

## A Literature Review on Handwritten Character Recognition based on Artificial Neural Network

**Rajdeep Singh<sup>1\*</sup>, Rahul Kumar Mishra<sup>2</sup>, S.S. Bedi<sup>3</sup>, Sunil Kumar<sup>4</sup>, Arvind Kumar Shukla<sup>5</sup>**

<sup>1,2</sup>School of Computer Science and Applications, IFTM University, Moradabad, U.P, India

<sup>3</sup>Department of Computer Science & Information Technology, MJPRU Bareilly, U.P India

<sup>4,5</sup>School of Computer Science and Applications, IFTM University, Moradabad, U.P, India

*\*Corresponding Author: singh.rajdeep1@gmail.com*

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

Accepted: 18/Oct/2018, Published: 30/Nov/2018

**Abstract**— In current scenario, character recognition is the most important field of pattern recognition because of its application in numerous fields. Optical Character Recognition (OCR) and Handwritten Character Recognition (HCR) has specific domain to use. OCR system is most fitted for the applications like multi selection examinations, written communication address resolution etc. In returning days, character recognition system would possibly function a key issue to make paperless setting by digitizing and process existing paper documents. During this paper, we have planned the detail study on existing strategies for hand written character recognition based on ANN. This paper presents an in depth review within the field of handwritten Character Recognition.

**Keywords**— HCR, Features, classification, Optical Character Recognition.

### I. INTRODUCTION

The development of handwritten recognition system began in the 1950. The objective of hand written character recognition is to simulate the human reading capabilities so that the computer can edit and read the work as human do with documents. Handwriting recognition technique has been one of the most interesting and challenging research areas in field of the image processing and pattern recognition in the recent years. It contributes hugely to the innovation of automation process and improves the interface between man and machine in several applications.

Devanagari is the most well liked script in Asian country and also the most well liked Indian language Hindi is written in script. Nepali, Indo-Aryan and Sanskritic languages are written in script. Moreover, Hindi is that the national language of Asian country and also the third most well liked language within the world [1].

The OCR actually is a converter which translates hand written text images to a machine based text. In general, hand written recognition is classified in two ways: offline and online. Here, the writing is basically capture optically by scanner the completed writing text is scanned by a scanner in to digital format.

That brings increase speed & precision to the entire recognition process. Handwritten recognition has been one of the most fascinating & challenging research areas in the field of image processing & pattern recognition in the recent year. OCR is a field of study than encompass many different solving methods. ANN (Sandhu & Leon, 2009), support vector machines & statistical classifiers seem to be the preferred solutions to the problem due to their proven accuracy in classifying new data. [32]

#### 1. Off-line handwritten character recognition

Offline character recognition refers to the typed/handwritten characters area unit scanned then regenerates into binary or grey scale image. The feature extraction and recognition method is carried over the binary image. Offline character recognition could be a tough and difficult task as there's no temporal order data concerning character strokes is offered. So offline character recognition is taken into account as a tougher task than its on-line counterpart.

#### 2. On-line handwritten character recognition

On-line handwritten character recognition is additionally called real time recognition of character. During this case, writing and recognition are done at the same time. User can write character on any sensory space wherever sensors can acquire the pen movements so on the premise of these pen movements characters are recognized. On-line character recognition is far easier than offline character recognition; as a result of temporal arrangement data is accessible there.

Hindi script employed within the Indian landmass is employed by over four hundred million folks. The most application of Hindi Handwriting Recognition is automatic computer recognition of characters from optically scanned and digitized pages of written text. OCR is one of the foremost fascinating and challenging areas of pattern recognition with varied practical application potentials.

It is well understood that the image recognition of written text could be a troublesome issue attributable to the writing style and therefore the form of the Characters hugely. Unconstrained Hindi writing style is more advanced than English cursive because of form of constituent strokes. The subsequent is that the specific details of the Hindi alphabet: eleven vowels, thirty three consonants, ten numerals and three special characters. Together with the vowel consonant and consonant-consonant mixtures, the amount will increase simultaneously.

