INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165

Handwriting Character Recognition: Survey for Different Languages

Heena Patil[1], Vijay Shelke[2]

Master of Engineering Student, Dept. of Computer Engineering, Alamuri Ratnamala Institute of Engineering and Technology, Mumbai University, India[1].

Associate Professor, Dept. of Computer Engineering, Yadavrao Tasgaonkar College of Engineering and Management, Mumbai University, India[2]

ABSTRACT: Due to increased usage of digital technologies in all sectors and in almost all day-to-day activities to store and pass information. Handwriting recognition has been one of the most fascinating and challenging research areas in field of image processing and pattern recognition in the recent years. It contributes immensely to the advancement of automation process and improves the interface between man and machine in numerous applications, were human operators whose job was to convert data from various documents into electronic format, making the process quite long and often affected by errors.

Automatic text recognition aims at limiting these errors by using image preprocessing techniques that bring increased speed and precision to the entire recognition process. Here, we develop such a tool which takes an image as an input and extract characters such as alphabets, digits, symbols from it. The image can be of handwritten document or printed document. It can be used as a form of data entry from printed records. The implementation of such a tool depends on two factors – Feature extraction and classification algorithm.

This work discusses about a method for analyzing real world handwritten text samples with the aid of technology. This project is based on Machine learning, we can provide a lot of data set as an input to the software tool which will be recognized by the machine and similar pattern will be taken out from them. This survey provides a comprehensive review of existing works for various techniques of Handwritten Character Recognition (HCR) in different languages are reviewed and analyzed

KEYWORDS: Recognition, Deep Learning, CNN, Neural Network.

I. INTRODUCTION

Handwritten character recognition is a field of research in artificial intelligence, computer vision, and pattern recognition. A computer performing handwriting recognition is said to be able to acquire and detect characters in paper documents, pictures, touch-screen devices and other sources and convert them into machine-encoded form. Its application is found in optical character recognition and more advanced intelligent character recognition systems. Most of these systems nowadays implement machine learning mechanisms such as neural networks. Machine learning is a branch of artificial intelligence inspired by psychology and biology that deals with learning from a set of data and can be applied to solve wide spectrum of problems. A supervised machine learning model is given instances of data specific to a problem domain and an answer that solves the problem for each instance. When learning is complete, the model is able not only to provide answers to the data it has learned on, but also to yet unseen data with high precision.

Offline Character Recognition: Offline handwritten character recognition refers to the process of recognizing characters in a document that have been scanned from a surface such as a sheet of paper and are stored digitally in grey-scale format. The storage of scanned documents has to be bulky in size and many processing applications as searching for a content, editing, maintenance are either hard or impossible.

Online Character Recognition: The online mode of recognition is mostly used to recognize only handwritten characters. In this the handwriting is captured and stored in digital form via different means. Usually, a special pen is used in conjunction with an electronic surface. As the pen moves across the surface, the two- dimensional coordinates of successive points are represented as a function of time and are stored in order. Recently, due to increased use of handheld devices online handwritten recognition attracted attention of worldwide researchers. This online handwritten recognition aims to provide natural interface to users to type on screen by handwriting on a pad instead of by typing using keyboard. The online handwriting recognition has great potential to improve user and computer communication.

II. RELATED WORK

Understanding the handwritten characters or typed documents is simple to the human beings as we have the ability to learn. The same ability can be induced to the Machines also by the use of Machine Learning and Artificial Intelligence. The field which deals with this problem is called as Optical Character Recognition. It's the area of study among various fields such as recognizing of pattern, also Image vision and also AI. This is the system for changing electronic and image text into the digital character to be read by the machines.

Step1: Image Acquisition
In Image acquisition stage, the input image is provided to the recognition system. The input can be either in an image format such as JPEG, BMT, etc. or scanned image, digital camera or any other suitable digital input device or one can draw on the canvas provided on the user interface.

Step2: Pre-Processing
The second method called as Pre-processing is the entry method for recognition of character and very important in deciding the recognition rate. Preprocessing works to normalize the strokes and also remove variations that can reduce the rate of accuracy

- Noise Removing: To eliminate the unwanted or undesired patterns, a technique called noise removing is used. There's a technique like Uniform and non-uniform filtering that is to be used.

