

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 8, August 2019

## Handwriting Analysis Based Human Personality Prediction Using Machine Learning Mechanism

Miss Swapnali S. Patil, Mrs.Vina M. Lomte

M.E Student(CE), Department of Computer Engineering, R.M.D. Sinhgad School of Engineering Pune, India

Assistant Professor(CE), Department Of Computer Engineering, R.M.D. Sinhgad School of Engineering Pune, India

**ABSTRACT:** English Handwriting Character Recognition approaches have been studied largely in the last few years and its progress and success rate is quite high. Segmentation is one of the important steps of character recognition. This work mainly deals with the new methods for line segmentation and character segmentation of characters of Handwritten English text and shows word spacing and margin results. In process firstly the text is segmented into lines after that lines into words and then from characters. Technique is finding the header lines and base lines by estimating the average line height and based on it. This technique works efficiently on characters for different text images. Using word spacing and margin result we will predict the personality of person. We will use Java technology for this implementation.

**KEYWORDS:** Character Recognition, Header lines, Baselines, Word Segmentation, Handwritten documents;

### I. INTRODUCTION

#### A. BACKGROUND

From the past few years handwritten character recognition is a front area of study and research on handwritten documents. No complete hand written text recognition approach is available in Indian plan and it is critical due to large character set of Indian languages and the presence of vowel modifiers and compound characters in Indian language and also English language.

The segmentation is one the most important method that decides the success of character recognition technique. It is used to diffuse of an image of a sequence of characters into sub images of separate symbols by segmenting lines and words. Segmentation is a process to separate the line, word and character from the image.

In recent years there is much work has been done in automated human behavior prediction approach. It shows that there are many research studies applying various techniques mixed with automated human behavior prediction approach graphology. If the graphology test is still done manually it takes a long time considering the features inspected in graphology very much. Accuracy of English handwriting analysis depends on how skilled the analyst is or how much knowledge to analyst have. Development in image processing and pattern recognition lead to analyzing of handwriting can be done using image processing techniques. The input is a human handwriting image, so that recognition can be done through the stages of conversion of images into numerical vector, image processing for quality improvement, followed by feature extraction and pattern recognition and more.

#### B. MOTIVATION

Segmentation is one the most important process which decides the success of character identification method. Segmentation is used to dissolve an image of a sequence of characters into sub images of individual symbols by segmenting lines and words of image. Image segmentation is the partitioned of image into multiple parts. With respect to the segmentation of human handwritten words into characters it is a critical task because of complexity of structural features and varieties in writing styles.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 8, August 2019

## C. OBJECTIVES

1. Increase the accuracy of segmentation.
2. To determine a persons personality based on structure of handwriting.
3. System takes minimum time for segmentation and personality prediction.

## II REVIEW OF LITERATURE

ShubhraSaxena, P. C. Gupta proposed a work in that, handwritten character recognition performs an important role in the modern world. It can resolve more complicated issues and makes humans job easier. The present paper represents a novel approach in recognizing handwritten Devanagari character through feed forward back propagation neural network[1].

SnehalS.Patwardhan, R.R. Deshmukh, proposed a work in that, Character Recognition is a process by which a computer identifies letters, numbers, or symbols and turns them into a digital form that a computer system can use and it is an important active part of research now days. It consists of Pattern Recognition and Image Processing. Character Recognition is generally categorized into Optical Character Recognition (OCR) and Handwritten Character Recognition (HCR) [2]

NehaSahu, R. K. Rathy, InduKashyap, proposed a work in that, Research in Optical Character Recognition (OCR) is very important particularly with an eye on its applications in banks, post offices, defense organizations, library automation, etc. Devnagari Optical Character Recognition needs more consideration as it is national language and there is less development in this field due to complexity in the script [3].

SaiprakashPalakollu, RenuDhir, Rajneesh Rani, proposed a work in that, mainly deals with the new approach for line segmentation and character segmentation of overlapping characters of Handwritten Hindi text. The text of image is segmented into lines, then lines into words and then from lines words header lines are detected and converted as straight lines. Each word is divided into three parts upper modifier, consonant and lower, so that character segmentation becomes easy[4].

Naresh Kumar Garg, LakhwinderKaur, M. K. Jindal, proposed a work in that, the main purpose of this paper is to provide the new segmentation technique based on structure approach for Handwritten Hindi text [5].

VikasJDongre, Vijay H Mankar, proposed a work in that a simple histogram based approach to segment Devnagari documents is proposed in this paper. Various challenges in segmentation of Devnagari script are also discussed[6].

