Character Revealing Handwriting Analysis based on Segmentation method using Support Vector Machine

Ankur M. Bobade Prof. Nikkoo. N. Khalsa

***Abstract*— Earlier character of person is revealed by spending lot of time with the person. As we know spending time with person is very easier said than done. Referring to this problem, in the present study a method has been proposed for the character behavior prediction of a person through automated handwriting analysis. Handwriting analysis is a method to reveal character of a Person .This is done by Image Processing in MATLAB. In order to reveal the character we are going to take the writing sample and from it we are going to extract different attributes i.e. slant of letters and words, pen pressure, spacing between letter, spacing between word, size of letters, baseline Segmentation method is used to extract the attributes of handwriting which are given to the SVM which reveals the character behavior of the writer sample. This gives optimum accuracy with the use of Redial Kernel function.**

***Index Terms*—Support Vector Machines (SVMs), Image processing, Segmentation, MATLAB, personal attributes, Psychology.**

I. INTRODUCTION

Handwriting analysis is also known as graphology which is a method of identifying the character related to an individual. Handwriting is a skill that differs person to person. It can reveal character such as emotional and mental instability, which can further lead an individual to engage in deviant behavior [4]. It helps in understanding personality character through the strokes and patterns revealed by handwriting.

Handwriting is often referred to as indication of personality attribute represented by neurological patterns in the brain [2]. Handwriting is brain writing, representing the mental status of the person [1]. Handwriting reveals the true personality including emotional outlay, fears, honesty, defenses and much other individual personality attributes [2]. Handwriting was developed a long time ago as a means to expand human facilitate communication.

Handwriting is unique to every individual regardless of the word formation of an individual's handwriting the shape of the character is will remain the same [4]. Handwriting analysis is a projection technique as the body language that profiles the human behavior in areas of the social skills, achievements, thinking styles, or work habits. Handwriting also depicts the possible ways of person's transactions with stress [1]. Methods proposed in literature involve the preliminary process of text extraction from the sample and then application of various algorithms or techniques to determine the characteristic attributes [2]. For this the first step is to teach the computer what handwriting is and how to segment it [1]. Segmentation method which involves splitting up of the handwriting sample into individual letters is another work available in literature [2]. Behavioral prediction by handwriting analysis with the aid of a computer has been studied previously by various researchers [2]. Handwriting analysis done by a computer is fast, accurate and indentifies the handwriting better than visual inspection [2]. Moreover computer assisted handwriting analysis is automated, efficient and devoid of human errors. Handwriting analysis is very fast and accurate as compared to manual document handwriting analysis [1].

The automated pattern recognition system also may need training on few samples so that they can get the scale and do the analysis for the next available scanned samples [4]. Collecting digital samples of handwriting and computer prediction is very low-cost and convenient method. One can easily give the digital sample of his/her handwriting to a computer and it calculates the attributes using the image processing techniques and reveals the nature of the writer [1]. 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 attributes in the handwriting simultaneously [2]. The attributes identified such as slant, size, pressure, upper zone or case (as in I, t, h, S, etc), lower zone (as in g, q, y, z, etc), word spacing, line spacing, page margins, middle zone or case (as in a, o, c, s, e, etc), arcade, garland, angle, thread. All attributes are extracted automatically from the digital image of handwriting. These samples are then input to the support vector machine for classification. The algorithm proposed is simple and easy to implement and use. MATLAB is the tool used for the same [1].

A. Literature summary and related work

Throughout history, scientists, philosophers, artists and others have been interested in the relationship between the handwriting and the writer. This interest appeared as early as 1622. Efforts at handwriting analysis began in 1872, with the work of the French abbe, Hypolite Michon, who gave graphology its name. Michon and his compatriot, Jules Crepieux-Jamin developed the school of isolated signs. The foundation of Support Vector Machines (SVM) has been developed by Vapnik and has gained popularity due to its many attractive, analytic and computational attributes. But he has used less attributes to reveal character or behavior because of these result is very poor. Barrick and Mount summarized the role of personality at work in seven divergent research streams to demonstrate that character matters because it predicts and reveals behavior at work. [5]

Shitala Prasad, et.al. used six attributes to reveal the character or behavior of person, i.e. size of letters, slant of letters and

words, baseline, pen pressure, spacing between letters and spacing between words. With these attributes they did not get the accurate result because of that they are unable to reveal the character or personality of person. Accuracy of result is optimum. [1]

