2009 10th International Conference on Document Analysis and Recognition

Graphological Analysis of Handwritten Text Documents for Human Resources Recruitment

Ricard Coll, Alicia Fornés, Josep Lladós
Computer Vision Center - Computer Science Department
Universitat Autònoma de Barcelona, Spain
{rcoll, afornes, josep}@cvc.uab.es

Abstract

The use of graphology in recruitment processes has become a popular tool in many human resources companies. This paper presents a model that links features from handwritten images to a number of personality characteristics used to measure applicant aptitudes for the job in a particular hiring scenario. In particular we propose a model of measuring active personality and leadership of the writer. Graphological features that define such a profile are measured in terms of document and script attributes like layout configuration, letter size, shape, slant and skew angle of lines, etc. After the extraction, data is classified using a neural network. An experimental framework with real samples has been constructed to illustrate the performance of the approach.

1. Introduction

Graphology, since 1871, is a science based on psychology and statistics that study concrete features from handwriting to find the psychological profile of the writer. Graphology holds that writing is a reflex action from our brain, so it shows our personality and the frame of our mind [12]. Graphological analysis of handwritings have proven to be effective in applications like employment profiling, medical diagnosis or jury screening. In recruitment processes, the graphological tests used by human resources departments usually are complementary and even more useful than interviews. They allow to map some features from writing samples to a personality profile, matching the congruency of the applicant with the ideal psychological profile of employees in the position. Graphology can show to the interviewer some hidden aspects of the applicant's personality and it allows to find any desired skill the job requires, like leadership, teamwork or organization capabilities.

A graphological analysis requires a complex interpretation process. Following the graphology premise that handwriting is personal and unique, there are many handwriting characteristics that phychologists analyze for a personal screening. Examples of features and related interpretation according to [4] are slant (emotional expressiveness), text skew (optimism and energy), shape of strokes (aggressiveness), shape of individual letters (position of t bar, or i dot is related to self esteem), pressure while writing (stress), etc.

An interesting observation is that Document Image Analysis (DIA) community provides algorithms to extract a number of features used in graphology. Outstanding approaches exist in the literature of handwriting recognition, writer identification and handwriting forensics that can be adapted to automatic graphological analysis. The analysis of graphometric features has been used in signature verification [8]. In writer identification, Bensefia *et al.* [1] use the morphology of the letter, or parts of it, to find individual graphemes in order to identify a writer, exploiting its redundancy characteristic. Some other researches have created a set of features related to the slant and skew angle, the letter body size and the behavior of the letter when fractal geometry is applied [7][5]. In [9], the authors use Gabor filters to analyze textural characteristics of the text.

In this work, we propose a model to map a number of standard handwriting recognition features to a personality screening of the writer. In particular, our model is validated in an employment profiling application for a human resources company. We have defined our model in an scenario in which after extracting a number of image features from handwriting samples, we make an hypothesis to map them into a profile in terms of active, inquisitorial personality, with working motivation and leadership, and team coordination skills. The proposed application, after extracting the image features, classifies the writer in the corresponding class, according to the personality skills of our model. A neural network approach is used for classification.

The rest of this paper is organized as follows. First, in section 2 we describe the graphological attributes considered in our employment profiling model. In section 3

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.213

we propose the hypothesis of mapping handwriting image features to graphological attributes, and the corresponding classification application to give the final writer profile. Section 4 provides experimental results. Finally, section 5 is devoted to conclusions and final discussion.

2. Graphological attributes

As stated above, the main contribution of this paper is to define a model that maps writing features from sample images to relevant graphological interpretations. These interpretations are considered in an application scenario of employment profiling for a human resources company. The skills that we want to automatically assess are active, inquisitorial personality, with working motivation and leadership, and team coordination skills. In table 1 we present the above skills and the corresponding graphological features that should be found in the person handwriting.

Personality Skills	Graphological Features
Without personality conflicts	Proportional writing. Readable writing. Without Squiggles and irregular strokes.
Integral personality	Letter size. Script slant. Writing speed.
Emotional health	Relation between ascendants and descendants. Line skew. Script slant. Normal letter size.
Organization, clarity of ideas	Good document layout organization, harmony. Readability.
Adaptability	Linked letters
Sincerity	Good document layout organization. Readability. Proportional margin (10% of the sheet).
Activity	Angular writing. High pressure. Medium or high letter size. Letter slant to the right. Horizontal or upward text line orientation.
Aggressiveness	Angular writing. Writing speed. High pressure.
Ambition	Uneven pressure. Writing speed. Upward text line orientation.
Self-confidence	Important pressure. Well organized document layout. Letter size between 3 mm. and 4 mm.