ParthaPratimRoy, Umapada Pal, JosepLlads, proposed a work in that, in this paper we propose a method mostly based on morphological operation and run-length smearing algorithm (RLSA) to segment individual text lines from unconstrained handwritten document images[7].

Ratnashil N Khobragade, Dr. Nitin A. Koli, MahendraS Makesar, proposed a work in that, work on a set of preprocessing, segmentation, feature extraction, classification and matching techniques, which perform very important role in the recognition of characters. Feature extraction of which we can identify characters uniquely and with high degree of accuracy. So many techniques have been proposed for pre-processing, feature extraction, learning/classification, and post-processing[8].

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 8, August 2019

Miguel A. Ferrer, Moises Diaz, Cristina Carmona- Duarte and Aythami Morales propose a development of this methodology in which we generate dynamic information and provide a unified comprehensive synthesizer for both static and dynamic signature synthesis. The dynamics are found out by lognormal sampling of the 8-connected continuous signature trajectory, which contains, as a novelty, the pen-ups. The forgery generation imitates a signature by extracting the most perceptually relevant points of the given genuine signature and interpolating them [9].

Sara Rosenblum and Gideon Dror proposed a work in that, approach for automatic identification and characterization of dysgraphia in third-grade children is explained in this work. The technique is based on analyzing the childswriting direction by sampling the pressure the pen apply on the paper as well as the pens position and adjustment by using a standard digital writing pad. Ninety-nine samples were collected from writers with dysgraphia and proficient writers in this work. A wide range of features covering dynamic assets of the writing and typographic (i.e., visual) assets were extracted for each participant[10].

## II. PROPOSED METHODOLOGY

This proposed system is works handwritten image for segmentation of words. Firstly preprocessed the input handwritten image i.e. image grayscale, image noise removal, image edge detection. Also then image thresholding and image morphology can be done. After that in image segmentation, firstly the segmentation of words and then shows word spacing and margin results. The segmented words are then stores in final result. Using the word spacing and margin result we predict the human personality.

### Steps:

1. Read Input image
2. Convert image to gray scale image
3. Remove noise of image
4. Convert image to black and white i.e. threshold is to convert image to change foreground pixels as background pixel and viceversa
6. Image Morphology
5. Apply bounding box on segmented words
6. Display the related result i.e. display the word spacing and margin of words to predict the personality.

### Classification:

Classification is based on following point

1. Classification using spacing between line structures.
2. Classification using spacing between word structures
3. Classification using page margin structure.

### Advantages of Proposed System:

1. System takes minimum time to segmentation.
2. It also helps you to identify the hidden talents and aptitudes, thus, helping you to choose a right career path for yourself.
3. Handwriting analysis plays an important role in motivating employees to get better performance. It also helps a manager to understand the needs of his employees and work on them. It can revise teamwork by helping the manager to understand individual team strengths and weaknesses.
4. Proposed approach work on segmentation of words, also proposed system give word spacing and margin results.

