The Graphology Applied to Signature Verification

Luiz S. OLIVEIRA [a] , Edson JUSTINO [a] , Cinthia FREITAS [a] and Robert SABOURIN [b]

[a] *Pontifícia Universidade Católica do Paraná, Curitiba, BRAZIL*

[b] *Ecole de Technologie Superieure, Montreal, CANADA*
{soares,justino,cinthia}@ppgia.pucpr.br, robert.sabourin@etsmtl.ca

Abstract. In this paper we discuss automatic signature verification in the context of the graphology. Graphology is claimed to be useful for everything from understanding health issues, morality and past experiences to hidden talents, and mental problems. It is not restricted to this, though. Forensic document examiners use the concepts of graphology to examine handwriting in order to detect authenticity or forgery. In this work, we describe some of the main features of the graphology and propose a set of features to automatic signature verification. They are evaluated in a database of 5,600 signatures using hidden Markov models.

Keywords: Graphology, Graphometry, Automatic Signature verification.

1. Introduction

The handwriting has been studied for almost 400 years. The first person that carried out systematic observations on the manner of handwriting was Camillo Baldi in 1622. He published the book entitled "Treated how, by a letter missive, one recognizes the writer´s nature and qualities", which is considered the first known graphological essay. The term "graphology" was coined by Abb Jean-Hippolyte Michon in Paris in 1897 by merging two Greek words *graphein*, to write and *logos*, science. He was also the founder of The Society of Graphology and the first one to give scientific bases to the analysis of handwriting. The Michon´s work was continued by one of his pupils, J. Crépieux-Jamin. He put of the order in Michon´s work and divided the writing into seven fundamental elements: speed, pressure, form, dimension, continuity, direction, and order [1].

A branch of the graphology is the psychometrical graphology or graphometry. This is the term used to describe the technique of picking up psychic impressions about a person from a specimen of their handwriting. Gobineau and Perron [2] elaborated a theory of graphometry, or more exactly a statistical method of the graphic elements. In their work, they propose more than 60 features but choose 14 which they deem essential and easy to extract.

Graphology is claimed to be useful for everything from understanding health issues, morality and past experiences to hidden talents, and mental problems. The person that uses the concepts of graphology to this end is known as graphologist. However, the graphology is not restricted to this. Forensic document examiners (FDE) use it to examine handwriting in order to detect authenticity or forgery. A type of handwriting that is subject of analysis very often is the signature. With the power of computers growing exponentially, researchers have tried to use the ideas of graphology and the expertise of FDE to automatically analyze and verify signatures. Some of the concepts of graphology have been intrinsically used to build automatic signature verification systems by several different authors [5, 4, 3, 6]. However, in most of the cases they do not establish a connection between features and graphology/graphometry.

In this paper, we first describe some of the main graphological and graphometrical features. The criterion used to select them was if they were feasible computationally. Then, we establish a relationship between features from these two fields in order to propose a set of features that can be applied to automatic signature verification. The performance of such features are evaluated on a dataset composed of 5,600 signatures (genuines, random and simulated forgeries). The classifiers used are the hidden Markov models. Finally, we discuss the advantages and drawbacks of using such features in context of signature verification.

