



Identification of Personality Trait by Handwriting Analysis Using SVM Classifier

Syeda Asra¹, Dr.Shubhangi DC

¹Department of Computer Science & Engineering APPA IET, Kalaburagi

²Department of PG Studies CSE VTU Regional Centre, Kalaburagi.

Abstract

Personality trait identification is of paramount importance. Personality Traits are stable over time. Individual's behavior naturally varies somewhat from occasion to occasion but there is a core of consistency which defines the true nature. Traits directly influence behavior. There are several methods for identifying traits. drawing out ample avenue for trait identification through hand writing and Psychometric analysis. Handwriting reveals the true personality including emotional outlay, fears, honesty, defenses and many others. A big five factor model is considered to find out individual differences. To identify person's extrovert introvert behavior based on spacing between words and characters in the handwritten text has been considered in this paper. The classifier SVM is being used. 90% accuracy is achieved.

Keywords: ANN, personality trait, human behavior.

1. INTRODUCTION

Human behavior [1][2] speak of the variety of behaviors revealed by humans and which are shaped by culture, attitudes, emotions, values, ethics, authority, rapport, hypnosis, coercion and genetics. Human behavior is faced and felt during a person's entire lifetime It includes the way they behave built on diverse aspects such as genetics, social norms, core faith, and attitude. Behaviour is impacted by firm traits each individual has. The traits differ from person to person and can yield altered actions or behaviour. Generically, notions of personality trait [2] make two crucial assumptions. These do not change over time. Most people agree that an individuals' behaviour varies slightly case by case, but would uphold the basic permanency which defines individual's factual nature like the unchangeable stripes of a zebra. Social models also influence behaviour. In 1936, psychologist Gordon Allport [3] categorized these traits into three levels. Cardinal traits dominate entire life time of an individual to the extent that a person is identified by this trait. Central traits shape the fundamental foundations of personality. These are found in varying degrees in every person such as loyalty, kindness, friendliness, intelligent, honest, shy and anxious are deliberated as central traits. Secondary traits are associated to attitudes and emerge only in specific circumstances. Some instances would be getting restless standing in a queue either in a bank or any public places .Getting nervous during a presentation etc. The system developed in this work is to identify the central traits.

2. REVIEW OF LITERATURE SURVEY

Gerrald Matthews et .al [1], states that traits directly influence behavior. He elaborates that, a person spontaneously breaks into cheerful song; the behavior is explained by saying that he or she has a happy disposition.

Jennifer Dodorico [2], proposed that, there are potential weaknesses in using self-reports to measure psychological constructs of personality, like structure of the question affects the reported information, whether it accurately measures the construct or not. Author further says that, self-reports are a fallible source of data, and minor changes in question wording, question format, or question context can result in major changes in the obtained results. He recognized that self-reports leave a lot of room for 'response biases', and involve "a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content what the items were designed to measure. For example, people often respond in such a way that presents them in a more favorable light, even if these responses do not reflect how they actually think or behave and lack of credibility due to biased responding is a major issue because it could impede the validity of the self-report as a measure.



Susan J. Stabile [3], states in her paper that, the costs of making bad hiring decisions and the difficulties of getting meaningful information from reference checks of prospective employees have led many employers to use personality tests as part of their hiring process. Employers choose from a wide variety of tests in an effort to both weed out job candidates with undesirable traits, such as dishonesty, or tendencies toward violence or tardiness, and to judge the "fit" between the prospective employee and the job by seeking to identify prospective employees possessing personality traits likely to predict success in the job in question. Employers swear by them and are convinced that they are hiring better workers as a result of their use. Further the author says that, the widespread use of personality tests as a means of determining which employees to hire raises a number of issues, ranging from the validity and reliability of the tests to concerns about invasion of privacy and discrimination against minorities. These issues raise the question whether the benefits of personality tests outweigh the costs of employing them.