### A. Architecture

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 8, August 2019

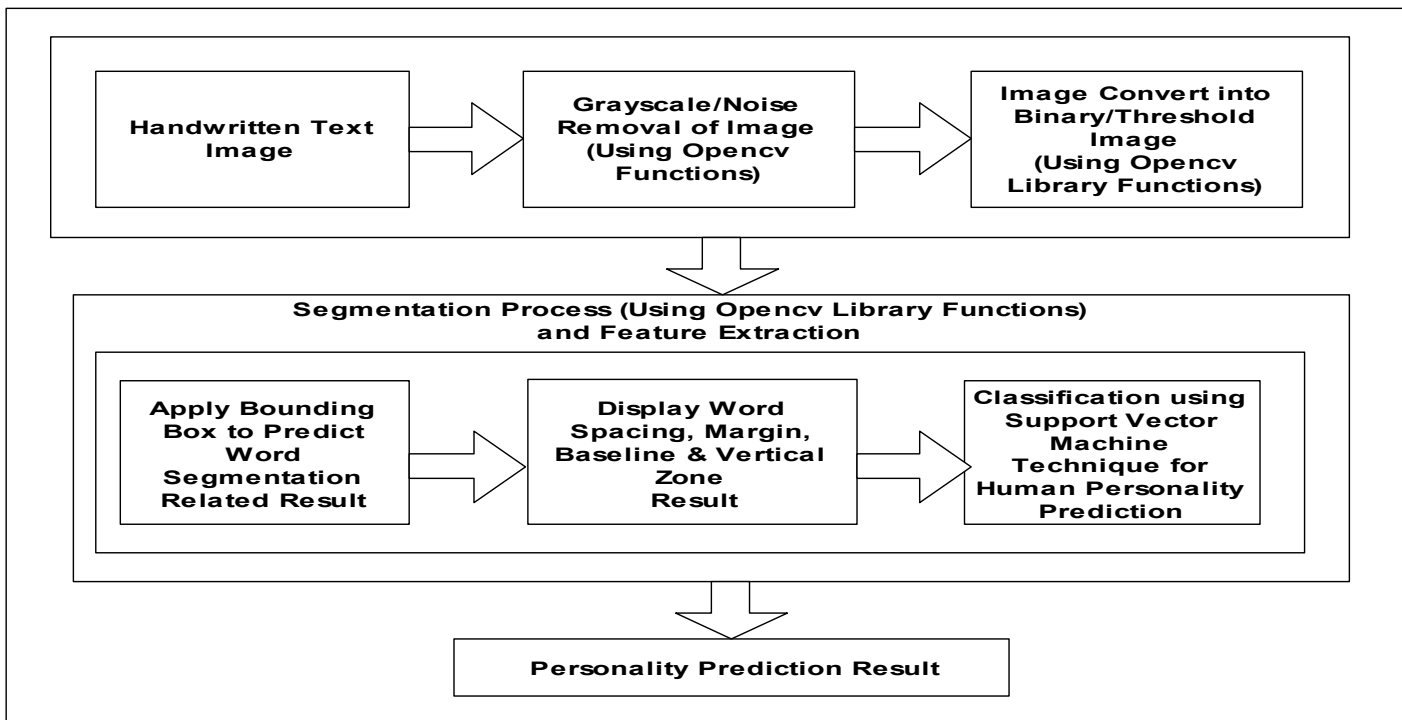


Fig. 1. Proposed System Architecture

## B. Algorithms

### A.1. Algorithm for Segmentation of Lines and words:

First step of segmentation process is segmenting the text region into lines, also called as line segmentation. First we need to calculate header lines and base lines for the Line segmentation. Header lines are rows with maximum number of black pixels and base lines are rows with minimum number of blackpixels.

#### Perform the following steps:

- 1) Find out the rows with minimum number of pixels and replace that row pixels with black pixels .So that we can separate text lines with blackrows.
- 2) Find out the height of the text lines using those blackrows.
- 3) Store the heights of lines in array and use sorting technique to sort the elements and take median as average height of the lines.
- 4) By using average line height calculate minimum height of the consonant in aline.  
Minimum height of the consonant = Average line height / 4.

#### Word Segmentation:

After lines are segmented from text, words are segmented from lines by vertical projection profile.

For each column of the line the number of black pixels is counted and the columns with zero black pixels are used as delimiters for wordseparation.To distinguish the character separation from the word separation, we have selected the delimiter as at least three continuous columns with zero black pixels for word separation.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 8, August 2019

## A.2. Support Vector Machine (SVM) Algorithm:

Support vector machine is supervised learning algorithm which can be used for both regression and classification.

In this algorithm, system plots each data item as point in  $n$ -dimensional space with the value of a particular coordinate. After that classification by finding the hyper-plane that differentiates the two classes very well. Support Vectors are simply the coordinates of individual observation. Support

vectors are simply the best segregates the two classes.

SVM is one of the best known methods in pattern classification and image classification. It is designed to separate a set of training images into two different classes,

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  where  $x_i$  in  $R^d$ ,  $d$ -

dimensional feature space, and  $y_i$  in  $\{-1, +1\}$ , the class label, with  $i=1..n$  [1]. SVM builds the optimal separating hyper planes based on a kernel function ( $K$ ). All images, of which feature vector lies on one side of the hyper plane, are belong to class  $-1$  and the others belong to class  $+1$ .

## C. Mathematical Model

### A.3. Algorithm for Segmentation of Lines and words:

We make following assumptions about the data:

1. The minimum height of character consonant in a line is eight pixels. Average line height is 30 pixels.
2. The skew in a text is not more than the height of a character consonant.

The algorithm for line segmentation has following mathematical steps:

Step 1: Initially, rough estimate of the header lines in whole of the text are made by the formula

$$pcol(i) > 15pcol(i) > pcol(i+8) \quad pcol(i+8) > 0$$

$$pcol(i) > floor(width(i)/7)$$

Where,

$pcol(i)$  : No. of pixel in row  $i$

Width ( $i$ ) : width of line  $i$  i.e. difference between last pixel and first pixel position of line  $i$ .

