*IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*
*e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 10, Issue 3, Ver. I (May - Jun.2015), PP 01-07*
*www.iosrjournals.org*

# Graphology for Farsi Handwriting Using Image Processing Techniques

Somayeh Hashemi[1], Behrouz Vaseghi[2,] Fatemeh Torgheh[3]
[1,2,3] *(Department of Electrical Engineering ,Abhar Branch, Islamic Azad University, Abhar, Iran.)*

***Abstract:*** *Handwriting Analysis or Graphology is a scientific method of identifying, evaluating and understanding personality through the strokes and patterns revealed by handwriting. Handwriting reveals the true personality including emotional outlay, fears, honesty, defenses and many others. Professional handwriting examiners called graphologist 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. Hence the proposed methodology focuses on developing a tool for behavioral analysis which can predict the personality traits automatically with the aid of a computer without the human intervention. The most predominant features of handwriting employed in graphological analyses include the shape of the page margins, line spacing, line skew, word slant, corner sharpness, size of letters, text density, writing speed and regularity of writing. In this paper, a number of methods are presented for automated extraction of these features from Farsi handwriting. Experimental results on 30 training and 150 test samples are presented and discussed.*

***Keywords:*** *Graphology; Farsi Handwriting; Image processing; Personality Traits; Human Behavior Analysis;*

## I. Introduction

Your handwriting develops right from childhood. When you write, your pen is under the control of the muscles of your fingers, hands and arm. All these body parts are under the control of your mind. The manner in which the words are eventually formed by the pen must bear a direct relationship to the mind that guides their formation. Each vibration of movement is unconsciously directed by the brain, so we can judge the mental state of the writer. It is a guide to the will power, intellect and emotions of a person. For an accurate analysis, written text should have been written in a natural manner and the effort should not be deliberate. The best samples are business letters or notes. Handwriting Analysis or Graphology is a scientific method of identifying, evaluating and understanding personality through the strokes and patterns revealed by handwriting. Handwriting reveals the true personality including emotional outlay, fears, honesty, defenses and over many other individual personality traits. Handwriting Analysis is not document examination, which involves the examination of a sample of handwriting to determine the author. 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 describe the corresponding personality trait.

However, most of the works dealt with the graphology of Latin scripts. However progress in Farsi (or Arabic) script graphology has been slow mainly due to the special characteristics of Farsi scripts. Farsi text is inherently cursive both in handwritten and printed forms and is written horizontally from right to left. Farsi writing, which this paper addresses, is very similar to Arabic in terms of strokes and structure. The only difference is that Farsi has four more characters than Arabic in its character set. Therefore, Graphology for Farsi Handwriting can also be used for Arabic handwriting graphology.

## II. Related Work

As described, handwriting Analysis or Graphology is a scientific method of identifying, evaluating and understanding personality through the strokes and patterns revealed by handwriting**.** Among the many aspects of handwriting that can serve as scheme to predict personality traits are baseline, size of letters, writing pressure, connecting strokes, spacing between letters, words and lines, starting strokes, end-strokes, word-slant, speed of handwriting, width of margins, and others [1],[2],[3]. Writer individuality rests on the hypothesis that each individual has consistent handwriting, which is distinct from the handwriting of another individual. However, this hypothesis has not been subjected to rigorous scrutiny with the accompanying experimentation, testing, and peer review [4] [5] [6] [7] [8] [9].

## III. Data Acquisition and Image Pre Processing

The research population consisted of the students in Islamic Azad University of Abhar. The input database includes 120 handwriting samples from 120 different writers. The writers were made to write the given text. The text is a simple paragraph that includes all possible characters of the Farsi alphabet. the samples were written on A4 size paper without any lines. The handwriting samples were scanned with the scanner whose resolution is 300dpi. In order to speed up the process, only the upper one-third of the page is used. The preprocessing steps include: pen width extraction, noise and scratch removal [10],[11].