Eric Garner [4], the author states that trying to understand the people's inner motivations is notoriously an uncertain science and further says that, personality type helps managers to know what motivates an individual. Personality gives clues to things a person loves to do and it is a clue to predicting how people will behave in certain situations such as stress, when coping with change and making important decisions. Personality types tell a manager how individuals will fulfill role, for example in teams, their unique way of leading others and their style of learning. Personality type is the key to the mystery of why people hit it off with others at work and why others have personality clashes.

Shitala Prasad et al [5] states that ,handwriting analysis is a method to predict personality of an author and to better understand the writer. He has used Allograph and allograph combination analysis as a scientific method of writer identification and evaluating the behavior. To make this computerized he considered six main different types of features: (i) size of letters, (ii) slant of letters and words, (iii) baseline, (iv) pen pressure, (v) spacing between letters and (vi) spacing between words in a document to identify the personality of the writer. Segmentation is used to calculate the features from digital handwriting and is trained to SVM which outputs the behavior of the writer. For this experiment 100 different writers were used for different handwriting data samples. The proposed method gives about 94% of accuracy rate with RBF kernel. In this paper an automatic method has been proposed to predict the psychological personality of the writer.

Rafael Gonzalez [6] , has described the steps of image processing. These steps form a fundamental level for each image processing related research work.

Beyerstein, Barry [7], says in this definitive study of graphology, the authors provide an evaluation of claims that personality, aptitude, and psychological and physical health can be determined through handwriting analysis. The contributors include both practicing graphologists, who present their case and describe their practices and methods, as well as critics from many fields, who evaluate graphology in terms of brain research; assess its accuracy through objective tests of validity; contrast graphology with verifiable psychological assessment techniques; and review the legality of using graphology in employee selection, psychological diagnosis, and the criminal justice system. He says that, graphology seems to be too accurate, too many personnel managers.

Charles H.Trafford et al [8], the authors in their book states that, the brain function research has proven the physical production process of making symbols is very potent. The Cursive hand writing leads to improved fluency skills that benefit the children in all written communication skills. The authors emphasizes cursive handwriting as the most ergonomic writing process and advocates that that teaching a young child ,cursive handwriting teaches the motor control process which is called anchoring. The authors say that the cursive is more relaxing because that motor pattern relies on opening muscles for 4 out of 6 legibility sub skills. Also the engineering of cursive lower case letters make them fit the muscles of the arm, wrist hand and fingers.

Brinda Petty [9], a certified hand writing expert says, when one prints, more of the conscious mind is revealed and less of the unconscious. When one writes in cursive, it is the subconscious mind that is revealing the most valuable information. For this reason, professional handwriting analysts prefer to work with cursive handwriting vs. block printing. She says based on her experience, cursive handwriting reveals 6 times more valuable personality information than printing alone.

Ruby Ernica Samy [10] says, compared with the bulk of the rest of a sample of writing, a lonely little "i" dot seems very insignificant. "i"-dots reveal a great deal. Their placement, shape, pressure, all of the same size. All these things

have to be considered. First, where the “i”-dot is placed, in relation to the rest of the letter “i.” The closer it is to the top of the “i,” the more attention to detail is shown.

Anand Kumar *et.al* [11], this paper proposes an automated method to predict the personality of an individual by his/her handwriting sample using SVM machine learning algorithm. SVM takes the size of the letters, pen pressure, baseline, letter spacing and word spacing and the most important slant of the letter and word in a document as an input and gives their psychological behaviour as the output for each individual writer. The various parameters are calculated by simple use of trigonometry and thresholding techniques .Further he states that it is a real-time handwriting analysis system which is writer independent. For future, the features can be increased for more correct results such as page margins, arcade, garland, thread, etc.