Step 2: After finding first header line, we skip 8 rows (equal to minimum height of consonant) to find the next header line. Step 3: From  $(i+8)^{th}$  row to  $(i+22)^{th}$  row, we find the  $m^{th}$  row with minimum of pixels.

Step 4: We skip the rows up to  $m^{th}$  row and go to step 1 to find the next headerline.

## 2. Mathematical Equations of Support Vector Machine:

We have  $k$  sub-spaces so that there are  $k$  classification results of sub-space, called  $CLSS_1, CLSS_2, \dots, CLSS_k$ . Thus the problem is how to integrate all of those results. The simple integrating way is to calculate the mean value:

$$CL = \frac{1}{k} \sum_{i=1}^k SS_i$$

Or weighted mean value:

$$CL = \sum_{i=1}^k W_i SS_i$$

Where  $W_i$  is the weight of classification result of subspace,  $SS_i$  and satisfies:

$$R$$

$$i=1$$

$$W_i = 1$$

The centroid is calculated as follows:

$$\bar{X} = \frac{\sum_{i=1}^k X_i}{k}$$

$$\bar{Y} = \frac{\sum_{i=1}^k Y_i}{k}$$

Where  $(\bar{X}, \bar{Y})$  represents the centroid of the hand,  $X_i$  and  $Y_i$  are x and y coordinates of the  $i^{th}$  pixel in the hand region and  $k$  denotes the number of pixels that represent only the hand portion.

In the next step, the distance between the centroid and the finger tip was calculated. For distance, the following Euclidean distance was used:

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Where  $(x_1, y_1)$  and  $(x_2, y_2)$  represent the two co-ordinate values.

### III. RESULT AND DISCUSSION

Experiments are done by a personal computer with configuration: Intel (R) Core (TM) i5-6700HQ CPU @ 2.60GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and Jdk 1.8. The application is desktop application used tool for design code in Eclipse. Some functions used in the algorithm are provided by list of jars like Opencv2.4.7.

Experiments for word segmentation are carried out with images which are scanned from the handwritten Data set. Handwritten data is collected from the 30 different writers from various backgrounds such as students, teachers, doctors, lawyers and business men. Data of different sizes, i.e. writers handwriting is different so the size of characters varies from 15-60 pixels and different resolutions. As data is collected from the people from various backgrounds.

Following result shows the image noise removal, image grayscale conversion, image thresholding, image morphology and word segmentation of proposed approach using english handwritten. So our work gives the better accuracy as compared to existing work.

Depending on this result will find the features contains wordspacing, line spacing, top margin, bottommargin, rightmargin, left margin, baseline, vertical zone etc.



