A Review on Recognition of Indian Handwritten Numerals

Geetika Mathur
Mody University of Science and Technology, Rajasthan
geetika.mathur123@gmail.com

Suneetha Rikhari
Mody University of Science and Technology, Rajasthan
suneetha.rikhari@gmail.com

Abstract: This paper presents a detailed review of various handwritten numeral recognition of various Indian languages. Character recognition is a field of pattern recognition which is in active research for the past few decades for its vast fields of application and still remains a challenging topic for research. Handwriting recognition is one the applications of Character Recognition. Various handwriting recognition for Indian languages has been under research in the past few years. The major challenges with handwritten character recognition are that the non-uniformity in the writing style of character, different styles of writing the same character, smoothness of the curves of the character and similarity between two characters. Several techniques have been developed to recognize Indian Numerals and have varying recognition rate. Some of the techniques are discussed below.
Keywords: Neural Network, OCR, Multilingual Documents, Handwritten Documents.

I. INTRODUCTION

Character recognition is a prominent field of pattern recognition which is in active research for the past few decades for its vast fields of application and still remains a challenging topic for research. Historically, optical character recognition was considered a technique for solving certain pattern recognition problems [1]. Character recognition by definition means a process to classify the input characters as per a pre-defined class of characters. The development of character recognition is very remarkable and many methods have been developed for the same. [2]The development of character recognition in last decade is remarkable and the methods for character detection are vast. The advancements of Character Recognition are evident in Optical Character Recognition (OCR), Document Classification, Computer Vision, Data Mining, Shape Recognition, and Biometric Authentication [2].Character recognition has its application in identification of text in images.

According to Census of India, India has 122 major languages as well as 1599 other languages. But it is very challenging task to recognize any Indian script character and digits for over a half century research in this area is ongoing and character recognition rate in modern OCR is above 99% on a high-quality document and 85% of handwritten documents. For degraded documents and books, the efficiency of OCR comes down to 80%. Various handwritten numerals recognition has been developed to recognize the handwritten numerals in Indian languages.
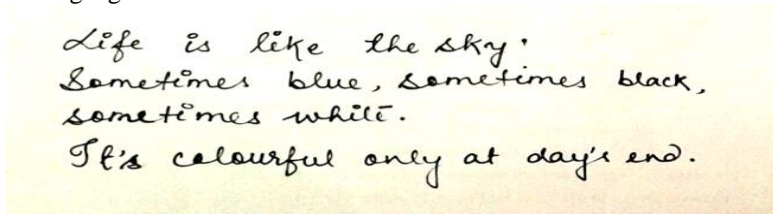

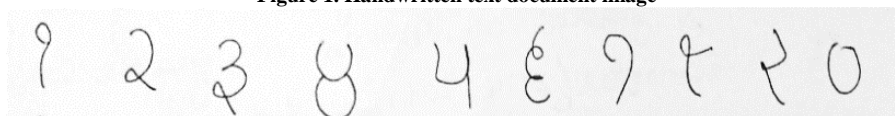**Figure 1. Handwritten text document image**

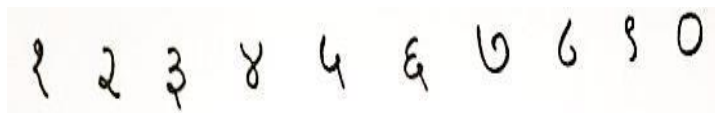**Figure 2. Samples of Handwritten Punjabi number series**

