Text Recognition Techniques

S J Fiona G Sathiaraj[1], Grishma S. Pingale[2], Souvik Majumdar[3], Saniya J. Shaikh[4], Dr. Bhushan S. Thakare[5]

[1, 2, 3, 4, 5] Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India

[1]sjfiona.gsathiaraj@gmail.com [2]grishmaspingale97@gmail.com [3]majumdar.souvik762@gmail.com [4]saniya.shaikhactive@gmail.com
[5]bhushan.thakare86@gmail.com

***Abstract*— Text recognition is an analytical field of emerging research and development, and is indeed instrumental in converting the text obtained from the underlying documents, images or other input sources into digital text. Primarily, there are two methods of text recognition, viz. online recognition method and offline recognition method depending on the sources of input. The two major problems that often arise or occur in the process of text recognition include: text location identification within the document or image and text recognition and identification. Moreover, the complexity increases while dealing with handwritten text. A number of techniques to recognize type-written as well as hand-written text have been developed till date. This paper presents the state of the art and a model is proposed that can yield better results.**

Keywords—neural network; text recognition; machine learning; deep learning; OCR;

1. Introduction

Text recognition is extensively being applied and used in various fields. This process involves the extraction and also the conversion of any text, both typewritten text and handwritten text, from images or documents into machine editable forms. This makes it possible to store the text in a resultant format that allows effective data retrieval and update operations efficiently. OCR systems have two prime benefits or advantages, viz. the capacity/ability to upsurge throughput by decreasing the total amount of staff involvement in the process and also the ability to store the text information obtained efficiently [1].

Two aspects that must be well-thought-out in this process of text recognition are speed and accuracy. Speed is measured depending on the number of characters recognized by the technique per second whereas accuracy measures the total number of actual characters or words that are correctly recognized out of the total of these characters or words in the document or image. The recognition systems are expected to provide high speed and high accuracy with low error rates.

The development of text recognition methods can be traced several years ago i.e. the early 1900's. An idea was proposed to develop devices that would assist the visually impaired in reading. A few years later an assistive system was developed that was able to interpret Morse code and read out text loudly. In the later years of the 1980's, various companies developed software that was able to read any kind of text document. In recent years, however, text recognition software is made available for free by different organizations. The text that is to be recognized can be typewritten or handwritten [16]. While recognizing typewritten text is relatively easy, handwriting recognition is still a challenging task. Moreover, old documents and ancient manuscripts are torn and some portion of text is eventually lost. This makes it even more difficult in extracting important information from the underlying document or manuscript. The existing recognition techniques, therefore, do not suffice or fall short when the text is handwritten or when the text involves a variety of styles, fonts, etc. [2] In addition to this complexity, there is also a crucial requirement to produce most accurate results irrespective of the complications involved.

I. CLASSIFICATION

Text recognition techniques can be classified based on the input source of the text data. The results thus obtained differ according to the technique that is used. They can be categorized into the following divisions:

A. ONLINE RECOGNITION SYSTEMS

In online recognition systems [13] [14] [15], the input data is made available from the strokes of the pen or stylus while the user is writing. This technique requires a digital tablet and special hardware to record the stroke information [6]. By using online recognition systems, the spatial and also the temporal information about the input text is recorded. It includes information about the position, velocity and direction of each stroke [10]. The data is stored as a set of points (x, y) which is a function of time.

Advantages

- Ability to recognize real time data efficiently.
- Less pre-processing required
- Segmentation is easy
- Low data requirement in the order of few hundreds of bytes

- Higher recognition accuracy

Limitations

- Cannot recognize text from physical documents.

B. OFFLINE RECOGNITION SYSTEMS

In offline recognition systems, the input is obtained from a physical document or image that is scanned to extract the text in it. The input undergoes various pre-processing stages. The document or image is scanned and converted into machine editable or digital form text.

