

Prediction of Human Character through Automated Script Analysis

Ankur M. Bobade, Prof. N. N. Khalsa, Dr. S. M. Deshmukh

Abstract— In the contemporary study a method has been suggested for the behavioral prediction of a person through automated script analysis. The current work classifies the psychological individuality in the writing namely size, slant and pressure, baseline, number of breaks, margins, speed of writing and spacing between the words. The script is analyzed through Image Processing in MATLAB. The behavioral pattern of the person is predicted from the above individuality of the script. The developed system identifies handwritten script strictly which may not be attainable by a graphologist. It is real time system and involves less image preprocessing. The proposed system is calibrated with manual analysis. The results obtained through the system are in good concord to more than 80 percent of the cases with ideal manual analysis.

Index Terms— Script analysis, MATLAB, image processing, handwritten script, Personality attributes, predict behavioral, graphology.

1 INTRODUCTION

Handwritten script is habitually referred to as indication of personality attribute represented by neurological patterns in the brain. In other words our brain or subliminal mind actually forms the characters as an outcome of habit. Script analysis also known as graphology, which is a pseudo-scientific study of script next to human psychology. Graphology can be used for identifying, evaluating and understanding personality of a person through the strokes and patterns revealed by script. The main applications of Graphology include behavior analysis, forensic evidence and disease analysis. Script reveals the exact personality including emotional status, fears, honesty, defenses and much other individual personality attribute [2]. The authenticity of a person's signature or script in the suicide note is frequently subjected to forensic document examination during investigation in order to verify authorship. [1-3]. 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 [6-5].

Person's script analysis made by a computer is fast, accurate and identifies the script superior than visual inspection. Moreover computer assisted script analysis is automated, efficient and devoid of human errors. Behavioral prediction by script analysis with the help of a computer has been studied earlier by various researchers [7-10].

Proposed methods in literature involve the prelude process of text extraction from the sample and then application of various method / 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 script text. The slope of the alphabet/text is found via the coordinates of the polygon. Generalized Hough Transform is a second method used to detect any random 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 method for behavioral analysis [2]. Segmentation method splitting up of the script sample into individual letters is one more work availa-

ble in literature [18]. Discussed methods are not very simple to automate and therefore require exists for a simple method which could be automated effortlessly.

The existing work focuses on development of an automated method for determining the characteristic individuality of a person through Image Processing called Automated Script/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 script and identifies eight features in the script parallelly. The features identified are: size of the letters, baseline, pressure of writing, slant of the script, number of breaks, spacing between the words, margins and speed of the writing in the sample. The system is designed to straight indicate the behavior of the person from the above mentioned features. The system can be used in various applications such as detection of diseases like Parkinson's disease or Alzheimer's disease, lie detection and forensic document examination.

2 METHODOLOGY

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

The person's written script image is taken from a Sony cyber-shot DSC H55 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 separated from the overall handwriting sample after applying an adequate RGB (Red, Green Blue) threshold. The characteristic attributes are quantified by comparing them with extremities.

The cropped image from the handwritten script sample is shown in Figure-2. As blue ball point pen is used for all samples, the threshold for blue is diverse slightly for each sample. The image is subjected to a threshold to highlight the ink image from the background.

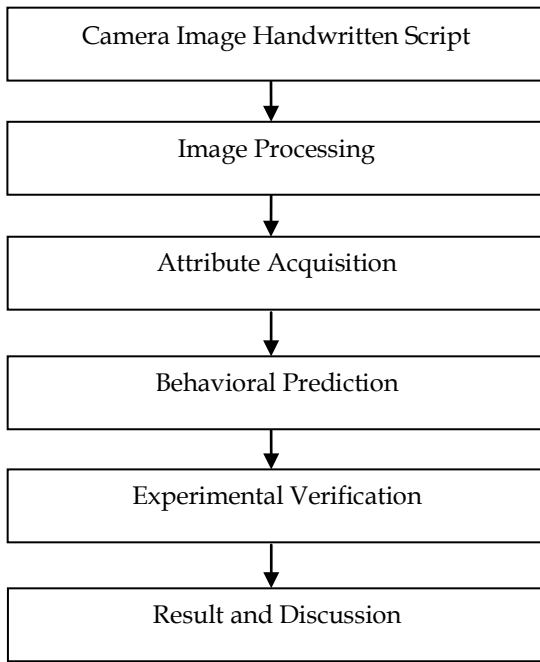


Figure 1. Block Flow diagram of the proposed system.

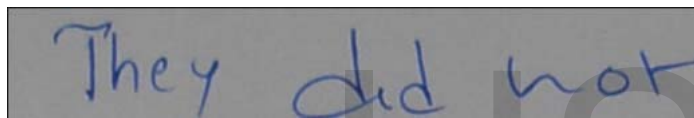


Figure 2. Cropped image from the writing sample.

