1 + F_3 + F_4$	86.83%	89.85%
$F_2 + F_3 + F_4$	86.96%	89.71%
$F_1 + F_2 + F_3 + F_4$	86.71%	89.61%

- Step III These 19 sub-feature values are averaged to form a single value which is then placed in the corresponding zone as its feature
- Step IV Corresponding to the zones whose diagonals do not have a foreground pixel, the feature value is taken as zero

These steps will again give a feature set with n elements.

5.3c Transition features (F_3): Extraction of transition feature elements depends on the estimation and area of transitions from background to foreground pixels and vice versa. To ascertain transition information, image is examined from left to right and top to bottom. This procedure, will give n features for a character image [22].

The following steps have been implemented for extracting these features.

- Step I Divide the thinned image of a character into n ($=100$) zones, each of size 10×10
- Step II Calculate number of transitions in horizontal and in vertical directions for each zone

5.3d Peak extent based features (F_4): In this section, we have presented peak extent based features proposed by Kumar *et al* [23]. They have achieved noticeable recognition accuracy with these features of offline handwritten Gurmukhi character recognition for 56-class problem. In this technique, features are extracted by taking into account the sum of the peak extents that fit successive black pixels along each zone. Peak extent based features can be extracted horizontally and vertically. In the horizontal peak extent features, they considered the sum of the peak extents that fit successive black pixels horizontally in each row of a zone, whereas in vertical peak extent features they considered the sum of the peak extents that fit successive black pixels vertically in each column of a zone. So, using this technique, we will obtain $2n$ features corresponding to each character.

The steps that have been used to extract these features are given below.

Figure 2. Feature wise writer identification accuracy using k -NN and linear-SVM classifier.

- Step I Divide the bitmap image into n ($=100$) number of zones, each of size 10×10 pixels
- Step II Find the peak extent as the sum of successive foreground pixels in each row of a zone
- Step III Replace the values of successive foreground pixels by peak extent value, in each row of a zone

Table 2. Writer wise identification accuracy using $F_1 + F_3 + F_4$ features with linear-SVM classifier.

Writer (W) number	Recognition accuracy	Writer (W) number	Recognition accuracy
W ₁	100%	W ₄₆	80.97%
W ₂	97.14%	W ₄₇	100%
W ₃	98.29%	W ₄₈	63.34%
W ₄	54.57%	W ₄₉	100%
W ₅	100%	W ₅₀	100%
W ₆	86.43%	W ₅₁	86.93%
W ₇	86.11%	W ₅₂	63.57%
W ₈	100%	W ₅₃	70.00%
W ₉	76.43%	W ₅₄	73.36%
W ₁₀	100%	W ₅₅	100%
W ₁₁	64.46%	W ₅₆	98.86%
W ₁₂	77.96%	W ₅₇	97.14%
W ₁₃	100%	W ₅₈	94.93%
W ₁₄	76.43%	W ₅₉	69.71%
W ₁₅	100%	W ₆₀	99.05%
W ₁₆	46.14%	W ₆₁	72.57%
W ₁₇	82.87%	W ₆₂	91.44%
W ₁₈	100%	W ₆₃	86.68%
W ₁₉	100%	W ₆₄	59.98%
W ₂₀	100%	W ₆₅	82.86%
W ₂₁	70.29%	W ₆₆	84.64%
W ₂₂	83.83%	W ₆₇	92.14%
W ₂₃	90.60%	W ₆₈	83.83%
W ₂₄	84.77%	W ₆₉	99.43%
W ₂₅	100%	W ₇₀	100%
W ₂₆	76.21%	W ₇₁	86.93%
W ₂₇	100%	W ₇₂	63.57%
W ₂₈	61.81%	W ₇₃	70.00%
W ₂₉	100%	W ₇₄	73.36%
W ₃₀	97.14%	W ₇₅	100%
W ₃₁	77.17%	W ₇₆	98.86%
W ₃₂	100%	W ₇₇	97.14%
W ₃₃	83.57%	W ₇₈	94.93%
W ₃₄	80.71%	W ₇₉	69.71%
W ₃₅	73.71%	W ₈₀	99.05%
W ₃₆	100%	W ₈₁	72.57%
W ₃₇	77.14%	W ₈₂	91.44%
W ₃₈	97.14%	W ₈₃	86.68%
W ₃₉	73.89%	W ₈₄	59.98%
W ₄₀	85.29%	W ₈₅	82.86%
W ₄₁	79.29%	W ₈₆	84.64%
W ₄₂	77.14%	W ₈₇	92.14%
W ₄₃	64.75%	W ₈₈	83.83%
W ₄₄	100%	W ₈₉	99.43%
W ₄₅	74.86%	W ₉₀	100%