These systems, however, do not have access to the temporal information about the text to be recognized. Offline recognition systems require more data as compared to the online systems in the order of few hundreds of kilobytes. Recent trends in offline text recognition involve the usage of neural networks and artificial intelligence techniques like machine learning.

Advantages

- Can recognize text from physical documents like historical manuscripts, banking checks, invoices, bills, etc.

Limitations

- Data requirement is high
- Segmentation is difficult
- Cannot recognize real time data
- Lower accuracy as compared to online recognition systems.

Table 1 summarizes the essential features of both the recognition systems.

II. METHODS

A. USING DIAGONAL-BASED FEATURE EXTRACTION

This method is based on the method diagonal-based feature extraction. This method is mainly used to recognize and identify the input handwritten text from images.

It uses the pixel information obtained from the two diagonals. The neural network recognizes handwritten numerical digits, English characters and Tamil vowels.

This method extracts relevant features from the diagonals and the characters and words based on the extracted features are classified as well as recognized.

OBTAINED RESULTS

The system shows decent results while recognizing digits. The network can also efficiently and accurately distinguish between similarly structured digits like '3', '6' and '8' with less input information. However, the proposed diagonal-based feature extraction technique or method does not classify Tamil vowels efficiently. The classification of the characters improves as more number of pixels are fed to the neural network of the system [3].

B. USING BACKPROPAGATION NEURAL NETWORK

This recognition system uses back propagation neural network to recognize input character set. The proposed system is mainly divided into two different stages: training and recognition stage. The initial pre-processing stage performs digitization of the image, noise removal, and line segmentation followed by character extraction. Once the characters are extracted, the character matrix obtained is normalized into a 12 X 8 matrix. The normalized matrix thus obtained is then converted into a feature vector of size 96 X 1.

NETWORK SPECIFICATION

The input layer consists of 96 input neurons, and 62 output neurons. Each output neuron in the neural network corresponds to one character that the system is expected to recognize. The below figure shows the 3 layers of the network specified.

OBTAINED RESULTS

The system recognizes numeric digits efficiently with a recognition rate of 99%. The recognition rate is 97% for capital alphabets and 96% for small alphabets. However, for inter-class characters, the recognition rate drops down to 93% [19].

C. USING DEEP LEARNING

This method is used to recognize text from multimedia documents like videos and also different images. It employs a deep learning method to extract and recognize text from given input.

OBTAINED RESULTS

The system uses both Convolutional Neural Network (CNN) and Back Propagation Neural Network (BPNN). If only CNN is used, the accuracy lies in the range of 90 to 100%. However, as the system uses both CNN and Feed Forward Network, the accuracy improves [20].

D. USING SUPPORT VECTOR MACHINES

The system puts forth a technique for the pre-processing stage and normalization of the dataset and also offers optical character recognition (OCR) based on the SVM classifier. It concretely emphases on the recognition of different styles of handwriting used for the same set of characters. This technique is widely used for forensic resolves and author identification of a document [4].

OBTAINED RESULTS:

The table below shows the obtained accuracy for each model that is trained. The best accuracy obtained is with the use of the Radial Basis Function (RBF) type of kernel. The application of the max-pooling method or technique on the dataset, gives worse results than that obtained by using all the 625 features with accuracy as 92.86% [4]. As a result, the max-pooling technique does not contribute to the improvement of accuracy.

E. KNOWLEDGE TRANSFER USING NEURAL NETWORK BASED APPROACH

This system/work, shows how structural depiction learned from a common writer-independent handwriting recognition model can be adapted to individual authors. Use of Convolutional Neural Network gives great performance in learning image-based depiction that are used for the purpose of classification [5].

RESULTS:

The model is trained on both the IAM and Washington datasets. This model performs worse than a model that has been trained using the IAM dataset too.