3 FEATURE DETECTION

Once the certain image of the script sample is preprocessed, eight characteristic attributes of handwritten script are determined. Features detection is nothing but a reduction of high dimension data input.

3.1. Size of the Letters

Size of the letter is judged by the vertical height. The size of handwritten script is judged by a benchmark of 3mm as normal writing and full height of 9mm. If we found the bold letters, author wants to notice me. The letters in a script are divided into three zones: lower case or zone (e.g. g, y), upper case or zone (e.g., I, t), middle case or zone (e.g. a, c, e).

3.2. Baseline

The emotional steadiness and disposition of writer is judged by the baseline in the handwritten script as in Figure 3. To calculate this equation (1) is used. In this, line spacing is also

considered, which shows that the writer wishes to stand back and take a large view of the life.

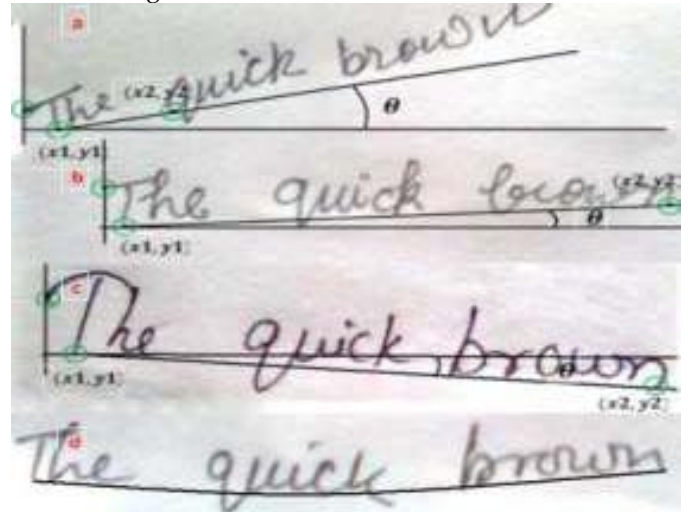


Figure 3. The size and the baseline of the writing.

3.3. Pressure of the writing

Usually it is understood that the pressure of the script in the image can be determined by the width of the stroke, this is in line with what graphologists would do when analyzing the script, 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 amount of black pixels at each intersection with the letters is calculated and divided by the amount of intersections to give the pressure of the writing; this is shown in Figure-4.

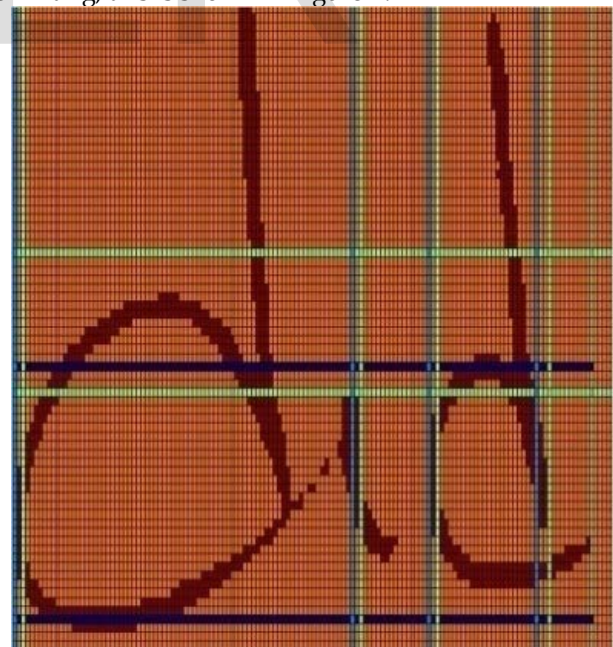


Figure 2. Pressure determination of the writing.

3.4. Slant of the writing

The slant is typically 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 let-

- Ankur M. Bobade is currently pursuing masters degree program in Digital Electronics at Prof. Ram Meghe Institute of Technology and Research, Badnera, India, ankurbobade@gmail.com
- Prof. N. N. Khalsa is currently working as a Asst. Professor in Department of Electronics and Telecommunication Engineering at Prof. Ram Meghe Institute of Technology and Research, Badnera, India, nnkhalsa@mitra.ac.in
- Dr. S. M. Deshmukh is currently working as a Professor and Head in Department of Electronics and Telecommunication Engineering at Prof. Ram Meghe Institute of Technology and Research, Badnera, India.

ter, and determining the slant of the line using Equation 1.

