

The main objectives of Pre-processing methods are:

- In preprocessing technique we perform 2 operation
- Binarization
- Thinning

After pre-processing phase, a cleaned image is available that goes to the segmentation phase. The raw data, depending on the data acquisition type, is subjected to a number of preliminary processing steps to make it usable in the descriptive stages of Script analysis.

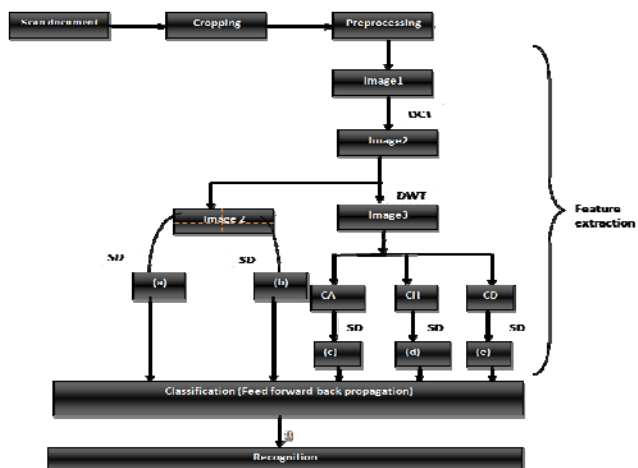


Figure 1: Block Diagram of Script Identification

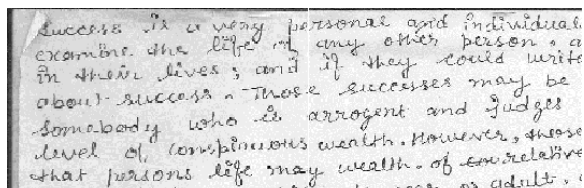


Figure 2: Script Sample of English Language

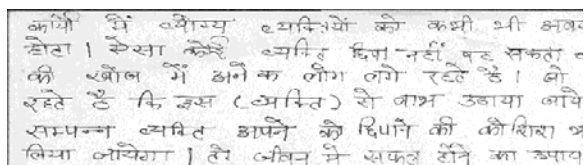


Figure 3: Script Sample of Hindi Language

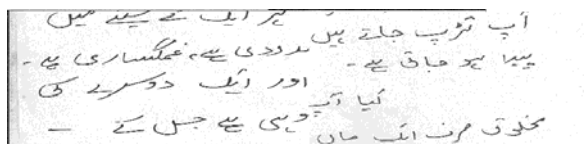


Figure 4: Script Sample of Urdu Language

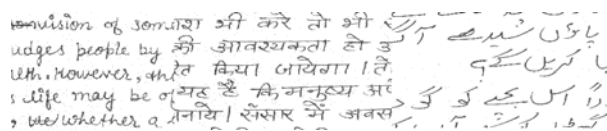


Figure 5: Combined Sample of multi script

C. Segmentation

It is intended for the decomposition of an image is divided into a configuration script for each symbol sub-image operations. Segmentations are key condition script controls

the effectiveness of the straight system. Some approaches can be used followed by a square-like segmentation strategy according to the text and to identify the type of cut based segmentation and classification methods for classification. To achieve wide use, it is a segmentation method has the following properties are important:

1. Capture group or sensibility significant area, which often reflects the image of a global problem. Two central issues are what are important to provide accurate Securizations perception, and you can specify the same for a given partition. Performance should be precisely defined division caused better facilitate a better understanding of the method and different methods. The actual method to be used in edge detection quickness' anatomical or other similar low-level visual processing technology, which means close to the linear operation with a low constant factor. For example, the video processing operations may be used in several applications per frame division.

