Development of an automated handwriting analysis system

Article *in* Journal of Engineering and Applied Sciences · September 2011

CITATIONS
15

READS
3,158

5 authors, including:

Navin Karanth
National Institute of Technology Karnataka
11 PUBLICATIONS **51** CITATIONS
SEE PROFILE

Vijay Desai
National Institute of Technology Karnataka
75 PUBLICATIONS **204** CITATIONS
SEE PROFILE

S. M. Kulkarni
National Institute of Technology Karnataka
129 PUBLICATIONS **1,176** CITATIONS
SEE PROFILE

VOL. 6, NO. 9, SEPTEMBER 2011 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

DEVELOPMENT OF AN AUTOMATED HANDWRITING ANALYSIS SYSTEM

Vikram Kamath, Nikhil Ramaswamy, P. Navin Karanth, Vijay Desai and S. M. Kulkarni
Mechanical Engineering Department, National Institute of Technology Karnataka, Surathkal, Srinivasanagar, Mangalore, India
E-Mail: nrghatta@gmail.com

ABSTRACT

In the present study a method has been proposed for the behavioral prediction of a person through automated handwriting analysis. The present work identifies the psychological traits in the writing namely size, slant and pressure, baseline, number of breaks, margins, speed of writing and spacing between the words. The handwriting is analyzed through Image Processing in MATLAB. The behavioral pattern of the person is predicted from the above traits of the handwriting. The developed system identifies handwriting closely which may not be possible for a graphologist. It is real time and involves less image preprocessing. The proposed system is calibrated with manual analysis. The results obtained through the system are in good agreement to more than 80 percent of the cases with ideal manual analysis.

Keywords: handwriting analysis, image processing, behavioral prediction, MATLAB, automated.

1. INTRODUCTION

Handwriting is often referred to as indication of personality trait represented by neurological patterns in the brain. In other words our brain or subconscious mind actually forms the characters as a result of habit. Handwriting analysis also known as Graphology, which is a pseudoscientific study of handwriting in relation to human psychology. Graphology can be used for identifying, evaluating and understanding personality of a person through the strokes and patterns revealed by handwriting. The main applications of Graphology include behavior analysis, forensic evidence and disease diagnosis. Handwriting reveals the true personality including emotional outlay, fears, honesty, defenses and many other individual personality traits [1]. The authenticity of a person's signature or handwriting in the suicide note is frequently subjected to forensic document examination during investigation in order to determine authorship. [2-4]. In the medical field, it can be used as an aid in diagnosis and tracking of diseases like Parkinson's disease, Alzheimer's disease, and even cancer through Kanfer Test's [5-7].

Handwriting analysis done by a computer is fast, accurate and indentifies the handwriting better than visual inspection. Moreover computer assisted handwriting analysis is automated, efficient and devoid of human errors. Behavioral prediction by handwriting analysis with the aid of a computer has been studied previously by various researchers [9-15].

Methods proposed in literature involve the preliminary process of text extraction from the sample and then application of various algorithms / techniques to determine the characteristic traits. Polygonalization method is one such technique which involves a closed polygon produced around a line in the scanned image of the handwritten text. The slope of the text/alphabet is found using the coordinates of the polygon. Generalized Hough Transform is a second technique used to detect any arbitrary shape in an image by creating a table for storing all the edge pixels of the target shape. Template matching with certain predefined templates is also used as a technique for behavioral analysis [1]. Segmentation method which involves splitting up of the handwriting sample into individual letters is another work available in literature [16]. Methods discussed are not very simple to automate and therefore a need exists for a simple method which could be automated easily.

The present work focuses on development of an automated technique for determining the characteristic traits of a person through Image Processing called AHWAS (Automated Handwriting Analysis System). The proposed work involves lesser image preprocessing of the image as it crops the given sample automatically and uses a RGB filter to extract the text in the handwriting and identifies eight features in the handwriting simultaneously. The features identified are: size of the letters, baseline, pressure of the writing, slant of the handwriting, number of breaks, spacing between the words, margins and speed of the writing in the sample. The system is designed to directly indicate the behavior of the person from the above features. The system can be used in various applications such as detection of diseases like Parkinson's disease or Alzheimer's disease, forensic document examination and lie detection.

2. METHODOLOGY

The flow diagram of the proposed method is shown in Figure-1. It has primarily three steps namely image preprocessing, feature extraction and prediction.

