

A Study of Handwritten Character Recognition Techniques

Ravleen Kaur Chadha¹, Neeraj Mehta²

Department of Computer Science & Engineering, IPS, Indore, India¹

Assistant Professor, Department of Computer Science & Engineering, IPS, Indore, India²

Abstract: In this paper we give a survey of handwritten character recognition methods. These methods are based on neural network, self organizing map, statistical methods etc. We also give an overview of fundamentals of handwritten character recognition. We describe modern and popular approaches for handwritten character recognition. Their strengths and weaknesses are also compared.

Keywords: Character recognition, feature extraction, neural network, optical character recognition.

I. INTRODUCTION

CRS stands for character recognition system. It is heavily used in data entry, text entry, process automation, large scale data processing e.g. postal address reading, cheque sorting, shorthand transcription, medical transcription, reading aids for visually handicapped people etc.

Character recognition is an art done to detect; segment and identify the characters from an image. More precisely Character recognition is procedure of detecting and recognizing the characters from input image and converts it into ASCII or other corresponding machine editable form (K. Ding et al. [6]). It contributes immensely to the development of computerization process and enhancing the edge between man and machine in many applications (J. Pradeepa et al. [5]).

Character recognition is one of the most interesting and attractive areas of pattern recognition and artificial intelligence (L. Cheng- Lin et al. [7]). Character recognition is getting additional consideration while last decade due to its large collection of applications. Translation of handwritten characters is important for converting numerous important documents related to our past, such as manuscripts, into machine editable form so that we can access it easily. Independent work is done in Optical Character Recognition to process the printed/computer generated document and handwritten or manually created document processing i.e. handwritten character recognition.

II. LITERATURE SURVEY

Picture handling and recognition of patterns or design has influence in handwritten character recognition. In [12], R. Tokas, A. Bhadu have communicated distinctive sorts of arrangement of feature extraction ways e.g. statistical feature based techniques, structural feature based techniques or global transformation methods. Factual routines which depend on arranging of in what manner ought to information be chosen. It uses the measurable dissemination of pixels present in a picture. It can be basically arranged in three classes: 1). Partitioning in regions, 2). Profile generation and projections

3). Distances and crossing. Auxiliary elements can be separated from structure and geometry of characters like number of horizontal and vertical lines, perspective proportion, number of cross focuses, number of circles, number of branch focuses, number of strokes, number of bends and so forth. Global elements are considered by changing over pictures in recurrence space.

Feature separation can be either a low level component or an abnormal state highlight. Where low level elements can incorporate width, tallness, waviness, angle proportion and so on of the characters of a dialect. These alone can't be utilized to separate one character from another in the character set (A. Sampath et al. [2]). In this means, there are a variety of other abnormal state elements, for example, number and position of circles, straight lines, features, bends and so forth which ought to likewise be incorporated.

T. Dash have proposed HCR utilizing cooperative memory net (AMN) in their paper [14]. They worked specifically at pixel level. Dataset was planned in MS Paint 6.1 with ordinary Arial text style of size 28. Measurement of picture was kept 31 X 39. When characters are extricated, their paired pixel qualities were specifically used to prepare AMN.

I. K. Pathan et al have talked about a logged off methodology for manually written and separated Urdu characters in their work [4]. Urdu character may contain one, two, three or four section. In which one part is perceived as essential (for the most part speaks to substantial persistent stroke) and rest of all are perceived as optional segments (for the most part speaks to little stroke or specks). Creators utilized minute invariants (MI) highlight to recognize the characters. MI elements are surely understood to be invariant under turn, interpretation, scaling and reflection. MI components are determination of the pixel dissemination around the focal point of gravity of character and it speaks to the worldwide character shape data. On the off chance that character picture is single segment then it is made uniform in 60 X 60 pixels and on a level plane separated in

equivalent 3 sections. 7 MI are extricated from every zone and 7 MI are computed from general picture, so aggregate of 28 elements are utilized to prepare SVM. What's more, if picture is having multi part than 28 MI are extricated from essential segment (60 X 60) and 21 MI are separated from auxiliary segment (22 X 22). Separate SVM are prepared for both and choice is taken taking into account rules fulfilling some criteria. Proposed framework can get most noteworthy 93.59 % precision.

In [5], J. Pradeepa et al have anticipated neural system based arrangement of handwritten character recognition framework. Every individual character is resized to 30 X 20 pixels for preparing. They are utilizing paired components to prepare neural system. However such elements are not strong. In post preparing stage, perceived characters are changed over to ASCII design. Info layer has 600 neurons equivalent to number of pixels. Yield layer has 26 neurons as English has 26 letter sets. Proposed ANN utilizes back proliferation calculation with force and versatile learning rate.