2. Features From Graphology and Graphometry

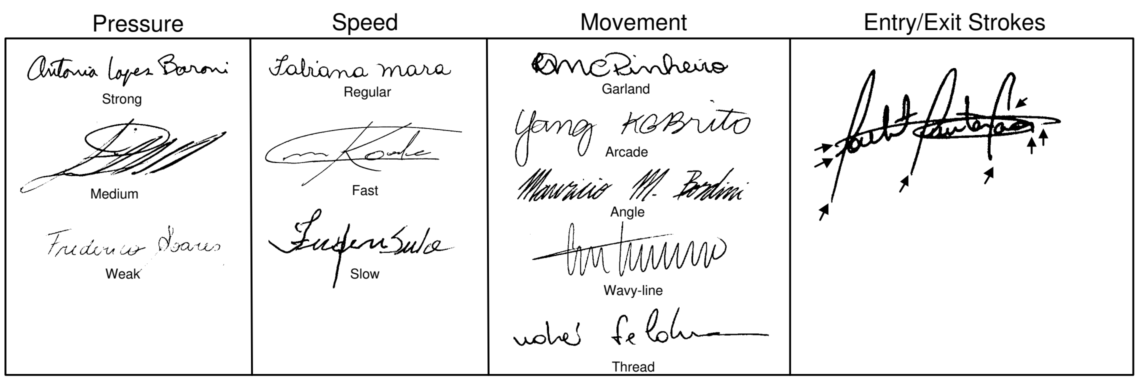
Nowadays we can find two different schools of graphology. One is called the mimic school and tries to identify a person's character based on holistic features of the handwriting such as height, width, slant, and regularity. The other school is called symbolic and it is based mainly on the study of the interpretation of the symbols. The main features in this case are: order, proportion, dimension, pressure, constancy, form, characteristic gestures, and occupation of the space. We can visualize better some of these features looking at Figure 1.

The order refers to the distribution of the graphical elements. It can be clear, confusing, concentrated, and spaced (Figure 1a). The Proportion is related to the symmetry of the writing. Figure 1b shows proportional, unproportionate, and mixed. Dimension shows the enthusiasm of the writer. Basically, we can classify dimension into high-dimension when the height of the letters are bigger than the width and

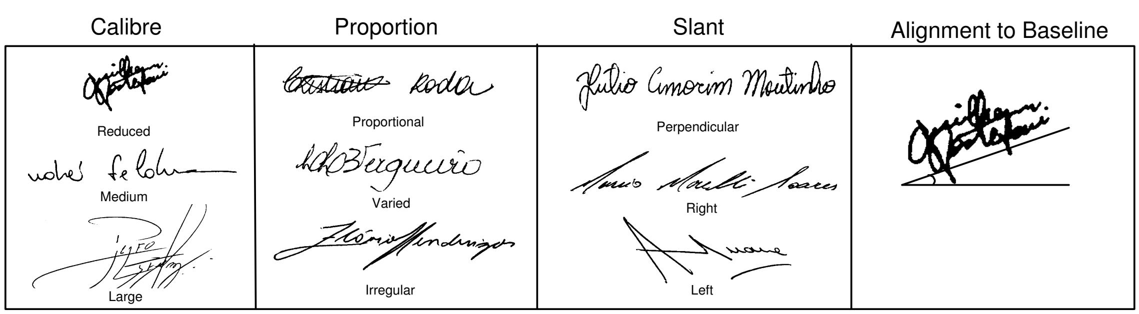
Fig. 1. Features from graphology.

low-dimension, otherwise (1c). Pressure is related to the changing width of a line as pen pressure varies. Constancy refers to the speed and intensity of the writing. The Form in graphology concerns to the graphical models employed, i.e., the kind of stroke that prevail over the image. We can have rounded, vertical, and horizontal strokes. We can also have a calligraphical model (Figure 1d). As the name says, the Characteristic Gestures are gestures that the writer repeat periodically, e.g., the way the writer makes a t bar, the way he/she starts/finishies writing, etc. Occupation of the space regards the way the writer uses the space available for the writing. This feature will be discussed in more details later.

The graphometrical features can be classified into genetic and generic. The genetic features are: minimal graphics (i dots, commas, cedillas, tildes, etc), pressure, speed, entry/exit strokes, and movement (Figure 2a). The generic features are: calibre, spacing between characters and words, proportion, slant, and alignment to baseline (Figure 2b).



(a)



(b)

Fig. 2. Graphometrical features: (a) genetic and (b) generic.

3. The Proposed Set of Feature for Signature Verification

Based on the features presented in the previous section, we defined a set of features that can be applied to signature verification. Table 1 shows some features we can adapt from graphology and graphometry for signature verification. Although some features have different names in graphology and graphometry they are exactly the same.