- Step IV Find the largest value of peak extent in each row. As such, each zone has 10 horizontal peak extent features
- Step V Obtain the sum of these 10 peak extent sub-feature values for each zone and consider this as a feature for the corresponding zone
- Step VI For the zones that do not have a foreground pixel, take the feature value as zero
- Step VII Normalize the values in the feature vector by dividing each element of the feature vector by the largest value in the feature vector

Similarly, for vertical peak extent features, we have considered the sum of the lengths of the peak extents in each column of each zone. These steps will give a feature set with $2n$ elements.

5.4 Classification

Classification is the final stage of the writer identification framework. Classification phase is a decision making phase, which is used to identify the writer based on the features extracted in the previous phase. For classification, we have considered two classification methods, specifically, k -NN and SVM, because these classification algorithms are the most widely used algorithms in the field of pattern recognition. k -Nearest Neighbour (k -NN) is a method for classifying unknown samples in view of neighbouring samples in the training feature space. Locations and labels of the training samples are used to divide the space into regions. Among the k nearest training samples, if a class is most frequent when it is assigned a position in the space. Usually Euclidean distance is used to calculate the distance between stored feature vector and candidate feature vector in k -Nearest Neighbour algorithm. SVM is an extremely valuable technique for data classification. The SVM is a learning machine, which has been broadly connected in pattern recognition. SVM depends on a measurable learning hypothesis that uses supervised learning. In supervised learning, a machine is trained rather than programmed to play out a given task on various information input/output sets. SVM classifier has additionally been considered with a linear kernel.

6. Experimental results and discussion

In this section, we have presented the outcomes of the proposed writer identification framework in view of the Gurmukhi script characters. The outcomes are produced by applying various feature extraction techniques and classification techniques. For experimental outcomes, we have utilized a collection of 31,500 samples of segmented characters of Gurmukhi script. In training dataset, we have considered 70% data (22,050 samples) and in the testing

Figure 3. Writer wise identification accuracy using $F_1 + F_3 + F_4$.

Table 3. Feature wise writer identification accuracy using k -NN and Linear-SVM classifier (10-fold cross validation accuracy).

Feature extraction technique	Classification technique	
	k -NN ($k=1$)	Linear-SVM
Zoning features (F_1)	92.68%	93.46%
Diagonal features (F_2)	91.64%	92.38%
Transition features (F_3)	92.06%	92.76%
Peak extent based features (F_4)	88.45%	91.52%
$F_1 + F_2$	91.72%	93.46%
$F_1 + F_3$	91.10%	91.24%
$F_1 + F_4$	90.28%	91.89%
$F_2 + F_3$	92.28%	93.48%
$F_2 + F_4$	89.98%	94.21%
$F_3 + F_4$	91.28%	94.24%
$F_1 + F_2 + F_3$	91.89%	92.24%
$F_1 + F_2 + F_4$	91.09%	92.15%
$F_1 + F_3 + F_4$	92.91%	94.76%
$F_2 + F_3 + F_4$	92.04%	94.46%
$F_1 + F_2 + F_3 + F_4$	92.17%	92.67%

dataset we have taken 30% data (9,450 samples) of the complete dataset of 31,500 samples. By using this proposed framework, we have accomplished an accuracy of 89.85% with a combination of zoning, transitions and peak extent based features with linear-SVM classifier, as depicted in table 1. These results are graphically presented in figure 2. For this case (a combination of zoning, transitions and peak extent based features with linear-SVM classifier) writer wise identification accuracy is given in table 2. Graphical representation of writer wise recognition accuracy is shown in figure 3.