## IV. The Set of Features

When processing handwriting, different variations of one individual's handwriting should be considered. The optimal sample is one, which has been written without prior knowledge of its usage and special effort. During the process of analyzing the personality of an individual, the graphologist must consider a number of parameters. The most important features of handwriting are as follows [12]:

- **Left and right page margins:** The margins of a handwritten sample can take different forms, each of which has a specific meaning. For example, large and equal margins on both sides of a page show a law-abiding personality and good management characteristics.
- **Word expansion:** In graphology, a text with expanded words represents an honest and trustworthy personality.
- **Letter size:** A text may have small size or large size letters. A text with large letters indicates an extrovert personality while a text with small letters represents an introvert personality.
- **Line and word spacing:** According to graphologists a text with small line spacing belongs to a more narrow-minded individual or a "collector". Large line spacing represents a person who can make open-minded and situation-specific decisions. In other words, word spacing shows the extent to which an individual is close to his/her social environment.
- **Line skew:** The lines with an upward orientation indicate an optimistic character. On the other hand, downward orientation belongs to pessimistic characters.
- **The ratio of vertical to horizontal elongation of words**: A text with a high vertical elongation in comparison to horizontal one represents an individual with high ideals. The opposite represents a self-satisfied personality.
- **Slant:** The slant to the left represents a warm and friendly disposition, whereas the slant to the right represents a pessimistic and shy disposition.

Methods for extracting the above mentioned features are presented in Sec. 5.

## V. Feature Extraction and Analysis

The main features are extracted as follows [13].

### 5.1 Right and Left Page Margins

Farsi is written from right to left. The right margin can be progressive, regressive, aligned, convex, concave or irregular. The left margin can be aligned or irregular. To speed up the process, the resolution is reduced to 40dpi. The end row points of the image are detected. These points are dilated 10 pixels upward and 10 pixels downward. Then, the side pixels of the dilated image are extracted and known as the page margin. The right and left margins are smoothed using the windows of sizes 3 and 10 respectively. In order to analyze the margin shape, the first and second derivatives are used. If the first derivative for the right margin is positive, negative or zero, the curve is ascending, descending or straight, respectively, and therefore the shape of the right margin is progressive, regressive or aligned. The convexity of the right margin's shape is determined from its second order derivative. If the first derivative for the left margin is zero or nonzero then the shape of the left margin is aligned or irregular respectively.

### 5.2 Right and Left Page Margins

Words in a text can be normal or expanded. In order to determine the degree of expansion, the connected components is labeled [14], then their area is calculated and those less than 20 times of the pen width square are eliminated. The mean of the remaining areas is divided by the pen width [13]. This is used as an index of the word expansion. Considering the tests taken, if the expansion index is higher than 300, the text is classified as expanded.

### 5.3 Letter Size

The size of letters in handwriting is large, small, or normal. To define a factor related to letter size, the gravity centers of the connected components are found. The image is divided into 14 equal horizontal bands in each band; the minimum distance between each center and other centers is calculated. The median of the minimum instances in all bands is used as an index for letter size [13]. Considering the tests taken, the index less than 57 represents handwriting with small letters. If it is more than 61 the letter size is large.

### 5.4 Line and Word Spacing

Text density is affected by line or word spacing.

#### 5.4.1 Line spacing

The line spacing can be narrow, wide, or normal. To calculate an index for the line spacing, the resolution is reduced to 20 dpi. Then the convex hull of the whole shape is drawn. The number of black pixels is divided by the product of the area of the convex hull and the pen width. The obtained value is considered as an index of line spacing [13]. If the low-resolution image has less than 10 black pixels, the line spacing will be considered “ambiguous”. Taking into account our handwritten samples, the line-spacing index for the texts with lines far apart is less than 1. This factor is higher than 3 for the handwritings with close lines.

#### 5.4.2 Word Spacing

The word spacing can be wide, narrow, or normal. To calculate an index for word spacing, the image is divided into 14 equal horizontal bands. In each band, the connected components are labeled. Then the median of distances between side-walls of neighboring bounding boxes of the connected components is calculated. The obtained value is divided by the expansion factor derived in Sec. 5.2. This value is used as an index for word spacing [13]. According to the tests taken, for the handwritings with words closed to each other, this factor is less than 0.2, while it is more than 0.26 for handwritings with words that are far apart.