Figure 3. Samples of Handwritten Hindi number series

Figure 4. Samples of Handwritten Telugu number series

Figure 5. Samples of Handwritten Bengal number series

II. Components of an OCR system

An OCR system consists of 5 phases that are Scanning of image, Pre-processing and Segmentation, Feature Extraction, Classifications and Recognition and Post Processing. [3]. In scanning step the image is digitalized. The quality of the image obtained depends highly on the scanner used. Generally, in practical applications, the image obtained are not perfect some unnecessary details in the image which can cause a disturbance in the detection of the characters in the given image. Preprocessing involves removal of noise by applying various filters and conversion of the image like an RGB image can be converted into grayscale or binary image for further processing of the image. Feature extraction involves extracting the feature required for the system to recognize the characters. Classifications and Recognition phase is the recognition phase of the process. After finishing the OCR process several post processing steps can be done depending on the application, e.g. tagging the documents with some secondary data like author, year, etc.or proof-reading the documents for correcting OCR errors and spelling mistakes [4].
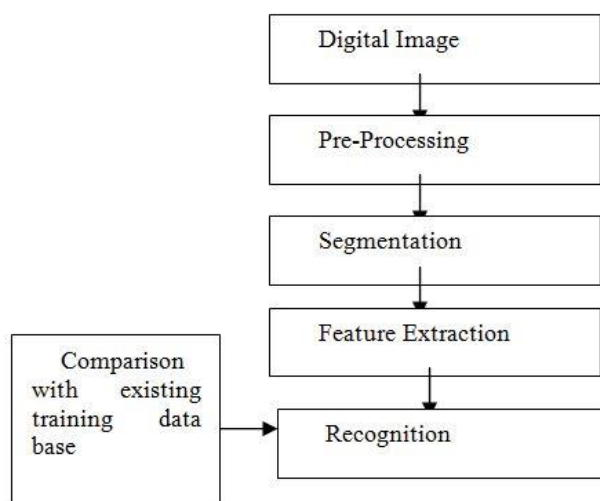

Figure 6. Components of an OCR system

III. Various approaches for OCR

1. Matrix Matching: Matrix Matching converts each character into a matrix, and then compares the pattern with a database of known characters. Its recognition rate is strongest on monotype and uniform single column pages. [3]
2. Fuzzy Logic: Fuzzy logic is a multi-valued logic that also allows the middle values to be defined between conservative assessments like yes or no, true or false, black or white etc. Fuzzy logic is used when answers do not ensure a distinct true or false value and there is some uncertainty involved.
3. Feature Extraction: This method defines each character by some key features, like its height, width, density, loops, lines and some other character traits. Feature extraction is a perfect method for OCR of magazines, laser print, and high-quality images.
4. Neural Networks: Neural networks process information in a similar way the human brain does. It samples the pixels of each image and matches them to a known catalog of character pixel patterns which were trained beforehand using some training images. The ability to recognize characters through this generalization method is great for faxed documents and damaged text. The algorithm of the artificial neural network has been applied successfully in the field of artificial intelligence, voice recognition, image processing etc. All the areas like Artificial intelligence, neural networks cognitive modeling, and are information processing structure inspired by the working of biological neural systems. [1][4]

IV. LITERATURE SURVEY

N.P. Banashree, D. Andhre, R. Vasanta and P.S. Satyanarayana[5] in 2007 has successfully implemented diffusion halftoning algorithm for Hindi numerals recognition using the neural network and 16-segment concept for feature extraction. Here they have achieved an accuracy level up to 98% for digits.

In 2008, four south Indian scripts were successfully classified by S.V. Rajashekararadhya and P.V. Ranjan in [6].They had considered back propagation neural network as a classifier and from experimental setup accuracy of 99% for Kannada and Telugu, 96% for Tamil and 95% for Malayalam was obtained.

Gujarati handwritten digit identification was done by A Desai [7] in 2010 which used the artificial neural network to recognize Gujarati digits and achieved approximately 82% of success rate.

In 2012, Marwan A. Abu-Zanona and Bassam M. El-Zaghmouri [8] developed an algorithm to recognize Arabic Hand Written Numbers using Segmentation and Artificial Neural Network. The recognition accuracy of this algorithm is 98%.

In 2014, Mahendra Chaudhary, M.Hasnine Mirja, N K Mittal,[9] recognized Hindi Numerals using Neural Network.The network used was fed forward network as classifier and segmentation for feature extraction. The success rate for the same was 80%.

Leo Pauly, Rahul D Raj, Dr.Binu paul,[10] in 2015 used the artificial neural network and HOG features to recognize various Hand written Digit Recognition System for South Indian Languages. The recognition rate for the same was 83% for Malayalam, 84% for Devanagari, 83% for Hindi, 85% for Telugu and 82% for Kannada. The overall Recognition rate for the same was 83.4%.

Abhishek Sethy, Prashanta Kumar Patra[11] recognized Off-line Odia Handwritten Numeral Recognition Using Neural Network where they used binarization as a training technique to recognize Odia Handwritten Numerals using feed forward neural network. The success rate for this technique was 80.2% as some Odia numerals look alike.