One will visualize a Hindi Language word in terms of three strips: a core strip, a prime strip and a bottom strip. Prime and bottom strips have solely modifiers and diacritic marks wherever the core strip contains all characters and therefore the vowel modifier '।'. {The prime the highest} and bottom strip could also be empty for a word; solely top could also be gift or simply the lowest. The core and prime strips area unit separated by shirorekha. Following figures1 and 2 are glimpse of vowels and their corresponding modifiers, and consonants in Hindi script.

Vowels:	अ	आ	इ	ई	उ	ऊ	ऋ	ॠ	ए	ऐ	ओ	औ
Modifiers:		।	ि	ी	ु	ू	ॄ	ी	े	ै	ो	ौ

Figure1: Vowels and corresponding modifiers in Hindi script

क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट
ठ	ड	ढ	ण	त	थ	द	ध	न	प	फ
ब	भ	म	य	र	ल	व	श	ष	स	ह

Figure2: Consonants in Hindi script

Optical Character Recognition is a powerful field of analysis in Pattern Recognition. The matter of character recognition is classified according to two criteria. One is predicated on the sort of the text that is written or hand written. The opposite is predicated on the acquisition method which may be on-line or off-line [2]. It's typically thought-about that the on-line technique of recognizing written text has achieved better results than its off-line counterpart. This might be attributed to the actual fact that the number of data is also captured within the on-line case cherish the direction, speed and therefore the order of strokes of the handwriting. On the opposite hand machine-printed character recognition can do

excellent results on good quality documents. Just in case of on-line character recognition, there's real time recognition of characters. On-line systems have higher data for doing recognition since they need temporal arrangement data and might avoid the initial search step of locating the character as within the case of their off-line counterpart. On-line systems get the position of the pen or written character as perform of your time directly from the interface. Offline recognition of characters is understood as a difficult downside owing to the complicated character shapes and nice variation of character symbols written in different modes.

Character recognition is Associate in nursing art of sleuthing segmenting and distinguishing characters from image. Different number of exact Character recognition is method of sleuthing and recognizing characters from input image and converts it into computer code or different equivalent machine editable kind [3-5]. It contributes vastly to the advancement of automation method and raising the interface between man and machine in several applications [6]. Character recognition is one in every of the foremost fascinating and interesting areas of pattern recognition and computing [7, 8]. Character recognition is obtaining a lot of attention since last decade because of its big selection of application. Conversion of written characters vital is very important for creating many important documents involving our history, cherish manuscripts, into machine editable kind so it is simply accessed and preserved [9].

## II. PHASES OF RECOGNITION

Normally handwritten recognition is divided into six phases which are image acquisition, pre-processing, segmentation, feature extraction, classification and post processing. The block diagram of the basic character recognition is shown in figure3.

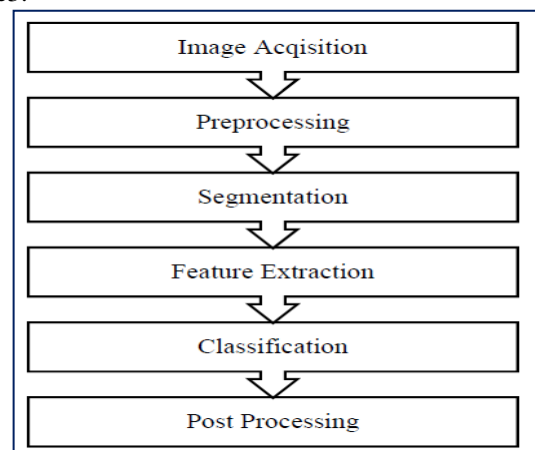


Figure3. Block diagram of handwritten character recognition system

### A. Image Acquisition

Digitized/Digital Image is at the start taken as input. The foremost common of those devices is that the electronic

medication or digitizer. These devices use a pen that's digital in nature. Input pictures for written characters can even be taken by exploitation different ways like scanners, images or by directly writing within the pc by employing a stylus.

### **B. Preprocessing**

Pre-processing is the basic part of character recognition at crucial permanent recognition rate. The most objective steps of pre-processing are to normalize strokes and take away variations that might otherwise complicate recognition and scale back the popularity rate. These variations or distortions embody the irregular size of text, missing points throughout pen movement collections, noise gift in text, left or right bend in handwriting and uneven distances of points from neighboring positions. Pre-processing includes five common steps, namely size normalization and centering, interpolating missing points, smoothing, slant correction and resampling of points.