### 5.5 Line Skew

In graphology the value of line skew is interpreted in the following intervals: more than 6 degrees, between 6 and 2 degrees, between 2 and -2 degrees, between -2 and -6 degrees, and less than -6 degrees [13]. There are many algorithms available for skew correction, most of which are based on baseline extraction [15],[16]. In this work a very simple algorithm is used, because it is not intended for character recognition. In the method used [16], the handwriting sample is rotated in the range of [-8...8] degrees with steps of 1 degree. For each rotation, the horizontal projection is calculated and the entropy of this projection is determined. The angle corresponding to minimum entropy is considered as the line skew [17].

### 5.6 The Ratio of Vertical to Horizontal Elongation of Words

Handwritings can be elongated in vertical or horizontal directions. On the other hand they can be relatively proportional.

To define an index for the ratio of vertical to horizontal elongation, first, the skew of the handwriting is corrected through the rotation of the image with the degree obtained in Sec. 5.5. Then the image is divided into 6 equal vertical bands. The horizontal projection is calculated for each band. With respect to zero points in this projection, each band is divided into a number of horizontal bars .The bars with heights less than 5 times the pen width are ignored. Each remaining bar is divided into upper and lower parts based on the maximum amount of its horizontal projection .The upper parts with the height of less than 5 times the pen width are eliminated and the median of the height of remaining parts is considered as the ascender length. The descender length is calculated in the same way from the lower parts. The sum of ascender and descender lengths is considered as vertical elongation index.

In order to calculate the horizontal elongation, the expansion index derived from Sec. 5.2 is divided by vertical elongation and pen width. In a different method, the bounding box for each connected component with an area of more than 20 times the pen width is drawn .The mean of the widths of the bounding boxes divided by the pen width is considered as the second index for horizontal elongation. The product of these two horizontal elongation indices determines the horizontal elongation factor that is used in the ratio of vertical to horizontal elongations. according to the tests taken, an elongation ratio of less than 0.14 shows high horizontal elongation. On the other hand, a ratio of more than 0.55 indicates a handwriting with high vertical elongation.

### 5.7 Slant

In graphology the value of slant is interpreted for the following intervals: more than 110 degrees, between 110 and 95 degrees, between 95 and 80 degrees, between 80 and 65 degrees, and less than 65 degrees [13].There are many algorithms available for slant calculation, most of which are based on contour extraction and chain code implementation, both of which require considerable preprocessing [14],[18],[19],[20]. In this paper the following algorithm is used for slant estimation [15]. First, in each row, the horizontal lines longer than 1.5 times the pen width are eliminated. Then the image is divided into 40 equal vertical bands. In each band, the horizontal projection is calculated. With respect to zero points in this projection, each vertical band is divided into some horizontal bars.

The bars with a height of less than 5 times the pen width are eliminated. Each of the remaining bars is equally divided into upper and lower zones. The angle between the centers of gravity of these zones is taken as the slant of that bar [15]. The mode of the slants of all bars is considered as the most prevalent slant.

## VI. Classification Using SVM

After the feature extraction, to predict the personality 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 the prediction of personality we have to train a classifier for each feature. SVM can be used to classify the data of unknown data class into the correct data categories.

### 6.1 Overview Of SVM

SVMs are set of related supervised learning methods used for classification and regression [21]. They belong to a family of generalized linear classification. A special property of SVM is, SVM simultaneously minimize the empirical classification error and maximize the geometric margin. So SVM called Maximum Margin Classifiers. SVM is based on the Structural risk Minimization (SRM). SVM map input vector to a higher dimensional space where a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyper plane that separate the data. The separating hyper plane is the hyper plane that maximizes the distance between the two parallel hyper planes. An assumption is made that the larger the margin or distance between these parallel hyper planes the better the generalization error of the classifier will be [21].

We consider data points of the form

$$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), ............(x_n, y_n)\}$$

Where $y_n = +1/-1$, a constant denoting the class to which that point $x_n$ belongs. n = number of sample. Each $x_n$ is p-dimensional real vector. The scaling is important to guard against variable (attributes) with larger variance. We can view this Training data, by means of the dividing (or separating) hyper plane, which takes

$$W.X + b = 0 \qquad (1)$$

Where b is scalar and w is p-dimensional Vector. The vector w points perpendicular to the separating hyper plane. Adding the offset parameter b allows us to increase the margin. Absent of b, the hyper plane is forced to pass through the origin, restricting the solution. As we are interesting in the maximum margin, we are interested SVM and the parallel hyper planes. Parallel hyper planes can be described by equation:

$$W.X + b = (+/-)1 \qquad (2)$$

If the training data are linearly separable, we can select these hyper planes so that there are no points between them and then try to maximize their distance. By geometry, we find the distance between the hyper plane is $2/|W|$ So we want to minimize $|W|$. To excite data points, we need to ensure that for all $i$ either

$$W.X_i - b \geq 1 \quad \text{or} \quad W.X_i - b \leq -1 \qquad (3)$$

This can be written as:

$$y_i(W.X_i - b) \geq 1 \quad , \quad 1 \leq i \leq n \qquad (4)$$

Samples along the hyper planes are called Support Vectors (SVs). A separating hyper plane with the largest margin defined by $M = 2/|W|$ that is specifies support vectors means training data points closets to it. Which satisfy:

$$y_i[W^T.X_i+b]=1 \quad , \quad i=1 \tag{5}$$

Optimal Canonical Hyper plane (OCH) is a canonical Hyper plane having a maximum margin. For all the data, OCH should satisfy the following constraints:

$$y_i[W^T.X_i+b]\geq 1 \quad , \quad i=1,2,....,l \tag{6}$$

Where l is Number of Training data point. In order to find the optimal separating hyper plane having a maximal margin, A learning machine should minimize $\|W^2\|$ subject to the inequality constraints equation(6). This optimization problem solved by the saddle points of the Lagrange's Function:

$$L_p=L_{(W,b,\alpha)}=\tfrac{1}{2}\|W\|^2-\sum_{i-1}^{l}\alpha_i(y_i(W^T.X_i+b)-1)=\tfrac{1}{2}W^TW-\sum_{i-1}^{l}\alpha_i(y_i(W^T.X_i+b)-1) \tag{7}$$

Where $\alpha_i$ is a Lagrange's multiplier .The search for an optimal saddle points ($W_0,b_0,\alpha_0$) is necessary because Lagrange must be minimized with respect to w and b and has to be maximized with respect to nonnegative $\alpha_i$ ($\alpha_i \geq 0$). This problem can be solved either in primal form (which is the form of w & b) or in a dual form (which is the form of $\alpha_i$ ).Equation number (6) and (7) are convex and KKT conditions, which are necessary and sufficient conditions for a maximum of equation (6). Partially differentiate equation (7) with respect to saddle points ($W_0,b_0,\alpha_0$).

$$\partial L/\partial W_0=0 \quad \text{ie. } W_0=\sum_{i=1}^{l}\alpha_i.y_i.x_i \tag{8}$$

And

$$\partial L/\partial b_0=0 \quad \text{ie. } \sum_{i=1}^{l}\alpha_i.y_i=0 \tag{9}$$

Substituting equation (8) and (9) in equation (7). We change the primal form into dual form.

$$L_d(\alpha)=\sum_{i=1}^{l}\alpha_i-1/2\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i.\alpha_j.y_i.y_j.x_i^Tx_j \tag{10}$$

In order to find the optimal hyper plane, a dual lagrangian ($L_d$) has to be maximized with respect to nonnegative αi (i .e. $\alpha_i$ must be in the nonnegative quadrant) and with respect to the equality constraints as follow:

$$\sum_{i=1}^{l}\alpha_i.y_i=0 \quad , \quad \alpha_i\geq 0 \quad , \quad i=1,2,....,l \tag{11}$$