R.L. Das et al have proposed Hidden Markov Model based framework for English HCR in their writing [11]. They have utilized global and in addition nearby element extraction routines. Global component includes four angle highlights, six projection highlights and four bend highlights. What's more, to concentrate neighborhood highlights, picture is partitioned into nine equivalent pieces and 4 inclination components are figured from every square, so aggregate of 36 elements are extricated.

So general component vector contains 50 highlights for each characters. $O = [G(4) P(6) C(4) L(36)]$, where G, P, C and L speaks to global angle, projection, arch and neighborhood slope highlights individually. Number in bracket speaks to number of particular elements. HMM is prepared utilizing these element and a trial is performed. Post preparing is similarly connected after recognition period of HMM to exceedingly confounded gathering of characters like N and M, O and Q, C and O and so forth. For every gathering new component is ascertained to separate characters inside of the gathering.

Angle highlights based technique is talked about in [1] by A. Aggarwal et al. Investigation is completed on Hindi, third most prominent dialect on the planet. The first and foremost research work done on handwritten Devnagari characters was distributed in 1977. 300 Million People utilize the Devnagari language for citations in central and north locale of India. In proposed system, Gradient Vector is computed at every pixel and picture is partitioned in 9 X 9 pieces. At that point quality of inclination is amassed in eight standard bearings in every sub square. 9 X 9 squares is further down examined to 5 X 5 piece utilizing Gaussian channel. Preprocessing ventures for proposed techniques are as per the following: Intensity estimations of picture were balanced and after that pictures were changed over to parallel with limit esteem 0.8. Associated segment with pixel thickness under 30 were expelled from further preparing. Middle channel was connected to uproot pepper and salt noise present in double pictures. Lastly, singular characters were removed by line and segment histogram

handling and standardized to 90 X 90 pixel square. Inclination highlight were separated utilizing sobel administrator.

V. Ganapathy, K. L. Liew had proposed multi-scale neural system based methodology in [16]. Neural systems like Feed forward back proliferation neural system requires long preparing time to remember and sum up all data highlight vectors. Furthermore, still there are great odds of misclassification. Speculation issue can be overcome by utilizing multi scale neural system. Proposed framework by T. Dash [14] first change over camera caught RGB picture to double picture. Width to Height Ratio (WH), Relative Height (RH) proportion, Relative Width proportion (RW) is figured to expel superfluous joined parts from picture. For multi scale neural system, recognized character is resized to 20 X 28 pixels, 10 X 14 pixels and 5 X 7 pixels. Parallel components of these diverse determination pictures are given to three layer feed forward back proliferation calculation.

In [16], T. Som and S. Saha have utilized fuzzy membership capacity based methodology for HCR. Character pictures are standardized into 20 X 10 pixels. Normal picture (melded picture) is produced using 10 pictures of every character. Using so as to hold box around character is controlled vertical and level projection of character. Subsequent to trimming picture to bouncing box, it is resized to 10 X 10 pixels size. After that, thinning is performed and diminished picture is set in one by one crude of 100 X 100 canvas. Closeness score of test picture is coordinated with combination picture and characters are grouped. In writing [10], R. K. Mandal, N R Manna has proposed single layer neural system based methodology for HCR to diminish preparing time. Characters are composed on A4 size paper in uniform box. Portioned characters are scaled to 80 X 80 pixels. Every 0 is supplanted by - 1 for better preparing.

A skew based element extraction work was enhanced by O.P. Sharma et al, said in [8]. They anticipated zone based mixture highlight extraction technique. Euler number idea is utilized to show signs of improvement rate and accuracy. Thresholding, sifting and diminishing operations are executed as a division of preprocessing. Division can be characterized into three general classes: Top down, Bottom up and Hybrid strategies. In anticipated system sectioned character is resized to 90 X 60. In the wake of ascertaining Euler number from this picture, character is isolated into 10 X 10 pixel 54 zones. Every zone quality is supplanted by typical force esteem and is utilized as highlight worth. 9 and 6 elements are figured by averaging values line insightful and segment astute, so it shapes complete 69 highlights.

In J. Du et al. [17] HDR approach, the preprocessing step incorporate, clamor decrease procedures and the skeletonization strategies were connected to annihilate commotion.