V. RESULT USING COMPARISON TABLES

S.No.	Year	Authors	Proposed Work	Accuracy
1	2007	N.P. Banashree, D. Andhre, R. Vasanta and P.S. Satyanarayana	Diffusion halftoning algorithm for Hindi numerals recognition using the neural network and 16-segment concept for feature extraction.	98%
2	2008	S.V. Rajashekararadhya and P.V. Ranjan	Back propagation neural network for four Indian Languages	99% for Kannada and Telugu, 96% for Tamil 95% for Malayalam was obtained
3	2010	A. Desai	Artificial neural network to recognize Gujarati digits	82%
4	2012	Marwan A. Abu-Zanona and Bassam M. El-Zaghmouri	Recognize Arabic Hand Written Numbers using Segmentation and Artificial Neural Network	98%
5	2014	Mahendra Chaudhary, M.Hasnine Mirja	Recognize Hindi Numerals using Neural Network	80%
6	2015	Leo Pauly, Rahul D Raj, Dr.Binu paul	Hand wrote Digit Recognition System for South Indian Languages using the artificial neural network and HOG features.	83% for Malayalam, 84% for Devanagari, 83% for Hindi, 85% for Telugu and 82% for Kannada. Overall Recognition rate = 83.4%.
7	2016	Abhishek Sethy, Prashanta Kumar Patra	Off-line Odia Handwritten Numeral Recognition Using Neural Network	80.2%

CONCLUSION

There has been a trend to digitalize data. The research in the field of OCR has been developed remarkably for the past decade. In this paper, we have reviewed the various Indian handwritten numeral detection schemes. Many techniques like HOG, 16-segment concept, Segmentation and much more are used for feature extraction. Feed forward neural network is generally used for recognition. The same techniques can be extended to do recognition of handwritten Indian language characters.

REFERENCES

[1]   Shalin A. Chopra, Amit A. Ghadge, Onkar A. Padwal, Karan S. Punjabi, Prof. Gandhali S. Gurjar,"Optical Character Recognition", *International Journal of Advanced Research in Computer and Communication Engineering* , Vol. 3, Issue 1, January 2014, ISSN (Online) : 2278-1021,ISSN (Print): 2319-5940, pp. 4956-4958.

[2] Suruchi G. Dedgaonkar, Anjali A. Chandavale, Ashok M. Sapkal,"Survey of Methods for Character Recognition", *International Journal of Engineering and Innovative Technology (IJEIT),* Volume 1, Issue 5, May 2012, ISSN: 2277-3754.,pp. 180 – 189.

[3] Sarika Pansare, Dhanshree Joshi," A Survey on Optical Character Recognition Techniques", *International Journal of Science and Research (IJSR)*, Volume 3 Issue 12, December 2014, ISSN (Online): 2319-7064, pp. 1247-1249.

[4] Sukhpreet Singh," Optical Character Recognition Techniques: A Survey", *Journal of Emerging Trends in Computing and Information Sciences,* Vol. 4, No. 6 June 2013, ISSN 2079-8407, pp. 545-550.

[5] N.P. Banashree, D. Andhre, R. Vasanta and P.S. Satyanarayana, "OCR for script identification of Hindi (Devnagari) numerals using error diffusion Halftoning Algorithm with a neural classifier," Proceedings of World Academy of Science Engineering and Technology 20, pp. 46–50, 2007.

[6] S.V. Rajashekararadhya and P.V. Ranjan, "Efficient zone-based feature extraction algorithm for handwritten numeral recognition of popular south Indian scripts", *Journal of Theoretical and Applied Information Technology 7 (1)*, 2009, pp. 1171–1180.

[7] A. Desai, "Gujarati handwritten numeral optical character recognition through the neural network," Pattern Recognition, vol. 43, 2010, pp. 2582-2589.

[8] Marwan A. Abu-Zanona, Bassam M. El-Zaghmouri, "Current Arabic (Hindi) Hand Written Numbers Segmentation and Recognition Advance ImageProcessing and Neural Network", *Journal of Emerging Trends in Computing and Information Sciences*, VOL. 3, NO. 6, June 2012, ISSN 2079-8407, pp. 936- 941.

[9] Mahendra Chaudhary, M.Hasnine Mirja, N K Mittal, "Hindi Numeral Recognition using Neural Network", *International Journal of Scientific & Engineering Research*, Volume 5, Issue 6, June-2014, ISSN 2229-5518, pp. 260-268.

[10] Gunjan Singh, Sushma Lehri," Recognition of Handwritten Hindi Characters using Backpropagation Neural Network", *International Journal of Computer Science and Information Technologies*,*2012*, Vol. 3 (4) , pp. 4892-4895.

[11] Abhisek Sethy , Prashanta Kumar Patra , "Off-line Odia Handwritten Numeral Recognition Using Neural Network: A Comparative Analysis", *International Conference on Computing, Communication and Automation (ICCCA2016)*, pp. 1099-1103, ISBN: 978-1-5090-1666-2/16.