HANDWRITTEN CHARACTER RECOGNITION USING NEURAL NETWORK AND FUZZY LOGIC

[1]Jay Fantin V P, [2]M.Vinothini, [2]D.Yuventha, [2]M.Shofia Nashrin, [2]R.S.Subhiksha

[1]Assistant Professor, [2]Bachelor of Engineering Student

Department of Electronics and Communication Engineering,

SSM Institute of Engineering and Technology,

Tamilnadu, India.

Abstract: Image based content extraction is one of the quickest developing examination territories in the field of sight and sound innovation. The extraction of content from intricate or increasingly vivid pictures was a difficult issue. Content information present in the pictures contains valuable data for routine clarification, ordering and organizing of pictures. Extraction of this data includes identification, confinement, extraction, improvement and acknowledgment of the content from the given picture. We have recommended that we convert the content from picture utilizing profound neural network. In this paper, an optical character acknowledgement framework dependent on Deep Neural Network (DNN) is used. The ANN is prepared utilizing the Back Propagation calculation. In the proposed framework, each composed English letter is represented by binary numbers that are utilized as contribution to a basic element extraction framework whose yields not resisting the information, are fed to a DNN. A while later, the feed forward algorithm gives knowledge about the activities of a neural network followed by the DNN algorithm which Bargains, Trains, Calculates error and modifies weights.

Keywords: Optical Character Recognition, Binarization, Thinning, Skewing.

I. INTRODUCTION

The goal of this project is to convert printed textual characters or handwritten characters recorded offline using both scanning system and cameras into a device usable textual content through simulating a Neural Network simply, so it would decrease the way of accumulating and storing statistics via human employees. Another motive is to provide an trade higher and quicker algorithm with better accuracy to recognize the characters. In this context we pick synthetic neural network and make it plenty greater tolerant to anomalies in the recorded picture or facts. Common optical person recognition duties include figuring out easy side detection and matching them with predefined patterns. In this research, characters are identified although noise along with inclination and skewness offers with the useful resource of education, the network to search for discrepancies in records and relate them with the use of vocabulary grammar and common place recurrences that may stand up after a person. Images are also masked in more than one method and processed in our own opinions to increase the self-perception stage of prediction.

II. LITERATURE SURVEY INTRODUCTION

2.1 Arabic Scripter Character Recognition

In 2002 Majid M. Altuwaijri and Magdy A. Bayoumi They develop system to recognize Arabic text using neural network used set of moment invariants descriptors(under shift, scaling and rotation) and artificial neural network(ANN) used for classification. The study has shown 90% of the high accuracy rate.

In 2015 Ashraf Abdel Raouf, Colin A. Higgins, Tony Pridmore and Mah-moud I. Khalil Haar studied approach for recognizing Arabic characters using Haar Cascade Classifier (HCC). These classifiers were trained and tested on some 2,000 images. To extract feature Haar- like feature extraction used and boosting of a classifier cascade. The system was tested with real text images and produces 87% accuracy rate for Arabic character recognition. In 2017 N. Lamghari, M. E. H. Charaf and S. Raghay. On this research the data are divided into three parts. From 34,000 characters 70% are used for training, 15% for testing phase and 15% for validation. To extract feature hybrid feature extraction used (pixel density, resize, freeman code, structural features, invariant) for recognition used feed forward-back propagation neural network. The system has achieved 98.27% high recognition rate.

In 2018 Noor A. Jebrila, Hussein R. Al-Zoubib and Qasem Abu Al-Haijac in addition to the preprocessing step includes in particular three levels. In the primary section, they employed word segmentation to extract characters. In the second section, Histograms of Oriented Gradient (HOG) are used for feature extraction. The very last phase employed Support Vectored Machine (SVM) for classifying characters. They have carried out the proposed method for the recognition of the Jordanian metropolis, city and village names as a case examine, similarly to many other phrases that offers the characters shapes that aren't included with Jordan cities. The set has cautiously been selected to include each Arabic character in its all forms. To the conclusion, they have got constructed their own dataset inclusive of greater than 43.000 handwritten Arabic phrases (30000 used for training and 13000 used for testing stage). Recognition results show 99% rate of accuracy.

