HABIT: Handwritten Analysis Based Individualistic Traits Prediction

Abdul Rahiman M   *rehman_paika@yahoo.com*
Director, AICTE
Ministry of HRD, Govt of India
New Delhi, INDIA

Diana Varghese   *dianavarghese100@gmail.com*
Business Analyst, Mu Sigma Business Soln Pvt Ltd
Bangalore, INDIA

Manoj Kumar G   *manojkumar_gg@hotmail.com*
Associate Professor, Dept of Computer Science
LBSITW,Trivandrum, Kerala, INDIA

Abstract

Handwriting Analysis is a scientific method of identifying, evaluating and understanding an individual's personality based on handwriting. Each personality trait of a person is represented by a neurological brain pattern. Each of these neurological brain patterns produces a unique neuromuscular movement that is the same for every person who has that particular personality trait. When writing, these tiny movements occur unconsciously. Strokes, patterns and pressure applied while writing can reveal specific personality traits. The true personality including emotional outlay, fears, honesty and defenses are revealed. Professional handwriting examiners called graphologists analyze handwriting samples for this purpose. However, accuracy of the analysis depends on how skilled the analyst is. The analyst is also prone to fatigue. High cost incurred is yet another deterrent. This paper aims at implementing an off-line, writer-independent handwriting analysis system "HABIT" (Handwriting Analysis Based Individualistic Traits Prediction) which acts as a tool to predict the personality traits of a writer automatically from features extracted from a scanned image of the writer's handwriting sample given as input. The features include slant of baseline, pen pressure, slant of letters and size of writing. The implementation uses Java and Eclipse-Indigo as tools.

Keywords: Handwriting Analysis, Feature Extraction, Patterns.

1. INTRODUCTION

Handwriting Analysis, also known as Graphology is a scientific method of identifying, evaluating and understanding personality through the strokes and patterns revealed by handwriting. Handwriting reveals true personality including emotional outlay, fears, honesty, defenses and many other individual traits. Handwriting is often referred to as brain writing. Each personality trait is represented by a neurological brain pattern. Each neurological brain pattern produces a unique neuromuscular movement that is the same for every person who has that particular personality trait. When writing, these tiny movements occur unconsciously. Each written movement or stroke reveals a specific personality trait. Graphology is the science of identifying these strokes as they appear in handwriting and describing the corresponding personality trait.

Personality identification of a human being by their handwriting is an old technique. Earlier this was done manually by spending a lot of time to predict the nature of the person. Handwriting is brain writing, representing the mental status of the person. Handwriting analysis is a projection technique that profiles the human behavior in areas of social skills, achievements, thinking styles

and work habits. Handwriting also depicts the possible ways of a person's transactions with stress.

In case of manual analysis, accuracy of the analysis depends on how skilled the analyst is. The analyst is also prone to fatigue when several samples are to be analyzed. High cost incurred in getting the aid of a well-experienced graphologist is yet another deterrent.

Compared to manual analysis, automated handwriting analysis is very fast and accurate in the prediction of human personality. Collecting digital samples of handwriting and using computer prediction is a very low-cost and convenient method. One can easily give the digital sample of his/her handwriting to a computer and it calculates the features using image processing techniques and predicts the nature of the writer.

In this paper, a system is proposed to predict the personal behavior of an individual from their handwriting analysis in digital form. To predict the actual personality of the individual there are various features, such as slant, size, pressure, upper zone (as **in l, t, h, S**), lower zone (as in **g, q, y, z**), word spacing, line spacing, page margins, middle zone or case(as in **a, o, c, s, e**). In the proposed system the features utilized are the slant of letters, slant of the baseline, pen pressure and size of the letters as they are enough to predict the behavior of the person.

2. DEVELOPMENT OF GRAPHOLOGY

Graphology is the study of a person's personality from their handwriting. It has a long and respectable history dating as far back as the ancient Romans. The historian Seutonius Tranquillus was among the first to notice in the handwritings of several emperors he was researching that the letters were formed differently in each case. He theorized that this could indicate different character traits. The 19thcentury work by Abbott Jean-Hippolyte Michon and Abbott Flandrin, in which they collected innumerable handwriting samples and matched styles of writing with various temperaments, was the beginning of modern graphology. It was Abbott Michon who coined the term 'graphology'.

