



REVIEW-3 WIN SEM 2018-19
M.TECH- (SOFTWARE ENGINEERING)
PROJECT BASED COMPONENT

SWE 1011-SOFT COMPUTING

SLOT: D2 + TD2

TITLE: HANDWRITTEN TEXT RECOGNITION
USING ARTIFICIAL NEURAL NETWORKS

TEAM MEMBERS:

16MIS0323 – T. DHANVIKAS

16MIS0426 – V. ACHYUTH KUMAR

16MIS0485 – T. HARSHA

16MIS0352 – SUMANTH VARMA

SUBMITTED TO:

PROF. CHIRANJI LAL CHOWDARY

CERTIFICATE

This is to certify that the Project work entitled'' Handwritten text Recognition using artificial neural networks'' that is being submitted by ''T. Dhanvikas, V.Achyuth Kumar, Sumanth AND T.Harsha'' in M. Tech (S.E) for SWE1011: SOFT COMPUTING is a record of bonafide work done under my supervision. The contents of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted for any other course.

TABLE OF CONTENTS

Contents

Introduction.....	3
Methods.....	4
Literature Survey.....	5
□ Handwritten Character Recognition using Neural Network.....	6
□ Handwritten Character Recognition Using Gradient Features	6
□ Character Recognition Using Matlab's Neural Network Toolbox.....	6
Process Algorithm.....	8
Block Diagram and Algorithm.....	9
Character Extraction Algorithm	11
Flow Chart.....	12
Architecture of the system.....	13
Home Page	13
Uploaded Image	14
Neural Network Training	14
Output File.....	15
Evaluation.....	17
Conclusion.....	17
Summary	17
References	18

Introduction:

The purpose of this project is to take handwritten English characters as input, process the character, train the neural network algorithm, to recognize the pattern and modify the character to a beautified version of the input.

This project is aimed at developing software which will be helpful in recognizing characters of English language. This project is restricted to English characters only. It can be further developed to recognize the characters of different languages. It engulfs the concept of neural network.

One of the primary means by which computers are endowed with humanlike abilities is using a neural network. Neural networks are particularly useful for solving problems that cannot be expressed as a series of steps, such as recognizing patterns, classifying them into groups, series prediction and data mining.

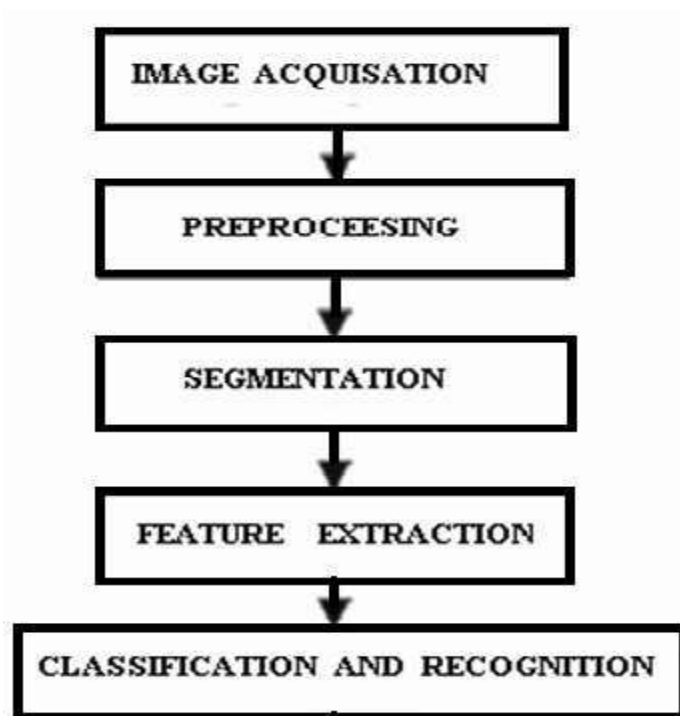
Pattern recognition is perhaps the most common use of neural networks. The neural network is presented with a target vector and also a vector which contains the pattern information, this could be an image and hand-written data. The neural network then attempts to determine if the input data matches a pattern that the neural network has memorized.

A neural network trained for classification is designed to take input samples and classify them into groups. These groups may be fuzzy, without clearly defined boundaries. This project concerns detecting free handwritten characters.