2.2 Devanagari Scripter Character Recognition

In 2011 Gyanendra K. Verma, Shitala Prasad, and Piyush Kumar curvelet present in approach for Hindi handwritten character recognition using curvelet transformer. The study used Dataset that contain 200 images of character. Feature extract using curvelet transform and for recognition k-nearest neighbor the experiment result show more than 90% accuracy. In 2013 Divakar Yadav, Sonia Sanchez-Cuadrado and Jorge Morato develop optical character recognition system using neural network for Hindi characters and trained with 1000 dataset. Feature extraction technique is Histogram of projection based on mean distance, on pixel values and vertical zero crossing. Then classify using back-propagation neural network with two hidden layers. Experimental results show 98.5% correct recognition. In 2015 Akanksha Gaur and SunitaYadav made this feature extraction using k-means clustering and classified used support vector machine using linear kernel and Euclidean distance. The evaluation shoe the SVM has better results using lines kernel giving 95.86%results.

In 2018 Nikita Singh presents system with the title "An Efficient Approach for Handwritten Devanagari Character Recognition Based on Artificial Neural Network" for the recognition of Hindi characters. For Feature extraction they used histogram oriented gradients (HOG) and recognition used artificial neural network (ANN) classifier. The system gets 97.06% high accuracy.

2.3 English Scripter Character Recognition

In 2004 N. M. Noor, M. Razaz and P. Manley Cooke proposed system using global geometrics feature extraction and geometric density classifier for feature extraction, then neural fuzzy logic used for classification. Evaluation of the system has achieved for Geometric Density 77.89% and Geometric Feature 76.44% accuracy rate.In 2010 Dewi Nasien, Habibollah Haron and Siti Sophiayati Yuhaniz. This study takes three datasets from NIST database considered lowercase letters 1.89,411, uppercase letters 2,17,812 and combination of lowercase and uppercase letters 4,07,223 samples are used. Those samples are divided into 80% for training and 20% for testing. For feature extraction used Freeman Chain Code (FCC). Support Vector System (SVM) is selected for recognition step. The method recognized for the first dataset 86% accuracy, second dataset 88% of accuracy and third dataset 73%of accuracy achieved. In 2011 Vijay Patil and Sanjay Shimpi develop systems that recognize handwritten English characters using neural network, for feature extraction system they used Character Matrix and for recognition back propagation neural network is used. The result indicate that hand back propagation network provide more than 70% accuracy rate.

In 2015 M.S. Sonawane and Dr. C. A. Dhawale this study compare and evaluate two classifier which is artificial neural network and nearest neighbor. Used grid method to extract feature and the result show nearest neighbor achieve 61.53% accuracy when neural network gives 57.69%. Math lab tool was used for features extracted and recognition. The evaluation outcome suggests Nearest Neighbor is a better recognizer comparing with artificial neural network when implemented to English character.

III. CHARACTER RECOGNITION USING ANN

Optical character recognition generally called OCR is a manner of converting the photograph acquired by using scanning a text or a report into gadget editable layout. Computer device equipped with such an OCR machine can improve the rate of enter operation and decrease a few viable human errors. Recognition of revealed characters is itself a difficult problem on account that there may be a version of equal character due to alternate of fonts or introduction of varies types of noises. Difference in fonts and sizes makes popularity task hard if pre-processing feature extraction and recognition not study. There can be noise pixels that are introduced because of scanning of the photographs. Besides identical fonts and length may have ambitious face individual in addition to normal one. Thus, width of the stroke is likewise a thing that affects popularity. Therefore an awesome person recognition technique need to do away with the noise after reading binary picture statistics, clean the photographs for better recognition, extract functions correctly, tech the gadgets and classify styles.

Lot of human beings now a days are seeking to write their very own OCR(Optical Character Recognition)system are to improve the high-quality of an existing one. This paper indicates how the usage of artificial neural community simplifies improvement of an optical man or women recognition software at the same time as reaching maximum satisfactory of popularity and true performance. OCR device is a complex venture and requires plenty of attempts. Such systems usually are virtually complicated and may hide plenty of common sense. The use of synthetic neural network in OCR programs and improve nice of recognition while attaining top performance. There are basic techniques used for OCR: Matrix matching and Function extraction are the two ways to approach characters. Matrix matching is a less difficult and extra common. Matrix matching compares what the OCR scanner sees as a man or woman with a library of individual matrices or templates. The photograph matches such a prescribed matrix of dots inside a given level of similarity. The computer labels that photo because of the corresponding ASCII characters.