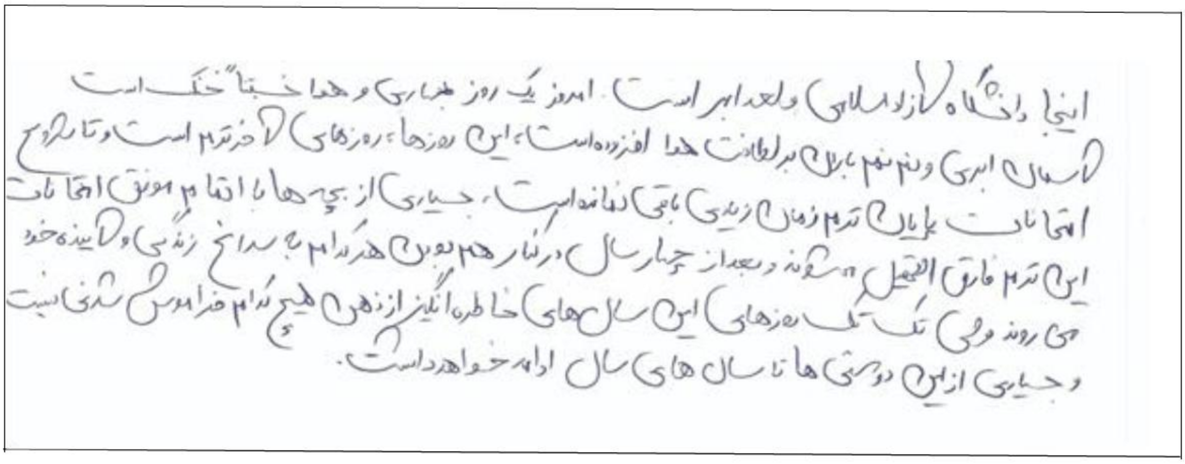
Note that the dual Lagrangian $L_d(\alpha)$ is expressed in terms of training data and depends only on the scalar products of input patterns ($x_i^T.x_i$).More detailed information on SVM can be found in Reference no.[21],[22].

### 6.2 A sample analysis

Decisions about the personality can be taken based on the rules defined in Sec.4 and features acquired in Sec. 5 by using SVM classifier. The classification can be performed in following three steps:

1. First, the input features 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.

A sample of handwriting, and its graphological analysis obtained from our system are presented in Fig. 1 and table1, respectively.



اینجا دانشگاه آزاد اسلامی واحد ابهر است. امروز یک روز بهاری و هوا نسبتاً خنک است
آسمان ابری و نم نم باران بر لطافت هوا افزوده است. این روزها، روزهای آخر ترم است و تا شروع
امتحانات پایان ترم زمان زیادی باقی نمانده است. بسیاری از بچه ها با اتمام موفق امتحانات
این ترم فارغ التحصیل می شوند و بعد از چهار سال در کنار هم بودن هر کدام به سراغ زندگی و آینده خود
می روند ولی تک تک روزهای این سال های خاطره انگیز از ذهن هیچ کدام فراموش شدنی نیست
و بسیاری از این دوستی ها تا سال های سال ادامه خواهد داشت.

**Fig. 1.** A sample input of the proposed system.

**Table 1.** The output of the proposed system to Fig. 1

| NO | Feature | personality review |
|---|---|---|
| 1 | page margins | law-abiding personality and good management characteristics |
| 2 | Word expansion | honest and trustworthy personality. |
| 3 | Letter size | extrovert personality |
| 4 | Line and word spacing | narrow-minded individual or a "collector". |
| 5 | Line skew | optimistic character |
| 6 | Slant | warm and friendly disposition |

## VII. Conclusion

From 120 handwriting samples, the training set of 30 samples was used to train SVM for determining the thresholds. The remaining samples were used in the test set. Learning samples have been selected so that they have all necessary characteristics. by using SVM an expert labeled the samples based on the left and right page margins, word expansion, letter size, line and word spacing, line skew, the ratio of vertical to horizontal elongation, and slant. The results of automatic analysis and comparison with the graphologist's opinion for 30 training samples and 118 test samples are presented in table 1.

After taking the handwriting sample image we have extracted different features. All these features are given to SVM which predict the personality of the individual writer.

Feature extraction from handwriting is difficult and entails a high degree of uncertainty. This paper showed that a number of graphological features for Farsi automatically extracted. An exact and precise extraction of features and their analysis can be very useful for graphologists and any other users of this system. This system showed promising results.

## Acknowledgements

This work has been supported by the Islamic Azad University, Abhar Branch, Abhar-Iran. Authors wish to thank students in the communication lab. In the Islamic Azad University of Abhar for cooperation and valuable development work.

## References

[1]. Champa H N, K R AnandaKumar, **"A Scientific Approach to Behavior Analysis through Handwriting Analysis"**, National Conference on Research Trends in Information Technology, S R K R Engineering College, Bhimavaram, Andhra Pradesh, 2008.

[2]. N Mogharreban, S Rahimi, M Sabharwal, "**A Combined Crisp and Fuzzy Approach for Handwriting Analysis**", IEEE Annual Meeting of the Fuzzy Information,2004, vol1, pp 351- 356,

[3]. Sung-Hyuk and Charles C Tappert , "**Automatic Detection of Handwriting Forgery**", Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition, 2002, vol1, pp.351-356.

