

State of the Art in Writer's Off-Line Identification

Carlos M. Travieso González

University of Las Palmas de Gran Canaria, Spain

Carlos F. Romero

University of Las Palmas de Gran Canaria, Spain

INTRODUCTION

Today, advances in Computer Science and the proliferation of computers in modern society are an unquestionable fact. Nevertheless, the continuing importance of orthography and the hand-written document are also beyond doubt.

The new technologies permit us to work with online information collecting, but there is still a large quantity of information in our society which requires using algorithms for samples off-line. Security in certain applications requires having biometric systems for their identification; in particular, banking checks, wills, postcards, invoices, medical prescriptions, etc, require the identity of the person who has written them to be verified. The only way to do this is with writer recognition techniques.

Furthermore, many hand-written documents are vulnerable to possible forgeries, deformations or copies, and generally, to illicit misuse. Therefore, a high percentage of routine work is carried out by experts and professionals in this field, whose task is to certify and to judge the authenticity or falsehood of handwritten documents (for example: wills) in a judicial procedure. Therefore nowadays research on writer identification is an active field.

At present, some software tools enable certain characteristics to be displayed and visualised by experts and professionals, but these experts need to devote a great deal of time to such investigations before they are able to draw up conclusions about a given body of writing. Therefore, these tools are not time-saving and nor do they provide a meticulous analysis of the writing. They have to work with graph paper and templates in order to obtain parameters (angles, dimensions of the line, directions, parallelisms, curvatures, alignments, etc.). Moreover, they have to use a magnifying glass

and graph paper in order to measure angles and lines. This research aims to lighten this arduous task.

BACKGROUND

Writer identification is possible because the writing for each person is different, and everyone has intrinsic characteristics. The scientific bases for this idea come from the human brain. If we attempt to write with our less skilful hand, there will be some parts or strokes very similar to the writing which we make using our skilful hand. This is because the brain sends the commands for carrying out the writing and not the hands.

Generally, this effect is projected toward the writing by two types of forces, which are:

- Conscious or known: because it is controlled by the individual's own free will.
- Unconscious: because it escapes the control of the individual's own free will. This is divided into forces of mechanical and emotional means, which behaviour feelings.

Everybody writes using their brain, and simultaneously the handwritten impulse, which is the symbolism of the space in order to obtain the dimensions of the writing, is adapted proportionally, the size of the text being maintained or modified depending on whether the individual is forced to write in a reduced space.

Nowadays, writer identification is a great challenge because such research work has not been as fully developed as that of identification based on fingerprints, hands, face or iris (other biometric techniques), due mainly to the fact that the operation of the brain is very difficult of parameterize. On the other hand, the

above-mentioned techniques use widely researched biometric information.

Most of the characteristics implemented offer information in the vertical and horizontal plane (Zhenyu, Bin, Jianwei, Yuan, & Xinge, 2005) (Zhenyu, Yuan, & Xinge, 2005) (Schlapbach, & Bunke, 2006) (Bulacu, & Schomaker, 2005). We have introduced a new parameter, the proportionality index, which projects in all directions, depending on the selected points.

OFF-LINE WRITER IDENTIFICATION SYSTEM

As with the majority of the works proposed to the present date on biometric recognition, the framework of the system depends on the basic steps showed in figure 1. The images acquisition is a previous step to this system; therefore, this system is an off-line system. The data have to be scanned or photographed in order to build our database.

Data Acquisition

The forensic analysis of hand-written documents requires an extensive database of a known writer's hand-written samples. Therefore samples are gathered of different writers' writing and in turn several samples are taken of each one owing to the temporary invariability.

The creation conditions of a database have to be normalized with different types of paper, pen, and similar place of support (for doing the writing) because our work is centred on the writing and the efficiency of proposed parameters. For these off-line systems, the documents have been generated, and therefore, for the building of the database, the system has to be scanned or a high resolution picture taken. 300dpi on grey scale (8 bits) is a good threshold.