IV. STEPS INVOLVED IN CHARACTER RECOGNITION

4.1 Image Acquisition

To start with an Optical Character Recognition, picture can be captured by the way of virtual camera additionally but after seeing the venture been confronted in previous work higher to use scanner therefore do not forget first need placing collectively an amazing optical scanner. With the help of this scanner, a photograph of original file or report id captured. It is commanding to pick out scanner with an amazing sensing g tool and transport mechanism.

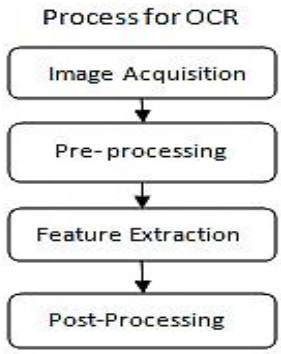

Fig.1. OCR Process Flow

4.2 Pre-Processing

Pre-Processing is performing unique operation at a scanned or entered photo. It facilitates to put off the noise from photograph make individual clear ant it basically complement the photo rendering appropriate for segmentation. Pre Processing has numerous process such as converting color image to gray scale, binarization, thinning, skewing and normalization.

4.3 Feature Extraction

This is one of the riskiest components in an OCR development. The principle role is to extract vital pattern from traits. The decided features are expected to contain pattern that differentiate one individual from different and applicable facts from the input records, in order that the type may be accomplished via the use of those pattern extraction from man or woman this instead of the whole authentic statistics.

4.4 Post-Processing

In this final manner, process like grouping, blunder detection and correction take vicinity. During grouping, symbols inside the text are associated with strings. However, it is no longer feasible to reap a 100% correct identification of characters, only a few of the errors can be detected and deleted as in step with the content.

V. PROPOSED SYSTEM

The work space window fig.2 has the space allocated for input image display and processed image display. It also contains the conversion steps as icons. The loading input image to the program is done by clicking the input icon. Then the conversion of the color image to binary image is done by the IM2BI which is shown in the fig.3. The optical recognition of the image is done by the OCR. The converted text from the processed image is displayed in a notepad file in the monitor which is shown in the fig.4. The fig.5 shows the input image generated from typing in computer and the binary image of the same. The fig. 6 shows the text recognized by the system of the fig.5 input image.

