Forensic Signature Verification

Apurva Sharma
Department of Computer Science
Georgia Institute of Technology
Atlanta, GA
asharma70@gatech.edu

Parth Parekh
Department of Computer Science
Georgia Institute of Technology
Atlanta, GA
parthparekh@gatech.edu

Urjit Bhatia
Department of Computer Science
Georgia Institute of Technology
Atlanta, GA
urjit.bhatia@gatech.edu

Abstract

This paper implements a system that addresses the issues in the area of "Forensic Signature Verification". Two main approaches exist in this field- signature verification and signature identification. Our efforts focus on offline signature verification - the task of identifying whether a signature is genuine or forged given a genuine copy of the signature. Working on offline (static images) is a tougher task because temporal information which can give key distinguishing factors is missing. A part of our research focuses on trying to determine which are the key features which can help us discriminate between genuine and forged signatures and then developing algorithms which are able to do so from images of known genuine signatures and forgeries. Another focus area is on evaluating existing machine learning techniques to the extracted data sets and making suggestions for the same.

1 Introduction

This paper attempts to address problems in the forensic handwriting examination domain. In particular, we focus our attention on verifying the authenticity of handwritten signatures. This research area holds great relevance in the judicial investigations. As the implications of this decision process is extremely critical; it is imperative that the false positives and false negatives are minimal. Because of such nature, currently the verification is manually done by signature experts having years of experience in this domain.

There are two key variations to the above problem: Signature authentication and Signature identification. Signature authentication is the task of authenticating whether the questioned signature is genuine or forged, given a genuine signature. In contrast, Signature identification accepts and signature and tries to predict the writer of that signature. This subtle difference is better explained through below diagram:

Since writer identification requires the handwriting characteristics of all the writers prior

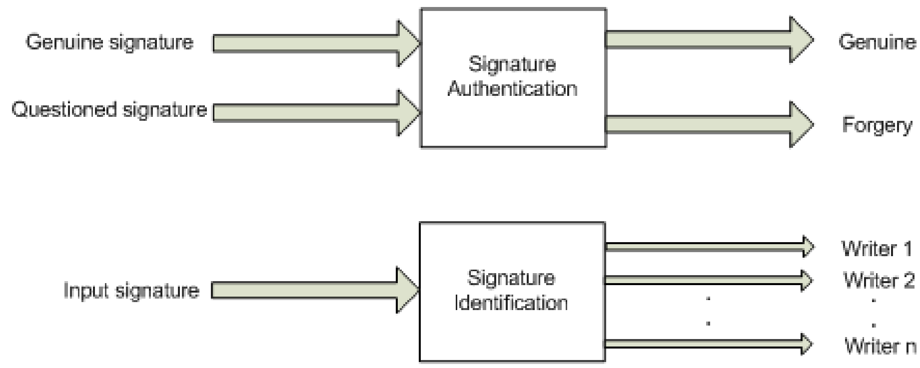

Figure 1: Comparison of Authentication and Identification models

the identifying one amongst them, we realize this is often not the case. Authentication model on the other hand provide results that have statistical inference [1]. Another distinguishing factor relies on the way the data is collected - online or offline. On-line signature capture approaches typically employ devices like digital tablets which not only capture handwriting characteristics but also temporal characteristics which often provide very useful information. Offline signature capturing is a lot easier as all it requires is a pen and a paper. However this makes authentication a harder problem as we lose the temporal information. In our study we focus on datasets which collect information in an offline manner.

We attempt to implement an end-to-end system which employs machine learning techniques to verify the authenticity of handwritten signatures. This system accepts images of genuine and questioned signature as input and outputs whether the questioned signature is authentic. Below diagram shows the architecture of the system:

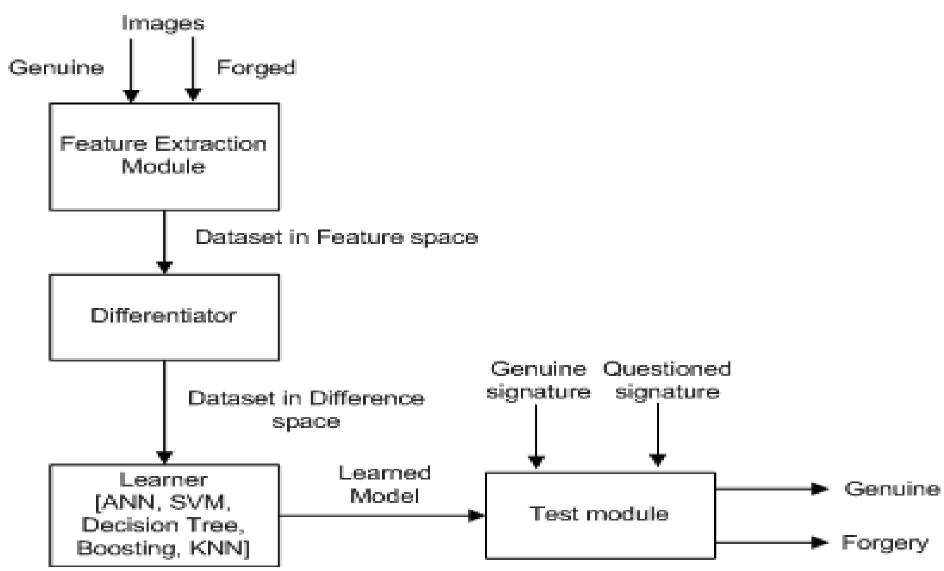

Figure 2: System architecture

2 Imageset Description

For our research we have used the image data set made publicly available by Cedar Labs [2]. The data set is in the form of images of genuine and forged signatures. The dataset consists of signature samples taken from 55 people. For each person, the 24 authentic and 24 forgeries, giving a total of 1320 genuine and 1320 forged signatures. All the images are assumed to have been written with the same pen as otherwise we run the risk of predicting whether the signatures were written from the same pen as opposed to whether the signatures were written by the same person. All background page for all the signature is plain white which reduces overhead tasks like line removal. The images are in png format and though they seem gray scale, actual image analysis revealed there was information in the RGB channels too. These images are first converted to pure grayscale to ease processing. Sample raw images for one pair of original and forged signatures and their corresponding processed images are shown in below Figure:

3 Feature selection and Extraction

This is one of our key research focuses. The end performance of the system depends greatly on the choice of these attributes and how discriminating they are. After going through a lot of literature in the field [1],[3],[4],[5],[6] (amongst many others). The features that represent the individual idiosyncrasies of the signatures can be divided into related groups:

1. Features based on the whole image Macro Features. 2. Features based on the character level changes Micro Features. 3. Features based on the DTW approach Dynamic Time Warping

We focus on macro features as they capture the most basic information required and many such features are utilized by existing Forensic Handwriting Experts too. DTW is again a very interesting approach that uses Zernike Moments to try and capture temporal like information for offline systems. However, these will be a part of future work as add on features to the macro features.

The macro features we decided upon were broadly classified into the categories of measures of pen pressure, measure of writing movement, measures of stroke formation, slant and proportion. A hierarchy of this feature set can be visualized in Figure 4.

These features are described briefly below with explanations of how these were extracted:

1. Entropy of gray values:

2. Grey Level Threshold:

3. Number of Black Pixels:

4. Number of Interior Contours:

5. Number of Exterior Contours:

6. Number of Vertical Slope Components:

7. Number of Horizontal Slope Components:

8. Number of Positive Slope Components:

9. Number of Negative Slope Components:

10. Slant

11. Height

4 Extracted Dataset Description

Once the macro features are extracted, our problem is still an N-class problem (for N writers) as at this point, we can still only make predictions by comparing questioned signature with N genuine signatures and predict which is the closest. However, there is a better approach which converts this to a 2-class categorization problem by making use of the fact that within writer distances are much less than between writer distances[1]. This though comes at the cost of explosion of the dataset. Till now, we had 2640 instances. However when we migrate the dataset to difference space, the number of instances increases to 46,860! (55* (24C2 +24*24)). This includes feature wise differences between all original signatures and all original and forged signatures. We keep the data set size in check by ignoring the intra class distances for the forged signatures as this is of no interest to us- we need not know how much two forgeries differ from an original signature.

5 Headings: first level

First level headings are lower case (except for first word and proper nouns), flush left, bold and in point size 12. One line space before the first level heading and 1/2 line space after the first level heading.

5.1 Headings: second level

Second level headings are lower case (except for first word and proper nouns), flush left, bold and in point size 10. One line space before the second level heading and 1/2 line space after the second level heading.

5.1.1 Headings: third level

Third level headings are lower case (except for first word and proper nouns), flush left, bold and in point size 10. One line space before the third level heading and 1/2 line space after the third level heading.

6 Citations, figures, tables, references

These instructions apply to everyone, regardless of the formatter being used.

6.1 Citations within the text

Citations within the text should be numbered consecutively. The corresponding number is to appear enclosed in square brackets, such as [1] or [2]-[5]. The corresponding references are to be listed in the same order at the end of the paper, in the **References** section. (Note: the standard BIBTEX style `unsrt` produces this.) As to the format of the references themselves, any style is acceptable as long as it is used consistently.

6.2 Footnotes

Indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).[2]

---

[1] Sample of the first footnote

[2] Sample of the second footnote

Table 1: Sample table title

PART	DESCRIPTION
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

6.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. Figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

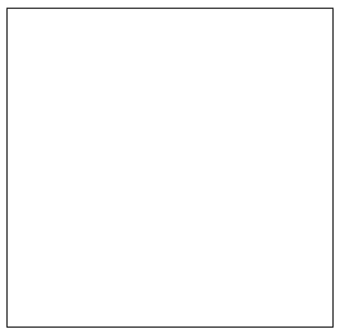

Figure 3: Sample figure caption

6.4 Tables

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. Table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

7 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular: do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Leave pages unnumbered.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper.

References

References follow the acknowledgments. Use unnumbered third level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to 'small' (9-point) when listing the references.

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D. S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609-616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer-Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.