© 2017 IJSRSET | Volume 3 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

A Character Based Handwritten Identification Using Neural Network and SVM

Rohitash Kumar, Mandeep Kaur
Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, India

ABSTRACT

Handwritten identification is carried out using handwritten text. The Research work on investigates highly discriminating features for handwritten identification for off-line handwritten multiple text lines and passages. Five categories of features are tested: slant and slant energy, skew, pixel distribution, curvature, and entropy. These features support high recognition rates and are competitive with other state of the art methods for handwritten identification. The directional element features are first extracted from the handwriting character scripts, then the dimensions of the features is reduced using PCA in order to cope with the small sample size problem. The most discriminative features are extracted from the reduced feature space using Fisher's Linear Discriminant Analysis. The Euclidian distance is Research for classification. Experimental result using SVM and NN verified the effectiveness of the Research work. For the implementation of this Research work we use the Image Processing Toolbox under MATLAB software.

Keywords: PCA, SVM, MATLAB, DEF, PCA, LDA, IAM, ICDAR, GUI, CCR, PSNR

I. INTRODUCTION

Handwritten identification is the process of confirming a writer's identity by comparing some specific attributes of his/her handwriting with those of all the writers enrolled in a reference database. According to the script used, the identification methods can be classified into two types: text-insensitive (or text-independent) approaches and text-sensitive (or text-dependent) ones. Text-insensitive method viewed the handwriting as textural image and extracted textural features based on Gabor filters. For this method, whole pages of handwritten texts are needed.

The problem of Handwritten identification addresses whether or not a classifier can be created that is able to identify the Handwritten of some handwritten text with a high accuracy given a set of handwritten documents from many writers. Handwritten identification is an area of great interest and is especially important for verification purposes. For example, validating signatures on checks and legal documents such as wills are two applications of Handwritten identification as a means of validation. Likewise, Handwritten identification of documents is extremely important in legal matters where handwritten documents are often used as evidence.

Writer identification is a well-studied problem. In general, Handwritten discrimination and verification approaches based on handwritten text are hardly found in the literature. Security reasons or specific law restrictions have prevented serious results of significant importance on the topic from publicity. To the knowledge of the authors few publications are related to Handwritten discrimination and especially to feature extraction. Feature extraction from handwritten text can be carried out using approaches that resemble those of signature verification. However, features which contain information of the trace of each word are usually preferable.

II. METHODS AND MATERIAL

PCA and LDA for Handwritten Identification

Although DEFs have great classification ability, in order to further increase identification accuracy rate and overcome small training sample size problem compared with the DEFs dimension, methods for reducing the dimensionality of the DEFs are required. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two widely used methods for dimension reduction.

Brief introduction of various features selected :- Slant and Slant Energy

Slant is the measure of deviation in the vertical direction in a hand written textl. Sobel operator is applied to calculate the slant in the grayscale image of a handwriting sample in order to detect edges. Canny operator is used to approximate the gradient of an image at each pixel.

Skew

Skew is the measure of the difference of the text lines being parallel with the horizontal. For calculating the skew, a document or line is rotated by increments of 0.05 degrees.

Pixel Distribution

Text image is binarised and corrected for skew. Next, the horizontal projection profile is calculated, and the location of the largest peak of the profile is determined. This peak approximates the centerline of a text line. The numbers of pixels above and below the centerline are counted.

Curvature

Curvature is the measure of deviation of line from being straight. Many features relating to curvature are investigated.

Entropy

Entropy is a measure of the randomness of an image. Entropy is calculated as follows:

$$E = -\sum P \,.\log_2 P$$

Where; P is a vector containing the probabilities of each pixel value (0 or 1 for binary and 0 to 255 for grayscale) in the image.

A. Classifiers To Be Used For Proposed Work

Artificial Neural Networks

Artificial neural networks are composed of interconnecting artificial neurons Artificial neural networks may either be used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system. It is used to reduce the amount of computation required to simulate artificial neural networks, so as to allow one to experiment with larger networks and train them on larger data sets.

In this proposed work our classifier will work using feed-forward neural networks. Feed-forward ANNs allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. They are extensively used in pattern recognition. This type of organisation is also referred to as bottom-up or top-down. Single-layer perceptron, multilayer perceptron and radial basis function are types of feed forward neural networks.