Methods

The proposed method comprises of 4 phases:

1. Pre-processing.
2. Segmentation.
3. Feature Extraction.
4. Classification and Recognition.



Literature Survey:

A few state-of-the-art approaches that use hand written character recognition for text identification have been summarized here:

□ **Handwritten Character Recognition using Neural Network**

Chirag I Patel, Ripal Patel, Palak Patel.

Objective of this paper is to recognize the characters in a given scanned documents and study the effects of changing the Models of ANN. Today Neural Networks are mostly used for Pattern Recognition task. The paper describes the behaviours of different Models of Neural Network used in OCR. OCR is widespread use of Neural Network. We have considered parameters like number of Hidden Layer, size of Hidden Layer and epochs. We have used Multilayer Feed Forward network with Back propagation. In Preprocessing we have applied some basic algorithms for segmentation of characters, normalizing of characters and De-skewing. We have used different Models of Neural Network and applied the test set on each to find the accuracy of the respective Neural Network.

□ **Handwritten Character Recognition Using Gradient Features**

AshutoshAggarwal, Rajneesh Rani, RenuDhir.

Feature extraction is an integral part of any recognition system. The aim of feature extraction is to describe the pattern by means of minimum number of features that are effective in discriminating pattern classes. The gradient measures the magnitude and direction of the greatest change in intensity in a small neighbourhood of each pixel. (In what follows, "gradient" refers to both the gradient magnitude and direction). Gradients are computed by means of the Sobel operator. In this paper an effort is made towards recognition of English Characters and obtained recognition accuracy of 94%. Due to its logical simplicity, ease of use and high recognition rate, Gradient Features should be used for recognition purposes.

□ **Character Recognition Using Matlab's Neural Network Toolbox**

Kauleshwar Prasad, Devvrat C. Nigam, Ashmika Lakhotiya and Dheeren Umre.

Recognition of Handwritten text has been one of the active and challenging areas

of research in the field of image processing and pattern recognition. It has numerous applications which include, reading aid for blind, bank cheques and conversion of any hand written document into structural text form. In this paper we focus on recognition of English alphabet in a given scanned text document with the help of Neural Networks. Using Mat lab Neural Network toolbox, we tried to recognize handwritten characters by projecting them on different sized grids. The first step is image acquisition which acquires the scanned image followed by noise filtering, smoothing and normalization of scanned image, rendering image suitable for segmentation where image is decomposed into sub images. Feature Extraction improves recognition rate and misclassification. We use character extraction and edge detection algorithm for training the neural network to classify and recognize the handwritten characters.

- **Neural based handwritten character recognition** *Hanmandlu M, Murali Mohan K.R, Kumar H.*

This paper explores the existing ring based method (W.I.Reber, 1987), the new sector based method and the combination of these, termed the Fusion method for the recognition of handwritten English capital letters. The variability associated with the characters is accounted for by way of considering a fixed number of concentric rings in the case of the ring based approach and a fixed number of sectors in the case of the sector approach. Structural features such as end points, junction points and the number of branches are used for the preclassification of characters, the local features such as normalized vector lengths and angles derived from either ring or sector approaches are used in the training using the reference characters and subsequent recognition of the test characters. The recognition rates obtained are encouraging.

- **A feature extraction technique based on character geometry for character recognition.**

Dinesh Dileep.

This paper describes a geometry-based technique for feature extraction applicable to segmentation-based word recognition systems. The proposed system extracts the geometric features of the character contour. These features are based on the basic line types that form the character skeleton. The system gives a feature vector as its output. The feature vectors so generated from a training set were then used to train a pattern recognition engine based on Neural Networks so that the system can be benchmarked.

- **A Review of Gradient-Based and Edge-Based Feature Extraction Methods for Object Detection.**

Sheng Wang.

In computer vision research, object detection based on image processing is the task of identifying a designated object on a static image or a sequence of video frames. Projects based on such research works have been widely adapted to various industrial and social applications. The field to which those applications apply includes but not limited to, security surveillance, intelligent transportation system, automated manufacturing, and quality control and supply chain management. In this paper, we are going to review a few most popular computer vision methods based on image processing and pattern recognition. Those methods have been extensively studied in various research papers and their significance to computer vision research has been proven by subsequent research works. In general, we categorize those methods into to gradient-based and edge based feature extraction methods, depending on the low level features they use. In this paper, the definitions for gradient and edge are extended. Because an image can also be considered as a grid of image patches, it is therefore reasonable to incorporate the concept of granules to gradient for a review.