The image is taken from a Nikon coolpix S610 camera. The image is stored in JPEG format. The region of interest is cropped and the image is the input to AHWAS. The region of interest is segregated from the overall handwriting sample after applying a sufficient RGB (Red, Green Blue) threshold. The characteristic traits are quantified by comparing them with extremities.

The cropped image from the handwriting sample is shown in Figure-2. Since blue ball point pen is used for all samples, the threshold for blue is varied slightly for

every sample. The image is subjected to a threshold to highlight the ink image from the background.

3. FEATURE DETECTION BY AHWAS

Once the given image of the handwriting sample is preprocessed, eight characteristic traits of handwriting are determined. Algorithms are developed for detecting the different features in the same way as is done by graphologists. The procedure for extraction of features is explained in detail below.

3.1 Size of the letters

Size of the letter is judged by the vertical height. It is essential to split the sentence into words so that each word can be analyzed individually, to do so; a loop is run to find the number of white pixels between the end of the previous word and beginning of the next word. If the spacing is more than a set threshold value the column indicates the space between the words.

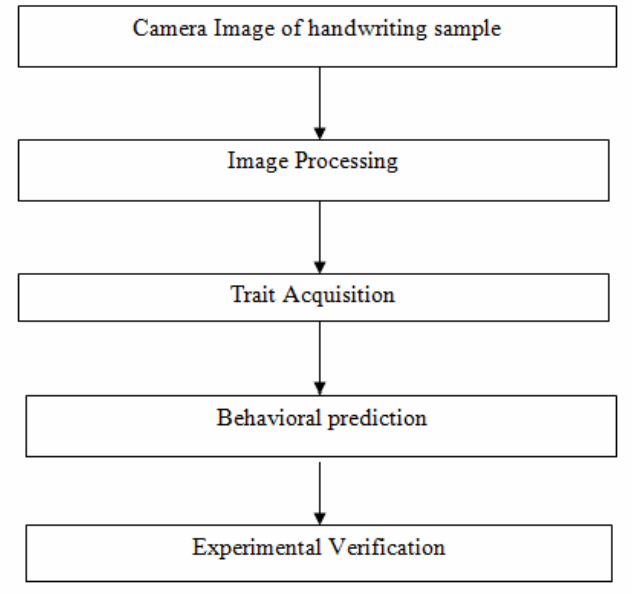

Figure-1. Flow diagram of the proposed system.

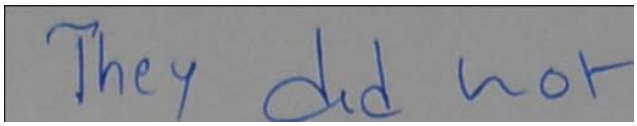

Figure-2. Cropped image from the writing sample.

The array of black pixels for the entire sentence is obtained column wise. A line joining the average of all the highest points and all the lowest points is drawn as shown by the two lines with circles (green) in Figure-3. The distance between the two lines with circles gives the average size of the entire sentence. The same algorithm for determining the size of the sentence is applied for every word which is shown by the two lines in Figure-3. The distance between the two blue lines gives the size of the writing.

3.2 Baseline

The baseline is an imaginary line on which the bottom of the middle zone letters aligns. The lower array of black pixels obtained for every word is joined to give the baseline of the handwriting as shown by the line with rectangles (red) in Figure-3.

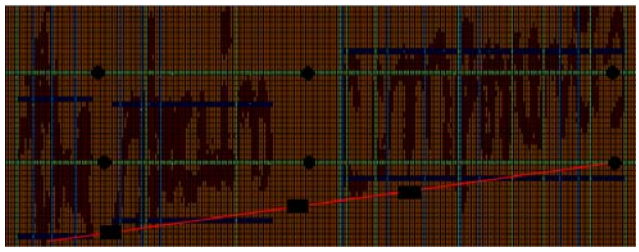

Figure-3. The size and the baseline of the writing.

3.3 Pressure of the writing

Generally it is assumed that the pressure of the writing in the image can be determined by the width of the stroke, this is in line with what graphologists would do when analyzing the handwriting, especially when written from a fountain pen. As explained in the algorithm for size of the writing, a line (green) which cuts across the word is obtained. The number of black pixels at all intersections with the letters is calculated and divided by the number of intersections to give the pressure of the writing; this is shown in Figure-4.