D. John Antony, O.F.M. Cap [12] writes in his book that people write small letters for various reasons. First of all, the person wants to be a hermit withdrawn from the external world. Secondly, the person writes small letters for concentration. As one concentrates on what he/she writes, the movements of the hand and fingers are restricted and so the writing becomes small. Thirdly, it may be due to their low self-image that people write small letters. Graphologists say that between 5 and 10 % of people on account of their inferior feeling write small letters. Their sense of inadequacy in the world and life makes them write small letters. They feel as small as their letters appear to be. Yet another reason for writing small letters is due to the sense of economy. These people want to save paper. They may tend to be economical in all the facets of their life. It may be also due to the limited space provided. He says that, Scientists, composers, authors, and mathematicians write small-sized letters most of the time. Their profession requires deep concentration on account of which the letters become small. When the writing is small and well formed, it means that the person can concentrate on small details for a long time as well.

Bernardete , Noel Joao [13], proposed a system of signature identification using graphics processing units based SVM classifier. This was used to identify the fake signatures from genuine ones .It employs a GPU- based SVM classifier using Multi-Threaded parallel CPU standalone SVM version. This system provides excellent result on the identification of an individual’s signature both in terms of computational cost and performance.

3. METHODOLOGY

The work has been divided into two phases testing and training as shown in the figure 1. Initially work begins with the pre-processing of the given input image by performing color conversion as RGB to gray scale. The training data set consists of images of words of any length, complete or incomplete left or right inclined and also word with no inclination.

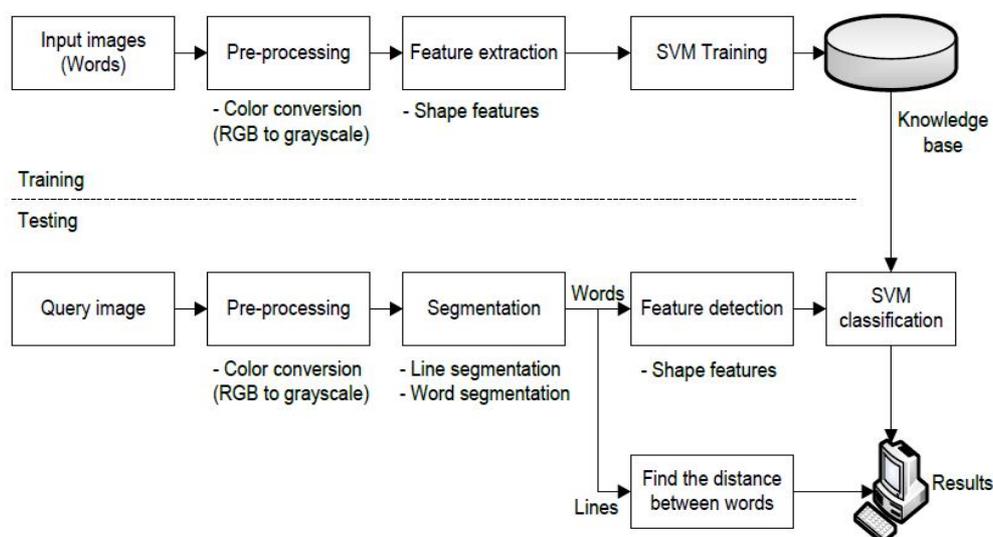


Figure 1 Architecture of Proposed System

3.1 Shape Features

The shape features mentioned are obtained on binary image utilizing MATLAB inbuilt function regionprops. Initially color images are converted to gray scale images and then into binary images by using MATLAB built-in functions rgb2gray and im2bw. Rgb2gray changes RGB values to grayscale values via forming weighted sum of the R, G, and B components: $0.2989 * R + 0.5870 * G + 0.1140 * B$ im2bw changes the grayscale image I to a binary image. The output image BW will replaces all pixels for input image by means of luminance greater than level with the value 1 (white) as well as it replaces every other pixels with the value 0 (black). There will be specification level in the range [0,1]. The range is closer to signal levels that are likely for image's class. As a result, a level value of 0.5 is in the middle among black and white, despite of class. We may take help of the function graythresh to compute the level argument. If we not specify level, im2bw uses the value 0.5.

Area – Actual number of valid (pixel value=1) pixels in the region.

Perimeter –Perimeter is the distance around the boundary for region. Function regionprops will find the perimeter through calculating distance among each adjoining pair of pixels around border of the region. When the image has any discontinuous regions, regionprops will give unexpected results.

