Pattern Recognition using Convolution Neural Network for Handwritten Gujarati Numerals.

PROF.KALAM NARREN
PROF.V.VINAY KRISHNA
Dept of Computer Engineering,JSPM RSCOE,Savitribai Phule University,Pune
Prof.Department of Computer Engineering

Abstract *— Languages have played a main role in Indian history and they continue to influence the lives of Indians till date. Various researches on optical character recognition (OCR) technique for Indian languages such as Hindi, Tamil, Bangla, Kannada, Gurumukhi, Malayalam and Marathi have already been carried out. Research efforts on Gujarati digit recognition are few and still have to gain momentum. This project intends to bring Gujarati digits recognition in attention. Methods based on CNN classifier are exercised for handwritten Gujarati numerals recognition. Experiments are carried out on datasets using different kinds of features and their fusion.*

Keywords--- *Gujarati handwriting, Gujarati numerals, handwritten numerals, Gujarati character recognition*

I. INTRODUCTION

Digit Recognition means to detect or identify either the printed or a handwritten digit. Character set includes all the numeric values i.e. from 0 to 9. Recognizing the printed digit is still an easier task as the font styles and families are limited while in case of handwritten text recognition becomes strenuous as there is no restriction on the styles of writing the text by a human being. Other than this distinction text may be recognized during the process of writing or afterwards.

The process of digit recognition involves several sub-routines that is sample collection from different dataset, applying pre-processing on the collected samples to make them ready to extract the features, an effective and efficient algorithm is needed to be applied to extract feature and for feature vector creation, once the feature vector is created now is the turn to make use of some classification algorithm to assign particular character to a specific class.

A number of feature extraction techniques have been proposed by the researchers to mine the specific features. Feature extraction methods for handwritten character recognition are based on either structural features or statistical features. For classification and training purpose several algorithms are available like CNN.

1. Goals and Objectives

The objective of this research is to give the optimum recognition accuracy of handwritten numerals using deep learning algorithm. Convolutional Neural Network (CNN) is the class of deep, feed forward neural network, most commonly applied to analyzing visual imagery. CNN requires little pre-processing as compared to other image classification algorithm. So we are quite sure that we will get better accuracy using CNN as our classification algorithm in recognizing handwritten Gujarati digits.

II. LITERATURE REVIEW

1: Classification of Handwritten Gujarati Numerals [3]

Journal: IEEE

Year: 2015

Authors: (1) Archana.N.Vyas Information Technology Dept. Dharamsinh Desai University, Nadiad, Gujarat, India

(2) Mukesh M.Goswami Information Technology Dept. Dharamsinh Desai University, Nadiad, Gujarat, India

Keywords: OCR;Gujarati Handwritten Numeral recognition; Modified Chain code; KNN;ANN,SVM

Dataset: Sample were collected from 300 different age group and different background people, So overall 3000 digits, 900 were collected with thick marker pen and 2100 digits were written with think or regular pen.

Summary: This paper addresses the problem of recognizing handwritten numerals for Gujarati Language. Three methods are presented for feature extraction. One belongs to the spatial domain and other two belongs to the transform domain. In first technique, a new method has been proposed for spatial domain which is based on Freeman chain code. In second and third method, 85 dimensional Fourier descriptors and Discrete Cosine Transform coefficients were computed and treated as feature vectors. These methods are tested with three different classifiers namely K-Nearest Neighbour, Support Vector Machine and Back Propagation Neural Network. Experimental results were evaluated using 10 fold cross validation.

2: Handwritten Hindi Character Recognition using K Means Clustering and SVM [4]

Journal: 4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (IEEE)

Year: 2015

Authors : (1) Akanksha Gaur Department of Compute Science, AKGEC Ghaziabad, India

(2) Sunita Yadav Department of Compute Science, AKGEC Ghaziabad, India

Keywords: Hindi characters; Feature Extraction; Classification; K-means clustering; Support vector machine (SVM).

Dataset: Character image set of size 430 is taken for implementation. In which 140 characters are used for training and rest 290 images of characters are taken as test data for classification using SVM.

Summary:

In this paper recognition of Hindi characters is done by using a three step procedure. First step is preprocessing, the next step is feature extraction in which region based k-means clustering is used and the feature vector is created and used in classification phase as input. Third step is classification process, for which support vector machine with linear kernel function is used. Support vector machine uses hyper-plane for classification.