- Binarization: In these all-typed characters are translated into grey-scale picture. Each and every image of character is to be catches vertically after translating the gray scale image into the binary matrix.
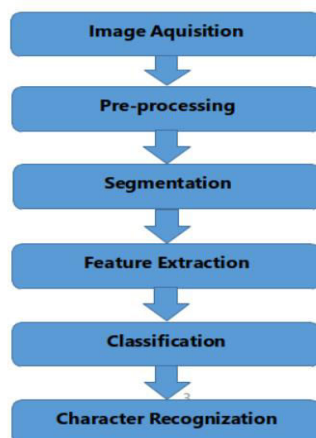

Fig.1 Schematic Diagram of Handwritten Recognition

Step3: Segmentation
Segmentation is used to convert input image consisting of many characters into the individual characters. The techniques used are word, line and character segmentation. It is generally performed by dividing single characters from the word picture. Moreover, a content is processed in a way that is tree like. In initial scenario, row histogram is used to segment the lines. Then after, every level, characters are retrieved by technique called histogram and then finally getting it retrieved.

Step4: Feature Extraction
The aim of feature extraction is to allow the extraction of pattern which is most important for the classification. Some of the Feature extraction techniques like Principle Component Analysis , Scale Invariant Feature Extraction, Linear Discriminant Analysis, Histogram, Chain Code , zoning and Gradient based features can be applied to extract the features of individual characters. All of these features are used to train the given system.

Step5: Classification
The decision making is done in the classification phase. For recognizing the characters, the extracted features are used. Different classifiers like SVM and Neural Networks are used. The classifiers sorts the given input feature with reserved pattern and find the best matching class for input, for which Soft Max Regression is used.

Step6: Character Recognition

The final and last phase of the character recognition is the Post-processing. It is the procedure for correcting the misclassified output by using natural language. It processes output by getting it after the shape have been recognized. If the shape is recognized purely then the accuracy can be improved in accordance with the knowledge of language. Shape recognizers behave differently for different handwriting input. For the few, it results in a individual characters of string while also including a few number of alternates in second case, by including the measure of confidence in every alternative

III. PROPOSED WORK

A. *Challenges in handwritten character recognition*

The Perils of Poor Handwriting - Perhaps the most obvious problem when processing handwritten forms during the data capture process is poor quality or illegible handwriting. We all know the old stereotype about doctors' handwriting, so trying to perform accurate data capture and validation on this type of form-filling may result in little meaningful data being extracted.

Good Handwriting is not all it's cracked up to be either - It is not just poor handwriting that can cause an issue, but also joined / cursive handwritten text is known to cause similar issues. During the data capture and validation stages of any forms processing activity, all required text fields are processed which involves recognising and extracting the written characters. The problem that cursive handwriting can cause is that letters may not be as easily recognised and possibly cause false and incorrect information being processed.

Error Rate - As shown in the literature various algorithms have been designed to solve the problem of handwritten character recognition, but accurate detection is still a challenging issue. Figure6 also depicts the same scenario of Bangla character.

Speed of Writing - Any script written with high speed will have some missing points and gaps between the points. These missing points can be calculated using various techniques which help to interpolate points among fixed number of points. In addition, re-sampling of points is required to keep the points in the list at equal distances, as far as possible. For any pair of points in the list having a distance greater than one, new point between such pairs should be added. Any pair having distance less than one is untouched. The list obtained after the re-sampling of points is pre-processed.

Detection Speed - Advance algorithms and deep networks take time in training so to process multiple images, detection time automatically increases.

Handwriting Size - Size of the input script depends on how user moves the pen on writing surface. Script is not generally entered when the pen is moved along the border of writing surface. Size normalization and centring of stroke is a necessary process that should be performed in order to recognize a script. This can be achieved by comparing input script border frame with assumed fixed size frame and further can be moved along with the assumed centre location

Scalable Detectors - Development of scalable detection algorithms that can detect the expanding data properly is a burning issue of handwritten character recognition.