D. Feature Extraction

Each script has some features, which play a significant role in pattern recognition. English, Hindi and Urdu script has many special features. Description, such a classification model feature extraction task becomes very easy to contain information about a pattern in the shape of a proper driver. These handwritings splinter scrutiny system in HMSR feature extraction stage, and selected a customary of landscapes that can be castoff to categorize abnormal script section. Mainly, this is the heart HMSR stage system, since the results be contingent on these topographies. Article abstraction are assumed to family, is included in the program for measuring information related to the shape of a pattern, the sort pattern so that the task is facilitated through the formal name of the program. Tangled in building recognition system in which different design issues, is perhaps one of the most extensive set of features to choose from. Feature extraction for exploratory data projector so that the concept of high-dimensional data to better understand and clustering data structure. The computational requirements are reduce to quotation of the characteristics of great discriminate dimensionality reduction, in the feature extraction of the image. However, feature extraction rule, projection timings exploratory objective is to minimize the error function of data, such as mean square error or difference from the inter mode, the purpose of feature extraction and classification is divided into classes the better enhancement. Therefore, the best feature extraction (for specific Standard) for exploratory data prediction is not automatically the best in the class can be divided into functions, and vice versa. In particular, two or more classes can have primary function is similar. In addition, article abstraction for examining data prognosis for two or believable data visualization, and classification typically require more than two or three characteristics. Therefore, not mostly for classification, and vice versa feature extraction paradigm exploratory data projection.

3. Representation of Script Features

Currently, in India, India handwritten script standard database is unavailable. Therefore, the training and test data classification scheme is to collect from different sources. Are in English, Hindi and Urdu script handwritten documents belonging to different people in different industries to collect. The document scanned at 300 dpi and gray scale image storage. Size 512 × 512 pixel image block, and then manually extract files from different regions of the image. It should be noted that the handwritten text block may contain two or more lines, rows, different font sizes (large and small) and variable bit between words and characters. We do not perform any processing, homogenization parameters. It ensures that the area of at least 50% block of text containing the text. These blocks are equivalent to a part of the handwritten document, and then the two values, so that the text and the background on behalf of a representative value of 0. In the proposed system, using morphological opening around the boundary noise is removed. This operation also removes non-contiguous pixel level.

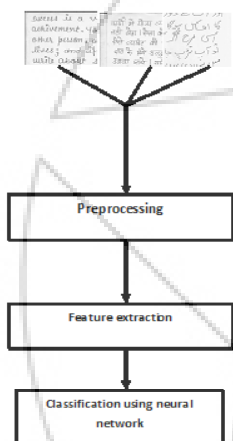


Figure 6: Block diagram of Methodology

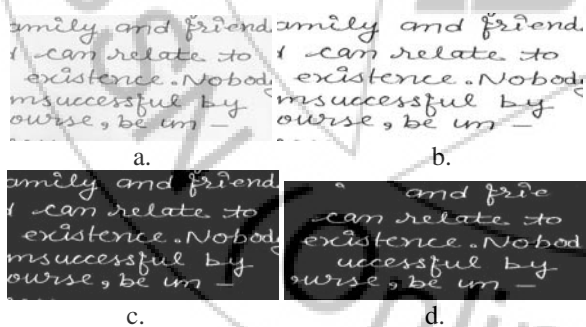


Figure 7a: Original Cropped Image of English Script b. Black & White Image c. Invert color d. Clear component clear border

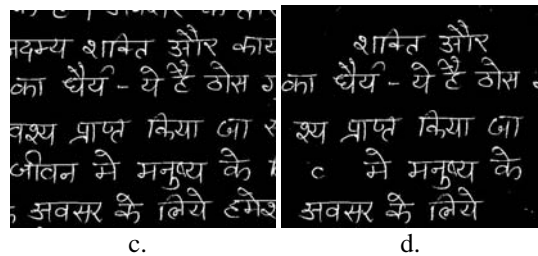
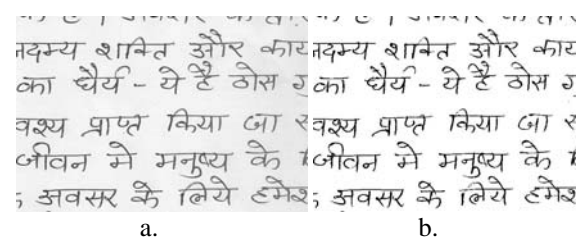