3: An Efficient Handwritten Devnagari Character Recognition System Using Neural Network [5]

Journal : Proceeding Conference IEEE

Year : 2013

Authors : (1) Ms. Neha Sahu Computer Science Department, ITM University, Gurgaon, Haryana, India.

(2) Mr. Nitin Kali Raman Electronic and commincation Department, Shri Baba Mast Nath Engineering College,

Rohtak, Haryana, India

Keywords: Devnagari Character Recognition, Off-line Handwriting Recognition, Segmentation, Feature Extraction, OCR.

Dataset : 5 samples of each of the 33 consonants making a total of 165 samples collected.

Summary :

This paper presents a system created for recognizing Devnagari handwritten characters using Neural Network Toolbox of Matlab and this has made it easier to construct best suitable Neural Network based character recognition system. Their Primary goal to show precision and accuracy of character recognition depending on the parameters of the implemented neural network. It communicates with the user through a GUI created by using Matlab's tools.

4: Combined Horizontal and Vertical Projection Feature Extraction technique for Gurmukhi Handwritten Character Recognition [6]

Journal: Proceeding conference IEEE

Year: 2015

Authors : (1) Manoj Kumar Mahto, Department of Computer Science Gurukul Kangri Vishwavidyalaya,

Haridwar, India

(2) Karamjit Bhatia, Department of Computer Science Gurukul Kangri Vishwavidyalaya, Haridwar, India

(3) R.K. Sharma School of Mathematics and Computer Applications, Thapar University, Patiala, India

Keywords: Character Recognition, Off-line Handwritten Gurmukhi Character Recognition, Horizontal Projection, Vertical Projection, Feature Extraction.

Dataset:

It is collected from 10 different persons belonging to different age groups, Each writer contributed 10 samples of each character of 35 different Gurmukhi alphabets and thus yielding a data set of 3500 samples.

Summary:

In this paper, They focussed on off-line recognition of handwritten characters of Gurmukhi, they have used a combined horizontal and vertical projection feature extraction scheme for recognition of Gurmukhi characters and For classification, they used SVM with linear and polynomial kernel.

5: Recognition of Handwritten Hindi Characters using Backpropagation Neural Network[7]

Journal : International Journal of Computer Science and Information Technologies

Year : 2016

Authors : (1) Gunjan Singh Faculty of Management & ComputeApplications, R.B.S. College, Agra, India

(2) Sushma Lehri Institute of Computer & Information Sciences, Dr. B. R . Ambedkar University, Agra, India.

Keywords: Neural Networks, Binarization, Normalization, Thinning, Feature Extraction.

Dataset: Total 1000(200 for 5 character). Out of 1000 samples, 60 % used for training , 20% for validation and rest 20 % were used for testing.

Summary:

An offline handwritten Hindi character recognition system using neural network is presented in this paper. The paper proposes the approach to recognize Hindi characters in four stages 1) Scanning, 2) Preprocessing, 3) Feature Extraction and, 4) Recognition. Preprocessing includes noise reduction, binarization, normalization and thinning. Feature extraction includes extracting some useful information out of the thinned image in the form of a feature vector. The feature vector comprises of pixels values of normalized character image. A Backpropagation neural network is used for classification.

6: Use of Gradient Technique for extracting Features from Handwritten Gurmukhi Characters and Numerals [8]

Journal : International Conference on Information and Communication Technologies

Year : 2015

Authors : (1) Ashutosh Aggarwal Department of computer Science & Engg. Khalsa College of Engg. And Technology, Amritsar-143001, India.

(2) Karamjeet Singh Department of computer Science & Engg. Rayat & Bhara College of Engg. And BioTechnology for Women, Mohali-140104, India.

(3) Kamalpreet Singh Department of computer Science & Engg. Lovely Professional University, Jalandhar - 144411, India.

Keywords: Handwritten Gurmukhi Script Recognition; Indian Script Recognition, Gradient Features, Pattern Recognition, OCR

Dataset : 7000 images of gurmukhi character and 2000 images of gurmukhi numeral of 63 X 63 size

Summary :

In this paper a recognition system for offline handwritten Gurmukhi Characters and Numerals using Gradient information as mode of feature extraction technique is used, two ways of feature extraction using gradient information are explained. Both methods operate by accumulating gradient information from an image by dividing it into sub-images and finally concatenating the obtained gradient feature obtained from each block to form a vector of feature value with dimensionality 200. The efficiency of this feature vector is tested on two separate handwritten databases of Gurmukhi characters and Gurmukhi Numerals .And they have used SVM with RBF kernel as classifier.