[4]. K Toraichi, T Horiuchi, R Haruki, "**Observation Method for Mathematical Graphology**", Proceedings of the Third International Conference on Document Analysis and Recognition, 1995, vol2, pp.656-659.

[5]. Sung-Hyuk Cha, Sargur N Srihari, "**Apriori Algorithm for Sub-category Classification Analysis of handwriting",** Proceedings of the Sixth International Conference on Document Analysis and Recognition, 2001, pp. 1022-1025.

[6]. H.E.S.Said, T.N.Tan and K.D.Baker, "**Writer Identification Based on Handwriting",** IEEE Third European workshop on Handwriting Analysis and Recognition, vol.33, no.1, 2000, pp 133-148.

[7]. Philip Jonathan Sutanto, Graham Leedham and Vladimir Pervouchine, "**Study of the Consistency of some Discriminatory Features used by Document Examiners in the Analysis of Handwritten letter 'a' ",** Proceedings of the seventh International Conference on Document Analysis and Recognition, 2003.

[8]. Saragur N Srihari and Zhixin Shi, "**forensic Handwritten Document Retrieval System**", Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL'04), 2004.

[9]. Ameur BENSEFIA, Ali NOSARY, Thierry PAQUET, Laurent HEUTTE,"**Writer Identification by Writer's Invariants",** Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition", 2002, pp 274-279.

[10]. B.Vaseghi , S.Hashemi 'Offline signatures Recognition system using Discrete Cosine Transform' Australian Journal of Basic and Applied Sciences, 6(12): 423-428, 2012 ISSN 1991-8178

[11]. Vaseghi.B., Alirezaee.Sh., "Off-line Farsi/Arabic Handwritten word recognition using vectorquantization and hidden markov model," Proceedings of the 12th IEEE International Multitopic Conference, 978-1-4244-2824-3/08/$25.00

[12]. G. Beauchataud, "Apprenez la graphologie", Farsi translation by A. Yalda, Ketabsara Publishers, 1998.

[13]. A. Bahramisharif, "Computer aided graphology for Farsi handwriting", M.Sc thesis, Dept. of Electrical Engineering, Tarbiat Modarres Univ., Winter 2005. (in Farsi)

[14]. R. C. Gonzalez, R. E. Woods, "Digital image processing", Prentice-Hall Inc., Pearson Education, USR New Jersey, 2002.

[15]. Alessandro Vinciarelli, "A survey on off-line cursive script recognition", IDIAP Research Report, IDIAP-RR00- 43, 2000.

[16]. R.M.Bozinovic, S.N.Srihari, "Off-line cursive script word recognition", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 11, no. 1, pp 68-83, 1989.

[17]. M.Cote, E.Lecolinet, M.Cheriet, C.Y.Suen, "Automatic reading of cursive scripts using a reading model and perceptual concepts-the PERCEPTO system", Int. J. of Document Analysis and Recognition, IJDAR,vol. 1, no. 1, pp 3-17, 1998. J.Cai, Z.Q.Liu, "Off-line unconstrained handwritten word recognition", Int. J. of Pattern Recognition and Artificial Intelligence, vol. 14, no. 3, pp 259-280, 2000.

[18]. J.Cai, Z.Q.Liu, "Off-line unconstrained handwritten word recognition", Int. J. of Pattern Recognition and Artificial Intelligence, vol. 14, no. 3, pp 259-280, 2000.

[19]. A.El-Yacoubi, M.Gilloux, R.Sabourin, C.Y.Suen, "An HMM based approach for off-line unconstrained handwritten word modeling and recognition", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 21, pp 759-760, 1999.

[20]. E.Kavallieratou, N.Fakotakis, G.Kokkinakis, "A slant removal algorithm", Pattern Recognition, vol. 33, pp 1261- 1262, 2000.

[21]. V. Vapnik. The Nature of Statistical Learning Theory. NY: Springer-Verlag. 1995.

[22]. Boser, B. E., I. Guyon, and V. Vapnik (1992). A training algorithm for optimal margin classifiers . In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pages. 144 -152. ACM Press 1992.