Vikram Kamath, et.al. proposed a new method to reveal character using Automated Handwriting Analysis System. They have used eight attributes to detect the personality i.e. Size, Baseline, Pressure, Slant, Breaks, Word spacing, Margins, Speed. With these attributes they did not get the accurate result because of that they are unable to predict the personality of person. Accuracy of result is not optimum as required. [2]

B. Improved working

In this paper we used handwriting image samples of different individuals which is digitally collected by scanning the handwritings of different writers. To each of them we told to write a text document of simple 50-60 words in running hand. The samples were written on a plane paper without any margin. In this paper ,we have used more number of attributes such as slant, size, speed, baseline, breaks, margin ,pressure, upper zone or case(as in I, t, h, S, etc), lower zone (as in g, q, y, z, etc), word spacing, line spacing, page margins, middle zone or case(as in a, o, c, s, e, etc), arcade, garland, angle, thread, wavy line (written by authors who are mentally mature and are skilful), and many others. An automated method is used to reveal the character of an individual by his/her handwriting sample using SVM machine learning algorithm. The various parameters are calculated by simple use of trigonometry and thresholding techniques. All these attributes are given to SVM to get more accurate and correct result. This gives optimum accuracy with the use of Kernel function i.e. (RBF).

II. METHODOLOGY

The level of accuracy of handwriting analysis in the result is totally depending on the knowledge and experience of the graphologist. Handwriting analyst called graphologist analyzes. To reveal character of person, the handwriting on a piece of paper written by the individual is taken and it is analyzed by graphologist which is very time consuming. [1] Hence Support Vector Machine is used to reveal character of person, which is not a time consuming method. SVM has three main steps: pre-processing, feature/ attributes extraction, and classification.

Following figure-1 shows different step to be followed while revealing the character and behavior of an individual.

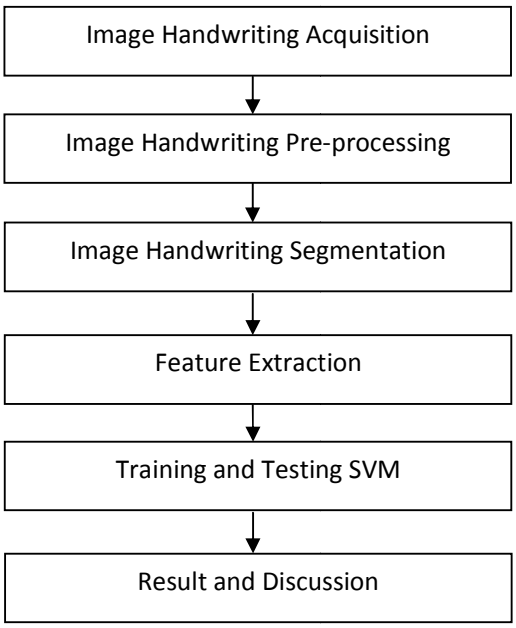

Figure 1. Block Flow Diagram of system

A. Image Handwriting Acquisition

Detailed Image acquisition means to capture the image digitally of related person whose character we have to reveal. We have taken different handwriting of different person on plain paper and asked to write 50-60 words on a plane paper. The image is taken from a Sony Cybershot DSC H55 camera. The image is stored in JPEG format. The region of interest is cropped. [1-2]

B. Pre-Processing of Image Handwriting

As we know pre-processing means smoothing the image and smoothing means removing the noise i.e. dots and other parameter from the captured image. Thus we have to apply image pre-processing to the image for which personality detection has to be done. [1-2]

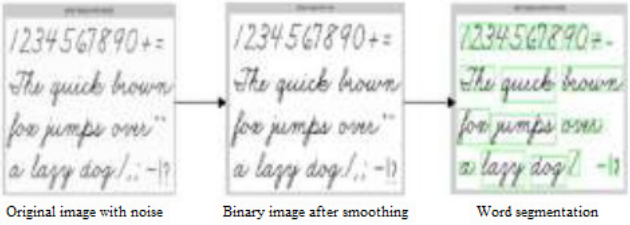

Figure.2

C. Segmentation of Image Handwriting

Once the given image of the handwriting sample is preprocessed segmentation is done. Segmentation means to break the image in to number of parts. Digitally collected image handwriting is segmented in to three parts i.e. line segmentation, word segmentation, letter segmentation.