Form factor – is computed by utilizing Area and Perimeter. $Formfactor = \frac{4 \times \pi \times Area}{Perimeter^2}$

Major axis length – This will give the length for major axis of the ellipse which has same normalized second central moments as the region. The 2-D input label matrices will only support this property.

Minor axis length – This will give the length of the minor axis of the ellipse which have the same normalized second central moments as the region. The 2-D input label matrices will only support this property.

Roundness – is computed using Area and Major axis length.

$$Roundness = \frac{4 \times Area}{\pi \times Major\ axis\ length^2} \tag{1}$$

Compactness is computed as

$$Compactness = \frac{\sqrt{4 \times Area / \pi}}{Major\ axis\ length} \tag{2}$$

Density – is formulated by utilizing BoundingBox. A smallest rectangle which contains the region can be defined as BoundingBox, a 1-by-Q *2 vector, where Q is the number of image dimensions.

Black pixels each line – Considering each row and each column of the binary image the number of white pixels can be computed. And inverting this value with the width of the image gives us the number of black pixels present in each line. By efficient utilization of content-based image features such as shape features in our proposed system, the semantic gap can be minimized.

3.2 Log Gabor Filter

Filters are very helpful to code natural images better which will have Gaussian transfer functions when they are looked on the logarithmic frequency scale. On linear frequency scale log-Gabor function will have a transfer function which is of the form

$$G(w) = e^{(-\log(w/w_0)^2) / (2 (\log(k/w_0))^2)} \tag{3}$$

3.3 Training

Images provided in this phase contain only words. Words can be of any length, complete or incomplete, left or right inclined. Images are arranged in such a way that every folder contain image of unique character, for e.g. one folder contain complete words with left inclined, another folder containing incomplete words with left inclined, contain

complete words with right inclined, incomplete words with right inclined, and another folder containing complete words with no inclination. Colour conversion of all the images from RGB color to grayscale is converted in pre-processing step. A grayscale digital image is one in which the values for all pixel is an only model, which means that it carry simply intensity particulars. This kind of images are even called to be black-and-white, they are made up of totally gray shades. Following this the shape features of all these images are extracted and trained using SVM algorithm. And these feature vectors are stored in database known as Knowledge base. The extracted features are predictable to have the related information from the input data, so that the preferred work can be performed by means of this summary demonstration instead of the complete initial information.

3.4 Testing

Query image is provided to be input for system. RGB image is then changed to grayscale in pre-processing step. Then segmentation is performed to separate each line from other line in a single image or document. Segmentation[4] is performed using morphological operations dilation, erosion, opening, and closing on binary image. In mathematical morphology, dilation is one of the basic operations. Its operation basically makes use a structuring element for searching as well as increasing the dimension present for given input image. This helps to know the accurate result by dilation for given input image . After having the this process each segmented line is used to separate each word to evaluate the space between words. Following the line segmentation cropping of the line is done, cropping helps to get a rectangular area of interest through given original image and later on word segmentation is performed. Distance between words is calculated using MATLAB[5] built-in function (regionprops, mainly Bounding box parameter). By calculating the distance between words the prediction of behaviour[6] [7]is made. Then similar to training phase, shape features of each word is extracted, and these features are classified using SVM classification step by comparing the features with knowledge base. Finally results are displayed, for e.g. with regard to document “Left inclined, complete words”, “Left inclined, incomplete words”, “Right inclined, complete words”, “Right inclined, incomplete words”, and “No inclination in words” as well as less or more space among words. The inclination shows emotional indicators of the individual as an behaviour, right inclination shows the writer reacts strongly to emotional situations and left inclination predicts the writer has a balanced emotional state. Spacing between the words is calculated so as to indicate the social attitude to others of an individual, close space between the words indicates a symbol of cooperation and large gap through words shows an individual being restful alone, as well as they might even disbelieve others. And finally the complete and incomplete words in the handwriting indicates that an individual wish to impart information and consideration for others. With incomplete words that can deep up promises last till the end same energy till the work gets accomplished. This is accomplished by handwriting analysis[8][9][10][11][12][13]. Classification is done by means of SVM[14] classifier.

