Online Detection of Multi-Language Handwritten Document

Neeta R. Panagale
Dept. of Computer Engg.
Shatabdi Institute of Engineering & Research
MH,INDIA

Sonali A.Boraste
Dept. of Computer Engg.
Shatabdi Institute of Engineering & Research
MH,INDIA

Anuradha K.Chaudhari
Dept.of Computer Engg.
Shatabdi Institute of Engineering & Research
MH,INDIA

Mahesh T.Dhande
Dept. of Computer Engg.
Shatabdi Institute of Engineering & Research
MH,INDIA

Abstract: Handwriting has dependably been an essential apparatus of communication. New innovations have been created keeping in mind the end goal to record this transcribed information. Special procedures portray nature of manually written scripts and have executed techniques to change over it into electronic data. Recognition of Handwritten content has been one of the dynamic and testing zones of research in the field of picture handling and example recognition. It has various applications which incorporate, perusing help for visually impaired, bank checks and transformation of any written by hand archive into auxiliary content shape. We attempted to perceive written by hand characters by anticipating them on various measured networks. The initial step is image obtaining which procures the checked image took after by noise filtering, smoothing and normalization of filtered picture, rendering picture appropriate for division where picture is disintegrated into sub pictures. Highlight Extraction enhances acknowledgment rate and misclassification. Handwriting acknowledgment is in research for more than four decades and has pulled in numerous specialists over the world.

Keywords: Handwritten Character Recognition, Feature Extraction, Neural Network, Offline Character Recognition, Online Character Recognition.

I.   INTRODUCTION

Handwriting recognition

Is the capacity of a computer to get and translate clear manually handwritten input from sources, for example, paper documents, photos, touch-screens and different gadgets.The image of the written text might be detected "offline" from a bit of paper by optical scanning (optical character recognition) or insightful word recognition. On the other hand, the developments of the pen tip might be detected "on line", for instance by a pen-based computer screen surface.

Offline Handwriting Recognition

In this type of recognition, the text is not recognized in the meantime as it is delivered. Rather it is recognized after the user has completed the process of writing i.e. for this situation, the content is initially written on a surface, for example, paper and from that point on it is recognized by the computer by scanning the surface. The checked handwriting is initially put away carefully in dark scale arrange e.g. bitmap image, and afterward additionally preparing is done on it to have a good recognition accuracy. Highlights for recognition are upgraded and extracted from the put away bitmap image by using digital image processing. This kind of recognition is additionally some of the time called as "Optical Character Recognition" [7].Recognition of machine printed characters is additionally a piece of Optical Character Recognition. Offline methods are less reasonable for man-machine communication because no real time interactivity is present. They are suitable for programmed transformation of paper records to electric documents which then might be translated by computers.

Online Handwriting Recognition

Rather than the d offline method of handwriting recognition, online handwriting recognition is done progressively in real time i.e. at the same time as the handwriting is produced. The surface used for handwriting is usually a digitized tablet and it is utilized alongside an advanced pen likewise once in a while called "Stylus", keeping in mind the end goal to compose at first glance. As the pen moves over the surface, the two-dimensional co-ordinates of progressive focuses are gathered and put away as a function of time.

International Journal of Science Technology Management and Research
Volume 2 , Issue 1 , January 2017
www.ijstmr.com

On-line handwriting recognition HWR is the undertaking of changing a language that is spoken to in its spatial type of graphical imprints into its symbolic representation. On-line HWR, as opposed to offline HWR, alludes to the circumstance where the acknowledgment is performed simultaneously to the composition procedure. This refinement has a few outcomes, with respect to further separations between on-line and offline HWR. Among others, these distinctions influence imperative angles like data acquisition, data representation, recognition methods and typical applications. On-line handwriting data are an element, digitized representation of an (digital) pen development, for the most part depicting successive data about position, velocity, and acceleration.

Goals and Objectives

The Main goal is to recognize specific client's handwriting. Nobody can do extortion utilizing any individual's name ,since we are preparing are framework in that way where transcribed report in any language can be perceive by framework furthermore that client can be identify.

