GRAWIS: GRONINGEN AUTOMATIC WRITER IDENTIFICATION SYSTEM

Marius Bulacu Lambert Schomaker

AI Institute, Groningen University, The Netherlands

Abstract

In the last several years, we developed new and very effective techniques for automatic writer identification that use probability distribution functions extracted the handwriting images to characterize writer individuality. The demonstration will present a html-based visualization tool that we recently built in order to be able to directly see and assess the results generated by our writer identification and verification software.

1 Introduction

After 9/11 and the anthrax letters, research in writer identification and verification receives renewed interest. A *writer identification* system performs a *one-to-many* search in a large database with handwriting samples of known authorship and returns a likely list of candidates. This list is further scrutinized by the forensic expert who takes the final decision regarding the identity of the author of the questioned sample. *Writer verification* involves a *one-to-one* comparison with a decision whether or not the two samples are written by the same person. The decidability of this problem gives insight into the nature of handwriting individuality.

Our research has been performed in the framework of the *Wanda* project financed by Fraunhofer Institute, Berlin. In the future, the *Trigraph* project financed by NWO will continue and extend the research on this topic.

2 Feature extraction and matching

Our features used to encode individual handwriting style operate at two levels of analysis: the texture level and the character-shape (allograph) level. They have been designed to be independent of textual content. In our approach the computer is completely unaware of what has been written in the samples. The handwriting is merely seen as a texture characterized by some directional probability distributions or as a simple stochastic shape-emission process characterized by a grapheme occurrence probability. Feature matching is performed using the χ^2 distance. In writer identification searches, all the samples in the dataset are ordered with increasing distance from the query sample. In writer verification trials, if the distance between two chosen samples is smaller than a predefined threshold, the samples are deemed to have been written by the same person. Otherwise, the samples are considered to have been written by different writers.

On a dataset with 250 writers (2 samples / writer), the best performance achieved by our system is: 81% Top-1 identification rate, 92% Top-10 identification rate and 5% Equal-Error-Rate in verification trials. A detailed description of our methods and their performance was given at BNAIC 2004 [1] and in [2].

[0]This paper was published as: Marius Bulacu, Lambert Schomaker, *GRAWIS: Groningen Automatic Writer Identification System*, Proc. of 17th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2005), 2005, pp. 413-414, 17-18 October, Brussels, Belgium

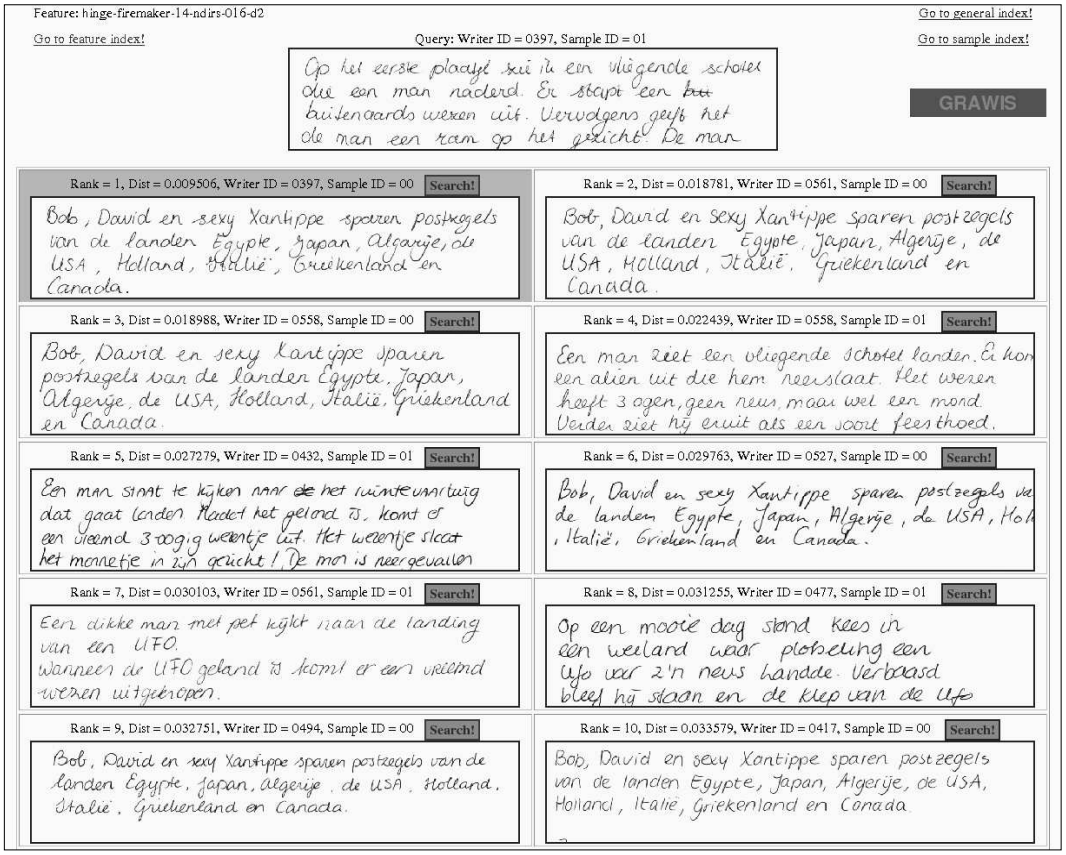

Figure 1: A successful writer identification search using GRAWIS: the query sample is in the top-center position, the best-matching sample (rank 1) was written by the same writer. A uniform handwriting style can be observed across the query sample and at the top of the hit list.

3 Demonstration

After feature extraction, feature matching and performance calculation, our programs generate html files containing numerical results (distances, ranks, writer/sample identity codes, thresholds) and hyperlinks to the handwritten samples (see Fig. 1). A web browser can then be used to visualize these html files. This approach allows a quick development of the visualization tool without the considerable programing effort needed to construct a complete GUI. For a chosen query sample, writer identification searches can be run using a battery of different features. Every sample from a hit list can in turn become the query and this allows a very handy navigation. We will invite the users to navigate in the space of individual handwriting styles. It will also be interesting to see the verification errors and to visually judge the resemblance between the handwritings being compared.

References

[1] Marius Bulacu and Lambert Schomaker. Analysis of texture and connected-component contours for the automatic identification of writers. In *Proc. of BNAIC 2004*, pages 371–372, Groningen, The Netherlands, 21-22 October 2004.

[2] Lambert Schomaker and Marius Bulacu. Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *IEEE Trans. on PAMI*, 26(6):787–798, 2004.