The next great leap in the scientific evaluation of handwriting came from Lugwig Klages who was the first to create a complete and systematic theory of 'graphology'. Klages classified personality traits by evaluating the up and down strokes of handwriting. For example, certain handwriting rhythms would indicate someone's intellectual passion, whereas certain letter forms would display someone's sense of inferiority. A 'normal' person would have a balance of contraction and release, whereas unstable people wouldhave an unbalanced rhythm. Even though the system was developed, there did not exist an objective way of rating someone's entire personality.

Joseph Zubin and TheaLewinson built upon Klages work and developed a system of scales, called the L-Z scales that evaluated the quantitative and qualitative aspects of handwriting. According to Lewinson-Zubin, there are four special components of handwriting – vertical, horizontal, depth, and form –by which each written letter can be evaluated. These four components yield the following dimensions of personality: rational, social, emotional and instinctual.

The vertical component concerns the height of the middle zone of a letter which emphasizes self-importance. The direction of the vertical lines belies the individual's mood level. The horizontal component is measured by the distance between letters and words. Right/left slants are also included in this measurement. Horizontal traits measure the relationship between the individual and his or her environment. The depth component is the pressure of the writing which represents one's instinctual drives. The form component measures the contour of the writing which can signify the degree of one's creativeness.

3. SYSTEM DESCRIPTION

Professional handwriting examiners called graphologists often identify the writer with a piece of handwriting. Accuracy of handwriting analysis depends on how skilled the analyst is. Although human intervention in handwriting analysis has been effective, it is costly and prone to fatigue.

The proposed system HABIT focuses on developing a tool for behavioral analysis which can predict the personality traits automatically with the aid of a computer. The system is designed to analyze scanned images of handwritten documents. The images are converted into binary black and white pixel images. It is difficult for the program to identify different letters, words and lines by pixel analysis without human intervention. Therefore, human decision process guides the system to achieve higher accuracy. Here the aim is to create software that allows a user to analyze handwriting samples, with the aim of making the process faster and more objective.

HABIT is an off-line, writer independent handwriting analysis system that predicts the personality of a person. The input to the system is a scanned image of a handwriting sample of the writer .The behavioral analysis is done from the baseline slant, the pen pressure, the slant of the writing and size of letters. The output is a set of personality trait of the writer. The entire system is depicted in figure 1.
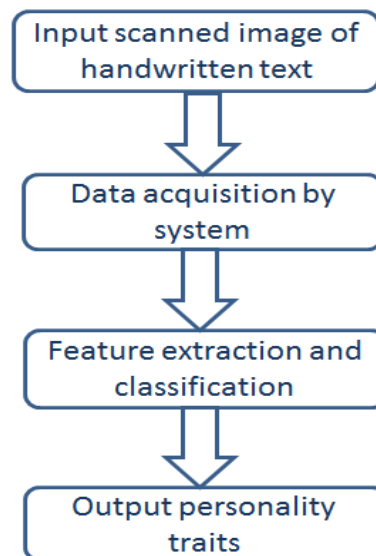

FIGURE 1: HABIT System- An overview.

4. IMAGE ACQUISITION & PREPROCESSING

Images of handwritten samples are uploaded to the system. The system requires preprocessing work from the user in order to begin calculating the scales from horizontal and vertical category. The uploaded image is preprocessed and resized to the correct orientation. Steps in Image acquisition and feature extraction are shown in figure 2.