### **C. Segmentation**

Segmentation is completed by separation of the individual characters of a picture. Typically document is processed in a very hierarchal manner. Initially level lines are divided victimisation row bar chart. From every row, words are extracted from victimisation column bar chart and at last characters are extracted from words.

### **D. Feature Extraction**

The main aim of feature extraction section is to extract that pattern that is most pertinent for classification. Feature extraction techniques like Principle part Analysis (PCA), Linear Discriminant Analysis (LDA), Chain Code (CC), Scale Invariant Feature Extraction (SIFT), zoning, gradient based mostly options, bar chart could be applied to extract the options of individual characters. These options area unit are used to train the system.

### **E. Classification**

When input image is conferred to HCR system, its options square measure extracted are given as an input to the trained classifier like artificial neural network or support vector machine. Classifiers compare the input feature with hold on pattern and resolve the most effective matching category for input.

### **F. Post Processing**

Post-processing refers to the procedure of correcting misclassified results by applying linguistic data. Post process is process of the output from form recognition. Language data will increase the accuracy obtained by pure form recognition. For handwriting input, some form recognizers yield one string of characters, whereas others yield variety of alternatives for every character, typically with a live of confidence.

## **III. RELATED WORK**

*Bag et al.* [10] propose a method to improve classification performance on Bangla Basic characters using topological features derived from the convex shapes of various strokes.

*Das et al.* [11] describes handwritten Bangla Character recognition using a soft computing paradigm embedded in a two pass approach. More specifically highly misclassified classes were combined to form a single group in the first pass or coarse classification. In the second pass, group specific local features were identified using Genetic Algorithm based region selection strategy to classify the appropriate class from the groups formed in the earlier pass. They used two different sets of features –

- a) Convex hull based features
- b) Longest run based features with Support Vector Machines (SVM), a well known classifier for this purpose.

They reported a recognition accuracy of 87.26% on a dataset of handwritten Bangla characters consisting of basic characters, compound characters and modifiers.

*Sarkhel et al.* [12] approached the issue from a perspective of multi-objective based region selection problem where the most informative regions of character samples were used to train SVM classifiers for character recognition. Two algorithms for optimization, specifically a Non-dominated Sorting Harmony Search Algorithm and Non-dominated Sorting Genetic Algorithm II were used to select the most informative regions with the objective of minimal recognition cost and maximal recognition accuracy. A recognition accuracy of 86.65% on a mixed dataset of Bangla numerals, basic and compound characters was reported.

*Das et al.* [13] described a Quad tree based features used for recognition of 55 frequently occurred compound characters covering 90% of the total of compound characters in Bangla using an Multilayer Perception (MLP) classifier.

*Pradeep et al.* [14] proposed neural network based classification of handwritten character recognition system. Each individual character is resized to 30 X 20 pixels for processing. They used binary features to train neural network. However such features are not robust. In post processing stage, recognized characters are converted to ASCII format. Input layer has 600 neurons equal to number of pixels. Output layer has 26 neurons as English has 26 alphabets. Proposed ANN uses back propagation algorithm with momentum and adaptive learning rate.

*Ashutosh et al.* [15] Gradient features based method was carried out on Hindi, third most popular language in the world. The first research work on handwritten Devnagari characters was published in 1977. Three hundred Million People use the Devnagari script for documentation in central and north region of India.

*I. K.Pathan et al.* [16] proposed offline approach for handwritten isolated Urdu characters in their work [13]. Urdu character may contain one, two, three or four segment. In which one component is known as primary (generally represents large continuous stroke) and rest of all are known as secondary components (generally represents small stroke or dots). Authors have used moment invariants (MI) feature to recognize the characters. MI features are well known to be invariant under rotation, translation, scaling and reflection. MI features are measure of the pixel distribution around the center of gravity of character and it captures the global character shape information. If character image is single component then it is normalized in 60 X 60 pixels and horizontally divided in equal three parts. 7 MI's are extracted from each zone and 7 MI are calculated from overall image, so total of 28 features are used to train SVM. If image is having multi component then 28 MI's are extracted from primary component (60 X 60) and 21 MI's are extracted from secondary component (22 X 22). Separate SVM are trained for both and decision is taken based on rules satisfying some criteria. Proposed system claim to get highest 93.59 % accuracy.