III. CONCLUSION

Literature review produce that neural network is the major option in the midst of specialists for training exercise.

Whereas, an assortment of adaption has been anticipated in feature extraction techniques. Work can be made far reaching in numerous ways e.g. extension from single character or set of characters to entire document handling. Different applications are in anticipation of the improvement in character recognition to be able to adopt it totally. Hybrid methods can be anticipated which utilizes more than one feature extraction methods to categorize characters correctly. Post processing is too vital to categorize comparable structured characters like Q and O, M and N resembling pairs in English language. It is not achievable to put forward completely automated system which can minimize all kinds of inconsistencies. HCR system have to be domain specific as well as restricted to limited flexibility in writing style for more correctness. Handwritten character recognition is still a volatile research part of pattern recognition. Each and every step is responsible straightly to the correctness of system, like preprocessing, segmentation, feature extraction, training methods etc. So all these parts are open for autonomous research. A lot enhancement can be made in each step.

REFERENCES

- [1.] Aggarwal, R. 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.
- [2.] A. Sampath, Tripti C, Govindaru V, "Freeman code based online handwritten character recognition for Malayalam using Back propagation neural networks", Advance computing: An international journal, Vol. 3, No. 4, pp. 51-58, July 2012.
- [3.] Bhatia, N. and Vandana, "Survey of Nearest Neighbor Techniques," International Journal of Computer Science and Information Security, Vol. 8, No. 2, (2001), 302-305.
- [4.] I.K. Pathan, A. A. 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.
- [5.] 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.
- [6.] K. Ding, Z. Liu, L. Jin, X. Zhu, "A Comparative study of GABOR feature and gradient feature for handwritten Chinese character recognition", International Conference on Wavelet Analysis and Pattern Recognition, pp. 1182-1186, Beijing, China, 2-4 Nov. 2007.
- [7.] L. Cheng-Lin, Nakashima, Kazuki, H, Sako, H.Fujisawa, "Handwritten digit recognition: investigation of normalization and feature extraction techniques", Pattern Recognition, Vol. 37, No. 2, pp. 265-279, 2004.
- [8.] O. P. Sharma, M. K. Ghose, K. B. Shah, "An Improved Zone Based Hybrid Feature Extraction Model for Handwritten Alphabets Recognition Using Euler Number", International Journal of Soft Computing and Engineering, Vol.2, Issue 2, pp. 504-58, May 2012.
- [9.] P. K. Charles, V.Harish, M.Swathi, CH. Deepthi, "A Review on the Various Techniques used for Optical Character Recognition", International Journal of Engineering Research and Applications, Vol. 2, Issue 1, pp. 659-662, Jan-Feb 2012.
- [10.] R. K. Mandal, N R Manna, "Hand Written English Character Recognition using Row- wise Segmentation Technique", International Symposium on Devices MEMS, Intelligent Systems & Communication, pp. 5-9, 2011.
- [11.] R. L. Das, B.K. Prasad, G. Sanyal, "HMM based Offline Handwritten Writer Independent English Character Recognition using Global and Local Feature Extraction", International Journal of Computer Applications (0975 – 8887), Volume 46– No.10, pp. 45-50, May 2012.
- [12.] R. Tokas, A. Bhadu, "A comparative analysis of feature extraction techniques for handwritten character recognition", International Journal of Advanced Technology & Engineering Research, Volume 2, Issue 4, pp. 215-219, July 2012.
- [13.] S. Deshmukh, L. Ragha, "Analysis of Directional Features - Stroke and Contour for Handwritten Character Recognition", IEEE International Advance Computing Conference, pp.1114-1118, 6-7 March, 2009, India.
- [14.] T. Dash, "Time efficient approach to offline hand written character recognition using associative memory net.", International Journal of Computing and Business Research, Volume 3 Issue 3 September 2012.
- [15.] T.Som, S. Saha, "Handwritten Character Recognition Using Fuzzy Membership Function", International Journal of Emerging Technologies in Sciences and Engineering, Vol.5, No.2, pp. 11-15, Dec 2011.
- [16.] V. Ganapathy, K.L. Liew, "Handwritten Character Recognition Using Multi scale Neural Network Training Technique", World Academy of Science, Engineering and Technology, pp. 32-37, 2008.
- [17.] J.Du , J.S. Hu,B. Zhu, S. Wei, L. R. Dai ,” A Study of Designing Compact Classifiers using Deep Neural Networks for Online Handwritten Chinese Character Recognition”, 22 International Conference on Pattern Recognition,2014.