F. USING ONE NEURAL NETWORK FOR BOTH TEXT DETECTION RECOGNITION

A STN_OCR single semi-supervised deep neural network a text recognition technique mainly consists of two main components:

1. Spatial Recognition network.
2. Text Recognition network

The text part of the images are detected by the Spatial Recognition network and further using the Text Recognition network, hence by using both techniques the text area from image is extracted and recognized. This detected text region is analyzed and identified for text content. This technique is an end to end text recognition technique which is capable of detecting variety of texts arranged in different lines in images [11].

G. USING CONVOLUTIONAL NEURAL NETWORKS

The technique focuses on the problem of complete end to end text recognition in natural/usual images, by taking a different method or way and combining the power of representation of the large, multilayered neural network together with the development of unsupervised feature learning, that provides a common context to train accurate text detectors and also the character recognition modules.

For low-level data representation, only unsupervised feature learning algorithm is sufficient and hence used as it can automatically extract features from the given data. Further integrating simple random existing methods into a complete end-to-end text recognition system helps achieve highest level of performance on standard benchmarks [12].

III. PROPOSED WORK

After a detailed study of the existing methods used for text recognition and Optical Character Recognition (OCR), a model is proposed that can generate better ad efficient results. By taking into consideration the features of the existing methods, deep learning has proved to be a technique that generates optimal results.

The proposed model transliterates the handwritten text available in documents into digital text. This can be achieved with the help of neural networks using the deep learning approach [18] [19]. The neural network contains the following layers:

- Convolution Neural Network (CNN)
- Recurrent Neural Network (RNN)
- Connectionist Temporal Classification (CTC)

Given below are the steps that are used in the proposed model to recognize the handwritten text:

1. The image is pre-processed
2. The input is the fed to the CNN to get a sequence of features or a feature vector [7] [8].
3. The obtained feature sequence will be given to the recurrent neural network as an input. This will in turn generate the character probability matrix.
4. The character probability matrix that was generated in the previous step will be given as an input to the CTC layer.
5. Based on the current step, the CTC layer can either recognize the text or it may calculate the loss value.

IV. PREDICTED RESULTS

Based on the proposed model, the input to the system could be either an image or any document that contains handwritten text. The model is expected to outperform the existing models in terms of accuracy and efficiency. It will give better results, as the RNN is used to remember the character sequence of the past that makes it easier to recognize the characters in the latter part of the text.

SR.NO.	REFERENCE	TECHNIQUE	CATEGORY	CHARACTER SET	ACCURACY	HIGHLIGHTS	LIMITATIONS
1.	Sergey Tulyakov et.al [5], 2018	Deep Learning (Transfer Learning)	Online	Handwriting (Cursive) Alphabetic Symbols	-	Capable to study and get familiarized with a target dataset with restricted training data.	The model does not always produce proportional performance expansion.
2.	Wei et. al.[2], 2018	Deep Learning (Transfer Learning)	Offline	English Alphabets	Poor Quality Image - 78% Good Quality Image - 90.6%	Recognizes text from good quality text images efficiently and with good accuracy.	Accuracy of detecting text from low quality images is less.
3.	Hui Li[9],2018	Convolutional neural network, RNN with CTC	Offline	License plate images	Average detection ratio-99.73%	License plates can be documented simultaneously with good efficiency and accuracy.	Only an image of 700 pixels in shorter side can be used to test.
4.	Martin et.al[4],2017	Support Vector Machines	Offline	Comenia script of the Czech Republic	92.86 %	Combining more models for recognition can give better results.	Better accuracy than other models
5.	Wibowo et. al. [17], 2017	Deep learning	Offline	Javanese Characters	Accuracy - 94.57%	Good accuracy for characters with simple structures	Complex structure characters give relatively lower accuracy rate.

REFERENCES

[1] Tan Chiang Wei, U. U. Sheikh and Ab Al-Hadi Ab Rahman, “Improved Optical Character Recognition with Deep Neural Network” in 14th International Colloquium on Signal Processing & its Applications, 2018.

[2] Renuka Kajale, Soubhik Das, Paritosh Medhekar, “Supervised machine learning in intelligent character recognition of handwritten and printed nameplate”, IEEE 2017.