Support Vector Machine

The Support Vector Machine (SVM) is a state-of-the-art classification method. The SVM classifier is widely used in bioinformatics due to its highly accurate, able to calculate and process the high-dimensional data. SVMs belong to the general category of kernel methods. In this, the dot product can be replaced by a kernel function which computes a dot product in some possibly high dimensional feature space.

In the proposed work the Linear SVM Classifier will be used. When working with a linear classifier the only hyper parameter that needs to be tuned is the SVM soft-margin constant. For the polynomial and Gaussian kernels the search space is two-dimensional. The standard method of exploring this two dimensional space is via grid-search; the grid points are generally chosen on a logarithmic scale and classifier accuracy is estimated for each point on the grid [12]. classifier is then trained using the hyperparameters that yield the best accuracy on the grid. The accuracy landscape has an interesting property: there is a range of parameter values that yield optimal classifier performance. Furthermore, these equivalent points in parameter space fall along a ridge in parameter space. This phenomenon can be understood as follows. If we decrease the value , this decreases the curvature of the decision boundary; if we then increase the value of C the decision boundary is forced to curve to accommodate the larger penalty for errors/margin errors. The approach of explicitly computing non-linear features does not scale well with the number of input features.

B. Problem Statement

In this synopsis, we propose “Discriminating Features for Handwritten Identification using Neural Networks and Support Vector Machine” In the previously developed method experiments were conducted using two publicly available datasets: the IAM Handwriting Database and ICDAR 2011 Handwritten Identification Contest dataset using a nearest neighbor as classifier. The proposed method addresses the problem of Handwritten identification for off-line handwritten text lines and passages. Five categories of text independent features are to be examined: slant and slant energy, skew, pixel distribution, curvature, and entropy using the Neural Network and SVM as classifiers. These features results in high recognition rates competitive with other state of the art methods for Handwritten identification but still needed some improvement. Our experience suggests that no single feature is dominant.

C. Methodology

Our proposed method for Handwritten Identification will be developed by using the sequence described in the Flowchart given below.

The work will be done in following phases.

Phase 1: First code is developed for opening GUI for this implementation. After that, code is developed for loading the text image in MATLAB database.

Phase 2: Develop a code for the document processing for the loaded text image.

Phase 3: Develop a code for the edge detection and apply on the loaded text image. Develop the code for the classification using the discriminate feature like slant and slant energy, skew, pixel distribution, curvature, and entropy.

Phase 4: Develop a code for NN and SVM. After that code is developed for the verification and identification for the writer.

Phase 5: Finally develop code to analyze our result by calculating various parameters like PSNR, CCR (Correct Classification Rate) and Recognition rate.

III. RESULTS AND DISCUSSION

The results of Handwritten Indentification using Neural Network and SVM using handwritten loading images with higly accuracy rate using various kinds of structural and statistical features of and options performed by us are presented in Figures5.1–5.10, which show the results for only enhanced binarized samples. Parameters used for testing are Accuracy (values has been normalized to [0 1]), Correct Classification Rate, and Squared Correlation Coefficient.