Problem Definition

The purpose of this project is to take handwritten English characters as input, process the character, train the neural network algorithm, to recognize the pattern and modify the character to a beautified version of the input.

This project is aimed at developing software which will be helpful in recognizing characters of English language. This project is restricted to English characters and numerals only. It is also helpful in recognizing special characters. It can be further developed to recognize the characters of different languages. It engulfs the concept of neural network.

One of the primary means by which computers are endowed with humanlike abilities is using a neural network. Neural networks are particularly useful for solving problems that cannot be expressed as a series of steps, such as recognizing patterns, classifying them into groups, series prediction and data mining.

Pattern recognition is perhaps the most common use of neural networks. The neural network is presented with a target vector and also a vector which contains the pattern information, this could be an image and hand-written data. The neural network then attempts to determine if the input data matches a pattern that the neural network has memorized.

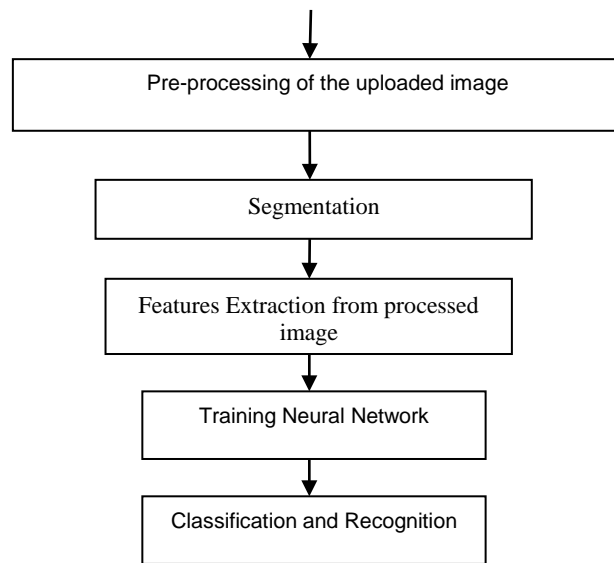
A neural network trained for classification is designed to take input samples and classify them into groups. These groups may be fuzzy, without clearly defined boundaries. This project concerns detecting free handwritten characters.

Block Diagram and Algorithm

The proposed methodology uses some techniques to remove the background noise, and features extraction to detect and classify the handwritten text.

The proposed method comprises of 4 phases:

1. Pre-processing.
2. Segmentation.
3. Feature Extraction.
4. Classification and Recognition.

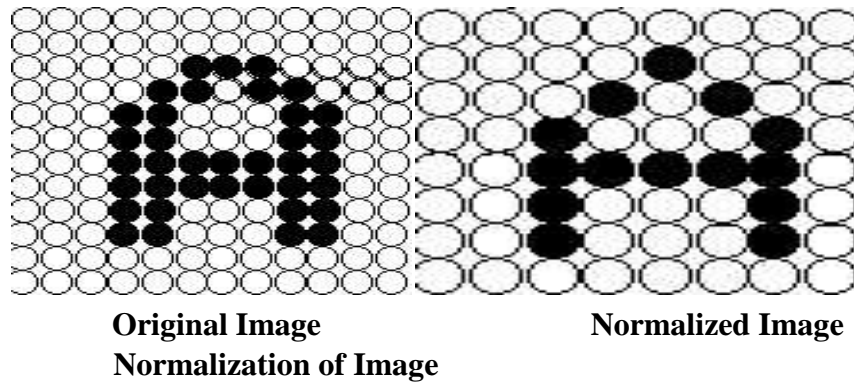


Pre-processing

The pre-processing is a series of operations performed on scanned input image. It essentially enhances the image rendering it suitable for segmentation. The role of preprocessing is to segment the interesting pattern from the background. Generally, noise filtering, smoothing and normalization should be done in this step. The pre-processing also defines a compact representation of the pattern. Binarization process converts a gray scale image into a binary image. Dilation of edges in the binarized image is done using sobel technique.

Segmentation

In the segmentation stage, an image of sequence of characters is decomposed into sub-images of individual character. The pre-processed input image is segmented into isolated characters by assigning a number to each character using a labelling process. This labelling provides information about number of characters in the image. Each individual character is uniformly resized into pixels. Normalization: After extracting the character we need to normalize the size of the characters. There are large variations in the sizes of each Character hence we need a method to normalize the size.