Table 1
The Proposed Feature Set

Feature	Name in Graphology	Name in Graphometry
Calibre	Calibre	Height, Width, Dimension
Proportion	Proportion	Regularity, Proportion
Spacing	Spacing	–
Alignment to baseline	Alignment to baseline	–
Progression	Speed	Constancy
Pressure	Pressure	Pressure
Gesture	Entry/Exit Strokes	Characteristic gestures
Occupation of the graphical space	–	Occupation of the graphical space
Minimal Graphics	Minimal graphics	–
Slant	Slant	–

Signature verification has several different applications, but our work was carried out in the context of bank cheque processing. In light of this, some of the aforementioned features are difficult to extract or computationally expensive:

- Pressure: In the case of bank cheques, the signature can be pre-printed in the form, so that the information about pressure is not available.
- Minimal graphics: We have verified that small fragments of images, such as i dots, periods, and commas, are very often eliminated due to pre-processing steps.
- Occupation of the graphical space: Since the area reserved for the signature in bank cheques is small and well delimited, there is no meaning in using this kind of feature. In Section 4. we discuss this issue in more details.
- Characteristic Gesture: This feature can be located anywhere in the writing, what makes it very difficult to find by means of computer program. A simplification of this feature of the graphology is the entry/exit stroke of the graphometry.

Therefore, the following are the features we propose for signature verification: Calibre, Proportion, Spacing, Alignment to Baseline, Progression, Form, and Slant. The first four are called static features, while the last three classes are pseudo-dynamic features. As we can observe from Figure 3, the static features are related basically to the occupation of the graphical space. The calibre describes the relationship between height and width, the Proportion refers to the symmetry of the signature, the Spacing shows when the writer put pen lifts and breaks between specific letter/stroke combinations, and Alignment to Baseline is simply the relationship of the writing to a baseline.

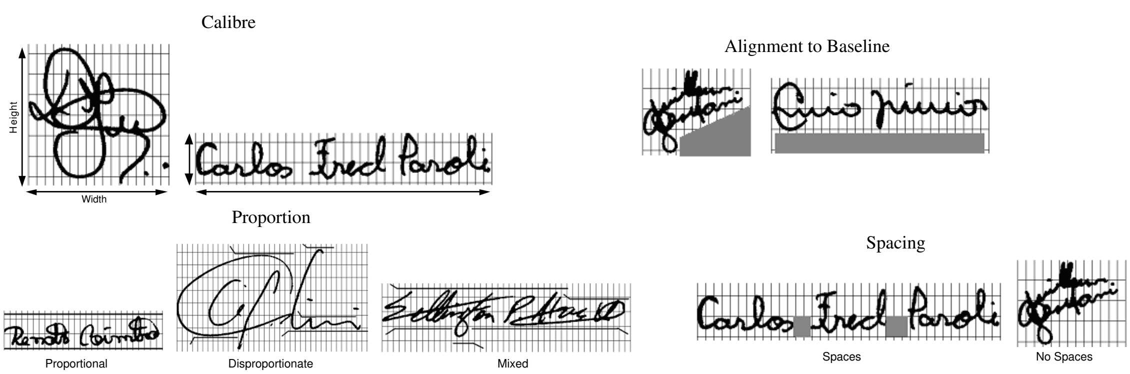


Fig. 3. Static features.

The pseudo-dynamic features also contain rich information about the signature, since they are directly related to the strokes of the signature. The Progression can be represented by three set of features: density of pixels, distribution of pixels, and progression. The density is what we call apparent pressure, since it describes the width of the strokes. In order to compute it, we put a grid over the image and count the number of black pixels in each cell (see Section 4.). The distribution of pixels is based on four measures as depicted in Figure 4a. In this case, each cell is divided in four zones. Then, the width of the stroke is computed in four direction (limited to the zones). These values are represented by the letters A, B, C, and D in Figure 4a. A more complex approach, but based on the same idea was proposed by Sabourin et al [7].

The third feature set based on progression is the progression itself. It is based on the level of tension in each cell and gives some vital information about the strokes, such as, the dynamics, speed, continuity, and uniformity. To determine this, we select the most significative stroke of each cell (i.e, the longest one), and them compute the number of times the stroke changes direction. When few directions are changed, we have a tense stroke, otherwise it is classified as a limp stroke (see Figure 4b).