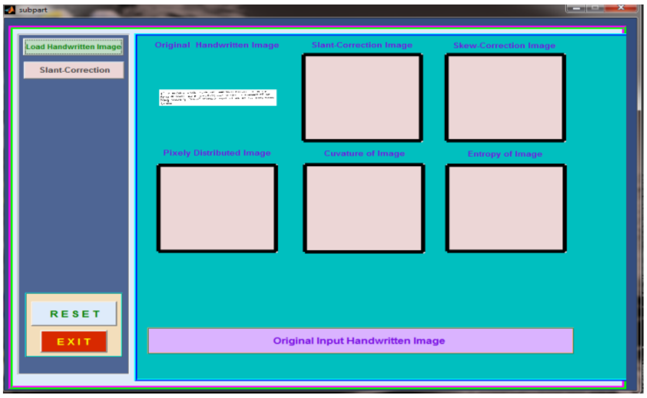

Figure 1. Load the Images

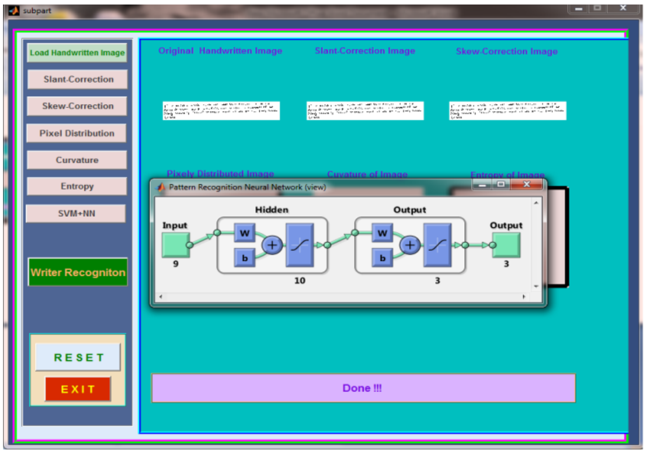

Figure 2. Input hidden and output

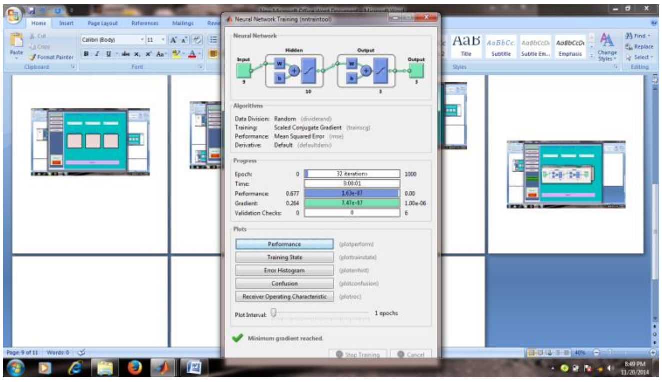

Figure 3. Discriminate identification correction

Table 1. Recognition of SVM Parameters values for various Feature extraction Techniques of Enhanced Images

S.No.	Odd trained Even tested	Parameters		
		Accuracy	Mean sq. ratio	Sq. col. cof.
1.	Projection (F1)	0.94	0.42	0.950454
2.	Slant and Slant Energy (F2)	0.84	1.66	0.812099
3.	Skew Correction (F3)	0.92	0.14	0.983196
4.	Gradient (F4)	0.84	4.1	0.555911
5.	Pixel distribution (F5)	0.86	2.64	0.707483
6.	Curvature (F6)	0.84	1.66	0.812099
7.	Vector Machine(F7)	0.9	3.74	0.598306
8.	all above combined (F8)	0.92	1.1	0.875458

Table 2. Recognition of SVM Parameters values for various Feature extraction Techniques of Enhanced Images

S.No.	Last trained First tested	Parameters		
		Accuracy	Mean sq. ratio	Sq. col. cof.
1.	Projection (F1)	0.9	2.16	0.752278
2.	Slant and Slant Energy (F2)	0.72	5.72	0.409129
3.	Skew Correction (F3)	0.84	1.9	0.782995
4.	Gradient (F4)	0.82	4.12	0.56307
5.	Pixel distribution (F5)	0.82	3.98	0.569884
6.	Curvature (F6)	0.72	5.72	0.409129
7.	Vector Machine(F7)	0.82	4.12	0.558121
8.	all above combined (F8)	0.95	2.12	0.75262

IV. CONCLUSION

This research work is to implement the character based handwritten identification using neural network and svm. It is based upon GUI (graphical user interface) in MATLAB. It is an effort to further grasp the fundamental of MATLAB and validate it as a powerful application tool. MATLAB (Matrix laboratory) is a numerical computing environment. It is fourth generation programming language. Mathworks develop MATLAB; this dissertation presents handwritten identification images using Neural Network and SVM. Mainly two stages including feature extraction and classification are carried out in detail. Major problem in this recognition from above described problems, which are been undertaken for recognition of characters. The structural and statistical features selected for recognition of handwritten images, which are used such as Slant and Slant energy, Skew Correction, Pixel Distribution and Curvature. SVM and NN have been used for classification purpose.

V. REFERENCES

[1]. Zachary A. Daniels and Henry S. Baird," Discriminating Features for Handwritten Identification". In the Document Analysis and Recognition (ICDAR),2014 .International Conference on,pages 1417-1421 ,IEEE 2014.