7 : Handwritten Devanagari Character Recognition using Wavelet Based Feature Extraction and Classification Scheme [9]

Journal : Annual IEEE India Conference (INDICON)

Year : 2014

Authors : (1) Adwait Dixit, Departmemt of Electronic and Telecommunications, Vishwakarma Institute of Engineering and Technology, Pune, India

(2) Ashwini Navghane Assistant Professor Department of Electronics and Telecommunications Vishwakarma Institute of Information Technology, Pune India

(3) Yogesh Dandawate Assistant Professor Department of Electronics and Telecommunications Vishwakarma Institute of Information Technology, Pune India

Keywords: OCR, Devanagari, wavelet features, neural networks.

Dataset: 20 Handwritten characters from 100 people have been collected resulting 2000 character used for experimentation

Summary :

In this paper, handwritten characters written on paper are scanned ,preprocessed and on every individual characters wavelet transform is applied so as to get decomposed images of characters. Statistical parameters are computed over the decomposition to form feature vector. The feature vectors serve as input to back propagation neural networks for classification into one of 20 classes and on based classes they are recognized.

8 : Devanagari Character Recognition Using Artificial Neural Network[10]

Journal : Internation Joural of Engineering and Technology

Year : 2017

Authors : (1) Vasu Negi Departmemt of Information Technology, Maharaja Surajmal Institute of Technology,

Guru Gobind Singh Indraprastha University, New Delhi,India

(2) Suman Mann Departmemt of Information Technology, Maharaja Surajmal Institute of Technology, Guru Gobind Singh Indraprastha University,New Delhi,India

(3) Vivek Chauhan Departmemt of Information Technology, Maharaja Surajmal Institute of Technology, Guru Gobind Singh Indraprastha University, New Delhi,India

Keywords: Devanagari, Character Recognition, Keras, Theano, Python

Summary :

In this paper, They have used Keras along with Theano which are python libraries for building neural network. They constructed a simple artificial neural network using keras to recognize isolated Devanagari characters and also analyze the impact of variations in parameters in learning phase.

9: RECOGNITION OF HANDWRITTEN DEVANAGARI CHARACTERS USING MACHINE LEARNING APPROACH [11]

Journal: International Journal of Computer Science and Information Technologies

Year: 2016

Authors: (1) PANKAJ KALE Department of Electronic and Telecommunications, Vishwakarma Institute of Information Technology, Pune, India

(2) ARTI V.BANG Department of Electronic and Telecommunications, Vishwakarma Institute of Information Technology, Pune, India

(3) DEVASHREE JOSHI Department of Electronic and Telecommunications, Vishwakarma Institute of Information Technology, Pune, India

Keywords: Devanagri, OCR, ANN Classifier, Feature Extraction, Neural Network Model. **Dataset :** 50 handwritten character from 10 people resulting 500 characters.

Summary :

The handwritten characters are scanned, preprocessed and on every individual character feature extraction is applied. The feature vectors are given as an input for training to back propagation neural network. Testing is carried out using individual characters as well as sentences.

10 : Multi-Layer Perceptron (MLP) Neural Networkd Technique for Offline Handwritten Gurmukhi Character Recognition [1]

Journal : IEEE

Year : 2016

Authors : (1) Gurpreet Singh SLIET, Longowal,Punjab, India

(2) Manoj Sachan Associate Professor SLIET, Longowal, Punjab, India

Keywords: Digitizing documents, Offline recognition, Gurmukhi Script, MLP, Feed Forward topology, Unicode.

Dataset : Samples collected from 25 different people , resulting 481 character images.

Summary :

In this paper, they present a technique based on Multi Layer Perceptron (MLP) Neural Network model. Here we consider isolated handwritten Gurmukhi characters for recognition. MLP is used because it uses generalized delta learning rules and easily gets trained in less number of iterations. The proposed method in this paper detect graphical symbols by identifying lines and characters from the image. After that it analyzes the symbols by training the network using feed forward topology for a set of desired unicode characters.

Recognition Accuracy :

The performance rate of proposed system maximum up to 98.96% for recognition of symbols by using MLP neural network.