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

The group of black pixels at the top and bottom of each word is plotted to get 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 a group 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° to 180°.

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

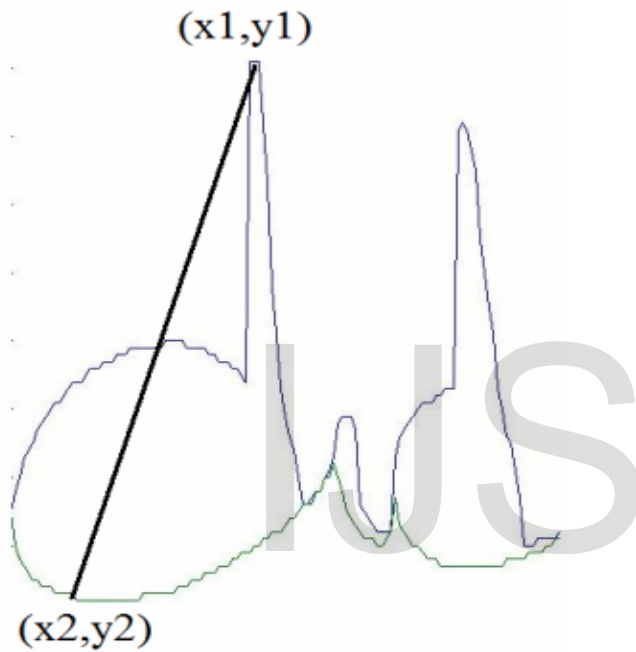


Figure 5. Method adopted to determine slant of the writing.

3.5. Breaks in the writing

Breaks represent the connectivity within a word in the script. It is signified by blob analysis. The amount of breaks in the sentence is identifies by assessing the amount of blobs, which are white patches obtained in the sample. The total amount of blobs excluding the amount of words gives the amount of breaks in the sentence.

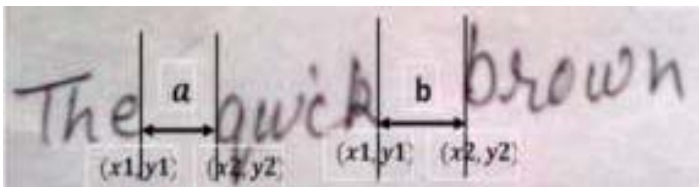


Figure 6. Breaks in the sentence.

3.6 Spacing between the words

This feature in a sample of handwritten script is obtained by the amount of pixels between the end of the one word and the start of the next word.

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
	Disconnect ed	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

Table-1. Handwritten script attributes and behavioral explanation.

3.7 Margins

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 is nothing but the margin. 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

How fast the writer has written the script sample is the speed of writing. To efficiently assess speed of the writing we use two algorithms one for measuring the ink density and the behavior of stroke length at the last 20 pixels. The ink density is obtained by assessing the amount 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 handwritten script attribute and their equivalent behavioral justification are shown in Table-1. The category for each attribute is explained based on information from [17-20]. It can be seen that for each of the eight attributes 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 Automated Script/ Handwriting Analysis System to effectively assess the characteristic behavior of the person.

4 TESTING

The Handwritten script samples were taken from 50 people of different ages between 20 to 40, belonging to both genders. The groups of writers selected were all right handed and were physically and mentally sound. These people were asked 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 handwritten script obtained is shown in Figure-7.

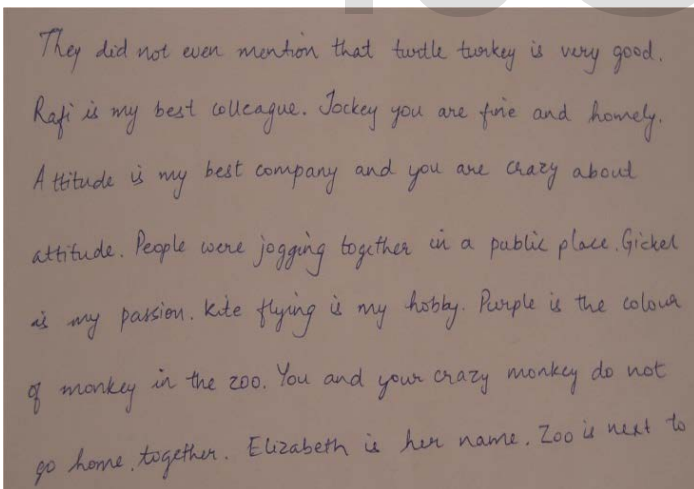


Figure-7. Handwritten script sample obtained.

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

The testing is done in two steps namely calibration and experimental verification. Attributes/ features are extracted from the script as explained in the previous section.

