FORENSIC HANDWRITING RECOGNITION

A Project Report

Submitted for the Partial Fulfillment of the Requirements for the Degree

of

Bachelor of Technology

In Computer science and engineering

By

MINUPALA KARTHIK (14CS01022)

Under the guidance of

Dr. Niladri Bihari Puhan (Asst. Professor)

SCHOOL OF ELECTRICAL SCIENCES

INDIAN INSTITUTE OF TECHNOLOGY BHUBANESWAR

ARUGUL -752050, ODISHA

CONTENTS

TITLE OF THE PROJECT

1. INTRODUCTION

Handwriting is a personal biometric that has long been considered to be unique to a person**.** For many centuries signature verification has been used for authentication purposes. Experts in forensic document analysis all around the world daily perform examination of handwritten documents to determine the authorship of a questioned document or detect evidence of forgery or disguise.

To represent a handwritten document, a set of features, extracted from the image of the document, is used. The features of handwriting are divided into two classes. Those that are used by forensic experts are called document examiner features, whilst those that are measured by computer algorithms are called computational features.

Over the past 30 years there has been a limited amount of research into using computers to enhance and automate the analysis performed by forensic document examiners. Much of the research centred on pattern recognition techniques for extracting static and dynamic features from handwriting and hand written signatures as well as enhancing document images and ESDA(electrostatic detection apparatus) lifts.

Most of the research papers on forensic handwriting recognition are based on usage of machine learning and deep learning techniques.

Machine learning:

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

2. METHODS:

1)Character Recognition

2)Manual Recognition (which is done offline)

3)Online Recognition.

Ways to achieve character recognition

1)Collection of characters (collecting data from given sample)

2)Extraction of features (extracting the optimal features from the input which are required to get details).

3)Analysis of features (analysing the optimal features)

4)Optimization of future vector for better pattern recognition.

5)Training of a classifier.

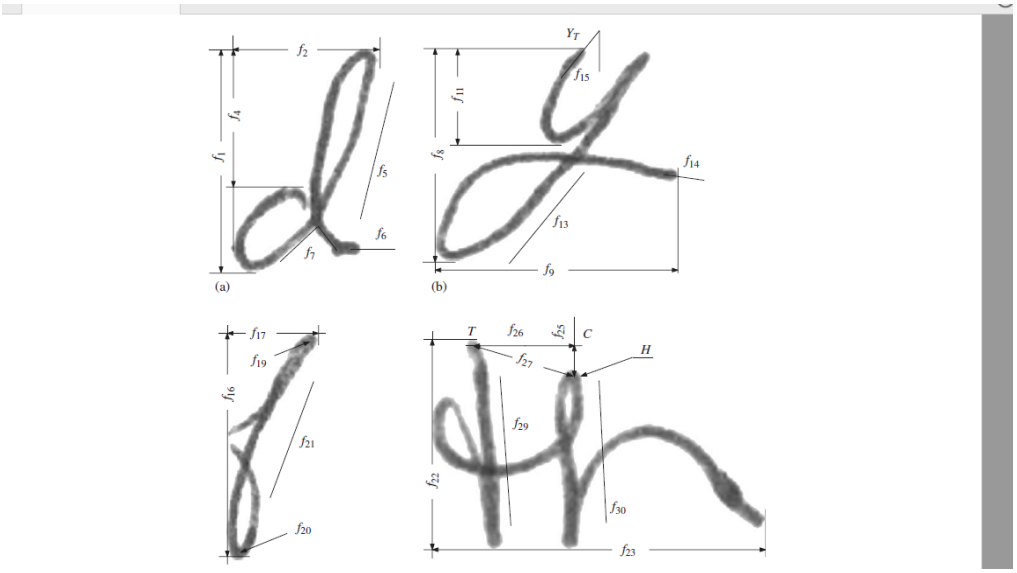
6)Output.

3. Approach

The purpose of the study is to determine whether it is possible to distinguish people by their handwriting using a subset of document examiner features. We assume that an experienced person, that is, an expert in forensic document analysis: (i) may effectively utilise more features than our system does, for it is very hard to express in strict mathematical terms many of the document examiner features, and (ii) the person may be able to determine which features should, and which should not, be used in a particular case (that is, having looked at handwritten samples, an expert is able to select only the important features for handwriting comparison). As a consequence, an expert may distinguish writers or establish authorship of questioned documents better on average, than our system.

Choice of features

Features are selected on basis considering the number of strokes a character or pattern has. Based on some considerations characters “d”, “y”, “f”, and grapheme “th” were chosen for study.



4. Feature extraction

Some of these elements, like height and width of a character, were strictly defined and thus were easy to extract from a character image. The features of the selected characters are stored and optimized using genetic algorithm and used for next checking with the document (sample) to be tested.

5. Analysis of feature usefulness:

First, all the features that may have discriminating power were extracted. Second, after extraction it was necessary to decide which features have discriminating power and which do not.

The features extracted are stored in a vector. Using genetic algorithm, the fitness values of the vector are calculated the one with the maximum fitness value is given to the next generation. In genetic algorithm there are also crossover and mutation which helps in finding the optimal solution.

Classifier

In our study we used n-fold cross validation to evaluate a feature subset . The training data was divided into n approximately equal partitions and the induction algorithm was then run n times each time leaving one subset for test and using the other n − 1 parts for training. The classification accuracy obtained from n tests was then averaged and associated with the corresponding feature subset.

DistAl(Distance algorithm), a constructive learning algorithm based on the multi-layer perceptron with spherical threshold units, was chosen as a classification system . There were several reasons for this choice over other possibilities:

DistAl is based on a distance metric between patterns which means it can easily be adapted to handle patterns with missing feature values. Experiments conducted on both artificial and real data demonstrated results of classification comparable to those obtained by other commonly used learning algorithms

All feature values were treated as real numbers. Having performed several experiments we chose the normalised Manhattan distance as a distance measure for DistAl because this measure was shown to be suitable for the problem at hand.

measure was shown to be suitable

$$d(\vec{F}^1, \vec{F}^2) = \frac{1}{k} \sum_{i=1}^{k} \frac{|F_i^1 - F_i^2|}{\max_i - \min_i},$$

where k is the number of feature

where k is the number of features, mini and maxi are the minimum and maximum values of the ith feature in the data set, respectively.

. The highest achieved classification accuracy was 58% when the optimal subset of all four character features was used.

Accuracy can be increased using convolution neural networks or deep learning methods. I am trying to use convolutional neural networks which may also make extracting features of characters easy.