Slant and Skew of the Handwriting - Slant of scripts is one of the most obvious measurable factors of different handwriting styles is the angle between most higher strokes in a word and the vertical direction referred to as the script slant. Slant correction for a single stroke becomes complex as no headline can be assumed in a stroke. Headline is the defined area where end of each character is considered

Zigzag Shape - Due to speed and shaking hand while writing, the script may contain some irrelevant points and a considerable amount of noise that complicates the work in next stages as illustrated. Applying some smoothing techniques are necessary to eliminate duplicated data points by forcing a minimum distance between consecutive points

Wild Points - Wild points are the occasional spurious points detected by digital devices due to hardware problems. Moreover, wild point's reduction is performed to reduce these points to enhance the performance of the system. Major improvements to digital devices have reduced these kinds of imperfections, but software processing is still required to completely eliminate this problem

Any advancement in software? - With computer systems and applications becoming more intelligent, it makes sense to focus on the automated capture and processing of incoming documents. Over the years much investment has been made in the ability to recognise and understand machine print, so that automated capture and processing can be achieved. However due to the complexities and differences in handwriting, the same level of automation is far more difficult to achieve.

B. *Observations and Comparisons':*

Handwriting character recognition is one of the research fields in computer vision, artificial intelligence, and pattern recognition. A computer application that performs handwriting recognition can be argued to have the ability to acquire and detecting characters in pictures, paper documents, and other sources and convert them into electronic format or machine-encoded form. The system may obtain Handwriting sources from a piece of paper through optical scanning or intelligent word recognition. Also, the system may be designed to detect the movement of the pen tip on the screen. In other words, handwriting recognition may involve a system detecting movements of a pen tip on the screen to get a clue of the characters being written.

In this case, the system sensors movement of pen-tip to detect characters and words. Different methods and techniques are used to ensure that computer systems can read characters from Handwriting images and documents. Among the existing techniques that are used to model, and train Handwriting character recognition include neural network, Hidden Markov Model (HMM), Machine Learning, and Support Vector Machine, to mention a few

Approaches	Description	Advantages	Disadvantages
Hidden Markov Models (HMM)	HMM is a statistical Markov model which is used in a system that is supposed to assume the Markov process	-Strong statistical foundation. -It allows a flexible generalization of sequence profiles	-Have many unstructured parameters. -Algorithms are expensive in terms of computational time and memory -Training requires repeated iterations, and this can be time-consuming.
Machine Learning	Machine learning-powered systems rely on patterns and inference instead of explicit instruction to read text and characters [	-No human intervention needed -Allows continuous improvement.	-Requires massive data to train. -Expensive in terms of time and resources. -High-error susceptibility.
Neural Network	A neural network can be considered as a large parallel computing system comprising of many interconnected nodes	-Can learn complex non-linear input relationships. -Has self-organizing capability. -Ability to work with	Different training may damage the capability of the system Overreliance on hardware.

		incomplete knowledge -Parallel processing capability -Ability to make machine learning	
Support Vector Machines (SVM)	Classifies the data using a hyperplane	Unlike neural networks, SVM approach relies on learning examples and structural behaviour. Has better generalization due to structural risk minimization	It is difficult to select a "good" kernel function Difficult to understand and interpret It is hard to visualize the impact of SVM models.

IV. SIMULATION RESULTS

Classification is the process of assigning the data to their corresponding class with respect to similar groups with the aim of discriminating multiple objects from each other within the image. Its goal is to predict the categories of input image using its features. It is carried out on the basis of stored features in the feature space such as structural features, global features etc. Some classification techniques used in character recognition systems are Template Matching, Statistical Techniques, Structural Techniques, Neural Network, and Support Vector Machine

