International Journal of Advanced Technology in Engineering and Science
Vol. No. 10, Issue No. 05, May 2022
www.ijates.com
ijates
ISSN 2348 - 7550

Character Segmentation of Handwritten Text of Support Vector Machine(SVM)Using Machine Learning

B.Pandu RangaRaju[1] ,P.Ganesh Reddy[2] ,S.Sushanth Kumar Raju[3]
A.Pranathi[4] ,V.Siva Jyothi[5]

[12345] *Assistant Professor, Department of AI&DS, AITS-Rajampet, AP, INDIA*

ABSTRACT

The importance of handwritten documents in everyday transaction technologies like optical character recognition will be a valuable addition to the new set of technologies. This technology allows the translation of different documents and images into editable, analyzable, and searchable information. Researchers have successfully integrated technologies like machine learning and artificial intelligence to automatically analyze printed or handwritten documents for converting them into electronic formats. At the time of recognizing a text, one needs to process the input image, extraction functionality, and classification schemes. This is the training stage of the system for acknowledging specific text. In this phase, the system is trained to find out similarities and differences between handwritten sample documents. Technology makes use of images of hand transcription and then transforms these images into a digital copy. The primary aim of this research report is to understand and explain the procedure for the development of character recognition systems.

Index Terms: *Character Recognition, Image Processing, Recurrent Neural Network (RNN), CNN Character Segmentation, Machine Learning*

I. INTRODUCTION

Handwriting recognition technology allows interpreting handwritten documents from sources like documents, screens of other devices, or other digital formats [1]. A handwriting recognition system can handle formatting and also performs character segmentation. It is also capable of recognizing the most plausible words. The translation of handwritten characters into digital formats is becoming more and more popular. Over time, the hard copies may get spoiled, but anything that has been stored in the system will remain accessible [2]. However digital files can also get lost if deleted or gets corrupted. Storing handwritten documents in digital format is gaining immense popularity.Optical character recognition technology is a tool that can convert text into a machine-encoded form. At present, OCR technology is being used for the digitization of handwritten scripts [3]. It also facilitates the conversion of typewritten text into digital format. This has also simplified the retrieval of information from huge piles of documents. Organizations can now easily access historic data from digital files [4]. This technology is also being used in domains like academics, law, etc.

The performance of an OCR system primarily depends on its level of extraction functionality and categorization of patterns. Handwritten OCR is gaining huge popularity as a subfield of this technology. The offline system comprises of static mechanism form using scanned images in an online system [5]. The information that is put into the system is not static and can vary as per the motion of the pen tip with specific projection angle, velocity, position, andlocus point. Online systems have a more complicated layout with advanced features. It is capable of dissolving overlapping issues of input data available in offline systems.After the texted document is fed into the system and an HD image is obtained, it starts by segmenting every character from the image for identifying the letters. In the next stage, the letters are recognized by the system, and the image is used for the detection of words. Machine Learning Algorithms are applied to go through these processes that are based on the data obtained in the training fees [6]. The machine sends the expected output in a word file format. The system can be easily trained for processing large sets of data that comprise different shapes and styles..

1.Research Objective

Setting up various kinds of forms in the system by creating software that will allow it is the central problem. The capability to perceive English handwritten numbers and alphabets concerning the form which is set up should be the main feature of the solution. It should also be able to save those as information. To highlight weaknesses in the research process and eliminate potential issues affecting further research.

1.1. Summarizing the existing research work in different languages with the help of a handwritten character recognition system.