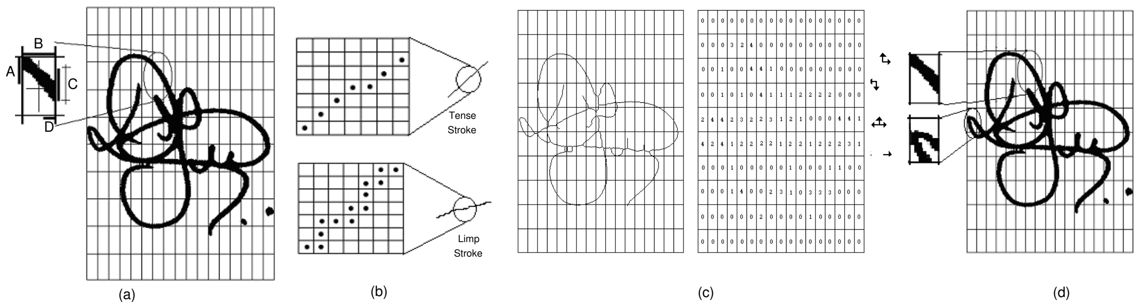


Fig. 4. Pseudo-dynamic features: (a) Distribution of pixels, (b) Progression, (c) Slant, and (d) Form.

In order to compute the slant we have applied the concept presented by Hunt and Qi [4], which determines the slant in two steps. First, a global slant is computed over the entire image and then the slant for each cell is computed as well. In this way, each cell has a slant value (Figure 4c) and the final local value is the most frequent value in the matrix. Finally, the final overall slant will be a combination of both global and local slants.

The last pseudo-dynamic feature we consider is the form. This is probably the most basic of individual characteristics. Form is the pictorial representation of a letter or writing movement. Computationally speaking, the concavities are very interesting way to get such pictorial representation of the handwriting. Therefore, we extract concavities measures of each cell, as depicted in Figure 4d.

4. Experiments

In bank cheques, usually the writer has a restricted space to sign. In light of this, we have made some experiments to verify how the writer behaves to space constraints. In other words, does he/she change the way of sign due to such constraints? To verify this, we have built a form (Figure 5) with different constraints so that we could analyze whether the writers respect them or not. We have collected 1,316 signatures from 94 writes (14 samples per writer). These signatures are not the same that we have used to train the models. Firstly, the writer is asked to sign in the back of the sheet so that we can know his/her signatures when no constraints are imposed. Then, the writer is asked to sign 13 times in the front of the sheet (Figure5). It is worth of remark that the writers were not instructed to respect the constraints. After analyzing the forms (the forms were evaluated by three experts), we verified that about 89% of the writers do not change their way of signing, i.e., they keep their signature in the same scale. This justifies our choice of not using this feature, at least explicitly.

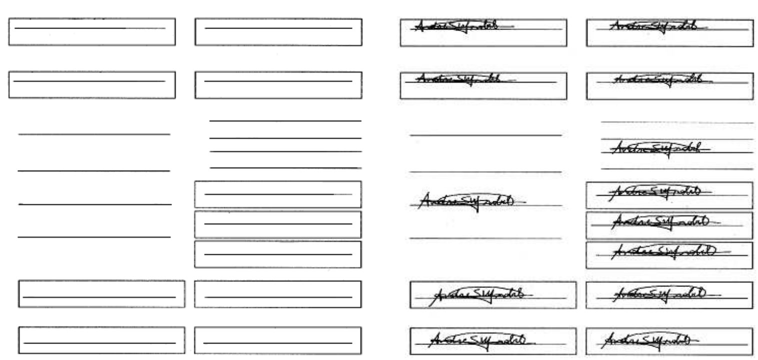


Fig. 5. Form proposed to the experiment on the occupation of the graphical space.