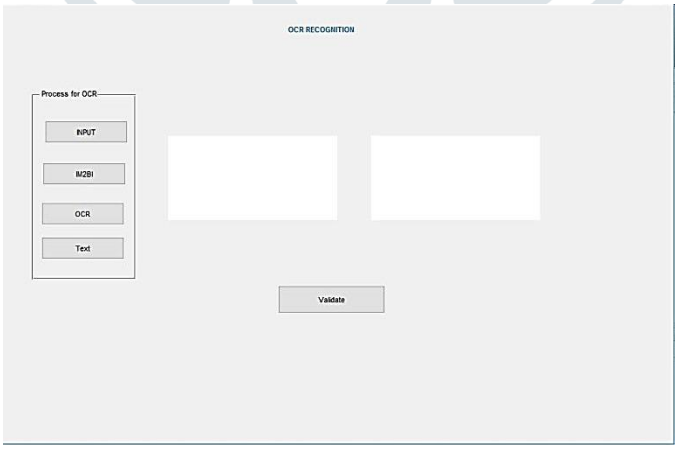

Fig.2. Work space

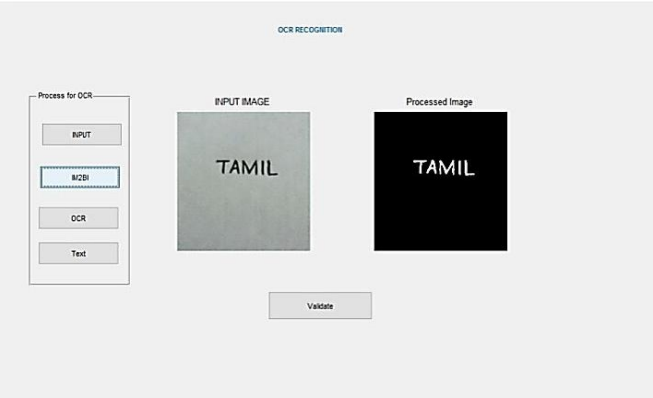

Fig.3. Image captured with proper lighting

VI. CHALLENGES OF OPTICAL CHARACTER RECOGNITION

For better and excessive character popularity, accuracy is such a lot of OCR techniques but nonetheless hard to gain 100% accurate popularity specifically for person that has similarity. Demanding situation I study in the course of review is a lot them related to the statistics collection and preprocessing if we can pick out and rid of these challenges then we can get high accurate recognition.

6.1 Conditions of Uneven Lighting

Many times photographs taken from in door or out of door are stricken by mild shadows. It make hard to stumble on and segment characters. This type of troubles makes scanning records greater preferable than capturing it by means of camera. Camera mild flash also may additionally help for added lights and create shadows in photos. The effect of uneven lighting is shown in fig.6 and fig.7.

6.2 Blurring and Degradation

This additionally caused by photographs thinking about digital camera. This happen while photographs are taken from distance, seeking to capture on actions and lack of focusing .Image taking on this and other circumstance face blurring and degradation. For accurate recognition sharpness of characters is needed.

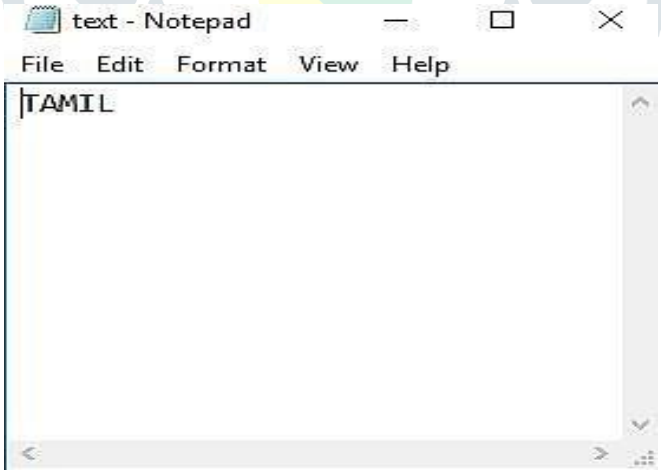

Fig.4. OCR output for image with proper lighting

6.3 Multilingual Environment

Characters that have multi environment in conjunction with large variety of individual languages like Ethiopian, Korean, Chinese, Japanese and other. Characters are written connectedly with every one-of-a-kind to another. It is difficult for computer to peer the distinction among most of them. Therefore the multi environmental characters are challenges for OCR to divide and extract individual characters and apprehend efficiently.

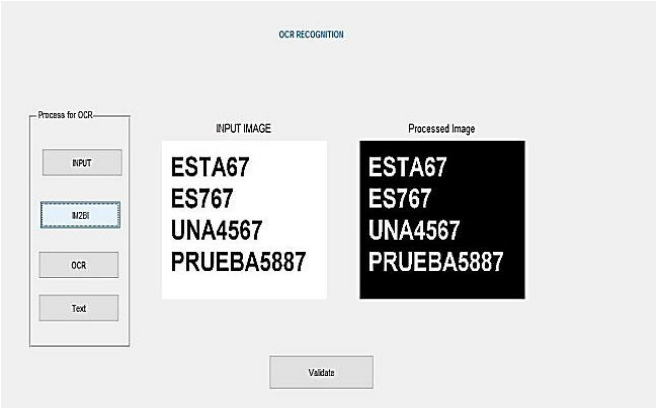

Fig.5. Input image generated in computer

6.4 Skewness (Rotation)

Image taking the use of digital camera also disturb by this difficulty. The attitude of the image is incorrect therefore when we fed this statistics to optical individual recognition system the outcome might be wrong. But there are strategies to resolve this trouble like fourier transformer, projection profile and so on.