Character Extraction Algorithm

1. Create a Traverse List: - List of pixels which have been already traversed.
This list is initially empty.
2. Scan row Pixel-by-Pixel.
3. Whenever we get a black pixel check whether the pixel is already in the traverse list, if it is simply ignore and move on else apply Edge-detection Algorithm.
4. Add the List of Pixels returned by Edge-detection Algorithm to Traverse List.
5. Continue the steps 2 - 5 for all rows.

Flow Chart

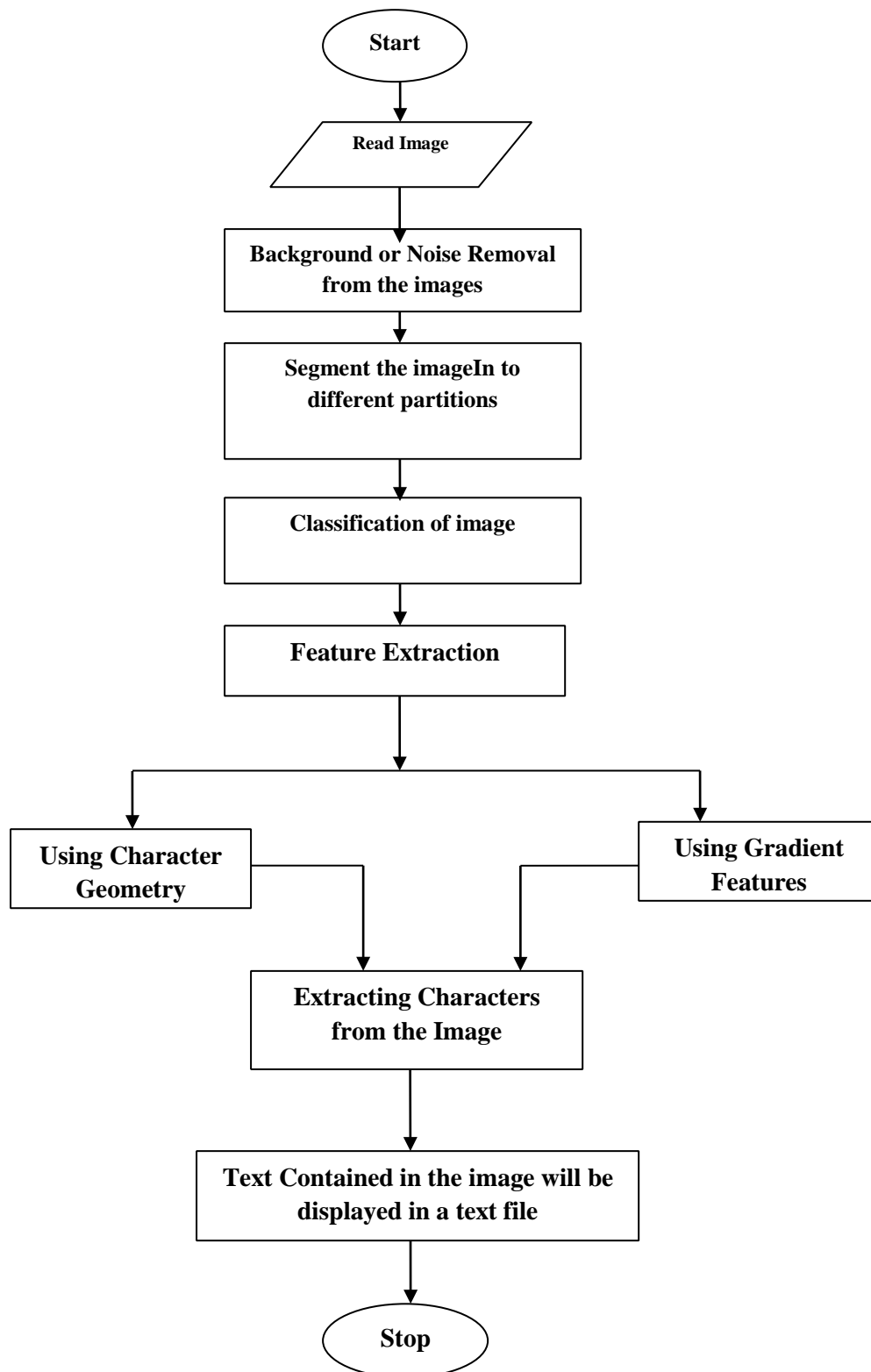
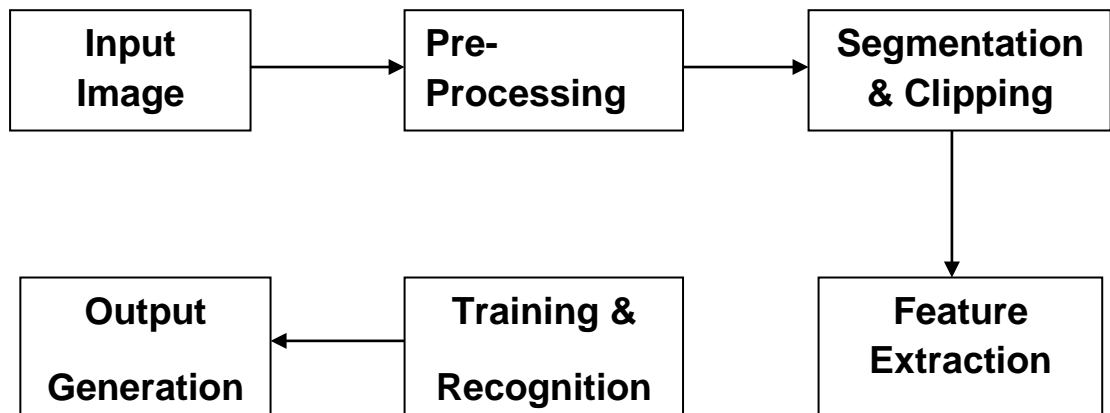