II.   Literature Survey

Handwriting recognition has been a standout amongst the most captivating and testing research territories in field of image preparing an example acknowledgment in the late years, this was expressed by S. Mori, C.Y.Suen and K. Kamamoto in 1992 [1]. It contributes enormously to the headway of automation process and enhances the interface amongst man and machine in various applications. All in all, Handwriting recognition is grouped into two sorts as disconnected and on-line Handwriting recognition techniques. The on-line techniques have been appeared to be better than their disconnected partners in perceiving manually written characters because of the worldly data accessible with the previous. Nonetheless, in the disconnected frameworks, similarly high recognition exactness levels .The recognition of characters from checked images of records has been an issue that has gotten much consideration in the fields of image preparing, design recognition and computerized reasoning. Recognition of manually written characters is an extremely complex issue. The characters could be composed in various size, introduction, thickness, organization and measurement. This will give infinite variations. This paper completes an investigation of different element based classification techniques for disconnected written by hand character recognition. After experimentation, it proposes an optimal character recognition technique. We have presented a radical new thought of acknowledgment of secluded written by hand digits which is known to be a difficult task and still does not have an agreeable specialized solution this was expressed by Qixiang Ye, David Doermann[10].

As of now specified, Handwriting recognition is a strategy by which a computer system can recognize characters and different images composed by hand in natural handwriting. At the largest amount, handwriting recognition can be broken into two classifications on the premise of how the raw data is acquired [3].Online handwriting character recognition is the way toward discovering what character is spoken to by a given online handwritten character information. Online handwriting character data is, all in all, an arrangement of the organize estimations of information focuses that are tested along the direction of the character as it is written. Future region of work will address utilizing portions at the word level for identifying the script. This will permit us to distinguish multi-script documents where different portions of the document can contain different scripts. The framework was prepared utilizing 500 samples handwritings given by both male and female participants of different age groups. Test system was trained on 500 samples other than tests for preparing that demonstrates that Fourier Description consolidated with back proliferation organize give good recognition accuracy of 94% for manually written English characters with less training time. For training, back-propagation algorithm has been implemented this was expressed by Verma B.K, and Dayashankar Singh [2, 5].The present paper proposes a novel approach for recognition of manually written digits i.e. neural system characterization. Back spread neural system is one of the most straightforward techniques for preparing multi layer neural systems. In this paper, we planned a back spread neural system and prepared it with an arrangement of manually written digits. The normal achievement rates of acknowledgment of all digits are 91.2%. An exploratory outcome demonstrates that ordinary components with back spread system yields good classification accuracy of 100% and recognition accuracy of 91.2%.

After a long period of concentrate on western and Asian scripts there is presently a general pattern in the on-line hand-writing recognition group to explore recognition of different scripts, for example, Arabic and different Indic scripts. In experiments two vast unconstrained handwriting databases, our approach accomplishes word recognition accuracies of 79.7% on online information and 74.1% on offline data.

International Journal of Science Technology Management and Research
Volume 2 , Issue 1 , January 2017
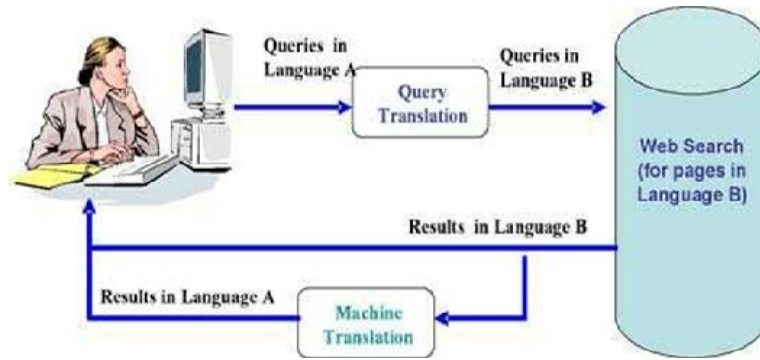**www.ijstmr.com**

III. EXISTING SYSTEM

Figure 1: Existing System

In existing system, there is a probability of recognizing handwriting which is composed in a single language. Google Translate is not perfect, particularly with the morphologically rich smaller languages, similar to Finnish and Hungarian. Use to perceive character E.g. (bolts, condition signs, star, heart shapes etc.)Google online handwriting recognition system that right now bolsters 22 scripts and 97 languages. We are executing neural system algorithm for proposed approach, which indicates higher productivity when contrasted with the current framework (i.e. Google translator).

Disadvantages
a) In existing framework, there is a probability of recognizing handwriting which is written in a single language.
b) Google Translate is not perfect, particularly with the morphologically rich smaller languages, similar to Finnish and Hungarian.