*Pritpal singh et al.* [17] Mentioned wavelet transforms based handwritten character and numeral recognition for Gurmukhi script. Color images are converted in gray scale and median filter is applied to remove noise. Binarized image is then normalized to 32 X 64 pixels size. Wavelets are localized basis functions which are translated and dilated versions of some fixed mother wavelet. The decomposition of the image into different frequency bands is obtained by successive low pass and high-pass filtering of the signal and down-sampling the coefficients after each filtering.

*Pal et al.* [18] proposed a scheme to slant correction and water reservoir concept for trilingual (Bangla, English and Hindi) city-name identification for Indian postal automation using lexicon information.

*Pal et al.* [19] used Zernike moment and histogram of gradient with support vector machine for trilingual (English, Hindi and Bangla) offline signature identification.

*Wshah et al.* [20] proposed a novel multilingual word-spotting framework based on hidden Markov model that works on the mass of multilingual handwritten documents and documents that contain more than one handwritten script.

*Rajput and Anita* [21] presented a multiple-feature-based approach to recognize the script type of the collection of handwritten documents, and eight popular Indian scripts are considered. The wavelets of Daubechies family, discrete cosine transform and Gabor filter are used to extract features

and obtained 100 % recognition accuracy for line-level biscripts.

*Wu et al.* [22] proposed a novel offline text-independent writer identification method based on scale invariant feature transform (SIFT) for three English datasets, one Chinese dataset and two hybrid language datasets.

*Chaudhuri and Bera* [23] proposed a text-line identification of handwritten Indian inscriptions, particularly of Malayalam, Bangla, English, Hindi, etc., and a new dual-method-based interdependency between inter-line gap and intra-line gap is proposed.

*Fiaz Hussain and John Cowell* [24] developed a system for Character Recognition of Arabic and Latin Scripts. They also discussed the main components used in the multi-stage system, paying particular attention to the normalization process used for orientation and size for a given bitmapped character.

*Anbumani et al.* [25] proposed an Optical Character Recognition of Printed Tamil Characters. Author used statistical parameters during recognition stage.

*K. Gaurav et al.* [26] this paper deals with the various pre-processing techniques involved in the character recognition with different kind of images ranges from a simple handwritten form based documents and documents containing colored and complex background and varied intensities. In this, different preprocessing techniques like skew detection and correction, image enhancement techniques of contrast stretching, binarization, noise removal techniques, normalization and segmentation, morphological processing techniques are discussed. It was concluded that using a single technique for preprocessing, we can't completely process the image. However, even after applying all the said techniques might not possible to achieve the full accuracy in a preprocessing system.

*Salvador España-Boquera et al.* [27] in this paper hybrid Hidden Markov Model (HMM) model is proposed for recognizing unconstrained offline handwritten texts. In this, the structural part of the optical model has been modeled with Markov chains, and a Multilayer Perceptron is used to estimate the emission probabilities.

*Sandhya Arora et al.* [28] used four feature extraction techniques namely, intersection, shadow feature, chain code histogram and straight line fitting features. Shadow features are computed globally for character image while intersection features, chain code histogram features and line fitting features are computed by dividing the character image into different segments. On experimentation with a dataset of 4900 samples the overall recognition rate was observed 92.80% for Devanagari characters.

*M. Hanmandlu, O.V. Ramana Murthy* [29] presented in their study the recognition of handwritten Hindi and English numerals by representing them in the form of exponential membership functions which serve as a fuzzy model. The recognition is carried out by modifying the exponential membership functions fitted to the fuzzy sets. These fuzzy sets are derived from features consisting of normalized distances obtained using the Box approach. The membership function is modified by two structural parameters that are estimated by optimizing the entropy subject to the attainment of membership function to unity. The overall recognition rate is found to be 95% for Hindi numerals and 98.4% for English numerals.

*Vedgupt Saraf and Rao* [30] have used genetic algorithm in character recognition of devnagari script. A flow chart for their entire process can be founded. Through their work using genetic algorithm, they claim to have an accuracy of around 97%-98%, although there are pairs that they found confusing.