The 10-fold cross validation writer identification accuracy has been also tested as shown in table 3. These results are graphically depicted in figure 4. Maximum accuracy of 94.76% has been achieved using 10-fold cross validation technique with a combination of zoning, transitions and peak extent based features and linear-SVM classifier for writer identification.

Figure 4. Feature wise 10-fold cross validation writer identification accuracy using k -NN and linear-SVM classifier.

Table 4. Comparative study of writer identification work in the light of the Gurmukhi script.

Factor	Existing work	Proposed work
Script	Gurmukhi	Gurmukhi
Number of writers	30	90
Number of classes	35	35
Dataset	10,500	31,500
Number of times each character writer by each writer	10	10
Feature extraction technique	Zoning, Intersection and open end points	Zoning, diagonal, transitions and peak extent based features
Classifier	k-NN and MLP	k-NN and SVM
Combination of features	No	Yes
Accuracy	52.40%	89.85%

Table 5. Comparative study of the proposed work with state-of-the-art work.

Author	Script	Data set	Feature Extraction methods	Classifier	Accuracy
Akbari <i>et al</i> [1]	Persian	50 writers	Wavelet transform and Co-occurrence matrix	K-NN	93.3%
Awaida and Mahmoud [2]	Arabic	500 paragraphs from 250 writers	Gradient distribution, windowed gradient, contour chain code distribution, etc.	K-NN	95.4%
Sreeraj and Idicula [11]	Malayalam	280 writers	Loop roundness, slant, direction angle, relative height, distance from centroid, etc.	K-NN	95.62%
Shahabi and Rahmati [25]	Farsi	40 people	Multi-channel Gabor filtering and Co-occurrence matrix	Weighted Euclidean distance	97.5%
Kalra and Rani [24]	Gurmukhi	30 writers	Zoning and open end point intersection based feature	K-NN and MLP	52.40% with MLP
Proposed work	Gurmukhi	90 writers	Zoning, diagonal, transition, and peak extent based features	K-NN and SVM	89.85% with SVM

7. State-of-the-art work

In this section, comparison between existing work and the proposed work has been presented. We have seen that a very few work has been presented for writer identification in light of the Gurmukhi script. In the proposed work, a comparative analysis performance of individual features and combination of various features has also been done. In the previous work of writer identification for the Gurmukhi script, only 52.40% accuracy was achieved by using zoning and intersection open end points based features [24]. But, using the proposed framework accuracy of 89.85% has been achieved using SVM classifier as compared in table 4. Comparative study of latest work presented by Kalra and Rani [24] and the proposed work has been presented in table 4. Comparison with other scripts has been depicted in table 5.

8. Conclusions and future directions

In this paper, we have presented a novel framework for writer identification based on isolated handwritten Gurmukhi characters. For experimental work, we have tested the proposed model with 31,500 samples in this work. Maximum accuracy of 89.85% has been acquired by using the proposed framework with a combination of zoning, transitions and peak extent based features and linear-SVM classifier when we considered 70% data as the training set and remaining 30% data as testing dataset. Using 10-fold cross validation technique, we have achieved accuracy of 94.76% with a combination of zoning, transitions and peak extent based features and linear-SVM classifier. This writer identification proposed framework can likewise be stretched out to other Indian scripts like Devanagari, Bengali, and Tamil, etc. which are similar to the Gurmukhi script.