The application allows users to crop images into lines, words and characters. Once the images are cropped, the cropped images will be displayed on the scratchpad. In the analysis window, a cropped image will be loaded.
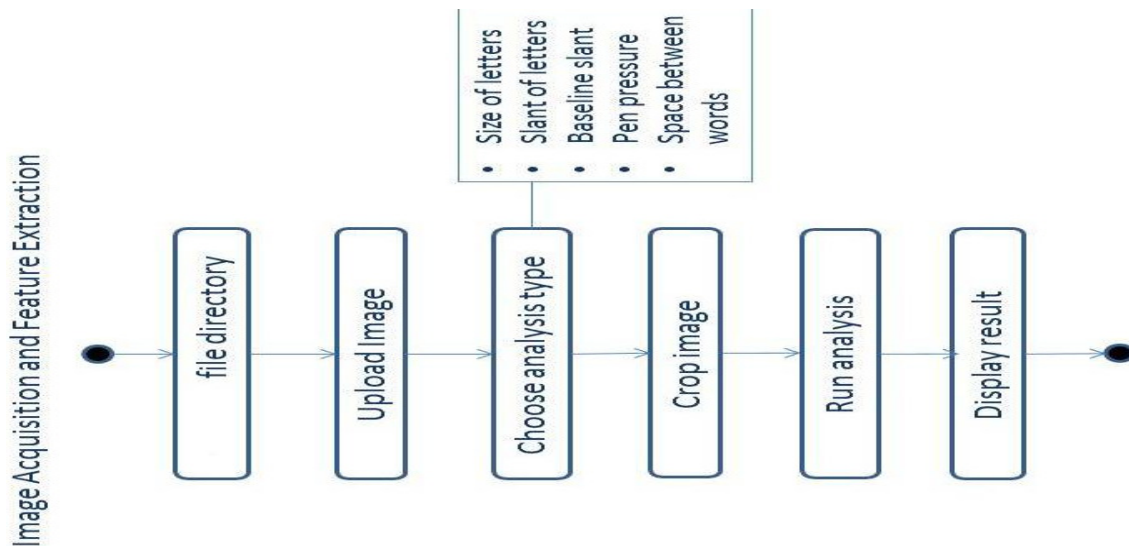
FIGURE 2: Image Acquisition & Processing.

Depending on the specimen chosen data points can be plotted. The user must plot some data points on the image that identify the top, middle, base and bottom lines of the handwriting. The collection of the data points themselves has a subjective nature because they depend on the user's expert eye to select points of interest. Due to this reason, linear regression approximation of least squares is a suitable method.

5. FEATURE EXTRACTION STAGE

5.1   Pen Pressure
One of the most important features in a handwriting sample is the pressure of writing. The amount of pressure exerted on the paper while writing indicates the depth of feeling, also called emotional intensity, of the writer. Based on the pen pressure, the writer can be classified as a light writer, medium writer or heavy writer. For the analysis of pen pressure, first the scanned image is converted into a grey scale image. Mean grey level value is computed using the grey level values of the image pixels. This mean grey level value of the image is compared with the pre-determined threshold value, th0. Higher value of the men indicates lighter pressure. If the mean is less than th0, then the writing pressure is considered to be high. Such a person has very deep and enduring feelings. This writer may forgive, but he will never forget. He feels situations intensely. If mean is greater than th0, then writing pressure is considered to be light. Such a person can endure traumatic experiences without being seriously affected. Emotional experiences do not make a lasting impression on him.
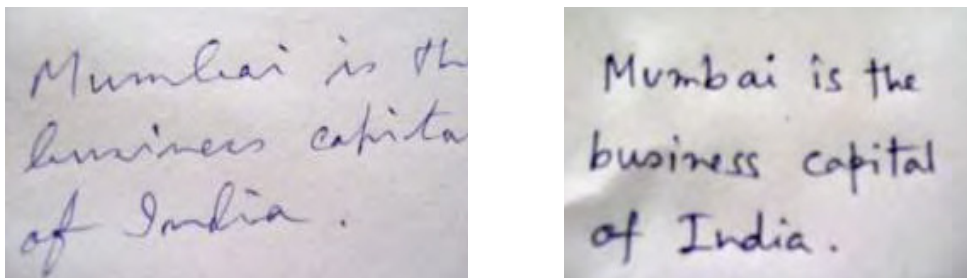
FIGURE 3: Light writer & Heavy writer

5.2   Slant of letters and Slant of baseline

The baseline in one's handwriting reveals a lot of accurate information about the writer. Baseline in one's handwriting is the line along which the writing flows. The three most common baselines found in any handwriting are ascending, descending and level. Given a set of data points, we need a trend line that passes closely among the points, especially if there is large number of data points. This trend line that we compute is called regression line. This regression line is computed using Least-Squares Linear Regression. To calculate the slant in drawn lines, the formula is