1.2. Identification of new technologies for character recognition and segmentation.

2. Research Motivation

It is not easy to read handwritten papers. Handwriting recognition technology is one of the most challenging research works in the field of pattern recognition and image processing. At present, it is one of the largest contributors towards the advancement of automated systems and helps in enhancing the interface between machine and man for different applications. Multiple research work has been conducted to focus on new methods and techniques for reducing the pre-processing duration to obtain higher accuracy [1]. Technologies like handwriting recognition are quite complex as it involves multiple phases of processing and segmentation. Besides, every people have a different style of writing. The system requires processing a large number of characters like small letters, capital letters, special symbols, and digits [8]. Incorporation of a large data set is necessary for training and OCR system. It is also known as an offline character recognition tool that can scan and recognize still images of handwritten characters

II. PROPOSED METHODOLOGY

In Proposed Methodology we are using the methodologies for handwritten text recognition systems involve several challenges for offline character recognition. Researchers are working towards exploring new techniques in this domain. The process of offline handwriting recognition comprises steps like text line segmentation, pre-processing, feature extraction, classification, and unit segmentation.The character recognition mechanism of an

image is also influenced by the composition of the image, preprocessing methodologies, selection techniques, and recognition model. This research report highlightsmethods of character recognition using different algorithms. It also promotes the use of deep learning forbetter recognition inaccuracy.

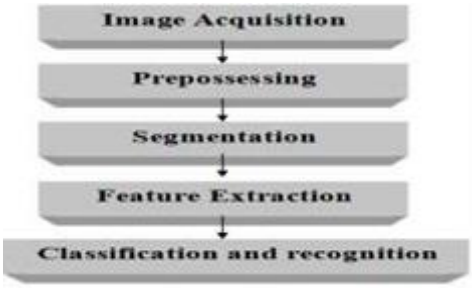

Fig. 1. Proposed Flow of Process [6]

III. IMPLEMENTATION

For regression analysis and classification, data analysis is done by supervised learning models with features for associated learning algorithms. These networks are called Support Vector Machines (SVM). An SVM training algorithm creates a model for assigning new examples to categories when provided with a set of training examples where each is marked as affiliated to any one of the two categories [12]. This makes it a non-probability binary linear classifier. Points in space can be seen as examples of SVM model representation. These are mapped so that a clear and wide gap can be used to divide examples from different categories. In that same space, new examples are mapped a gain and the category based on the side of the gap is predicted.

As can be seen in the diagram, sample collection is the very first step of the handwriting recognition mechanism. For offline handwriting text recognition tools, the writing style of sample text does not always follow writing rules as every person has their style and characteristic of writing [1]. To make the process a little more usable the following step involves a preprocessing where denoising, binarization, and normalization technique is necessary.

Fig. 2. Offline Handwriting Text Recognition Framework [1]

IV. LITERATURE GAP

The research work that was extracted for further analysis in this particular review was from various publications on the six most commonly spoken languages. It was noticed that there was a disparity between the techniques being used and the respective script as some techniques proved to be more accurate than the others [1]. For example, a technique called multilayer perceptron classifier was accurate f or Bengali as well as Devanagari numerals. The result was average f or the other languages in general. The type of dataset that is used for identification and the characteristics of it may be responsible for these contrasting results. The studies that are published in this particular research area suggest that there should be one technique for all languages

There are quite a few datasets that are available publicly that are well-aligned. However, they do not correspond well with the scenarios that are seen in real life. It is mainly concerning the thickness of characters, the style of writing and the distortion in strokes etcetera. Convolutional Neural Networks is the technique that has been most commonly used for recognizing both handwritten materials as well as characters printed in machines. The superior detection of images by CNN plays an important role for it is universally accepted and used if there was an object that needed to be identified from an image, CNN was routinely used for it. ResNet and AlexNet are some of the important architectures that are used by CNN for visually recognizing images.

V. RESULTS AND DISCUSSION

To recognize various kinds of handwritten text, a system can be developed with machine learning integrated with image processing knowledge [5]. The automated process of manual entry can get available with this implementation and this can be further enhanced with a huge amount of data sets and suitable training.The recognition of handwritten characters faced quite a few challenges. Obtaining solutions for the recognition of handwritten characters have several limitations.

1.Speed of detection- the rate at which multiple images are processed is increased considerably. I t is due to the time taken by advanced algorithms.
2.Rate of Error- even though several algorithms mostly can be used for the identification of handwritten texts; the accuracy of these algorithms is still questionable [9].
3.Detectors that are scalable- The data are always on the rise. Developing an algorithm that is scalable and detectable is important.