Table 1. Correspondences between personality characteristics and graphological features.

In our application scenario, the graphological features considered in the classification of a person profile are:

Document layout organization. To evaluate the goodness of document layout organization the following parameters are considered: margins size (between 10% and 25%), constant left margin, regular inter-line space, and regular line orientation.

Proportional and regular script. A well-proportioned script has letters whose body size is a third of the ascendant size and half of the descendant size. In addition, letters have normal size (between 2.5mm and 3.5mm) and regular in the whole document.

Document harmony. Without variation in inter-line space (around 1cm), slant and slope. Regular letter size (see above paragraph).

Links between letters. Non connected letters denote low adaptability to different situations, and a trend to the monotony [12]. Connected letters indicate a slow speed writing. A medium connectivity determines a right adaptability.

Script readability. A readable handwritten text means that the script does not have squiggles or irregular strokes. This feature is used for determining the clarity and sincerity of the person in front of the rest of the world [3].

Shape of the strokes. Circular handwriting indicates an agreeable, easygoing nature. In our application scenario, we analyze angular handwriting with sharp points, that indicates aggressiveness, directness, and high energy.

Pressure while writing. In our scenario, a moderate-to-heavy pressure is required. The heavier the pressure, the more intense the emotions of that person and his self-confidence.

3. Image Feature Extraction

Let us now describe the features extracted from the handwriting image, and how the attributes, presented in section 2, are measured from the handwriting image features.

Slant normalization and variations in script slant. In the preprocessing step, the slant normalization described by Vinciarelli in [11] is performed: First, a density histogram is computed, and using a shear function, we compute the square of the number of pixels belonging to a stroke in respect to the total of the column of the image. When the majority of the strokes are vertical, then, this value becomes a maximum. There are other local maxima values that correspond to the different variations that can appear in the writing, because not all the vertical strokes are drawn in the same direction. This feature is also important for evaluating the organization of the document.

Slope detection and variations in the slope. In the preprocessing step we also get information of the slope (text orientation). We have adapted the slope normalization technique described by Vinciarelli [11]. In this case, we are also interested in this feature, because it is used for evaluating the organization of the document and its harmony, detecting changes in slope in the different text lines. We also use as features the local maxima that appear in the text, which indicates the different slope orientations. This feature is indicative of personal moods.

Line spacing and variation of line spacing. Once the slant and slope of the image is normalized and in order to compute the line spacing, a horizontal projection of the image is done. Afterwards, we compute the mean and the standard deviation as important features for determining the organization and harmony of the document.

Size of margins, variation of margins, increasing or decreasing the sizes of margins. The upper and left margins are the most important size margins, because they provide graphological information. The upper size margin is computed at the same time that we compute line spacing, whereas the left margin is computed when we compute the word spacing, which is an important feature for evaluating the harmony, the text readability and the organization. In addition, we compute whether the left margin is increasing, decreasing or constant in the text. This latter feature shows that the writer has good planification and future vision skills.

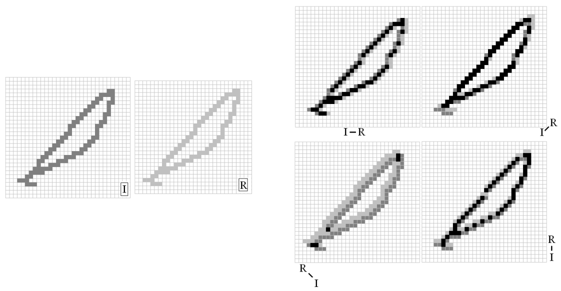
Size of the script: main body, ascendant and descendant. For the evaluation of the proportionality and regularity of the script, an horizontal projection of the text line is performed. The highest pixel's density corresponds to the main body of the script, whereas the upper and lower zone correspond to the ascendant and descendant, respectively. We are interested in the relation between the main body and ascendants and descendants of the script.

Image contours. If the letters are disconnected, the writer shows little adaptability to different situations, and in some cases, the writer tends to monotony [12]. If all the letters are connected, then the writing can be slow. A medium connectivity between letters shows a right adaptability and a good writing speed. In [10], the relation between the number of internal and external contours that can be extracted from the text shows the movement that has been done while writing. A speedy writing will have a lower number of internal contours than external contours because the letters will be probably longer, uncompleted and sometimes without links between letters of a word. Besides the writing speed, this relation between internal and external contours indicates the job adaptability degree of the applicant.