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

Standard reference angle (θo) is considered to be 90. θ is compared with θo to classify the slant height.
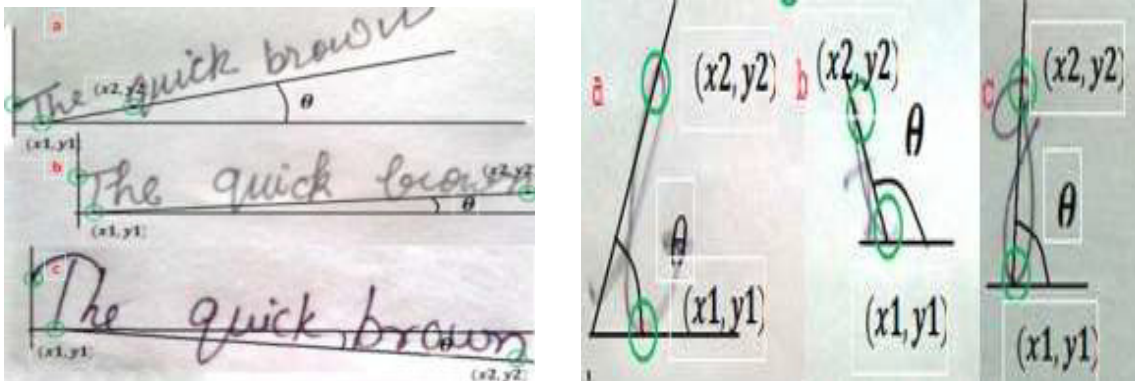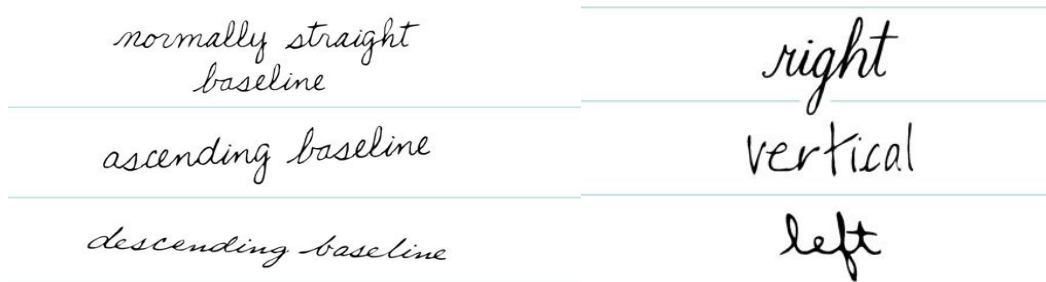

FIGURE 4: Slant of Letters & Trend line.

5.3   Size of Letters

The size of handwriting is judged by a benchmark of 3mm as normal writing and full height of 9mm. Other than this, writing is classified as large or small writing. The letters are divided into 3 zones: lower, upper and middle. Large writing can portray someone who is superior, takes pride, outgoing and extrovert, arrogant, boastful or they put on an act of confidence. Small handwriting can mean people who are respectful, tolerable, introvert, shy, deep thinkers or academic. If letters which reach into the upper zones are very extended then the person probably has unrealistic expectations of what they can truly achieve. Wide upper zone loops can show people who are deep thinkers and dreamers; then those whose letters go up then down directly over themselves are often un-imaginative. If the letters in the lower zone are straight it shows people who like to get the job done; those who loop them are often full of energy, good at making and investing money, or/and need security. The different zones are shown in figure 5. To find the size of letters we use the Pythogora's theorem to find the distance between the top point and baseline. For this we use the formula:

$$\sqrt{(\Delta x^2 + \Delta y^2)} = c^2$$ where Δx is the distance between x co-ordinates

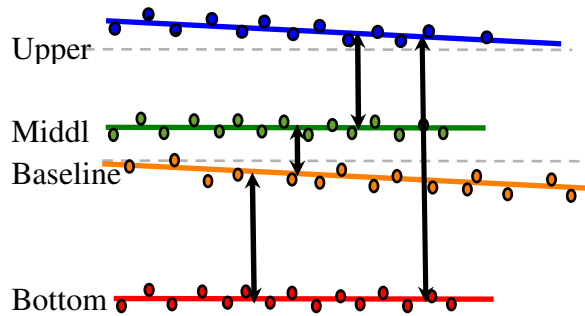Δy is the distance between y co-ordinates

FIGURE 5: Different Zones.