Fig: Flow chart

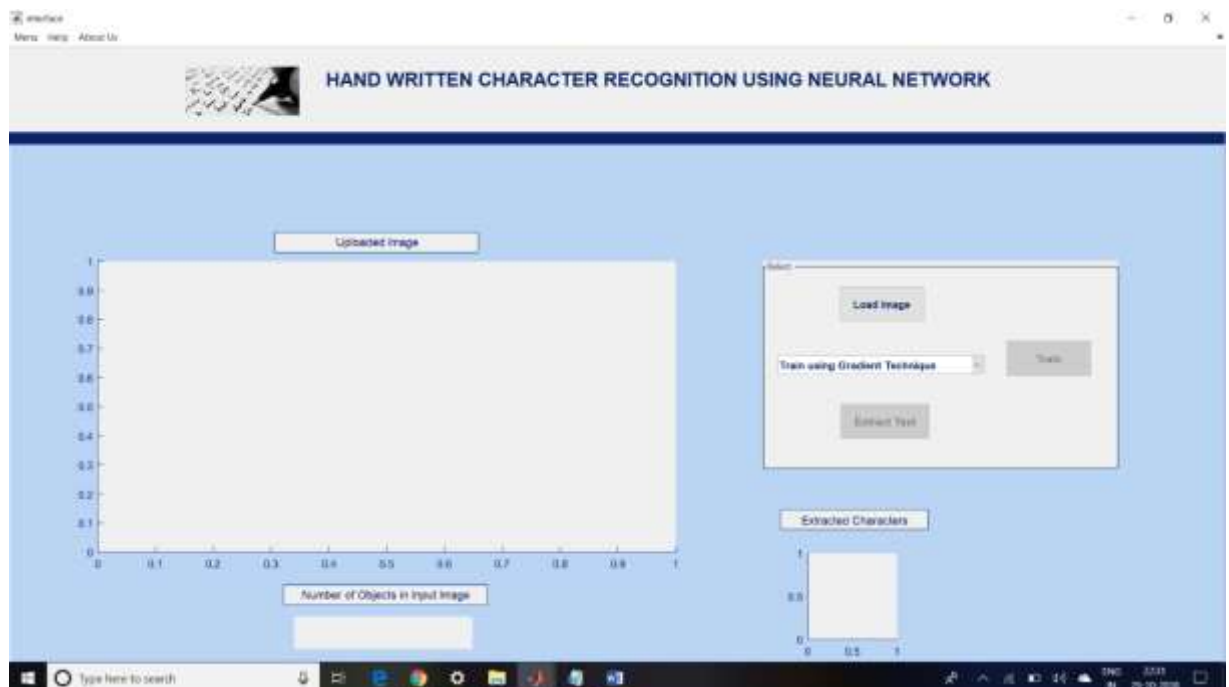
Architecture of the system



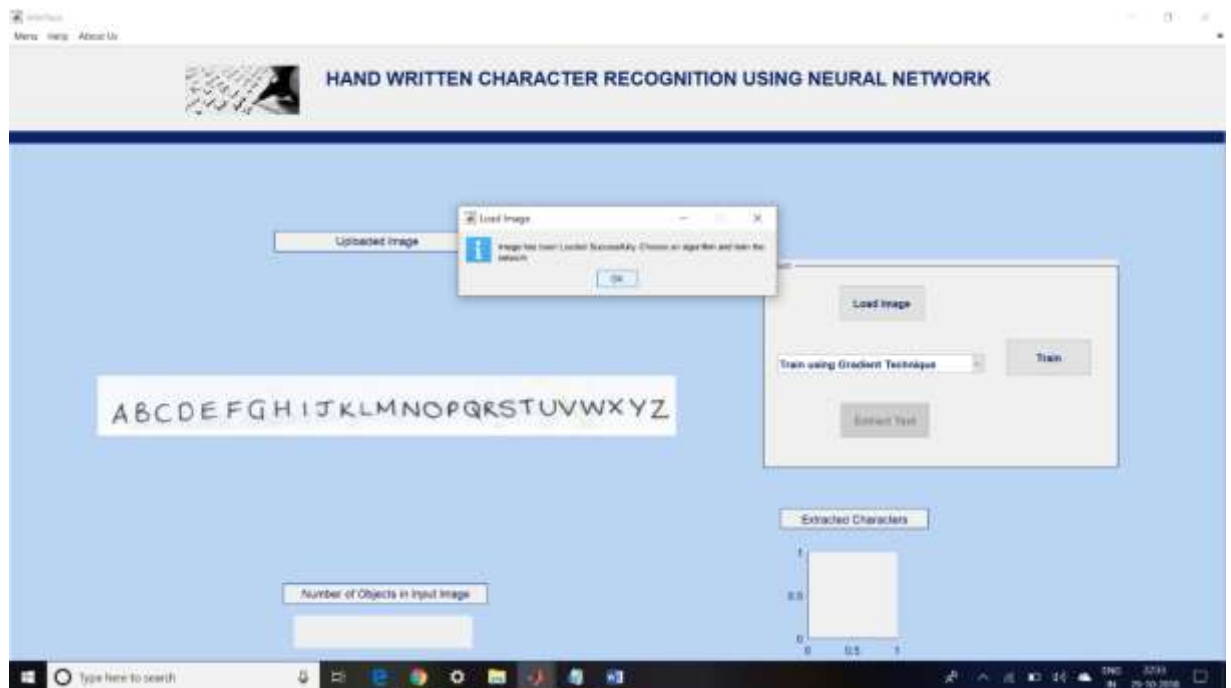
Architecture of the proposed system

