Signature Feature Extraction and Classification

Ankit Kumar - IRM2016002
Avinash Yadav - ITM2016004
Sashank Mishra - IIT2016515

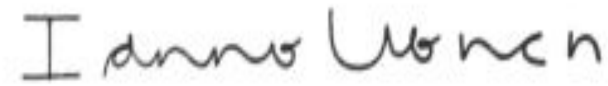
INTRODUCTION

In this project, we present a new model for signature classification, which uses some basic and advanced Image Processing techniques to extract certain features, that are then used to uniquely identify a person’s signature.
The extracted features are then input into some classification models to testify the accuracy and the resultant conclusions are derived.

DATASET DESCRIPTION

The dataset we used here is **GPDS**, which provides us with 300 users data with each user's data contains 54 images of signature in which 24 are original images and 30 are forged images by some skilled person.

Original signature

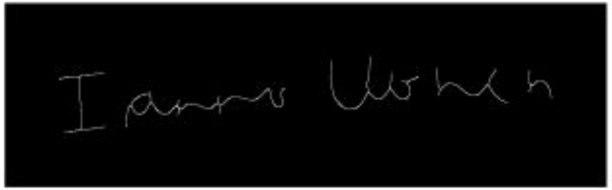


Forged signature

DATA PREPROCESSING

Since the width and stroke varies a great deal for every image, we cannot directly extract the feature set. So, we employed some data preprocessing techniques on all the signatures.

- ➔ Convert the image into binary image.
- ➔ Skeletonize the image to form single line image.
- ➔ Take the invert of the image.



Resultant image after Pre-processing

Feature Set Generation

LIST OF FEATURES SELECTED

- ➔ Center of Pixels (In both the directions)
- ➔ Total Number of Parts
- ➔ Number of Isolated Points
- ➔ Distance between Centers and Centers of Mass
- ➔ Average Height of the Signature
- ➔ Count of the Pixels
- ➔ Count of Harris Points
- ➔ Degree of Skewness
- ➔ Degree of Kurtosis

Center of Pixels
Just like in physics where the Center of mass is the measure of the group of objects (masses). Here we used the center of pixels instead with same ideology to map the collective pixels onto one single point.

Number of Isolated Points
This feature counts the points that are not connected to any other pixels. This feature is selected due to the fact that despite the name of many person, they draw some random dots (due to the flow) and this can be a prominent feature to find the distinction between two persons with same name as well as different names.

Average Height of the Signature
This feature was required to estimate the sleekness of the signature. Many people make the first letter of the signature huge then the rest follows a constant height. So, the centers can vary within one person’s data a bit, and this feature neutralizes the effect.

Total Number of Parts
Different person’s signatures are made of several number of disjoint parts due to one’s habit of lifting the pen. This property can be crucial while classifying the signature.

Distance between Effective Centers and Centers of Mass
Effective centers are defined as the centers of the image obtained, after drawing a tight rectangle across the signature. The center of masses is same as explained in the first feature. We calculated the distance between both type of centers.

Count of the Pixels
We have also included the count of pixels set in the skeleton of the image. Since the skeleton of a particular person’s signature cannot vary much and only single pixel lines are used to trace the original image. Therefore, this count can’t vary much, and hence is unique for a particular person.

Count of Harris Points

Harris Corner Detector is a corner detection operator that is commonly used in computer vision algorithms to extract corners and infer features of an image. This feature is a count that maps the curves and corners into an integer.

General Algorithm for Harris Point Counts:

1. Consider taking an image patch over the (x, y) and shifting it by (Δx, Δy).
2. The sum of squared differences (SSD) between these two patches, denoted f(x,y) , is given by:
   $f(x,y) = \sum_{(xk,\ yk)\ \epsilon\ W} ((I(x_k\ ,y_k\ ) - I(xk+\Delta x,yk+\Delta y))^2$
3. I(x+Δx, y+Δy) can be estimated using Taylor's Expansion as:
   $I(x+\Delta x,\ y+\Delta y) = I(x,y) + I_x(x,y)\Delta x + I_y(x,y)\Delta y$
   where I_x and I_y are partial derivatives.
4. So we can say
   $f(x,\ y) = \sum_{(xk,\ yk)\ \epsilon\ W} (I_x(x,y)\Delta x + I_y(x,y)\Delta y)^2$

Thus, we calculate f(x, y) for each location and add them for the image.

Degree of Skewness of image

In probability theory and statistics, skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive or negative, or undefined. This property can be used to find the skewness of the image. Since image is also a vector of vectors. Therefore, we first found out the skewness value of a single row and then the skewness value of those values.

Degree of Kurtosis of image

Just Like Skewness, kurtosis is also a property of statistics where we compare the distribution with normal distribution and try to find out the tail. Low kurtosis means high tail.

Classification Model

CLASSIFICATION MODEL FOR SIGNATURE DETECTION

After we found out the feature set of all images, we need a classification algorithm to complete the prediction model. Though, there are many classification algorithms available, we tried the following algorithms after looking at the properties of the dataset (smaller number of data per class and higher number of classes)

- ➔ KNN (K-Nearest Neighbours)
- ➔ SVM (Support Vector Machines)
- ➔ ANN (Artificial Neural Networks)

MODEL SUMMARY

Model / Number of Classes	***30***	***100***
KNN	62%	40%
SVM	87%	72%
ANN	24%	17%