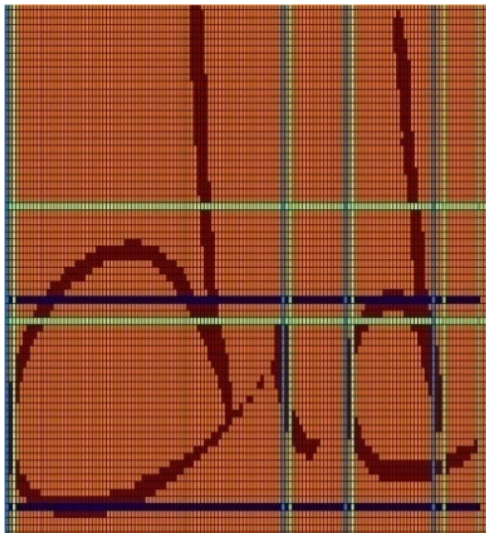

Figure-4. Determining pressure of the writing.

3.4 Slant of the writing

The slant is usually obtained for letters like t, h, d and l which have upper zones. The slant is obtained by joining the highest and lowest point surrounding a given point in the letter, and determining the slant of the line using Equation 1.

$$\theta = \tan^{-1}\frac{y_2 - y_1}{x_2 - x_1} \quad (1)$$

The array of black pixels at the top and bottom of every word is plotted to obtain the curve of word. The highest and lowest black pixel of near every point along the x axis is obtained and connected as shown by the black line in Figure-5. The slope of the line is obtained as an array and this gives the slant at every point in the word. The slope is classified into six subdivisions with angles varying from less than 90^o to 180^o.

The mean of the slopes gives the average slope of the word and the distribution of the slopes in the above sectors gives the erratic movements if any in the writing.

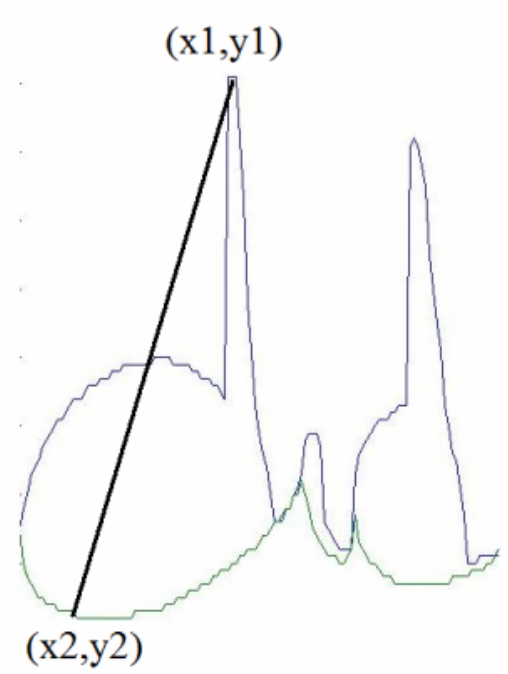

Figure-5. Technique adopted to determine slant of the writing.

3.5 Breaks in the writing

This represents the connectivity within a word in the handwriting. It is determined by blob analysis. The number of breaks in the sentence is found by assessing the number of blobs, which are white patches obtained in the sample. The total number of blobs excluding the number of words gives the number of breaks in the sentence.

Table-1. Handwriting traits and behavioral explanation.

Trait	Explanations	
Size	Small size	Ability to concentrate
	Large size	Ambition, farsightedness
	Medium size	Secure, traditional, realistic
	Variable size	Indecisive, moody
Baseline		Pessimistic, discouragement
		Optimistic, faith in future, joy
		Even temper, reason rules
Pressure	Light	Low determination, forgives rapidly
	Medium	Average level of emotional Intensity
	Heavy	Lasting memory of wrongs
	Variable	Concealing nature, erratic temperament
Slant	Right (BC)	Ruled by judgment
	Left (FA)	Cautious, introverted
	Vertical (AB)	Head controls heart
	Varying	Moodiness, unpredictability
	Right (CD)	Extroverted, future Orientation
	Right (DE)	Feels situations intensely
	Right (E+)	very expressive
Breaks	Connected	Objective, analytical, rational, Logical, compulsive
	Disconnected	Intuitive, sensitive, insecure
Word spacing	Very wide	separation from reality
	Narrow	lack of reserve, thriftiness
	Wide	Mental agility, objectivity
	Even	Consistent, systematic planning
Margins	Wide left	exhibits courage in facing life
	Wide right	Avoids future and is Reserved
	No margins	Insecure, talkative
	Wide upper	Formality, withdrawal
	Wide lower	Aloofness, superficiality
	Even	Self discipline, self Conscious
Speed	Fast	Smart, uncommunicative
	Slow	Lazy, clumsy, dishonest

3.6 Spacing between the words

This handwriting feature in a sample is obtained by the number of pixels between the end of the one word and the start of the next word.