5.4 Spacing Between Words

A person who has some words widely spaced are often open, honest but deep in thought and people with words narrow spaced may be unstable in either emotions or thinking. Spacing between the words is shown in figure 6. To find the distance between the two words perpendiculars are dropped from the end point of the first word and starting point of the second word. Length of the hypotenuse joining the two perpendiculars gives the distance between the two words. The equation to compute the length of the hypotenuse is
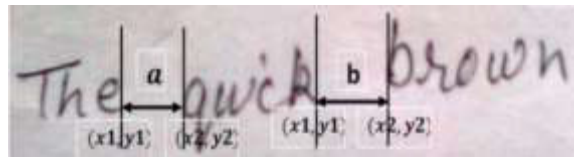
$$\sqrt{(\Delta x^2 + \Delta y^2)} = c^2$$

FIGURE 6: Word Spacing.

5.5 Handwriting Style Classification

The handwriting style classification is shown in Table 1 below.

Sl. No.	Writing Category	Sub-category Criteria
1	Size of letters	Large: size > normal size Small: size < normal size Medium : size = normal size
2	Slant of letters	Right: $\theta > \theta 0$ Left: $\theta < \theta 0$ Vertical: $\theta = \theta 0$
3	Baseline	Raising: $\theta > 0$ Falling: $\theta < 0$ Straight: $\theta = 0$ Erratic: otherwise
4	Pen Pressure	Light Pen: threshold> th0

		Heavy Pen: threshold < th0
5	Space between words	Far: space > average space Near: space < average space

TABLE 1: Handwriting Styles & Classification.

5.6   Personality Traits Predicted by Various Handwriting Styles

Writing Categories	Psychological Personality Behaviour
Large Letters	Likes being noticed, stands out in a crowd
Small Letters	Introspective, not seeking attention, modest
Medium Letters	Adaptable, fits into a crowd, practical, balanced
Right Slant	Sociable, responsive, interested in others, friendly
Left Slant	Reserved, observant, self-reliant, non-intrusive
Vertical	Practical, independent, controlled, self-sufficient
Light Pen Pressure	Can endure traumatic experiences without being seriously affected. Emotional experiences do not make a lasting impression
Heavy Pen Pressure	Have very deep and enduring feelings and feeels situations intensely.
Raising Baseline	Optimistic, upbeat, positive attitude, ambitious and hopeful
Falling Baseline	Tired, overwhelmed, pessimistic, not hopeful
Straight Baseline	Determined, stays on track, self-motivated, controls emotions, reliable, steady
Erratic Baseline	Wavering, lacks definite direction, emotionally unsettled, unpredictable
Far Spaced Words	Desires more space, enjoys privacy
Close Spaced Words	Closeness of sentiment and intelligence

TABLE 2: Personality Traits.

6. SYSTEM IMPLEMENTATION

The input to the system is a scanned image of a handwriting sample of the writer. The behavioral analysis is done from the baseline slant, the pen pressure, the slant of the writing size of letters and spacing between words. The output is a set of personality traits of the writer. The method used to implement this is simple linear regression which is an approach to modeling the relationship between a scalar dependent variable y and an explanatory variable denoted x.

6.1   Least Squares Linear Regression

Predictions for y from each value of x in the data will usually differ from the actual value of y that is being predicted. If you square the difference and add up these squared differences across all the predictions, you get a number called the residual or error sum or squares. With any two variables x and y, there exists one formula that will produce the best, or most accurate predictions for y given x. Any other equation would not fit as well and would predict y with more error. That equation is called the least squares regression equation. Step in running linear regression are:
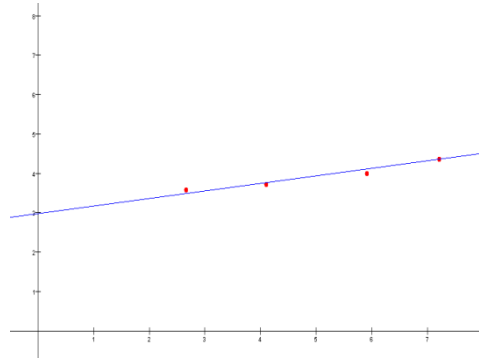
1. Specimen is displayed on the scratchpad
2. User plots data points on the sample
3. Regression line computed using selected data points

Formula for line-approximation using n points,

$$a_1 = \frac{n\sum_i^n x_i y_i - \left(\sum_i^n x_i\right)\left(\sum_i^n y_i\right)}{n\sum_i^n x_i^2 - \left(\sum_i^n x_i\right)^2}$$

$$a_0 = \bar{y} - a_1\bar{x}$$

where a_1 is the line slope and a_0 is the line intercept .
The regression line obtained is plotted in figure 7.