[2]. A. Bensefia, A. Nosary, T. Paquet, and L. Heutte," Handwritten identification by writer's invariants", International Workshop on Frontiers in Handwriting Recognition, pages 274–279, 2002.

[3]. A. Bensefia, T. Paquet, and L. Heutte, "Information retrieval based Handwritten identification" In Seventh Int. Conf. on Document Analysis and Recognition, pages 946–950, 2003.

[4]. G.Louloudis, B.Gatos, N.Stamatopoulos and A.Papandreou,"ICDAR2013 Competition on Handwritten Identification" In the Document Analysis and Recognition (ICDAR),2013 .International Conference on,pages 1429-1432 ,IEEE 2013

[5]. M. Bulacu, L. Schomaker, and L. Vuurpijl, " Handwritten identification using edge-based directional features", In Seventh International Conference on Document Analysis and Recognition, pages 937–941, 2003.

[6]. H. Bunke, " Recognition of cursive roman handwriting – past, present and future", Seventh International. Conference on Document Analysis and Recognition, pages 448–461, 2003.

[7]. S.H. Cha and S. Srihari, "Multiple feature integration for Handwritten verification"In L. Schomaker and L. Vuurpijl, editors, Proc. Seventh International Workshop on Frontiers in Handwriting Recognition, pages 333–342, 2000.

[8]. C. Hertel and H. Bunke., "A set of novel features for Handwritten identification", In J. Kittler and M. Nixon, editors, Audio and Video-Based Biometric Person Authentication, pages 679– 687, 2003.

[9]. F. Leclerc and R. Plamondon, " Automatic signature verification: The state of the art 1989–1993", In R. Plamondon, editor, Progress in Automatic Signature Verification, pages 13–19. World Scientific Publ. Co., 1994.

[10]. G. Leedham and S. Chachra, "Writer identification using innovative binarised features of handwritten numerals",In Seventh International Conference on Document Analysis and Recognition, pages 413–417, 2003.

[11]. U.V. Marti and H. Bunke, "Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system", International Journal of Pattern Recognition and Artificial Intelligence, 15:65–90, 2001.

[12]. U. V. Marti and H. Bunke. " The IAM–database: An English sentence database for off-line handwriting recognition", International Journal of Document Analysis and Recognition, 5:39–46, 2002.

[13]. S. Marukatat, T. Arti`eres, P. Gallinari, and B. Dorizzi "Rejection measures for handwriting sentence recognition" In Proc. of the Eighth International Conference on Frontiers in Handwriting Recognition, pages 25–29, 2002.

[14]. A. Nosary, L. Heutte, T. Paquet, and Y. Lecourtier, " Defining writer’s invariants to adapt the recognition task" ,In Fifth International Conference on Document Analysis and Recognition, pages 765–768, 1999.

[15]. R. Plamondon and G. Lorette, " Automatic signature verification and Handwritten identification – the state of the art",In Pattern Recognition, volume 22, pages 107–131, 1989.

[16]. R. Plamondon and S. Srihari, "On-line and off-line handwriting recognition: A comprehensive survey" , IEEE Trans. on Pattern Analysis and Machine Intelligence, 22:63–84, 2000.

[17]. L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", Proc. of the IEEE, 77(2):257–285, 1989.

[18]. A. E. Rosenberg, J. Delong, C. H. Huang, and F. K. Soong, "The use of cohort normalized scores for speaker verification", Proceeding International. Conference on Spoken Language Processing, pages 599–602, 1992.

[19]. H. E. S. Said, T. Tan, and K. Baker, "Personal identification based on handwriting", Pattern Recognition, 33:149– 160, Jan. 2000.

[20]. A. Schlapbach and H. Bunke, "Off-line handwriting identification using HMM based recognizers", Accepted for publication in Proc. 17th International Conference on Pattern Recognition, 2004.

[21]. S. Young, G. Evermann, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland. "The HTK book", 2002.

[22]. E. N. Zois and V. Anastassopoulos, " Morphological waveform coding for Handwritten identification", Pattern Recognition, 33:385–398, Mar. 2000.

[23]. S.N. Srihari, S.H. Cha, H. Arora, and S. Lee, "Individuality of handwriting: a validation study", In Document Analysis and Recognition, 2001. Proceedings, Sixth International Conference on, pages 106–109. IEEE, 2001.