3.7 Margins

This is the amount of space the writer leaves before the start of the writing from the edge of the paper or the amount of space left between the end of the last word and edge of the paper. The margins are obtained by the

space between the starting point of the sentence and the edge of the paper or the last point of the sentence and the edge of the paper.

3.8 Speed of the writing

It is how fast the writer has written the handwriting sample. To effectively asses speed of the writing we use two algorithms one for measuring the ink density and the behavior of stroke length at the last twenty pixels. The ink density is obtained by assessing the number of black pixels inside a box formed by the average upper and lower baselines of the word. This tells how densely packed the letters in the written word are.

The handwriting traits and their corresponding behavioral explanation are shown in Table-1. The category for every trait is explained based on information from [17-20]. It can be seen that for each of the eight traits determined namely size, slant and pressure, baseline, number of breaks, margins, speed of writing and spacing between the words there are variations which are determined by AHWAS to effectively asses the characteristic behavior of the person.

4. TESTING OF AHWAS

The Handwriting samples were taken from 30 people of different ages between 20 to 24, belonging to both genders. The selected groups were all right handed and were physically and mentally sound. They were made to write a given text of approximately 100 words which included almost all the English alphabets, which would ensure better analysis. The sample of the handwriting obtained is shown in Figure-6.

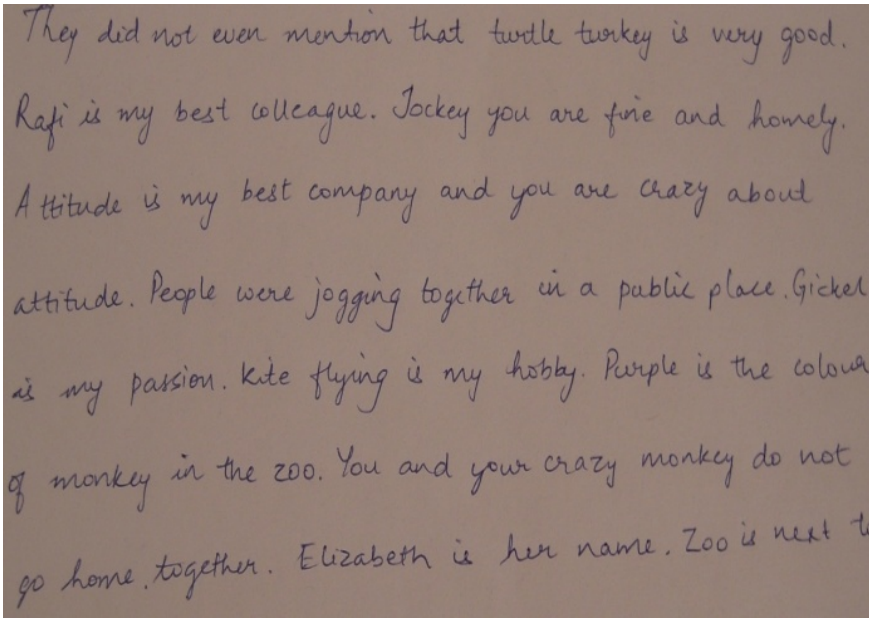

Figure-6. Handwriting sample obtained.

The handwriting samples obtained are analyzed manually by a handwriting analyst and the inferences are made. For the manual handwriting analysis size and slant gauges are used.

The testing is carried out in two steps namely calibration and experimental verification. Features are extracted from the handwriting as explained in the previous section.

The initial 50 percent of the handwriting samples analyzed are used for the calibration of the developed algorithm. For the purpose of calibration of pressure, people were asked to write with the same blue ball point pen on an A4 size sheet. The image of the samples are photographed and stored in digital form.

The primary step involved in calibration of the images to extract the maximum about of the handwriting, is to use a RGB (red green blue) filter to have a clear image of the handwriting. The width of an A4 sheet, which is 8.27 inches, is used for size calibration. This was done by determining the edges of the paper and passing the input to the program

Size of the letters is classified as small, medium and large based on the standard handwriting intervals of less than 1/16" (1.5875×10^{-3} m), between 1/16" - 3/16" (4.7625×10^{-3} m) and greater than 3/16" respectively.