III.   PROPOSED SYSTEM

In our multi-language handwriting recognition system,we have proposed two modules-Training and Testing. In the preparation module, as a matter of first importance a image will be transferred for preparing.Image processing will be completed and its components will be extracted.The examples of the specific handwriting will be distinguished and put away in a document. Next, for testing our system we will transfer another image in which its and composing must be recognized. Again by playing out some image pre-preparing steps, we will separate its elements. These extracting elements will be contrasted and the information put away in preparing set and on the off chance that it matches, it shows the writer of the written by hand message else not.

ADVANTAGES
A) GOOGLE TRANSLATE IS NOT PERFECT, PARTICULARLY WITH THE MORPHOLOGICALLY RICH SMALLER LANGUAGES, SIMILAR TO FINNISH AND HUNGARIAN.
B) USE TO RECOGNIZE CHARACTER E.G. (BOLTS, CONDITION SIGNS, STAR, HEART SHAPES AND SO ON.)
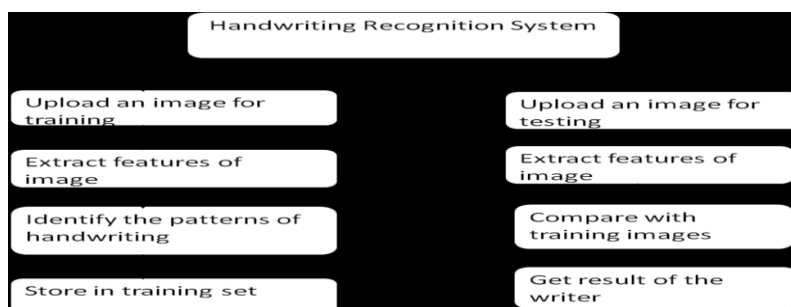
IV.   SYSTEM ARCHITECTURE

Figure 2: System Architecture

International Journal of Science Technology Management and Research
Volume 2 , Issue 1 , January 2017
www.ijstmr.com

Amlan Kundu et al. proposed different options for handwriting recognition [4]. Keeping in mind the end goal to distinguish a written by hand character, the image needs to experience some processing steps as described in Figure 2. The pre-preparing stage expelled undesirable noise, performs binarization and inclination angle correction.
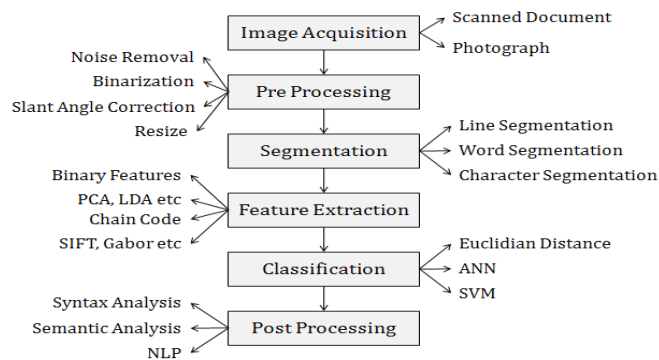

Figure 3: Basic stages in character recognition

1. Image Acquisition

Changing over paper report as advanced image with the assistance of filtering procedure is called image securing. Jon Almaz et al. had proposed a framework which tended to the issues of word spotting and word recognition on images [8]. Tijn van der Zant, et al., presented a strategy where the outcomes utilizing their framework as looked at standardized word-image coordinating were more proper [9].

2. Preprocessing

The advanced image acquired from filtering may contain some measure of noise depending on the nature of scanner. Reasoning of this noise from caught image is called preprocessing. Qixiang Ye and David Doermann had broke down and thought about specialized difficulties, strategies, and the execution of content identification and recognition look into in shading images [10]. Commotion may be a skewed or broken character. A detail clarification of commotion and content recognition in loud papers was proposed by Yefeng Zheng, et al. [11]. The strategy displayed was utilized for image upgrade to enhance page division precision of boisterous archives. After noise identification proof and removal, the zone division exactness expanded from 53% to 78% utilizing the Docstrum calculation. The key favorable position of pre-preparing a written by hand character was to systematize the data in order to make the recognition basic. Toru Wakahara and Kazumi Odaka set forward a versatile or classification subordinate standardization technique that standardizes an info design against every reference design utilizing worldwide/neighborhood relative change (GAT/LAT) [12]. It was demonstrated that GAT/LAT standardization on genuine, cursive info designs against right reference designs significantly diminished the level of penmanship vacillation inside a similar classification by around 20% and 40%. Other than noise lessening, standardization and division [13] of image is additionally done to make the further recognition prepare easy. In standardization, resizing of characters is accomplished for stroke width, incline, slant, stature of the characters. The standardization undertaking will trim down every character image to one comprising of vertical letters of uniform stature made up of one far reaching stroke [20],[14]. Zhenwen Dai and Jorg Lucke learned at expelling debasements from a solitary letter-estimate page construct just with respect to the data the page contains [15]. Shijian Lu, et al., exhibited a paper recovery procedure that was fit for looking record images without optical character acknowledgment (OCR) [16]. Their exploratory outcomes demonstrated that the proposed word image comment method was quick, strong, and equipped for recovering imaged records successfully.