VI. CONCLUSION

The method of recognizing handwritten characters from an image that is scanned is known as deep learning. It can be used to convert any handwritten information into an electronic form. This paper concludes that a sanguine tool or method needs to be developed that can help in this conversion of characters that are handwritten to a digital format. Deep learning may be immensely useful in this endeavor as image processing takes place excessively smoothly.

VII. REFERENCES

[1] P. Sunitha, S. Hegde, A. V. Shreyashree, B. Prasad and Chinmayi, "Machine Learning Approach for Translating Handwritten Document to Digital Form," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 2, pp. 11-16, 2021.

[2] S. Preetha, I. Afrid, K. Hebbar and S. K. Nishchay, "Machine Learning for Handwriting Recognition," International Journal of Computer (IJC), vol. 38, no. 1, pp. 93-101, 2020.

[3] S. S. Mor, S. Solanki, S. Gupta, S. Dhingra, M. Jain and R. Saxena, "Handwritten Text Recognition: with Deep Learning and Android," International Journal of Engineering and Advanced Technology (IJEAT), vol. 8, no. 3S, 2019.

[4] M. Kumar, S. Jindal, M. Jindal and G. Lehal, "Improved recognition results of medieval handwritten Gurmukhi manuscripts using boosting and bagging methodologies," Neural Process. Lett., vol. 50, p. 43–56, 2018.

[5] A. Singh and A. S. Bist, "A Wide Scale Survey on Handwritten Character Recognition using Machine Learning," INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING , vol. 7, no. 6, pp. 124-134, 2019.

[6] A. Shaikh, S. Arsalan, K. M. Altaf and S. Athar, "Handwritten Text Recognition of Document Form Using Machine Learning," International Research Journal of Engineering and Technology(IRJET), vol. 5, no. 2, pp. 2117-2120, 2018.

[7] S. Aqab and M. U. Tariq, "Handwriting Recognition using Artificial Intelligence Neural Network and Image Processing," (IJACSA) International Journal of Advanced Computer Scienceand Applications, vol. 11, no. 7, pp. 137-146, 2020.

[8] B. Rajyagor and R. Rakhlia, "Handwritten Character Recognition using Deep Learning," International Journal of Recent Technologyand Engineering (IJRTE), vol. 8, no. 6, pp. 5815-5819, 2020.

[9] S. R. Zanwar, U. B. Shinde, A. S. Narote and S. P. Narote, "A Comprehensive Survey On Soft Computing Based Optical Character Recognition Techniques," INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH , vol. 8, no. 12, pp. 978- 987, 2019.

[10] A. A. Ahmed Ali and M. Suresha, "Survey on Segmentation and Recognition of Handwritten Arabic Script," SN Computer Science, vol. 1, no. 192, pp. 1-31, 2020.

[11] Z. Roa, C. Zeng, M. Wu, Z. Wang, N. Zhao, M. Liu and X. Wan, "Research on a handwritten character recognition algorithm based on an extended nonlinear kernel residual network," KSII Transactions on Internet and Information Systems, vol. 12, no. 1, p. 413–435, 2018.

[12] M. Sharma and V. Dhaka, "Segmentation of handwritten words using structured support vector machine," Pattern Anal Applic, vol. 23, p. 1355–1367, 2020.

[13] R. Khan, A. Meyer, H. Konik and S. Bouakaz, "Saliency- based framework for facial expression recognition," Frontiers Comput. Sci., vol. 13, no. 1, p. 183–198, 2019.

[14] J. MEMON, M. SAMI, R. A. KHAN and M. UDDIN, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," IEEE Access, vol. 8, pp. 142642-142668, 2020.

[15] N. Shorim, T. Ghanim and A. AbdelRaouf, ". Implementing Arabic Handwritten Recognition Approach using Cloud Computing and Google APIs on a mobile application," in n Proceedings of the 14th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 2019.

[16] J. Pratikshaba, "A Review Character Recognition and Digitization of Sanskrit Characters," Journal of the Gujarat Research Society, vol. 21, no. 13, p. 551–560, 2019.