4. EXPERIMENTAL RESULTS

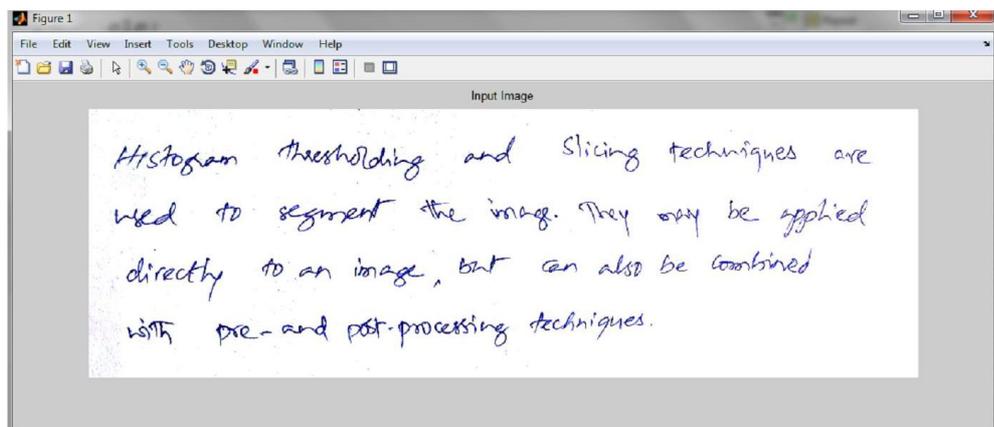


Figure 2 Input Image

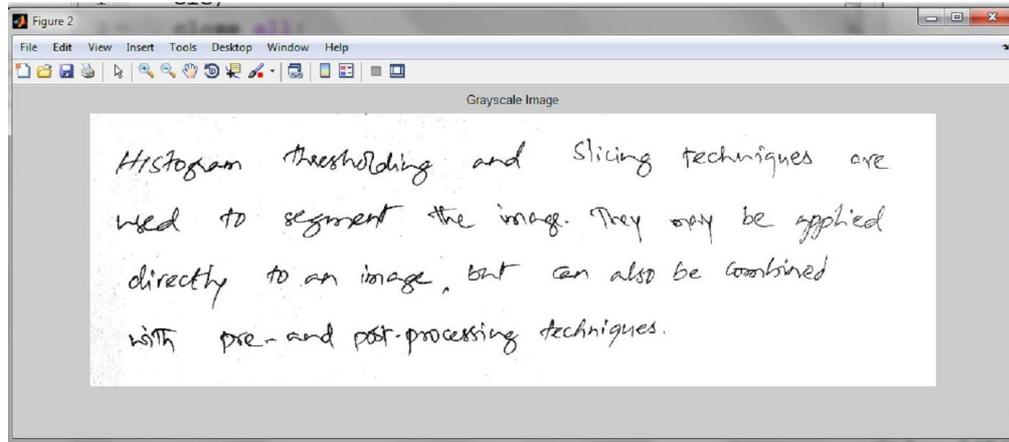


Figure 3 Gray scale Image

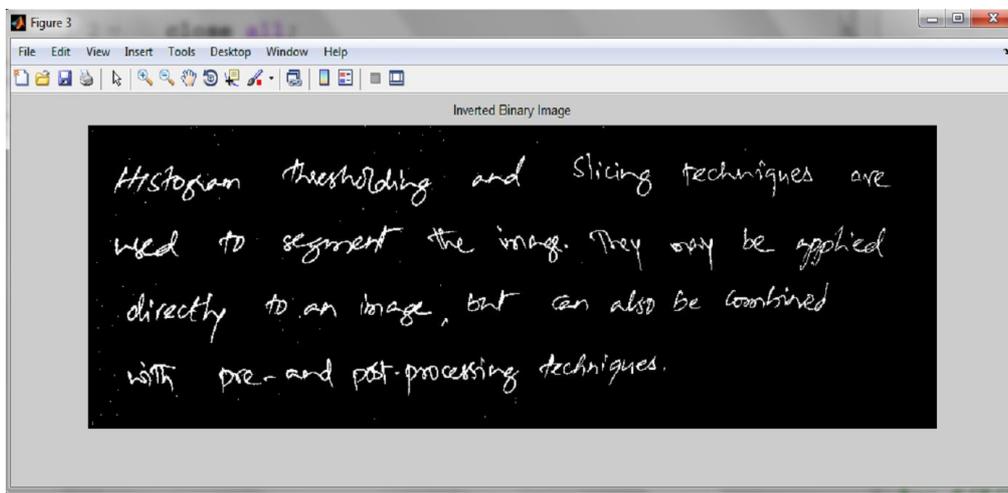


Figure 4 Inverted Binary Image

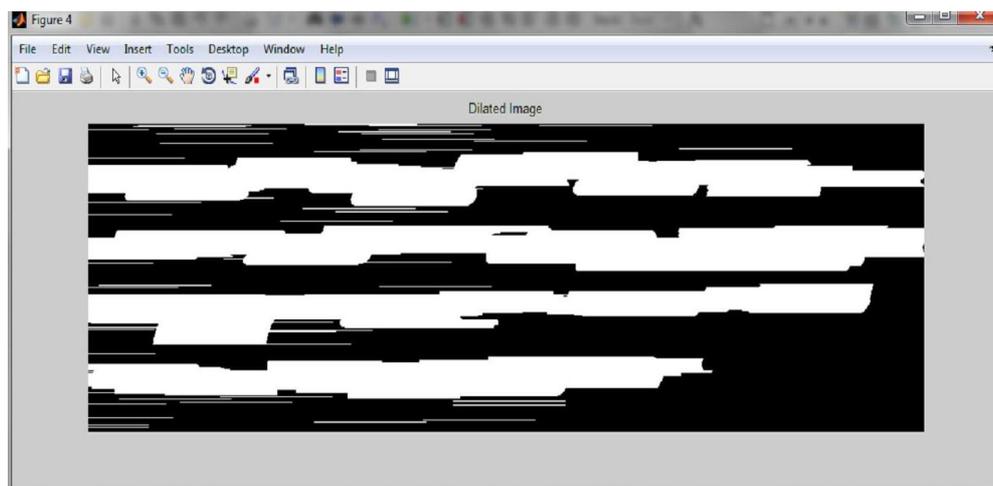


Figure 5 Inverted Binary Image

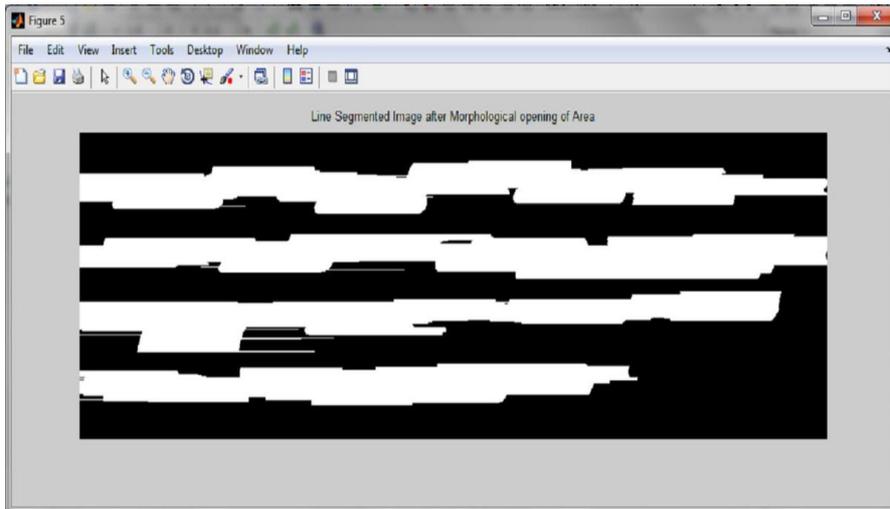