Baseline is determined by the slope of the line connecting the word. A negative slope indicates decreasing baseline and a positive slope indicates increasing baseline. A slope around zero indicates a straight line.

Pressure for the handwriting is calibrated by reducing the width of every stroke to millimeters. They are classified by setting certified pressure sample intervals as thresholds.

Slant for the sentence is classified into FA, AB, BC, CD, DE, E+ as per 90^o, 112^o, 125^o, 135^o and 150^o as the limits for each of the intervals respectively .The mode of the slant in these intervals is taken as the slant for the handwriting.

The number breaks in the sentence is determined by blob analysis; it is as explained in the previous section, so a limit of 5 breaks per word is taken as distinction between connected and disconnected handwriting.

The spacing between words is calibrated by the average width of the word. The spacing is classified as small if the spacing is less than the average width of the word, medium if it is in between average width and double the width of the word and wide if it's greater than double the width of word.

The left and the right margins are calibrated by comparing it with a standard reference [17]. The speed of the writing is calibrated by having a threshold value of 20% for the ink density and decreasing stroke width at the end of the word.

5. RESULTS AND DISCUSSIONS

The program is executed for all the samples of the writing when the calibration was completed and the results are tabulated for comparison with manual analysis. A sample set of results obtained through AHWAS for the handwriting sample shown in Figure-6 are compiled in Table-2. It is observed that the features determined by AHWAS are compared against a reference, which represent the criteria to classify various writing styles, which are selected as explained in the previous section and the trait category determined appropriately.

The calibrated algorithm is then used to analyze the remaining 50 percent of the handwriting samples. The percentage correct detection graph of the characteristic traits obtained is shown in Figure-8. It can be observed that the efficiencies of all the traits obtained are above 80 percent.

Table-2. Results of AHWAS and manual analysis.

Handwriting trait	Features determined by AHWAS	Reference	Trait category
Size	2.356 mm	1/16 " - 3/16"	medium
Baseline	4.67^o	> 0^o	upward
Pressure	0.0225 mm	< 0.025 mm	light
Slant	116^o	112^o -125^o	BC
Breaks	34	>5	disconnected
Word spacing	0.223 mm	> twice width of word	wide
Speed	0.3089	> 20%	fast

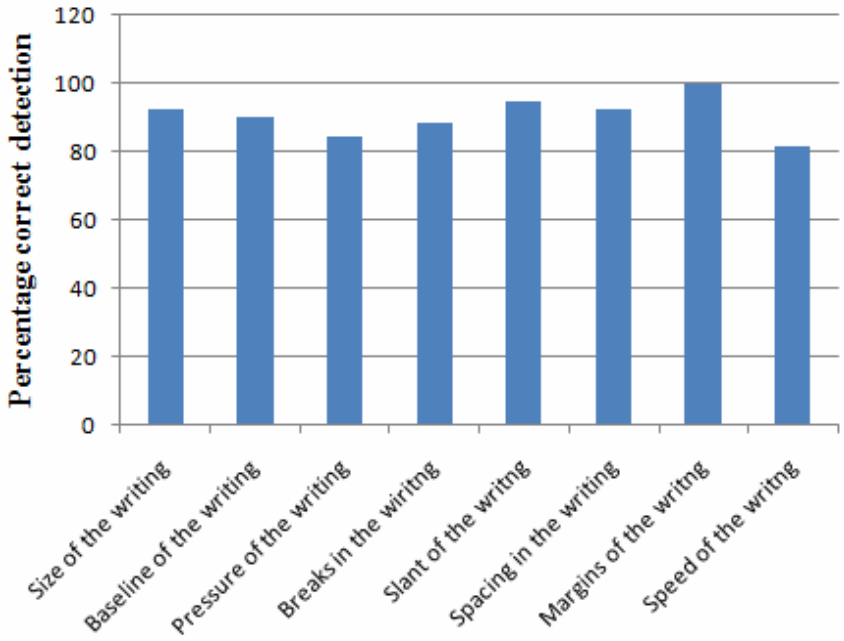

Figure-8. Percentage correct detection of the characteristic traits.