a. Segmentation of Line

To find baseline features line segmentation is used which reveal the writer's emotional stability and dispositions in the

baseline of writing. [1-2]

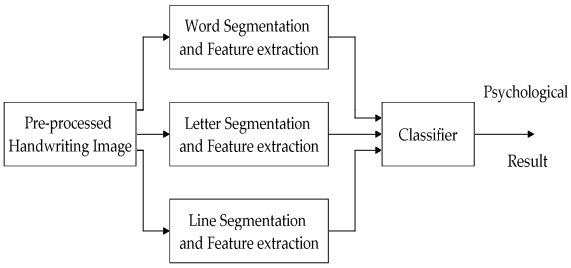

Figure 3. Segmentation and feature extraction.

b. Segmentation of Letter

To In letter segmentation each letter is segmented from the word to detect the future parameter. Figure 3, shows the letter segmentation which is future used to calculate various slants in letters indicating the openness of sentiment and consequently of the intelligence. [1-3]

Figure.4

c. Segmentation of Word of Letter

Word in the digitally collected handwriting image is segmented to calculate different attributes related to the human character.

D. Feature Detection

Once the given image of the handwriting sample is preprocessed, eight characteristic traits of handwriting are determined. Features detection is a reduction of high dimension data input.

a. Pen Pressure

Mental and physical stress on writer mind can be identified by the pen pressure. If the pen pressure is more, then he is showing his mental stress otherwise he talking normal with no mental stress. To determine the pen pressure some threshold value must be set (th_0). If the pen pressure is more than the threshold, then he has some mental stress otherwise it is normal. [1-2]

b. Baseline

The baseline is an imaginary line on which the bottom of the middle zone letters aligns. In this, line spacing is also considered which shows emotional stability. [1]

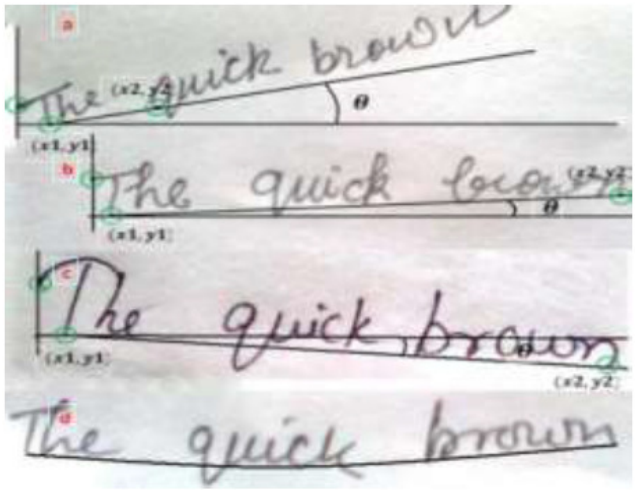

Figure.5

c. Size of the Letters

Size of the letter is judged by the vertical height. The size of handwriting 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 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). [1-2-4]

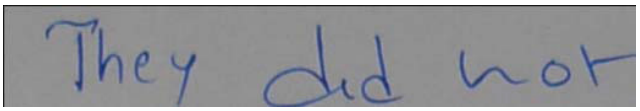

Figure.6

d. Spacing between Letters

It indicates the openness of sentiment and consequently of the openness of intelligence. [1-2]

e. Margins

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. [2]

f. Spacing between the words

It gives the same indication as like spacing between letters. [1-2]

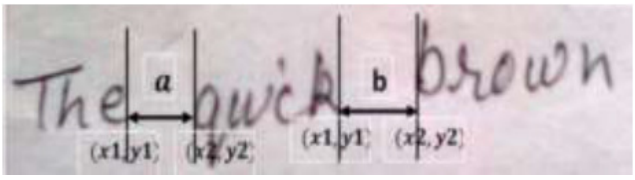

Figure.7

g. Slant of Words and Letters

The emotional interactions of the author are indicated by Slant in handwriting. The slant is obtained by joining the highest and lowest point surrounding a given point in the letter, and can be determine by following equation. [1-2-5]

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

Journal Impact Factor-2.02
Indexed by- ProQuest, DOAJ and many International Bodies.