3. Segmentation

Segmentation is a standout amongst the most imperative procedures that choose the achievement of character recognition system. It is utilized to rot an image of an arrangement of characters into sub images of individual images by fragmenting lines and words. R. Manmatha and Jamie L. Rothfeder actualized a review on segmentation of authentic archives [17]. This approach was tried on various diverse information sets and it was demonstrated that, on 100 tested papers of manually written archive images, an aggregate error rate of 17% was watched. Rakesh Kumar Mandal and N R Manna proposed a recognition framework utilizing CSIM in order to enhance its execution [18].

4. Feature Extractions

Components are data removed from the caught or info image. This data must be comparative for comparative images yet it must be unmistakable for different images. These components assume significant part in example recognition. Hence the determination of highlight extraction method turns into a vital calculates accomplishing high recognition execution. Kwok-Wai Cheung, et al. had actualized strategies in character extraction [19]. They had performed testing of by and large 633 written by hand city name images and accomplished a general exactness of around 92% for coordinating the main characters of the city names. Some component extraction techniques are projection, zoning, and

bordered transition and diagram coordinating [20]. The projection strategy performs pressure of information through a projection. Dark pixel tallies are brought parallel lines through the image zone to produce negligible conveyances. Border move system accepts that every one of the character is situated vertically. Every character is parceled into four equivalent quadrants. Zoning is a procedure that includes the segmentation of the character into littler part of ranges. Piotr Dollar and C. Lawrence Zitnick demonstrated that the outcomes got constant execution like requests of greatness speedier than many contending approaches [21]. Madhvanath, et al., led a review on shape representations of paired pictures of written by hand words [22]. The dictionary was lessened utilizing the elements of word length, ascenders, and descenders. With their framework, dictionary can be lessened into equal parts with an under 2% error rate. L. Anlo Safi and K.G.Srinivasagan exhibited an outline of highlight extraction systems for disconnected recognition of Tamil characters [23].

5. Classifications

Classification is done on the basis of components of the image. It is the way toward allotting information to their comparing class as for gatherings with homogenous attributes. Along these lines, it partitions the element space into a few classes in view of the choice run the show. Some order strategies utilized for recognition of written by hand character is neural system, support vector machine and so on.

6. Post Processing

Post processing implies gathering of images into string. The accuracy of optical character recognition can be expanded if the yield is compelled by a list that is permitted to happen in an archive. The yield stream might be plain content or record of characters [24].

CONCLUSION

In this paper, we learned about various strategies for recognition handwritten document. The system joins a couple existing and some new portions. For recognition handwritten document we use neural system algorithm. This paper discussed an approach that incorporated the use of Online Handwriting Recognition technique to lift the customer to have execution past the desires. We talked about variety of script- and language-specific elements that are used. The system described in this paper is at present utilized as a part of a few freely accessible items which include mobile and desktop variations of Google Translate, Google Handwriting Input for Android cell phones.