6. CONCLUSIONS

A new method is proposed for the automated behavioral analysis using Automated Handwriting Analysis System (AHWAS). Eight characteristic features of the handwriting are obtained to assess the personality of the writer. Behavioral analysis using proposed algorithms is compared with manual analysis. The result compares well with more than 80 % of the cases.

The proposed automatic system is real time and can be extended to be used for various applications of handwriting such as detection of diseases such as Parkinson's disease and cancer through the variation in features obtained over a period of time. Further it could be used in document authenticity and lie detection as the writer or the forger can never completely reproduce all the original traits

REFERENCES

[1] Champa H. N. and K. R. Ananda Kumar. 2010. Automated Human Behavior Prediction through Handwriting Analysis. First International Conference on Integrated Intelligent Computing.

[2] Peter Shaohua Deng, Li-Jing Jaw, Jau-Hwang Wang and Cheng-Tan Tung. 2003. Trace Copy Forgery Detection for Handwritten Signature Verification. Proceedings IEEE 37th Annual International Carnahan Conference on Security Technology.

[3] Madasu Hanmandlua, Mohd. Hafizuddin Mohd. Yusofb and Vamsi Krishna Madasuc. 2005. Off-line signature verification and forgery detection using fuzzy modeling. Pattern Recognition. 38: 341-356.

[4] Ball G.R., Stittmeyer R. and Srihari S. N. 2010. Writer verification in historical documents. Proceedings Document Recognition and Retrieval XVII San Jose, CA, SPIE.

[5] Judie Walton. 1997. Handwriting changes due to aging and Parkinsons symdrome. Forensic Science International. 88: 197-21.

[6] Jin H. Yan' Susan Rountree, Paul Massman, Rachelle Smith Doody and Hong Li. 2008. Alzheimer's disease and mild cognitive impairment deteriorate fine movement control. Journal of Psychiatric Research. 42(14): 1203-1212.

[7] Joel Engel. Early Cancer Detection through Hand writing Analysis. http://www.buzzle.com/articles/early-cancer-detection-through-handwriting-analysis.html.

[8] Plamondon R. and Srihari S.N. 2000. On-line and off-line handwriting recognition: A comprehensive survey. IEEE Trans. PAMI. 22(1): 63-84.

[9] Srihari S., Cha S.H., Arora H. and Lee S. 2001. Individuality of handwriting: A validity study. International. Proceedings ICDAR_01 Seattle (USA). pp. 106-109.

[10] Ameur Bensefia, Ali Nosary, Thierry Paquet and Laurent Heutte. 2002. Writer Identification by Writer"s Invariants. Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR"02).

[11] U.V. Marti, R. Messerli and H. Bunke. 2001. Writer Identification Using Text Line Based Features. Proceedings ICDAR"01 Seattle (USA). pp. 101-105.

[12] Champa H N and K R Ananda Kumar. 2008. A Scientific Approach to Behavior Analysis through Handwriting Analysis. National Conference on Research Trends in Information Technology S R K R Engineering College, Bhimavaram, A. P., India.

[13] Champa H N and K R Ananda Kumar. 2009. Writer's Personality Prediction through letter 'y' using Generalized Hough Transform (GHT). 3rd International Conference on Information Processing, organized by the Society of Information Processing Bangalore Defence Institute of Advanced Technology and University Visveswaraya College of Engineering.

[14] Champa H N and K R Ananda Kumar. 2009. Rule Based Approach for Personality Prediction through Handwriting Analysis. 2nd Intl. Conference on Biomedical Informatics and Signal processing. Sai's Bio Sciences Research Institute Pvt. Ltd.

[15] Champa H N and K R Ananda Kumar. Handwriting Analysis for Writer's Personality Prediction. Intl. Conference on Biometric Technologies and Applications- the Indian perspective. Biometrics India Expo New Delhi India. pp. 182-191.

[16] Shitala Prasad, Vivek Kumar Singh and Akshay Sapre. 2010. Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using Support Vector Machine. Intl. J. Computer Applications. 8(12), October.

[17] Dr. Ray Walker. 2005. The Science of Handwriting Analysis. http://www.HandwritingUniversity.com.

[18] Ruth Gardner. 1997. Instant Handwriting Analysis. P ustak Mahal.

[19] Jess E Dines. 2009. Handwriting Analysis made easy. Sterling Paperbacks.

[20] Andrea Mc Nichol. 1994. Handwriting Analysis: Putting It to Work for You. Contemporary Books.