Image Pre-Processing and Segmentation

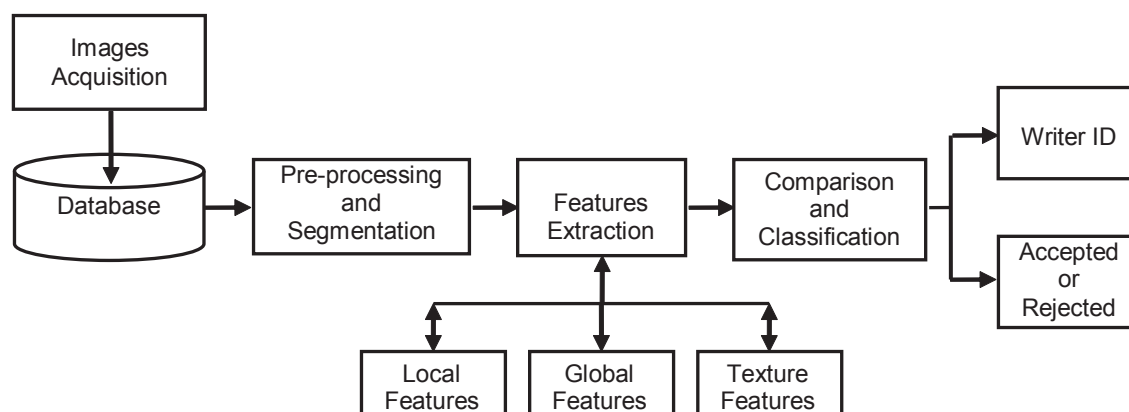
The first step of the image pre-processing consists of utilizing Otsu's method (or another method), which permits us to determine the necessary grey threshold value to carry out the binarization of the samples (Otsu, 1979).

As result of the binarization, in most cases, the line of writing remains with irregular appearance. For this reason, another pre-processing step is carried out, which enables the line to be smoothed out, thus remaining well defined. This also eliminates the existing noise in the images after the scanning process.

As previous step to the separation of words or connected components, the detection and elimination of the punctuation marks (full stops, accents and commas etc.) is carried out.

Finally, the words which compose the lines of writing are segmented (baselines) and for this, it is necessary to establish limits for each of the words. For this

Figure 1. System of writer identification



6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/state-art-writer-off-line/10429

Related Content

Named Entity System for Tweets in Hindi Language

Arti Jainand Anuja Arora (2018). *International Journal of Intelligent Information Technologies* (pp. 55-76). www.irma-international.org/article/named-entity-system-for-tweets-in-hindi-language/211192

Comparative Analysis of Supervised and Unsupervised Learning in Digital Forensics

Muthuvel Laxmikanthan, K. Siva Agora Sakthivel Murugan, Riddhi Chawla, Prasanta Chatterjee Biswas, Samandar Fayzievand Neeraj Chandnani (2026). *Trustworthy AI in Digital Forensics and Cybercrime Investigations* (pp. 173-202). www.irma-international.org/chapter/comparative-analysis-of-supervised-and-unsupervised-learning-in-digital-forensics/413406

An Investigation of Moroccan EFL Learners' Perceptions Towards Automated Writing Evaluation Tools

Mariam Elouirraoui, Cristina M. Gámez-Fernándezand Azize El Ghouati (2026). *International Journal of Artificial Intelligence (AI) in Teaching and Learning* (pp. 1-19). www.irma-international.org/article/an-investigation-of-moroccan-efl-learners-perceptions-towards-automated-writing-evaluation-tools/409035

Streaming Motherhood: A Digital Ethnography of Postpartum Voices on Social Media

M Siva Vidhyaand K Sunanda (2026). *AI and Digital Innovation in Healthcare Reform: From Policy to Practice* (pp. 227-262). www.irma-international.org/chapter/streaming-motherhood/408805

Dualistic Ontologies

F.A. Grootjenand Th.P. van der Weide (2005). *International Journal of Intelligent Information Technologies* (pp. 34-55). www.irma-international.org/article/dualistic-ontologies/2388