Once our goal is to build a system to automatically verify signatures in bank cheques, we have build a system based on hidden Markov models. The database used in our experiments contains 5,600 signatures (300 dpi, 256 gray levels) collected from 60 writers (60 samples per writing), and it is composed as follows: 20 genuine signatures for training, 10 genuine signatures for validation, and 50 (10 genuines, 10 simple forgeries, 10 simulated forgeries, and 20 random forgeries) for testing. The random forgery is usually a genuine signature sample belonging to a different writer, one who is not necessarily enroled in the signature verification system. The simple forgery is a signature written without any a priori knowledge of the genuine signature while the simulated forgery is a reasonable imitation of the genuine signature model.

In order to extract the aforementioned features we take into account a grid segmentation. It consists in putting a grid over the image and them computing the features for each cell. The size of the cell can vary, but we have found through experimentation that 16×40 pixels is a good configuration. This kind of segmentation extracts implicitly all static features described before, i.e., Calibre, Proportion, Spacing, Alignment to Baseline. In other words, when we extract the pseudo-dynamic features, the static ones are implicitly incorporated in the feature vector. Table 2 reports the results achieved on the test set for the pseudo-dynamic features. The error is divided into two classes. Type I error (false negative) occurs when the system rejects a signature correctly classified while Type II error (false positive) occurs and the system accepts a forgery. As stated before, a forgery can be simple, random, or simulated.

In order to compute those errors we have used local decision thresholds (one decision threshold for each author), which where determined by using random forgeries. In practice, frauds in banking industry are related in about 95% of the time with simple forgeries. The hypothesis we have made is that random forgeries seem to be a reasonable estimation of simple forgeries for the training of real signature verification systems. In other words, we estimate the decision thresholds with random forgeries and measure the overall performance considering only simulated forgeries for testing.

Table 2
Results on the test set.

Feature	Error I	Error II (%)			Average
	(%)	Random	Simple	Simulated	Error (%)
Density of Pixels	2.17	1.23	3.17	36.57	7.87
Distribution of Pixels	1.33	1.29	2.83	37.83	7.65
Slant	4.00	**0.72**	**2.50**	**32.33**	7.92
Progression	4.33	1.27	3.00	37.67	9.15
Form	6.20	0.93	2.63	35.45	11.30

We can observe from this table that the slant is the most robust feature against forgeries. This can be explained in part by the characteristic gestures from graphometry encoded in the slant. So, in spite of the fact characteristic gestures are not being used explicitly, it has been used implicitly into other features, such as the slant. On the other hand, the slant produces the highest error rate type I, since it can not absorb those signatures with high variability. This also explain the high error rate type I for the form feature.

5. Conclusion

In this paper we have discussed how the features of graphology and graphometry can be applied to automatic signature verification. Firstly, we have described the main features of graphology and graphometry and made a relationship between them. Thereafter, we have proposed a set of features that could be applied to automatic signature verification in the context of bank cheque processing. We have demonstrated thorough experimentation, using hidden Markov models as classifiers and a data set composed of 5,600 signatures, that these features reach interesting results in detecting simple, random, and simulated forgeries.

References

[1] J. Crepieux-Jamin. *L´ecriture et le caractere*. Presses Univ. de France, 14th edition, 1951.

[2] H. Gobineau and R. Perron. *Génétique de l´ecriture et étude de la personnalité: Essais de graphométrie*. Delachaux & Niestlé, 1954.

[3] K. Huang and H. Yan. Off-line signature verification based on geometric feature extraction and neural network classification. *Pattern Recognition*, 30:917–927, 1997.

[4] R. Hunt and Y. Qi. A multi-resolution approach to computer verification of handwritten system. *IEEE Trans. Image Processing*, 4:870–874, 1995.

[5] E. Justino, A. Yacoubi, R. Sabourin, and F. Bortolozzi. An off-line signature verification system using hmm and graphometric features. In *Proc. of the 4^{th} International Workshop on Document Analysis Systems*, pages 211–222, Rio de Janeiro, Brazil, 2000.

[6] F. Leclerc and R. Plamondon. Automatic signature verification: the state of the art 1989-1993. *IJPRAI*, 8(3):643–660, 1994.

[7] R. Sabourin, G. Genest, and F. Preteux. Off-line signature verification by local granulometric size distributions. *IEEE Trans. on PAMI*, 19:976–988, 1997.