11 : A Novel Hierachical Technique for Offline Handwritten Gurmukhi Character Recognition [2]

Journal : Springer

Year : 2014

Authors : (1) Munish Kumar Department of Computer Science, Punjab UniversityRural Centre Kauni,, Muktsar, Punjab, India

(2) M.K.Jindal Department of Computer Science, Punjab UniversityRural Centre Kauni,,Muktsar, Punjab, India

(3) R. K.Sharma Department of Electronic and Telecommunications, Vishwakarma Institute of Information Technology,Pune, India

Keywords: Character recognition · Feature extraction · Classification · Feature selection. **Dataset :** 3,500 samples of isolated offline handwritten Gurmukhi characters written by one hundred different writers.

Summary :

In present paper, Authors have presented a novel hierarchical technique for isolated offline handwritten Gurmukhi character recognition. A robust feature set of 105 feature elements is proposed under this work for recognition of offline handwritten Gurmukhi characters using four types of topological features, namely, horizontally peak extent features, vertically peak extent features, diagonal features, and centroid features. For classification Support Vector Machines (SVMs) classifier has been used in this work. SVMs classifier has been considered with four different kernels, namely, linear kernel, polynomial kernel, RBF kernel and sigmoid kernel.

II. EXISTING SYSTEM

The main objective of the project was to find an efficient algorithm for feature extraction with reduced the length of feature vector as well as which can increase the accuracy of recognition of Gujarati Numerals. For us it is very easy to recognize printed as well as handwritten character as we as do not recognize things on the basis of exact structure or pattern of any entity. We have a capability to derive the common patterns very quickly through the Intelligence gifted to us by God. But to write such an intelligent program for computer to show such characteristic like humans is a difficult and a tedious job because computer work on comparison if the two entities are exactly the same it can identify otherwise it will not. To make the computer intelligence it is necessary to first analyze and detect the procedure following which human gets able to remind or recall or identify things i.e. the uniqueness among them which is referred as feature vector. Then the next task is to make the computer learn it to get able to make judgments by adjusting the values

III. SURVEY OF PROPOSED SYSTEM

We have proposed an end-to-end trainable deep CNN . Although a great extent of work has already been done in Handwritten Character Recognition with different CNN algorithms using various feature extraction techniques. This is one of the biggest challenge with traditional CNN algorithms because a programmer needs to tell the computer in advance what features should be taken care of so that to get an accurate decision percentage to recognize Gujarati digit

using CNN algorithm.Normally handwritten character recognitionis divided into six phases which are image acquisition,pre-processing,segmentation,feature extraction,classification and post processing.Feature extraction plays an important role in getting the result of a model.The purpose behind CNN is to identify and analyze a document image by dividing it into characters.These characters are further compared with image pattern to predict the probable characters. With the advent of highest computing power and deep learning, applications of computer vision like handwritten character recognition took a little relief because deep learning models are itself responsible for what features to look upon without explicitly guided Convolution Neural Network.

ADVANTAGES OF PROPOSED SYSTEM:

- The device does not require a keyboard therefore the device could be smaller or it could have a bigger screen Some people find it quicker to input data by using handwriting than using a keyboard
- Efficient digit enhancement and feature extraction.
- Classify using neural network so time constraint will be reduced.
- Developing fast, automatic, efficient and accurate system.
- Accuracy achieved.