Table. 2 Comparison of various techniques based on different languages

Character Image	Phases of Character Recognition			Accuracy Claimed
	Pre-Processing	Feature Extraction	Classification	
Handwritten Numerals		Wavelet Transformation	Multilayer cluster nueral network	99.17
Devnagari Character	Median Filter, Binarization	Gradient	Support Vector Machine	86
Arabic	Noise Removal	SIFT	Templete Maching	88
Kannada	Morphological Operation	Normalized Chain Code	K-Nearest Neighbour classifier	95.07
Gujarati	Thining,Binarization	Wavelet	Euclidean distance Metric	92.11
English	Median filetr	Biorthogonal spline wavelet	Multilayer cluster nueral network	98.22

V. CONCLUSION AND FUTURE WORK

From this survey comparative study of various phases in character recognition has been carried out and it was concluded that Recognition of characters and digits is vital in today's digitized world, especially in organizations that deal with Handwriting documents that they need to analyze using computer systems. Systems that are used for classification and recognition of handwriting help organizations and individuals to solve complex tasks

This paper gives overview of various existing approaches for Handwritten Character Recognition. Many researchers have achieved good recognition rate for Indian scripts, but this rate is for an individual characters. Character with different modifiers and specially connected or joint characters are quite complex to identify. So, a lot more research is needed for Indian scripts to be identified accurately with all its complexities.

The work of the current research can be extended for character recognition in other languages. It can be used to convert books, newspapers, handwritten notes, and newspapers into digital text format using machine learning models used by the current research.

REFERENCES

[1]   Megha Agarwal, Shalika, Vinam Tomar, Priyanka Gupta; ”Handwritten Character Recognition using Neural Network and Tensor Flow”;International Journal of Innovative Technology and Exploring Engineering (IJITEE); ISSN: 2278-3075, Volume-8, Issue- 6S4, April 2019

[2]   Salma Shofia Rosyda and Tito Waluyo Purboyo.; “A Review of Various Handwriting Recognition Methods”, IInternational Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 2 (2018)

[3]   Renjini L , Rubeena B; ”A Comprehensive Study On Handwritten Character Recognition System”, IOSR Journal of Computer Engineering , Volume 17, Issue 2, Ver. IV (Mar – Apr. 2015)

[4]   Monica Patel, Shital P. Thakkar; “Handwritten Character Recognition in English: A Survey”, International Journal of Advanced Research in Computer and Communication Engineering,Vol. 4, Issue 2, February 2015

[5]   Mustafa Ali Abuzaraida, Akram M. Zeki; ”Problems of Writing on Digital Surfaces in Online Handwriting Recognition Systems”, 978-1-4799-0136-4/13/$31.00 ©2013 IEEE

[6]   Nikita Mehta, Jyotika Doshi ; ”A Review of Handwritten Character Recognition”, International Journal of Computer Applications (0975 – 8887), Volume 165 – No.4, May 2017

[7]   Sarkhel, Ritesh, et al ; "A multi-objective approach towards cost-effective isolated handwritten Bangla character and digit recognition." Pattern Recognition 58 (2016): 172-189.

[8]   Dinges, Laslo, et al. "Automatic recognition of common Arabic handwritten words based on OCR and N-GRAMS." Image Processing (ICIP), 2017 IEEE International Conference on. IEEE, 2017.

[9]   Acharya, Shailesh, Ashok Kumar Pant, and Prashnna Kumar Gyawali. "Deep learning based large scale handwritten Devanagari character recognition." Software, Knowledge, Information Management and Applications (SKIMA), 2015 9th International Conference on. IEEE, 2015.

[10]   Larasati, R., & KeungLam, H. (2017, November). Handwritten digits recognition using ensemble neural networks and ensemble decision tree. In 2017 International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS) (pp. 99-104). IEEE.

[11]   Ashay Singh, Ankur Singh Bist; ”A Wide Scale Survey on Handwritten Character Recognition using Machine Learning”, International Journal of Computer Sciences and Engineering,Vol.-7, Issue-6, June 2019

[12]   Sara Aqab, Muhammad Usman Tariq ; ”Handwriting Recognition using Artificial Intelligence Neural Network and Image Processing”, IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 7, 2020

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi® crossref

ISSN
INTERNATIONAL STANDARD SERIAL NUMBER
INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Scan to save the contact details