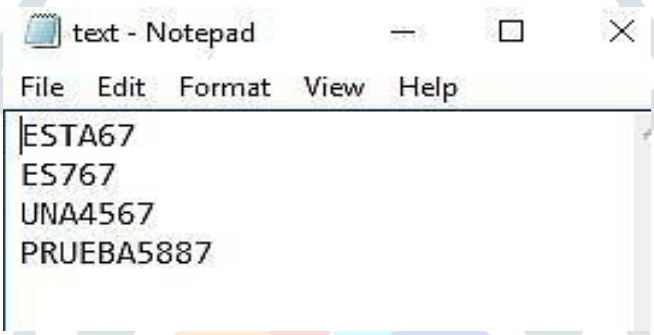

Fig.6. OCR output for image generated in computer

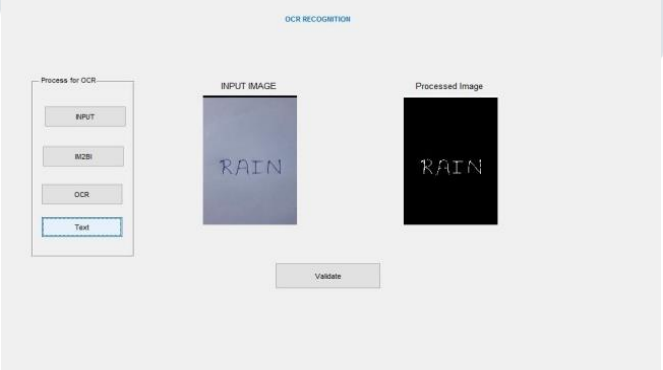

Fig.7. Image captured with uneven lighting

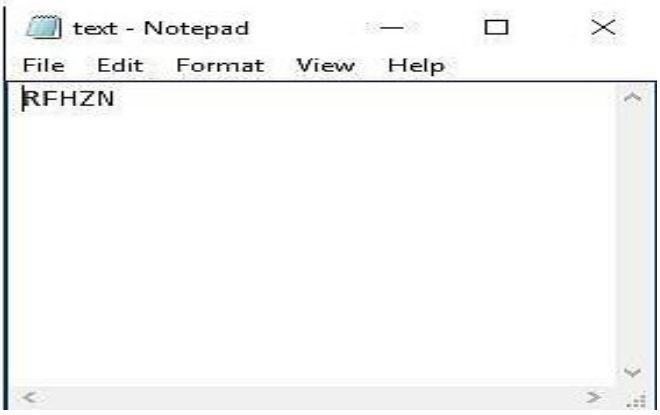

Fig.8. OCR output for image captured with uneven lighting

VII. CONCLUSION

In the studies work revised on this paper, person recognition machine use distinctive tactics and many of them get appropriate accuracy. What we are able to understand from this paper is characteristics extraction strategies need to be chosen consistent with the character you are working with because each scripts are alphabets has its own nature therefore want to locate strategies which match or appropriate for characters. The letter capable of extracting capabilities from man or women we are able to discover and recognize characters in maximum accuracy. For the images captured in uneven lighting, the accuracy is approximately 80% whereas for the images captured in proper lighting, the accuracy is almost 90%.

REFERENCES

[1] S. T. Javed and S. Hussain , "Segmentation Based Urdu Nastalique OCR," in Iberoamerica Congress on Pattern Recognition, Germany : Springer, 2013.

[2] B. Shi, X. Bai and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," IEEE Trans. Pattern Anal. Macj. Intell., vil. 38, Nov. 2017.

[3] S. Naz, A. Umar, R. Ahmad, S. B. Ahmad, S. H. Shirazi, I. Siddiqi, and M. I. Razzak, "Offline Cursive Urdu-Nastaliq script recognition using multi dimentional recurrent neural networks," Neurocomputing. Vol. 177. Pp. 228-241. Feb. 2016.

[4] N. H. Khan and A. Adnan, "Urdu optical character recognition systems: Present contributions and future directions, " IEEE Access, vol. 6, pp. 46019-46046, 2018.

[5] N. H. Khan, A. Adnan and S. Basar, "Urdu Ligature recognition using multilevel agglomerative hierarchical clustering, " Cluster Compute.., vol. 21 , pp. 503-514, 2018.