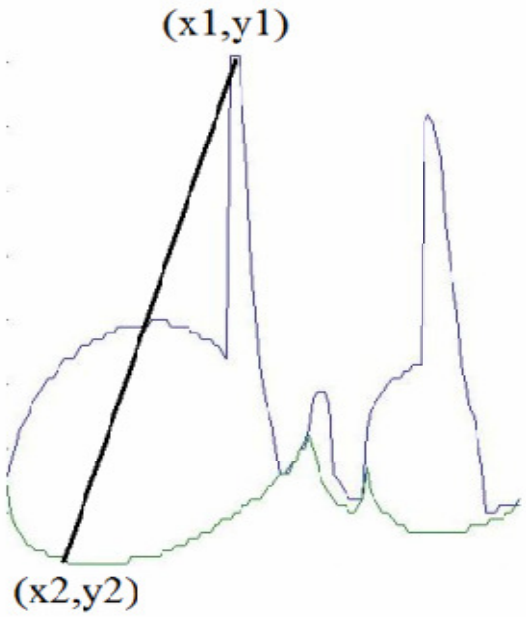

Figure.8

h. Speed of the writing

It is how fast the writer has written the handwriting sample. If the writing speed is fast then author is smart an uncommunicative otherwise he is lazy, clumsy, and dishonest. [2] Each steps or process explained above, gives some numerical value, which is -1, 0, or +1 depending on the threshold and the value obtained. Combining all these numerical value together in an SVM format and forwarded to the SVM classifier to classify the result and reveal the author's character.

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)	Cautions, 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

Table-1. Handwritten script attributes and behavioral explanation.

III. CLASSIFICATION

To reveal the character of human being we are going to use a classifier named as SVM (support vector machine). As compare to neural network SVM is more accurate and time efficient. For revealing the character we have to train the classifier. SVM can be used to classify the data of unknown data class into the correct data categories. Suppose that we are given a training data set

$$\{x_i, y_i\}, i = 1, \ldots, n \quad ; \quad x_i \in R^d; \quad y_i \in R^d \qquad (1)$$

Where x_i is a vector of input variables and y_i represents the corresponding scalar output (target) value. Now based on the training data, the goal of SVM is to develop a function f(x) that can predict output based on the training data. The linear SVM classification function is represented by,

$$f(x) = \omega \bullet x + b$$

Where • denotes the dot product, ω is the weight vector, and b refers to constant (bias). This relates to a separating hyper plane $\omega \bullet x + b = 0$. The hyper plane separating the training data sample with maximum margin such that closest point plane is $\frac{1}{\|w\|}$. [1-5-7]

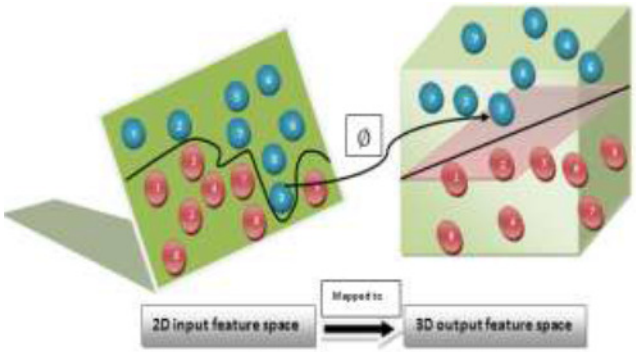

Figure 9. Principle of Support Vector Machine (SVM).

Even the SVM has already made outstanding results in classification research domains; the hyper-parameters and the kernel function of SVM fully influence the revealing accuracy. First, there are four kernel functions in SVM: linear, polynomial, radial basis function (RBF), sigmoid. According to Smola (1998), the RBF function is a general reasonable first choice. RBF kernel function uses the non-linear method to map samples to high dimensional space. [5]

RBF kernel is defined as,

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad \gamma > 0$$

Where K () is a kernel function mapping x_i & x_j of input space to high dimensional feature space and is the kernel parameter changing which may give better results.

The classification can be performed in following three steps:

1. First, the input attributes are formulated as input vectors in some feature space.

2. Map these feature vectors to the higher dimension feature space using RBF kernel function.

3. Then a division global hyper plane is computed to separate the feature space optimally to the classes of the training vector samples. [1]

IV. RESULTS

In this paper each writers was asked to write the test documents and were saved in digital form and we extract different features. Thus we observed that if the data base is more to train the machine, the result we obtained is more accurate as that of the graphologist analyzes. When we used linear, polynomial kernel function with SVM then it gives a poor accuracy. Hence it's better to use radial basis function (RBF) which will gives accuracy near about 90% with more training data based.