V. SYSTEM ARCHITECTURE

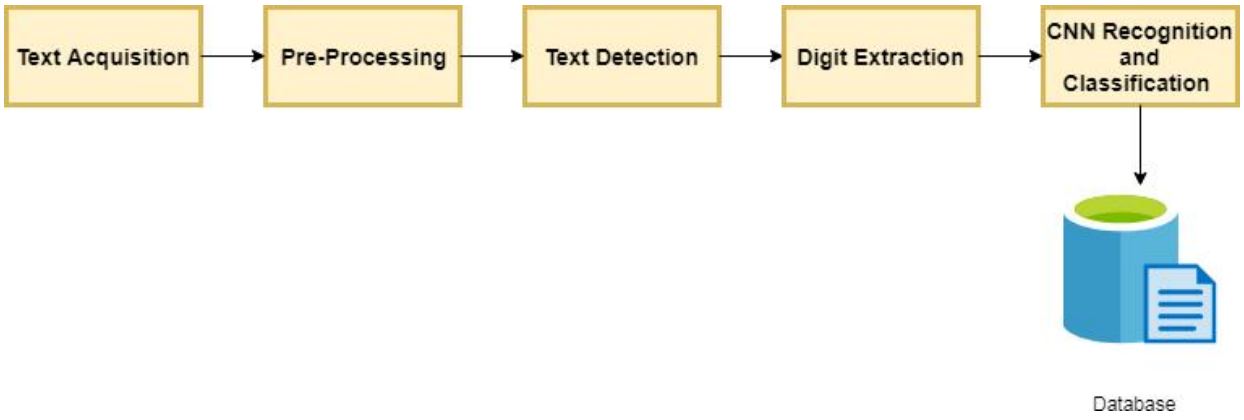

Fig.: System Architecture

VI. CONCLUSION AND FUTURE WORK

We conclude that we can produce better and optimum result in Handwritten numeral recognition with the help of deep learning. As deep learning is an emerging field and it has overcome the many disadvantages of previous algorithms and due to its ability to automatically fetch the features. This will obviously enhance the recognition

rate. We can even think of using a little bigger dataset so that recognition accuracy can be better judged with deep learning.

VII. REFERENCES

[1] Gurpreet Singh, Manoj Sachan, "Multi-Layer Perceptron(MLP) neural Network Technique for Offline Handwritten gurmukhi Character Recognition", 978-1-4799-3975-6/14/$31.00 ©2014 IEEE

[2] Munish Kumar , M. K. Jindal , R. K. Sharma," A Novel Hierachical Technique for Offline Handwritten Gurmukhi Character Recognition" 2014 Natl. Acad. Sci. Lett. 37(6):567–572.

[3] Archana N.Vyas, Mukesh m.Goswami, "Classification of Handwritten Gujarati Numerals", 978-1-4799-8792-4/15/$31.00 ©2015 IEEE

[4] Akanksha Gaur, Sunita Yadav, "Handwritten Hindi character Recognition using K-Means Clustering and SVM", 978-1-4799-5532-9/15/$31.00 ©2015 IEEE

[5] Neha Sahu, Nitin Kali Raman,"An Efficient handwritten Devanagari character Recognition System Using Neural Network", 978-1-4673-5090-7/13/$31.00 ©2013 IEEE

[6] Manoj Kumar Mahto, Karamjeet Bhatia, R.K.Sharma, "Combined Horizontal and Vertical Projection feature Extraction Technique for Gurmukhi Handwritten Character Recognition", 2015, International conference on Advances in Computer Engineering and Applications (ICACEA)

[7] Gunjan Singh et al, Recognition of Handwritten Hindi Characters using Backpropagation Neural Network 2012 / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (4) ,4892-4895\

[8] Ashutosh Aggarwal ,Karamjeet Singh, kamalpreet Singh,"Use of Gradient Technique for extracting features from Handwritten Gurmukhi Characters and Numerals", 2014, International Conference on Information and Communication Technologies (ICICT 2014)

[9] Adwait Dixit , Ashwini Navghane, ogesh Dandawate,"Handwritten Devnagari character Recognition using Wavelet Based Feature Extraction and Classification Scheme", 2014, Annual IEEE India Conference(INDICON).

[10] Vasu Negi, Suman Mann, Vivek Chauhan, " Devnagari Character Recognition Using Artificial Neural Network", 2017, International Journal of Engineering and Technology (IJET).

[11] *Pankaj Kale*, *Arti V. Bang*, *Devashree Joshi,* "Recognition Of *Handwritten Devanagari Characters Using Machine Learning Approach* " , International Journal of Industrial Electronics *and* Electrical Engineering , Volume-3, Issue-9 (Sep, *2015*)

[12] Munish Kumar, M.K.Jindal, R.K.Sharma,"SVM Based Offline Handwritten Gurmukhi Character Recognition", 2011, Conference SCAKD

[13] Seyyid Ahmed Medjahed, "A Comparative Study of Feature Extraction Methods in Image Classification",2015,I.J.Image,Graphics and Signal Processing, 2015, 3, 16-23

[14] Mohamed Cheriet, Nawwaf Kharma, Cheng-Lin Liu, Ching Y. Suen, "Character Recognition Systems: A Guide for Students and Practitioners",*Wiley Inter-Science,* 2007

[15] Li Z. et al. "Modified Binary Image Thinning Using Template-Based PCNN" (2013) International conference on information technology and software engineering volume 212 pp.731-74