Screen Shots of the Front End

Home Page

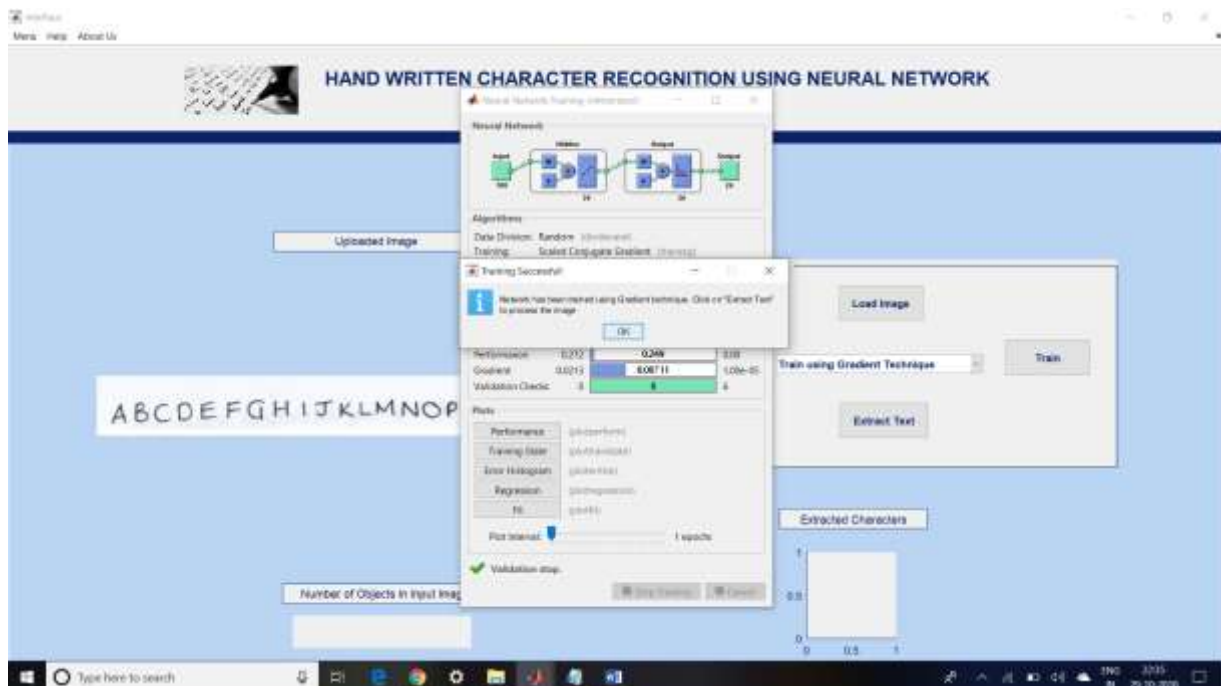


Uploaded Image



Uploading Image

Neural Network Training