FIGURE 7: Regression line of the plotted points

Hibernate has been used to perform operations on the database like select, insert, update and delete the records in the table. It automatically creates the query to perform these operations. To use Hibernate, it is required to create Java classes that represents the table in the database and then map the instance variable in the class with the columns in the database. The architecture is depicted in figure 8. Hibernate architecture has three main components:

- Connection Management: Hibernate Connection management service provides efficient management of the database connections. Database connection is the most expensive part of interacting with the database as it requires a lot of resources of open and close the database connection.

- Transaction management :Transaction management service provides the ability to the user to execute more than one database statements at a time.

- Object relational mapping: It is the technique of mapping the data representation from an object model to a relational data model. This part of Hibernate is used to select, insert, update and delete the records form the underlying table. When we pass an object to a Session.save() method, Hibernate reads the state of the variables of that object and executes the necessary query.
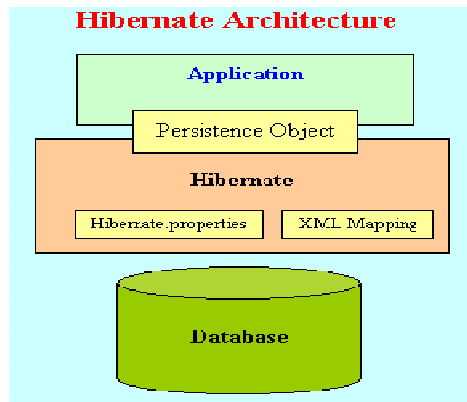
FIGURE 8: Hibernate Architecture.

7. CONCLUSION

HABIT is an off-line, writer independent handwriting analysis system that predicts the personality of a person. The input to the system is a scanned image of a handwriting sample. The output is a set of personality traits. To predict the actual personality of the individual there are various features, such as slant, size, pressure, upper zone loops, lower zone loops, word spacing, line spacing, page margins etc. HABIT utilizes the slant of letters, slant of the baseline, pen pressure and size of the letters as found in the person's handwriting sample. The implemented method here proves successful in analyzing handwriting irrespective of the language used. The analysis is completely devoid of user intervention. But stress has to be laid to the fact that the image acquisition process requires the aid of a graphologist. Enhancements possible on the implemented system include using the following features too in the traits prediction process.

- Position of 't' bar : Position of 't' bar indicates the self-esteem of a person. Using regression lines the distance between the 't' bar and bottom point or top point can be calculated. Space between letters:

- Perpendiculars are dropped from the end point of the first letter and starting point of the second letter. Length of the hypotenuse joining the two perpendiculars gives the distance between the two letters. Space between lines:

- Using Pythagoras theorem, distance between the perpendiculars can be calculated

8. 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, Volume 8– No.12, October 2010.

2 Champa H N, K R AnandaKumar, "Automated Human Behavior Prediction through Handwriting Analysis", IEEE First International Conference on Integrated Intelligent Computing, 2010.

3 Sunday Olatunbosun, Aaron Dancygier, Jayson Diaz, Stacy Bryan, and Sung-Hyuk Cha,"Automating the Lewinson-Zubin Handwriting Personality Assessment Scales", Proceedings of Student-Faculty Research Day, CSIS, Pace University,2009.

4 Srihari S.N., Sung-Hyuk Cha and SangjikLee,"Establishing handwriting Individuality using pattern recognition techniques",Proceedings of the Sixth International Conference on Document Analysis and Recognition, 2001.

5   AmeurBensefia , Ali Nosary, Thierry Paquet, Laurent Heutte,"WriterIdentification by Writer's Invariants",Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition", pp 274-279, 2002.

6   Srihari S.N., Sung-Hyuk Cha and Sangjik Lee, "Establishing handwriting Individuality using pattern recognition techniques", Proceedings of the Sixth International Conference on Document Analysis and Recognition, 2001.

7   N Mogharreban, S Rahimi, M Sabharwal, "A Combined Crisp and Fuzzy Approach for Handwriting Analysis", IEEE Annual Meeting of the Fuzzy Information, vol 1,pp 351-356, 2004.

8   OndrejRohlik, "Handwritten Text Analysis", Department of Computer Science and Engineering Laboratory of Intelligent Communication Systems, University of West Bohemia In Pilsen, 2001.