Figure 6 Line Segmentation

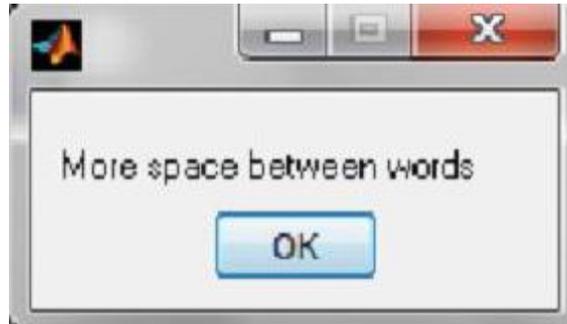


Figure 7 Results

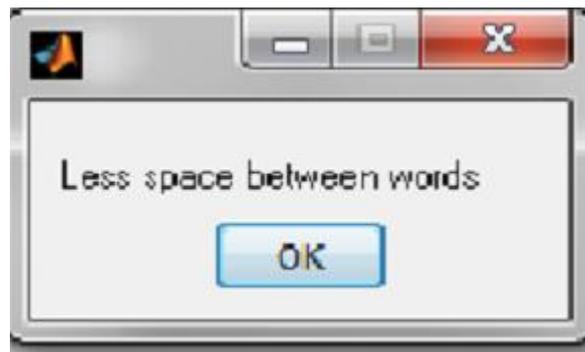


Figure 8 Results

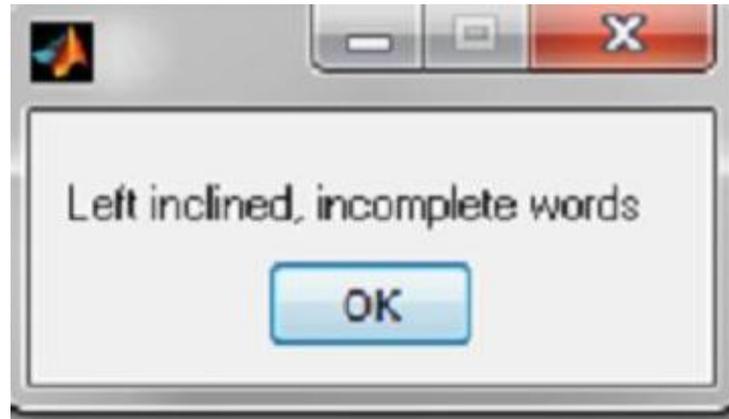


Figure 9 Results

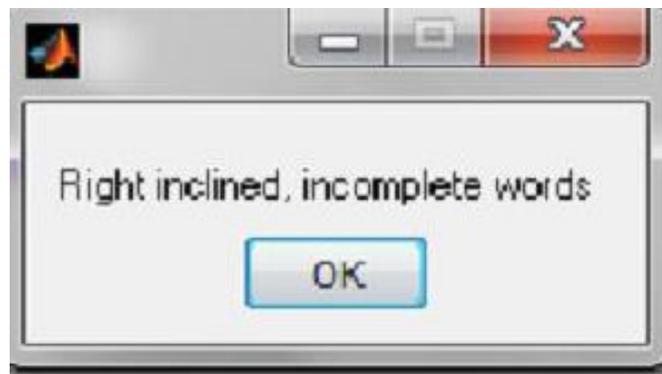


Figure 10 Results

Table 1: Handwriting analysis and psychological interpretation

Characteristic	Interpretation
More space between words	Isolated personality trait
Less space between words	Exhibits Socialite personality
Left inclined incomplete words	Lives in the past and cannot sustain energy for a long time to accomplish the task
Right inclined incomplete words	Lives in the future and cannot sustain energy for a long time to accomplish the task