Figure 7a: Original Cropped Image of Hindi Script b. Black & White Image c. Invert color d. Clear component clear border

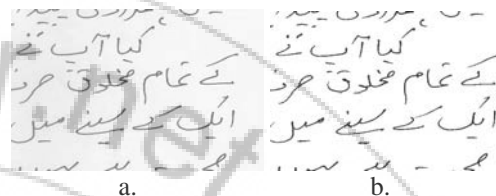


Figure 8a: Original Cropped Image of Urdu Script b. Black & White Image c. Invert color d. Clear component clear border

4. Results

For the reorganization of handwritten script that is prepared by different people in different location .The total dataset is 9862 there the Hindi sample is 3373, English sample 3269 and Urdu is 3220. In the total dataset is divided in to two parts one part is training purpose, other part is testing purpose. For recognition of each script. Features are calculated and safe for the training purpose. The neural network is having three type of the layer, one is the input layer, second is hidden layer and third is the output layer. If the increases the number of neurons in the hidden layer the result will be increases and decreases on given script.

In the back propagation algorithm one layer is behave like an input layer. Second one is the hidden layer and last one is the output layer. If increase the number of neuron in hidden layer then required memory allocation problem can be happened and also the required result are not acquired if the value of tolerance is increased can take more number of cycle for learning purpose to obtaining the results . But learning is not up to mark up the result not to acquire desire.

Scripts	No. of samples	Train/test	Recognition result
Hindi	3373		
English	3269		
Urdu	3220	4932/4930	92.70%