[3] K. Vijayalakshmi, S. Aparna, Gayatri Gopal and W. Jino Hans, “Handwritten Character Recognition Using Diagonal-Based Feature Extraction” in the IEEE WiSPNET, 2017.

[4] Martin Rajnoha, Radim Burgetand Malay Kishore Dutta, “Offline Handwritten Text Recognition Using Support Vector Machines”, 4th International Conference on Signal Processing and Integrated Networks (SPIN), 2017.

[5] Rathin Radhakrishnan Nair, Nishant Sankaran, Bharagava Urala Kota, Sergey Tulyakov, Srirangaraj Setlur, Venu Govindaraju, "Knowledge transfer using Neural network based approach for handwritten text recognition" , 13th IAPR International Workshop on Document Analysis Systems, 2018.

[6] Santosh K.C., Cholwich Nattee Institute National de Rechercheen Informatique et en Automatique (INRIA)," A comprehensive survey on on-line handwriting recognition technology and its real application to the Nepalese natural handwriting"

[7] J.Pradeep, E. Srinivasan, S.Himavati,"Neural Network based HCR system without feature extraction".

[8] S.V.Rajashekararadhya and P.Vanajranjan, "Efficient zone based feature extraction algorithm for handwritten numeral recognition of four popular south -Indian scripts", Journal of Theoretical and applied Information Technology, JATIT.

[9] Hui Li, Peng Wang, and Chunhua Shen "Toward End-to-End Car License Plate Detection and Recognition with Deep Neural Networks", 2018.

[10] Shoumorup Mukhopadhyay, 'Improving The Efficiency Of Tesseract OCR Through Super Resolution' (Iasc-Insa-Nasi, 2017)

[11] Christian Bartz, Haojin Yang, Christoph Meinel, "Stn-Ocr:A Single Neural Network For Text Detection And Text Recognition ",2017.

[12] Tao Wang, David J. Wu, Adam Coates, Andrew Y. Ng, "End-to-End Text Recognition with Convolutional Neural Networks".

[13] Salem Meftah Jebriel, Mustafa Ali Abuzaraida "The Detection of the Suitable Reduction Value of Douglas-Peucker Algorithm in Online Handwritten Recognition Systems" IEEE International Conference on Service Operations And Logistics, And Informatics (SOLI), 2015

[14] Douglas David and Peucker Thomas, "Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature," Cartographica: The International Journal for Geographic Information and Geovisualization, vol. 10.

[15] F. Biadsy, J. El-Sana and N. Habash, "Online Arabic Handwriting Recognition using Hidden Markov Models," In Proceeding of the 10th International Workshop on Frontiers in Handwriting Recognition, La Baule, France.

[16] R. Lienhart and A. Wernicke, "Localizing and segmenting text in images and videos," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 4, pp. 256–268, Apr. 2002.

[17] Mohammad Agung Wibowo, Muhamad Soleh, Winangsari Pradani, Achmad Nizar Hidayanto, "Handwritten Javanese Character Recognition using Descriminative Deep Learning Technique, 2nd International Conferences on Information Technology", Information Systems and Electrical Engineering (ICITISEE), 2017.M. J. Traxler and M. A. Gernsbacher, "Handbook of Psycholinguistics Amsterdam", Elsevier, 2006.

[18] Shyla Afroge, Boshir Ahmed, Firoz Mahmud, "Optical Character Recognition using Back Propagation Neural Network", 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), 2016.

[19] Usha Yadav, Satya Verma, Deepak Kumar Xaxa, Chandrakant Mahobiya, "A Deep Learning Based Character Recognition System from Multimedia Document", International Conference on Innovations in Power and Advanced Computing Technologies [i-PACT2017].

[20] Yi-Chao Wu , Fei Yin, Zhuo Chen, Cheng-Lin Liu, Handwritten Chinese Text Recognition Using Separable Multi-Dimensional Recurrent Neural Network, 14th AIPR International Conference, 2017.