5. CONCLUSION & FUTURE WORK

The work carried out, is a fusion of two domains, psychology and computer science engineering. Basically, the problem is identified in the field of psychology and realized in computer science engineering. Human behaviour is experienced throughout the life time of an individual and hence behaviours are influenced by traits. Understanding the personality trait is a complex phenomenon. They are unchangeable. In any circumstances the same set of behaviour/traits are revealed with few plus minus. The identification of these traits are of paramount, since these help in analysing handwriting. The analysis can be done by graphology. Later, once some specific patterns of handwritings are selected, that reveals personality trait of an individual, then by using image processing techniques the patterns can be feature extracted, de-noised and segmented followed by recognition using SVM Classifier or ANN. The Co-



relation is indicated between the pattern and the psychological characteristic. Hence this work could be used to understand person, holistically and provide ways for improvising the personality or human behaviour through handwriting to lead a better life professionally, personally, socially and hence build a better and a peaceful nation.

References

- [1] Gerrald Matthews, Ian J Deary, Martha Whiteman "Personality Traits" second edition 2003.
- [2] Jennifer Dodorico McDonald University of Cambridge" Measuring Personality Constructs the Advantages and Disadvantages of Self-Reports, Informant Reports and Behavioral Assessments" ENQUIRE Volume 1, Issue 1, June 2008.
- [3] Susan J. Stabile" The Use of Personality Tests As A Hiring Tool: Is The Benefit worth the cost? U Pa. Journal of Labor and Employment law vol.4 2002.
- [4] Shitala Prasad, Vivek Kumar Singh Akshay Sapre "Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using Support Vector Machine ,IJCA (0975 – 8887) Volume 8– No.12, October 2010
- [5] Eric Garner "Understanding Personality Types: Managing People Through Their Personality Types" ISBN 978-87-403-0000-0 2012.
- [6] Shitala Prasad, Vivek Kumar Singh Akshay Sapre "Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using Support Vector Machine ,IJCA (0975 – 8887) Volume 8– No.12, October 2010.
- [7] Rafael Gonzalez"Digital Image Processing"3rd Edition 2006.
- [8] Beyerstein, Barry L. (Ed),Beyerstein, Dale F. (Ed) Amherst, NY, US "Evaluations of graphology, the study of handwriting analysis "Prometheus Books. (1992) 515 pp
- [9] Charles H.Trafford and Rand H.Nelson , Peterson Directed Handwriting "Handwriting A Complete Guide To Instruction Teaching Physical Patterns For Reading and Writing Fluency"ISBN 1-890666-3 5-1, 2003.
- [10]Brenda Petty "Handwriting Analysis Training Site,The Online Community for Handwriting Analysis World Wide, Why Smart People Still Use Handwriting Analysis A Weekly News Letter, 2013. "
- [11]Ruby Ernica" Meaning of i dot in Hand Writing Analysis" Sciences 360 Published in 2010.
- [12]Champa H. N., Dr. K. R. Ananda Kumar, "Artificial Neural Network for Human Behaviour Prediction through Handwriting Analysis", International Journal of Computer Applications (0975-8887) Volume 2 – No.2, May 2010.
- [13]D. John Antony, O.F.M. Cap." Personality Profile through Handwriting Analysis" A Textbook of Handwriting Analysis, 2008.
- [14]Bernardete Ribeiro Noel Lopes and Jo~ao Gonc,alves," Signature Identification via Efficient Feature Selection and GPU-based SVM Classifier" International Joint Conference on Neural Networks (IJCNN) ,2014A. Bonnaccorsi, "On the Relationship between Firm Size and Export Intensity," Journal of International Business Studies, XXIII (4), pp. 605-635, 1992. (journal style)

AUTHOR

Syeda Asra received the B.E. and M.Tech. degrees in Electronic & Instrumentation Engineering and computer science & Engineering from PDA College of Engineering & KBN College of Engineering in 1999 and 2006, respectively. Currently working as an Associate professor at Appa IET, Gulbarga in the department of Computer Science & Engineering. Pursuing Ph.D. a research work in VTU since 2012 in Image processing.