*I K Sethi et al.* [31] described Devanagari numeral recognition based on the structural approach. The primitives used are horizontal line segment, vertical line segment, right slant and left slant. A decision tree is employed to perform the analysis based on the presence/absence of these primitives and their interconnection. A similar strategy was applied to constrained hand printed Devanagari characters.

#### IV. DISCUSSION & CONCLUSION

The paper discusses thoroughly all advances within the space of written character recognition. The foremost correct answer provided during this space directly or indirectly depends upon the standard similarity because the nature of the fabric to be scan. Varied techniques are precisely during this paper for character recognition in handwriting recognition system. From the study done to date, it's analysed that the choice of the classification similarity because the feature extraction techniques must be correct so as to achieve sensible rate in recognizing the character. A study within the paper reveals that there's still scope of enhancing the algorithms similarly as enhancing the speed of recognition of characters.

#### REFERENCES

- [1]. U. Pal, B. B. Chaudhuri, "Indian Script Character recognition: A survey", *Pattern Recognition*, vol. 37, pp. 1887-1899, 2004.
- [2]. Dholakia, K., A Survey on Handwritten Character Recognition Techniques for various Indian Languages, *International Journal of Computer Applications*, 115(1), pp 17-21, 2015.
- [3]. Nafiz Arica, Fatos T. Yarman-Vural, "An Overview of Character Recognition Focused On Off-line Handwriting", C99-06-C-203, IEEE, 2000.
- [4]. R. J. Ramteke, S. C. Mehrotra, "Recognition of Handwritten Devnagari Numerals", *International Journal of Computer Processing of Oriental Languages*, 2008.
- [5]. Harikesh Singh, R. K. Sharma, "Moment in Online Handwritten Character Recognition", *National Conference on Challenges & Opportunities in Information Technology (COIT-2007)* Gobindgarh. March 23, 2007.
- [6]. Aggarwal, A., Rani, R. and Dhir, R. (2012), "Handwritten Devanagari character recognition using Gradient features", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 2, 5, pp. 85-90.
- [7]. Shah, M. and Jethava, G. B. (2013), "A literature review on handwritten character recognition", *International Journal of Indian Streams Research Journal*, Vol. 3, 2, pp. 1-19.
- [8]. Das, N., Das, B., Sarkar, R., Basu, S., Kundu, M. and Nasipuri, M. (2010), "Handwritten BanglaBasic and Compound characterrecognition using MLP and SVM classifier", *Journal of Computing*, Vol. 2, 2, pp. 109- 115.
- [9]. Vaidya, S.A. and Bombade, B.R. (2013), "A Novel Approach of Handwritten Character Recognition using Positional Feature Extraction", *International Journal of Computer Science and Mobile Computing*, Vol. 2, 6, pp. 179-186.
- [10]. Bag, S., Bhowmick, P., Harit, G., 2011. Recognition of bengali handwritten characters using skeletal convexity and dynamic programming, in: *Emerging Applications of Information Technology (EAIT)*, 2011 Second International Conference on, pp. 265-268.
- [11]. Das, N., Sarkar, R., Basu, S., Saha, P.K., Kundu, M., Nasipuri, M., 2015. Handwritten bangla character recognition using a soft computing paradigm embedded in two pass approach. *Pattern Recognition* 48, 2054 – 2071.
- [12]. Sarkhel, R., Das, N., Saha, A.K., Nasipuri, M., 2016. A multi-objective approach towards cost effective isolated handwritten bangla character and digit recognition. *Pattern Recognition* 58, 172-189.
- [13]. Das, N., Basu, S., Sarkar, R., Kundu, M., Nasipuri, M., Basu, D., 2009. Handwritten bangla compound character recognition: Potential challenges and probable solution, in: *IICAI*, pp. 1901-1913.
- [14]. J. Pradeepa, E. Srinivasana, S. Himavathib, "Neural Network Based Recognition System Integrating Feature Extraction and Classification for English Handwritten", *International journal of Engineering*, Vol.25, No. 2, pp. 99-106, May 2012.
- [15]. Ashutosh Aggarwal, Rajneesh Rani, RenuDhir, "Handwritten Devanagari Character Recognition Using Gradient Features", *International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277-128X)*, Vol. 2, Issue 5, pp. 85- 90, May 2012.
- [16]. Imaran Khan Pathan, Abdulbari Ahmed Bari Ahmed Ali, Ramteke R.J., "Recognition of offline handwritten isolated Urdu character ", *International Journal on Advances in Computational Research*, Vol. 4, Issue 1, pp. 117-121, 2012
- [17]. Pritpal Singh, Sumit Budhi raja, "Handwritten Gurumukhi Character Recognition using Wavelet Transforms", *International Journal of Electronics, Communication & Instrumentation Engineering Research and Development*, Vol. 2, Issue 3, pp. 27-37, Sept 2012.
- [18]. Pal U, Roy RK, Roy K, Kimura F (2012) Multi-lingual city name recognition for Indian postal automation. In: 2012 International conference on frontiers in handwriting recognition (ICFHR), pp 169-173, 18-20 Sep 2012. ISBN: 978-1-4673-2262-1. doi:10.1109/ICFHR.2012.238
- [19]. Pal S, Alireza A, Pal U, Blumenstein M (2012) Multi-script offline signature identification. In: 12th International conference