Angular writing. We present in this paper a new method for determining the roundness factor of the writing. This method takes into account that if we have two identical circles one over the other, and we perform a slight translation in one of them, then we will find their common points. If we repeat this step for all possible directions, we obtain the same number of common points in each translation. However, if we have two straight lines, once we translate one over the other, we will only obtain a common points in case the orientation of the lines are the same. The steps of the method are the following:

1. Let's define V as a four directional vector, where we translate R over I.
   $V = \{(-1,-1),(0,-1),(1,-1),(1,0)\}$.
2. Take a $(N-2)*(N-2)$ region, being N the size of each side of the original image I. This new region is called R.
3. For each vector v_i, $v_i \in V$.
   3.1 Translate R over I in the orientation and module of the vector.
   3.2 Perform a logical AND operation to find the common points.
   3.3 Compute the number of active points in the image, which correspond to the common points of R over I. The result is normalized over the total number of active pixels of the image.
4. The roundness factor will be the variance of the normalized common points.

Whether the writing is rounded, the number of common points in all the orientations will have a low value, whereas in an angular writing, the variation of the common points ratio between the translations of R over I can have a higher value. We show in Fig.1 how the different translations are performed and the common points between both images (black pixels).



(a) (b)

Figure 1. Roundness factor: a) Image I and region R from a writing character and b) different transpositions of R over I

Fractal Dimension. Mandelbrot exposes in [6] that the fractal dimension is used for measuring the fragmentation and irregularity degree of objects. In [2] a method for computing the readability degree of a writing as a biometrical feature of a person is shown. This method consists in mesuring how the area grows when we apply a dilation operation to a binary image successively. First of all we choose a disk kernel of radius $\mu(\mu = 1..25)$. Every time we apply a new dilation over the writing the value of $\log(A(\mu)) - \log(\mu)$ is recorded as a function (called evolution graph) of $\log(\mu)$, where $A(\mu)$ is the area of the dilated image (*i.e.*, the number of active pixels). Three parts with different slopes can be observed in this evolution graph. The interest is focused in the second segment, specifically in its slope. When we apply this method on a bad formed writing, we obtain a greater slope value than if we apply on a well-formed writing, as figure 2 shows.

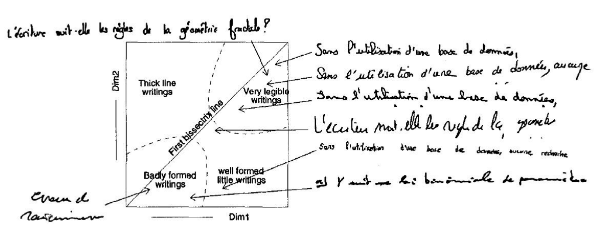


Figure 2. Distribution of family writings in a legibility graph [2]

Regularity in the writing frequency. To compute the auto-control, aggressiveness and impulsivity degree of the writer, a feature that computes the regularity in the writing frequency can be used. The regularity is not always the same (i.e. the width of the letters can vary in different locations of the writing, and also the letters' spacing and word's spacing...). For computing this feature, we apply the Discrete Fourier Transform (DFT) to each text line. Once the maximum magnitude of frequency for each line is computed, we extract the standard deviation λ. When λ has a low value, the text is very regular. Contrary, when λ has a high value, the width of the writing is very irregular.

Measure and variation of pen pressure. Although there are difficulties in obtaining parameters that can characterize the pressure of the writing in an offline system, there are some published works which use a combination of features that show the pressure of the writing (Three samples with different pressure are shown in figure fig:pressio). In [10] the pressure factor is computed taking into account three features: the threshold that has been used for binarizing the image, the distribution of the gray values using the entropy and the number of black pixels. We have used these three features and we have also added a feature that quantifies the pressure in the angular points of each word.

(a) (b) (c)

Figure 3. Pen Pressure: a) lighter, b) normal, uneven and c) higher pressure.

Backpropagation Neural Network. We have used that model to perform the classification. All the 25 extracted features from the image are the input pattern for a full-connected and feed-forward neural network. Once trained, for each image we want to analyze, the output of the net is a binary value that indicates if the input pattern belongs to the desired personality profile or does not.

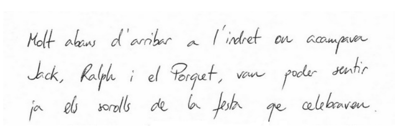
4. Experimental Results

For the experimentation stage, we took 98 samples from different people of ages between 12 and 85 years old, regardless of their gender or socioeconomic status. Samples were taken in three languages (Catalan, Spanish and English), according to the native language of the writer. The length of the text was about 90 words.