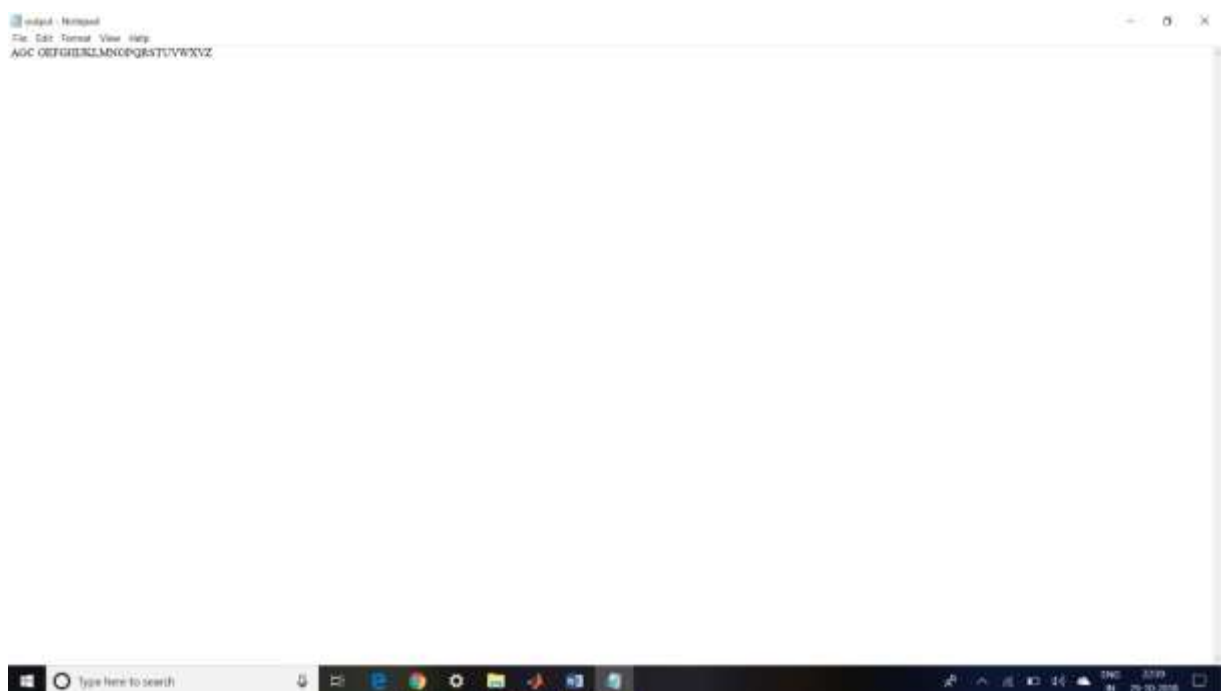
Neural Network Training

Character Extraction



Extracting Characters

Output File



Output File

Test Results:

Character Geometry

Epochs	Hidden layers	Config	Classification%
43	10	85-10-26	27.8
104	20	85-20-26	77.4
148	30	85-30-26	93.8
172	35	85-35-26	94.3
117	39	85-39-26	93.3
53	45	85-45-26	12.5
60	50	85-50-26	83.7
76	55	85-55-26	78.3
71	60	85-60-26	18
110	65	85-65-26	49.8
112	70	85-70-26	49.2

Using Character Geometry

Gradient Features

<u>Epochs</u>	<u>Hidden layers</u>	<u>Config</u>	<u>Classification %</u>
47	10	108-10-26	10
189	20	108-20-26	87.4
144	30	108-30-26	82
137	35	108-35-26	82.2
148	39	108-39-26	94.5
109	45	108-45-26	76.8
113	50	108-50-26	86.6
98	55	108-55-26	55.2
94	60	108-60-26	68.3
102	65	108-65-26	89.1
130	70	108-70-26	91.1

Using Gradient Features

Evaluation

- The Handwritten Character Recognition system was tested on several different scanned images containing handwritten text written with different styles and the results were highly encouraging.
- The proposed method performs preprocessing on the image for removing the noise and further uses feature extraction using gradient technique OR using character geometry which gives relatively good classification compared to OCR.
- The method is advantageous as it uses nine features to train the neural network using character geometry and twelve features using gradient technique. The advantage lies in less computation involved in feature extraction, training and classification phases of the method.
- The proposed methodology has produced good results for images containing handwritten text written in different styles, different size and alignment with varying background. It classifies most of the handwritten characters correctly if the image contains less noise in the characters and also in the background. Characters written with legible handwriting are classified more accurately.

Conclusion

Summary

- The effectiveness of the method that uses feature extraction using character geometry and gradient technique from scanned images containing handwritten characters is presented.
- The feature extraction methods have performed well in classification when fed to the neural network and preprocessing of image using edge detection and normalization are the ideal choice for degraded noisy images.
- The method of training neural network with extracted features from sample images of each character has detection accuracy to a greater extent.
- The proposed methodology has produced good results for images containing handwritten text written in different styles, different size and alignment with varying background
- The system is developed in MATLAB and evaluated for a set of sample images containing handwritten text on Intel dual core computer.

- The method is advantageous as it uses nine features to train the neural network using character geometry and twelve features using gradient technique.

References

- [1] Chirag I Patel, Ripal Patel, Palak Patel, "Handwritten Character Recognition Using Neural Networks", International Journal of Scientific & Engineering Research Volume 2, Issue 5, May-2011.
- [2] Kauleshwar Prasad, Devvrat C Nigam, AshmikaLakhotiya, DheerenUmre, "Character Recognition Using Matlab's Neural Toolbox", International Journal of u- and e- Service, Science and Technology Vol. 6, No. 1, February, 2013.
- [3] AshutoshAggarwal, Rajneesh Rani, RenuDhir, "Handwritten Character Recognition Using Gradient Features", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 5, May 2012.
- [4] Vinita Dutt, Sunil Dutt, "Handwritten Character Recognition Using Artificial Neural Network", Advances in Computing: 2011; 1(1): 18-23.
- [5] Rahul Kala, Harsh Vazirani, AnupamShukla, RituTiwari, "Offline Handwriting Recognition", International Journal of Computer Science issues, volume 7, March2010.
- [6] Dinesh Dileep, "A Feature Extraction Technique Based on Character Geometry for Character Recognition".
- [7] Alexander J. Faaborg, "Using Neural Networks to Create an Adaptive Character Recognition System", Cornell University, Ithaca NY, (May 14, 2002)
- [8] Swapnil A. Vaidya, Balaji R. Bombade "A Novel Approach of Handwritten Character Recognition using Positional Feature Extraction",IJCSMC, Vol. 2, Issue. 6, June 2013.
- [9] Sheng Wang "A Review of Gradient-Based and Edge-Based Feature Extraction Methods for Object Detection",Computer and Information Technology (CIT), 2011 IEEE 11th International Conference.