Table 1: Result of Multiple Scripts

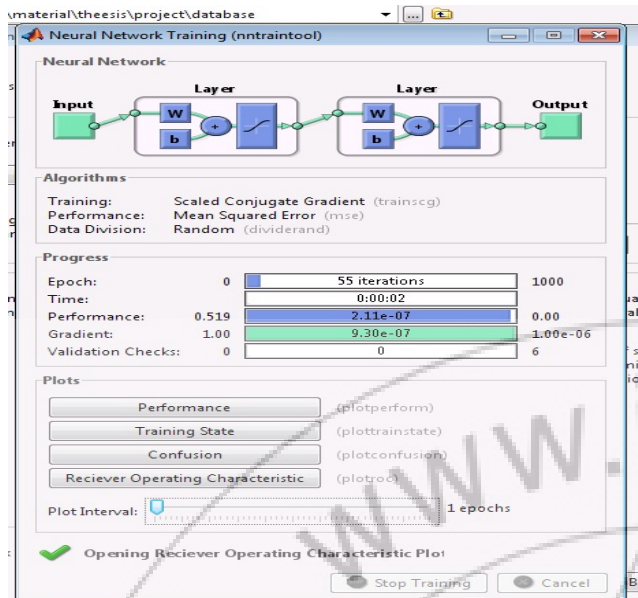


Figure 9: Diagram of NN Training

Reference

- [1] Nabin Sharma. With Co-Author with U. Pal, and R. Jayadevan, "Handwriting recognition in Indian regional scripts: A survey of offline techniques"
- [2] Ram sarkar, nibaran das, subhadip basu, mahantapas kundu, mita nasipuri and dipak kumar basu, "cmaterdb1: a database of unconstrained handwritten bangla and bangla-english mixed script document image", international journal on document analysis and recognition Volume 15, number 1 (2012), 71-83, doi: 10.1007/s10032-011-0148-6, 2012.
- [3] G. G. Rajput and Anita H. B., "Handwritten Script Recognition using DCT and Wavelet Features at Block Level", 2010.
- [4] Jayadevan, R. Pune Inst. of Comput. Technol., Pune, India Kolhe, S.R. ; Patil, P.M. ; Pal, U., "Offline Recognition of Devanagari Script: A Survey", Volume: 41 , Issue: 6, Product Type: Journals & Magazines, 2011.
- [5] AlKhateeb, J.H., "A new approach for off-line handwritten Arabic word recognition using KNN classifier", 18-19 Nov. 2009.
- [6] Assabie, Y., "HMM-Based Handwritten Amharic Word Recognition with Feature Concatenation", Document Analysis and Recognition, ICDAR '09. 10th International Conference, 2009.
- [7] Bahmani, Z., Alamdar, F., Azmi, R., Haratizadeh, S., "8) Off-line Arabic/Farsi handwritten word recognition using RBF neural network and genetic algorithm", Intelligent Computing and Intelligent Systems (ICIS), IEEE International Conference on 2010.
- [8] Pal, U., Roy, R.K., Kimura, F., "Handwritten street name recognition for Indian postal automation", Document Analysis and Recognition (ICDAR), International Conference on 2011.
- [9] Liangrui Peng, Changsong Liu, Xiaoqing Ding, Hua Wang, "Multilingual document recognition research and its application in China," dial, pp.126-132, Second International Conference on Document Image Analysis for Libraries (DIAL'06), 2006.
- [10] U. Pal and B. Chaudhuri. Automatic identification of English, Chinese, Arabic, Devnagari and Bangla script line. In International Conference on Document Analysis and Recognition, pages 790-794, 2001.
- [11] U. bhattacharya, T.K Das, A. Datta, S.K. Parui, B.B Chaudhuri, "A hybrid scheme for hand printed numeral recognition based on a self-organizing network and MPL Classifiers, Int.J. Pattern Recogniton Artificial Intelligence". 16(2002) 845-864.
- [12] K. H. Aparna, V. Subramaniam, M. Kasirajan, G. V. Prakash, V. S. Chakravarthy and S. Madhvanath, "Online handwriting recognition for Tamil", in the Proceedings of 9th International Workshop on Frontiers in Handwriting Recognition (IWFHR), pp. 438-443, 2004.
- [13] C. V. Lakshmi and C. Patvardhan, "A high accuracy OCR system for printed Telugu text", in the Proceedings of Conference on Convergent Technologies for Asia-Pacific Region (TENCON 2003), Vol. 2, pp. 725-729, 2003.
- [14] Lei Han, Jue Zhong, Arkady Voloshin, Image analysis and data processing of time series fringe pattern of PCBs by using moiré interferometry, in: Proceedings of HDP'04, 2004, pp. 141-145.
- [15] Ping Zhong, Chenjie Song, Nian Luo, Method of extracting high-resolution digital moiré fringe in warpage measurement, Physical and Failure Analysis of Integrated Circuits, IPFA, 2009, pp. 527-530.
- [16] V. Ablavsky and M.R. Stevens, "Automatic Feature Selection with Applications to Script Identification of Degraded Documents," Proc. Int'l Conf. Document Analysis & Recognition, Edinburgh, pp.750-754, Aug. 2003.
- [17] D. Dhanya, A.G Ramakrishnan and Peeta Basa pati, "Script identification in printed bilingual documents," Sadhana, vol. 27, part-1, pp. 73-82, 2002.

Authors Profile



Raushan Kumar Singh is a M.Tech Scholar in computer science department with information communication at Suresh Gyan Vihar university. He does work on pattern recognition in multiple language script and work with the help of MATLAB. His interest area is pattern recognition and Neural Network



Akhilesh Pandey is an Asst. Professor in department of computer science and engineering Suresh Gyan Vihar University, Jaipur. He did his MCA from IGNOU in 2002 and after that he worked as a faculty member in different engineering college. After that he acquired his M. Tech. (CSE) at Sharda University, Gr. Noida, India, His area of Interest is Pattern Recognition and neural network.