Fig. 2. Noise Removal Image

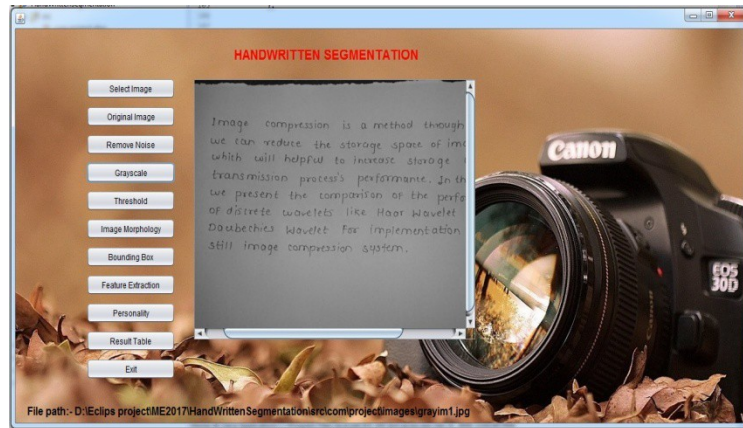


Fig. 3. Image Grayscale

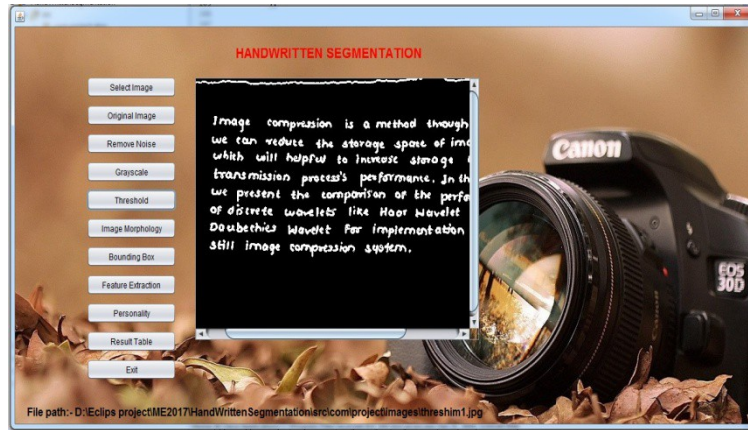
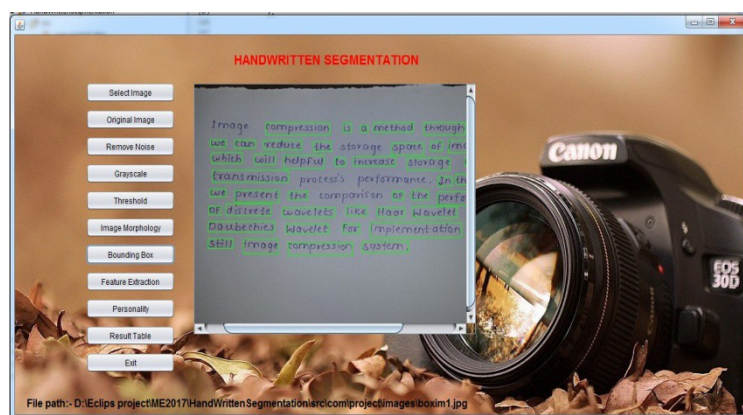


Fig. 4. Image Thresholding

Fig. 5. Image Morphology



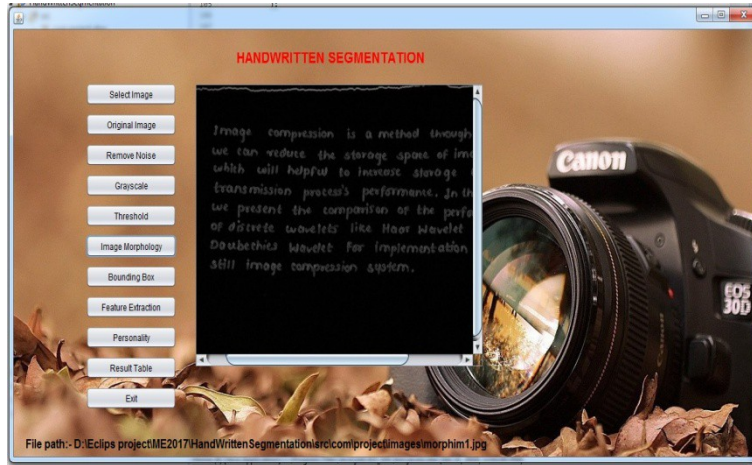


Fig. 6. Word Segmentation

#### IV. CONCLUSION

In this work, we have presented a primary work for segmentation of words of English Manuscript i.e. handwritten English script. We achieved successful segmentation result in word segmentation. It is difficult to identify exact connecting points for segmentation of upper and lower modifier, separate-inganuswara (.), full stop (.) from noise is critical as both resemble the same. After segmentation we have display the word spacing and word margin related result to predict the personality.

#### REFERENCES

- [1] ShubhraSaxena, P. C. Gupta, A Novel Approach of Handwritten De- vanagari Character Recognition through Feed Forward Back Propagation NeuralNetwork.
- [2] SnehaS.Patwardhan, R.R. Deshmukh, A Review on Offline Handwritten Recognition of DevnagariScript.
- [3] NehaSahu,R.K.Rathy,InduKashyap,SurveyandAnalysisofDevnagari CharacterRecognitionTechniquesusingNeuralNetworks.
- [4] SaiprakashPalakollu, RenuDhir, Rajneesh Rani, Handwritten Hindi Text SegmentationTechniquesforLinesandCharacters.
- [5] Naresh Kumar Garg, LakhwinderKaur, M. K. Jindal, Segmentation of Handwritten HindiText.
- [6] Vikas J Dongre, Vijay H Mankar, Devnagari Document Segmentation Using HistogramApproach.
- [7] ParthaPratimRoy, Umapada Pal, JosepLlads, Morphology Based Hand- written Line Segmentation Using Foreground and Background Information.
- [8] Ratnashil N Khobragade, Dr. Nitin A. Koli, Mahendra S Makesar, ASurveyonRecognitionofDevnagariScript.
- [9] Miguel A. Ferrer, Moises Diaz, Cristina Carmona-Duarte and AythamiMorales, A Behavioral Handwriting Model for Static and Dynamic SignatureSynthesis.
- [10] Sara Rosenblum and Gideon Dror, Identifying Developmental Dys- graphia Characteristics Utilizing Handwriting ClassificationMethods.