- on hybrid intelligent systems (HIS), pp 236–240, 4–7 Dec 2012. ISBN 978-1-4673-5114-0. doi:10.1109/HIS.2012.6421340
- [20]. Wshah S, Kumar G, Govindaraju V (2012) Multilingual word spotting in offline handwritten documents. In: 21st International conference on pattern recognition (ICPR), pp 310–313, 11–15 Nov 2012. INSPEC Accession Number: 13324552, ISSN 1051-4651
- [21]. Rajput GG, Anita HB (2012) Handwritten script recognition using DCT, Gabor filter and wavelet features at line level. In: Book title: soft computing techniques in vision science, pp 33–43. ISBN 978-3-642-25506-9. doi:10.1007/978-3-642-25507-6\_4
- [22]. Wu X, Tang Y, Bu W (2014) Offline text-independent writer identification based on scale invariant feature transform. IEEE Trans Inf Forensics and Secur 526–536. ISSN 1556-6013. doi:10.1109/TIFS.2014.2301274
- [23]. Chaudhuri BB, Bera S (2009). Handwritten text line identification in Indian scripts. In: 10th International conference on document analysis and recognition, 2009. ICDAR '09, pp 636–640, 26–29 July 2009. ISBN 978-1-4244-4500-4. INSPEC Accession Number: 10904634
- [24]. Fiaz Hussain, John Cowell. "Character Recognition of Arabic and Latin Scripts," iv, p. 51, Fourth International Conference on Information Visualisation, 2000.
- [25]. Anbumani Subramanian and Bhadri Kubendran, "Optical Character Recognition of Printed Tamil Characters", Department of Electrical and Computer engineering, Virginia Tech, Blacksburg, December 10, 2000.
- [26]. K. Gaurav and Bhatia P. K., "Analytical Review of Preprocessing Techniques for Offline Handwritten Character Recognition", 2nd International Conference on Emerging Trends in Engineering & Management, ICETEM, 2013.
- [27]. Salvador España-Boquera, Maria J. C. B., Jorge G. M. and Francisco Z. M., "Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, April 2011.
- [28]. Sandhya Arora, "Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition", IEEE Region 10 Colloquium and the Third ICIS, Kharagpur, INDIA, December 2008.
- [29]. M. Hanmandlu, O.V. Ramana Murthy, "Fuzzy model based recognition of handwritten numerals", pattern recognition, vol.40, pp.1840-1854, 2007.
- [30]. Vedgupt Saraf, D.S. and Rao, 2013. "Devnagari script character recognition using genetic algorithm for better efficiency", IJSCE, ISSN: 2231-2307, Volume-2, Issue-4, April 2013.
- [31]. I.K. Sethi, "Machine Recognition of Constrained Hand Printed Devanagari", Pattern Recognition, Vol. 9, 1977, pp. 69-75.
- [32]. E.Tautu & F.Leon , "Optical Character Recognition System Using Support Vector Machines " 2012 PP -1-13 , 2012.