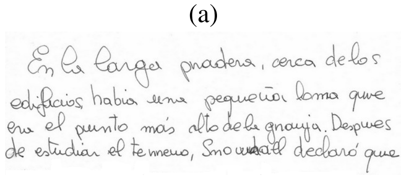
The image was acquired with a flatbed scanner, with a resolution of 300 dots per inch, and 8 bit grayscale. These samples were sent to a team of expert graphologists who advised us in order to create an acceptable ground-truth. Every sample was evaluated with a 0 or a 1 depending on the presence of each feature. The features observed in the writing were the following: well-proporcioned size of the letter, clear and harmonic writing, legibility, normal and regular size of the letter and margins, normal or high (preferably uneven) writting pressure, vertical or right-slanted words, linked letters, normal or fast writting speed, horizontal or growing line slope, well-organized text and angular-like letter form. Finally we split into suited or unsuited samples for the job, according to the score achieved in the sum of individual characteristics.

We used the backpropagation neural network to classify the samples after dividing them randomly among two groups aimed to training and testing (50 and 48 samples each). We got a correct classification ratio of 89%. Unfortunately we are not able to compare this result because it has not come to our knowledge any other previous work about graphological analysis. Nevertheless we can compare this ratio with the statistically proved rate of success -99%- that graphologists might have in their analysis. We want to highlight that selection criteria based on a set of features and not necessarily the same set for each writing. This could be the cause of our error rate. The learning of the net is weight-based: if the number of the samples of a concrete type of handwriting is not enough (i.e. there are not a normal distribution of all training patterns) this network will be wrong with its prediction.

In figure 4 we show the diferent pressure, slant and slope direction, harmony, writing regularity and proporcionality between two samples studied in this framework.



(a)



(b)

Figure 4. Two handwritten samples: sample a) is suitable for the hypothetical job, sample b) does not.

5. Conclusions

In the present paper we have presented a simple method to extract, analyze, correlate and classificate writtings from a psychological approach. This method mainly depends on the features which are looked for on the writtings. A valid ground-truth is essential for our decission algorithm to classificate the samples correctly since it is based on concrete and measureable graphometric features. We must also say there are some factors which are difficult to model on a quantitative way, such as intuitive decissions graphologysts make, fruit of their professional experience, when facing a writting. Besides we have proposed some features such as the roundness factor and frequencial analysis of word's core region, that can be useful in other fields as writer identification since they differ from one person to another.

The method presented provides HR professionals with a fast and useful tool, more reliable than personal interviews, that allows not only to improve selection filters but saving time and efforts in recruitment processes too.

6. Acknowledgements

The authors would like to thank Solid Selection S.L. for their advice on graphological matters.

This work has been partially supported by the Spanish projects TIN2006-15694-C02-02 and CONSOLIDER-INGENIO 2010 (CSD2007-00018).

References

[1] A. Bensefia, A. Nosary, T. Paquet, and L. Heutte. Writer identification by writer's invariants. *Proceedings of the 8th International Workshop of Frontiers in Handwriting Recognition*, pages 274–279, 2002.

[2] V. Bouletreau, N. Vincent, R. Sabourin, and H. Emptoz. Synthetic parameters for handwriting classification. *Proceedings of the 4th International Conference on Document Analysis and Recognition*, 1:102–106, 1997.

[3] M. R. Fiszbein. *Guía práctica de grafología*. Collecció integral - RBA, Barcelona, 2003.

[4] R. Gardner. *Instant Handwriting Analysis: A Key to Personal Success*. Llewellyn Publications, 2002.

[5] C. Hertel and H. Bunke. A set of novel features for writer identification. *Lecture Notes in Computer Science*, 2688:679–687, 2003.

[6] B. Mandelbrot. *Les objets fractals*. Flammarion, 1984.

[7] U. V. Marti, R. Messerli, and H. Bunke. Writer identification using text line based features. *Proceedings of the 6th International Conference on Document Analysis and Recognition*, pages 101–105, 2001.

[8] L. S. Oliveira, E. Justino, C. Freitas, and R. Sabourin. The graphology applied to signature verification. *Proceedings of the 2007 ACM symposium on Applied computing*, pages 637–638, 2007.

[9] H. E. S. Said, K. D. Baker, and T. T. N. Personal identification based on handwriting. *Proceedings of the 14th International Conference on Pattern Recognition*, 2:1761–1765, 2001.

[10] S. N. Srihari, S. Cha, H. Arora, and S. Lee. Individuality of handwriting. *Journals of Forensic Science*, 47(4), 2002.

[11] A. Vinciarelli and J. Luettin. A new normalization technique for cursive handwritten words. *Pattern Recognition Letters*, 22, 2001.

[12] M. Xandro. *Grafologia Elemental*. Herder, 1994.