References

- [1] Akbari M, Eslami R and Kashani M H 2012 Offline Persian writer identification based on wavelet analysis. In: *Proceedings of Fourth International Conference on Bioinformatics and Biomedical Technology*, pp. 180–186
- [2] Awaida S M and Mahmoud S A 2012 State of the art in off-line writer identification of handwritten text and survey of writer identification of Arabic. *Educ. Res. Rev.* 7(20): 445–463
- [3] Maadeed S A 2012 Text-dependent writer identification for Arabic handwriting. *J. Electr. Comput. Eng.* 13: 1–8
- [4] Brink A A, Smit J, Bulacu M L and Schomaker L R B 2012 Writer identification using directional ink-trace width measurements. *Pattern Recognit.* 45(1): 162–171
- [5] Ghiasi G and Safabakhsh R 2010 An efficient method for offline text independent writer identification. In: *Proceedings of Twentieth International Conference on Pattern Recognition*, pp. 1245–1248
- [6] Leng W Y and Shamsuddin S M 2010 Writer identification for Chinese handwriting. *Int. J. Adv. Soft Comput. Appl.* 2(2): 142–173
- [7] Pratama S F, Muda A Z and Choo Y H 2010 Feature selection methods for writer identification: a comparative study. In: *Proceedings of International Conference on Computer and Computational Intelligence*, pp. 234–239
- [8] Bulacu M, Schomaker L and Brink A 2007 Text-independent writer identification and verification on offline Arabic handwriting. In: *Proceedings of Ninth International Conference on Document Analysis and Recognition*, pp. 769–773
- [9] Gazzah S and Amara N B 2008 Neural networks and support vector machines classifiers for writer identification using Arabic script. *Int. Arab J. Inf. Technol.* 5(1): 92–101
- [10] Sreerej M and Idicula S M 2011a A survey on writer identification schemes. *Int. J. Comput. Appl.* 26(2): 23–33
- [11] Sreerej M and Idicula S M 2011b Identifying decisive features for distinctive analysis of writings in Malayalam. *Int. Mag. Adv. Comput. Sci. Telecommun.* 2(1): 13–20
- [12] Jain R, Doermann D 2013 Writer identification using an alphabet of contour gradient descriptors. In: *Proceedings of 12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 550–554
- [13] Fiel S and Sablatnig R 2015 Writer identification and retrieval using a convolutional neural network. In: *Proceedings of International Conference on Computer Analysis of Images and Patterns*, pp. 26–37
- [14] He S and Schomaker L 2017 Writer identification using curvature-free features. *Pattern Recognit.* 63: 451–464
- [15] Christlein V, Bernecker D, Honig F, Maier A and Angelopoulos E 2017 Writer identification using GMM supervectors and exemplar-SVMs. *Pattern Recognit.* 63: 258–267
- [16] Adak C and Chaudhuri B B 2015 Writer IDENTIFICATION FROM OFFLINE ISOLATED Bangla characters and numerals. In: *Proceedings of 12th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 480–490
- [17] Lehal G S and Singh C 1999 Feature extraction and classification for OCR of Gurmukhi script. *Vivek* 12(2): 2–12
- [18] Kumar M, Sharma R K and Jindal M K 2013 A novel feature extraction technique for offline handwritten Gurmukhi character recognition. *IETE J. Res.* 59(6): 687–692
- [19] Kumar M, Sharma R K and Jindal M K 2014 Efficient feature extraction techniques for offline handwritten Gurmukhi character recognition. *Natl. Acad. Sci. Lett.* 37(4): 381–391
- [20] Kumar M, Jindal M K and Sharma R K 2013 PCA based offline Handwritten Gurmukhi character recognition. *Smart Comput. Rev.* 3(5): 346–357
- [21] Zhang T Y and Suen C Y 1984 A fast parallel algorithm for thinning digital patterns. *Commun. ACM* 27(3): 236–239
- [22] Kumar M, Jindal M K and Sharma R K 2011 k-nearest neighbor based offline handwritten Gurmukhi character recognition. In: *Proceedings of International Conference on Image Information Processing, Jaypee University of Information Technology, Wagnaghat (Shimla)*, pp. 1–4
- [23] Kumar M, Jindal M K and Sharma R K 2014 A novel hierarchical techniques for offline handwritten Gurmukhi character recognition. *Natl. Acad. Sci. Lett.* 37 (6): 567–572
- [24] Kalra K and Rani S 2017 Writer identification from offline handwritten Gurmukhi characters. *J. Adv. Comput. Sci. Technol.* 10(5): 903–914
- [25] Shahabi F and Rahmati M 2009 A new method for writer identification of handwritten Farsi documents. In: *Proceedings of the 10th International Conference on Document Analysis and Recognition*, pp. 426–430