The initial 50 percent of the handwritten script 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 identical blue ball point pen on an A4 size sheet of paper.

The image of the script samples are photographed and stored in digital form. The very first step involved in calibration of the images to extract the maximum about of the handwritten script, is to use a RGB (red green blue) filter to have a clear image of the script. The A4 size paper is 8.27 inches in width and is used for size calibration. This was done via finding the edges of the paper and feeding the input to the program/code.

Letters Size is classified as small, medium and large based on the standard script intervals of less than 1/16" (1.5875x10⁻³ m), between 1/16" - 3/16" (4.7625 x 10⁻³ m) and larger than 3/16" respectively.

The slope of the line connecting the word is decided by baseline. A negative slope represents decreasing baseline and a positive slope represents increasing baseline. A slope around zero represents a straight line.

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

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

The amount breaks in the sentence is identified by blob analysis; it is as explained in the previous section, so a limit of 5 breaks per word is taken as peculiarity between connected and disconnected script.

The spacing between words is regulated 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 doubles the width of the word and wide if it's greater than double the width of word.

The left and the right margins are regulated by comparing it with a standard reference [16]. The speed of the writing is regulated by having a threshold value of 20% for the ink density and decreasing stroke width at the end of the word.

3 RESULTS AND DISCUSSIONS

The code/ program are executed for all the samples of the handwritten script when the calibration was finished and the outcomes are tabulated for evaluation with manual analysis. A sample set of results obtained through Automated Script/ Handwriting Analysis System for the handwritten script sample shown in Figure-6 are compiled in Table-2. It is examined that the features determined by the system are compared against a reference, which represent the criteria to classify various writing styles, which are preferred as explained in the previous section and the attributes category determined appropriately.

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

Handwritten script attributes	Features determined by AHWAS	Reference	Attribute category
Size	2.368mm	1/16"-3/16"	Medium
Baseline	4.69°	>0°	Upward
Pressure	0.0225mm	>0.025mm	Light
Slant	118°	112°-125°	BC
Breaks	33	>5	Discennected
Word spacing	0.224	>twice width of word	Wide
Speed	0.3078	>20%	Fast

Table-2. Results of AHWAS and manual analysis.

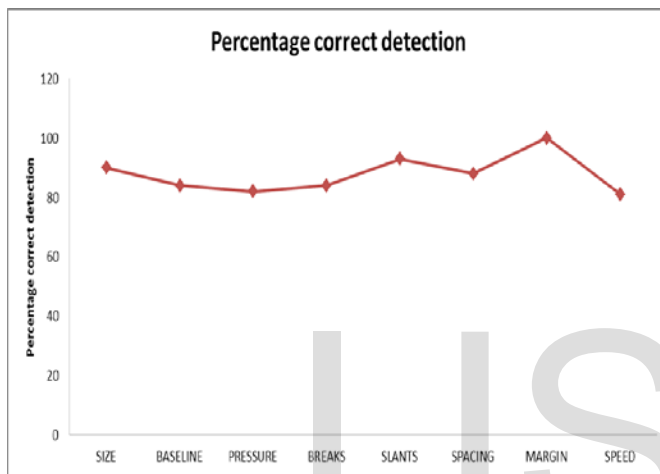


Figure-8. Percentage correct detection of the characteristic attributes.

6 CONCLUSIONS

A fresh method is proposed for the automated behavioral analysis using Automated Script/ Handwriting Analysis System. Eight characteristic features of the handwritten script are obtained to assess the personality of the writer. Behavioral analysis using proposed methods/ algorithms is compared with manual analysis. The result compares well with more than 80 % of the cases.

The proposed automatic method is real time and can be extended to be used for different applications of handwritten script such as document authenticity and lie detection as the writer or the forger can never completely reproduce all the original attributes. Furthermore it could be applied in detection of viruses such as Parkinson's virus and cancer through the deviation in features obtained over a period of time.

REFERENCES

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

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

[3] 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.

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

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

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

[7] 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.

[8] Jin H. Yan, Susan Rountree, Paul Massman, Rachele 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.

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

[10] 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.

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

[12] U.V. Marti, R. Messerli and H. Bunke. 2001. Writer Identification Using Text Line Based Features. Proceedings ICDAR '01 Seattle (USA). pp. 140-145.

[13] 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.

[14] 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.

[15] 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.

[16] Ruth Gardner. 1997. Instant Handwriting Analysis. Pustak Mahal.

[17] Vikram Kamath, Nikhil Ramaswamy, P. Navin Karanth, Vijay Desai and S. M. Kulkarni. 2011. Development of an Automated Handwriting Analysis System. ARPN Journal of Engineering and Applied Science. 6(9), Sept.

[18] 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.

[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.