V. CONCLUSION

In this paper we are going to reveal the character of human being by taking his/her handwriting image digitally. From above analysis various has been extracted such as 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. The various parameters are calculated by simple use of trigonometry and thresholding techniques. All these attributes are given to SVM which reveal the character of the individual writer. SVM with use of RBF kernel function gives optimum accuracy.

REFERENCES

[1] Shitala Prasad, Vivek Kumar Singh,Akshay Sapre, "Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using Support Vector Machine" International Journal of Computer Applications (0975 – 8887) Volume 8– No.12, October 2010.

[2] Vikram Kamath, Nikhil Ramaswamy, P. Navin Karanth, Vijay Desai and S. M. Kulkarni " Development Of An Automated Handwriting Analysis System" VOL. 6, NO. 9, September 2011 ISSN 1819-6608 ARPN Journal of Engineering and Applied Sciences.

[3] Rejean Plamondon, Sargur N. Srihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey" 1EEE Transaction On Pattern Analysis And Machine Intelligence.VOL. 22, NO. 1. JANUARY 2000

[4] Janet Fisher, Anish Maredia, Anita Nixon, Nerissa Williams, and Jonathan Leet, "Identifying Personality Traits and Especially Traits Resulting in Violent Behavior through Automatic Handwriting Analysis" May 4th, 2012.

[5] Yung-Ming Li, Cheng-Yang Lai, Chien-Pang Kao," Incorporate Personality Trait With Support Vector Machine To Acquire Quality Matching Of Personnel Recruitment".

[6] Luzheng Bi, Omer Tsimhoni, and Yili Liu, "Using the Support Vector Regression Approach To Model Human Performance" VOL. 41, NO. 3, MAY 2011.

[7] Enrique Frias-Martinez, Angel Sanchez and Jose Velez, "Support Vector Machines versus Multi-Layer Perceptions for Efficient Off-Line Signature Recognition"

[8] Tripathy N. and Pal U. 2006,"Handwriting segmentation of constrained Oriya text", Sadhna, Vol.31, Part 6, pp. 755-769.

[9] A. Srihari, S. Cha, H. Arora, S. Lee, "Individuality of Handwriting: A Validity Study", Proc. ICDAR"01, Seattle (USA), pp 106-109, 2001.

[10] Lei Zhang, Fuzong Lin, Bo Zhang, "Support Vector Machine Learning for Image Retrivel", Image Processing, IEEE 2001 International Conference, page(s): 721 - 724 vol.2.

[11] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[12] M. Cimen, "Estimation of daily suspended sediments using support vector machines," Hydrol. Sci. J., vol. 53, no. 3, Jun. 2008.

[13] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining Knowl. Discovery, vol. 2, no. 2, Jun. 1998.

[14] A. J. Smola and B. Scholkopf, "A tutorial on support vector regression," Statist. Comput, vol. 14, no. 3, Aug. 2004.

[15] C. L. Liu, K. Marukawa, Handwritten numeral string recognition: character level training vs. string level training, Proc.7th ICPR, Cambridge, UK, 2004,Vol.1.

[16] C. C. Chang and C. J. Lin, LIBSVM, A Library for Support-Vector-Machines.Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[17] J. Franke, Isolated handprinted digit recognition, Handbook of Character Recognition and Document Image Analysis.

[18] E. Asua, V. Etxebarria, and A. Garcia-Arribas, "Neural network-based micropositioning control of smart shape memory alloy actuators," Eng. Appl. Artif. Intell., vol. 21, no. 5, pp. 796–804, Aug. 2008.

[19] A. Cevik, M. A. Kutuk, A. Erklig, and I. H. Guzelbey, "Neural network modeling of arc spot welding," J. Mater. Process. Technol., vol. 202, no. 1–3, pp. 137–144, Jun. 2008.

[20] C. Wu, Y. Liu, and C. Walsh, "Queueing network modeling of a realtime psychophysiological index of mental workload—P300 in evoked brain potential (ERP)," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 38, no. 5, pp. 1068–1084, Sep. 2008.