REFERENCE

1. S. Mori, C. Y. Suen and K. Kamamoto, "Historical review of OCR research and development", Proc. of IEEE, vol. 80, July1992, pp. 1029-1058.
2. Verma B.K, "Handwritten Hindi Character Recognition Using Multilayer Perceptron and Radial Basis Function Neural Network", IEEE International Conference on Neural
   Network, 4, 1995, pp. 2111-2115.
3. Tappert, C. C., Suen, C. Y., and Wakahara, T.,"Online Handwriting Recognition- A Survey", International Conference on Pattern recognition, vol. 2, 1988, pp. 1123-1132.
4. Amlan Kundu, Yang He, and Mou-Yen Chen, "Alternatives to Variable Duration HMM in Handwriting Recognition", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 20, No. 11, November 1998, pp. 1275- 1280
5. Dayashankar Singh, Maitreyee Dutta and Sarvpal H.Singh, "Neural Network Based Handwritten HindiCharacter Recognition", ACM International Conference(Compute 09), Jan. 9-10, 2009, Bangalore, pp. 1-7
6. Choudhary, Rishi, "Improving The Character Recognition Efficiency of Feed Forward BP Neural Network", International Journal of Computer Science & Information Technology (IJCSIT),Vol 3, No 1, Feb 2011, pp. 85-96.
7. Tappert, C. C., Suen, C. Y., and Wakahara, T., "The State of the Art in Online Handwriting Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 8, 1990, pp. 787-808.
8. Jon Almaz_an, Albert Gordo, Alicia Forn_es, and Ernest Valveny, "Word Spotting and Recognition with Embedded Attributes", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 36, No. 12, December 2014, pp. 2552- 2566.
9. Tijn van der Zant, Lambert Schomaker, and Koen Haak, "Handwritten-Word Spotting Using Biologically Inspired Features", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 11, November 2008, pp. 1945- 1957.
10. Qixiang Ye, David Doermann "Text Detection and Recognition in Imagery: A Survey", IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 37, No. 7, July 2015, pp. 1480-1500.
11. Yefeng Zheng, Huiping Liand David Doermann, "Machine Printed Text and Handwriting Identification in Noisy Document Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 3, March 2004, pp. 337-353.
12. Toru Wakahara, Kazumi Odaka "Adaptive Normalization of Handwritten Characters Using Global/Local Affine Transformation", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 20, No. 12, December 1998, pp. 1332- 1341.
13. Thierry Artieres, Sanparith Marukatat, and Patrick Gallinari, "Online Handwritten Shape Recognition Using Segmental Hidden Markov Models", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 29, No. 2, February 2007, pp. 205- 217.
14. Yu Qiao, Mikihiko Nishiara, and Makoto Yasuhara, "A Framework Toward Restoration of Writing Order from Single-Stroked Handwriting Image", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 28, No. 11, November 2006, pp. 1724- 1737.
15. Zhenwen Dai, Jorg Lucke, "Autonomous Document Cleaning A Generative Approach to Reconstruct Strongly Corrupted Scanned Texts", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 36, No. 10, October 2014, pp. 1950-1962.
16. Shijian Lu, Linlin Li, and Chew Lim Tan, "Document Image Retrieval through Word Shape Coding", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 30, No. 11, November 2008, pp. 1913- 1918.
17. R. Manmathaand Jamie L. Rothfeder, "A Scale Space Approach for Automatically Segmenting Words from Historical Handwritten Documents", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 27, No. 8, August 2005, pp. 1212-1225.
18. Rakesh Kumar Mandal, N R Manna, "Hand Written English Character Recognition using Column-wise Segmentation of Image Matrix (CSIM)", Wseas Transactions On Computers, Issue 5, Volume 11, May 2012, pp. 148-158

International Journal of Science Technology Management and Research
Volume 2 , Issue 1 , January 2017
www.ijstmr.com

19. Kwok-Wai Cheung, Dit-Yan Yeung and Roland T. Chin, "Bidirectional Deformable Matching with Application to Handwritten Character Extraction", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 24, No. 8, August 2002, pp. 1133- 1139.
20. Prachi Mukherji, Priti Rege, "Shape Feature and Fuzzy logic based offline Devanagari handwritten optical character Recognition", International Journal of Pattern Recognition Research, Vol. 4, 2009, pp. 52-68.
21. Piotr Dollar, C. Lawrence Zitnick, "Fast Edge Detection Using Structured Forests", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 37, No. 8, August 2015, pp. 1558- 1570.
22. S. Madhvanath,G. Kim, V. Govindaraju, "Chaincode Contour Processing for Handwritten Word Recognition", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 21, No. 9, September 1999, pp. 928-932.
23. L. Anlo Safi, K.G.Srinivasagan,"Offline Tamil Handwritten Character Recognition using Zone based Hybrid Feature Extraction Technique", International Journal of Computer Applications (0975 – 8887) Volume 65– No.1, March 2013, pp. 1-5
24. Amit Choudhary, Nikita Walia, "Evaluation of Neural Network Accuracy for Handwritten CharacterRecognition", International Journal of Information and Computation Technology Volume 3, Number 3 (2013), pp. 119-124.