**READ ME:**

**INSTRUCTIONS FOR RUNNING THE SIGNATURE VERIFICATION SYSTEM**

**Following packages are required for running the system:**

**“Opencv, skimage, mahotas, sklearn, tkinter, PIL, numpy, scipy, os, sys, warnings, pandas”**

* Keep all the file in one folder and run final\_siganture.py file.
* A GUI will pop up which will guide you how to proceed further.
* Default directory is dataset4.
* In order to change directory, click on the appropriate click button in GUI.
* And always select genuine signature folder first then forged folder.
* And thus new classifier will be trained on new directory.
* Training will take time depending upon the number of image in the directory.
* And after training gets completed GUI will close and you need to re run the final\_signature.py.
* And now the directory has changed.

**Demo video is also present one can see it for understanding how to run the files.**

**MOTIVE:**

To make offline signature verification system which will not only perform good with large signature database but also perform well with small signature database as this will save a lot of memory, time and money as many existing model uses deep learning approach which requires a large dataset for training which causes memory problem and thus a lot of money is wasted in storing a large database, thus to prevent this more emphasis is given to feature extraction part and most important features from each image is extracted and then two model is created which takes a small database for training and gives a maximum accuracy score of 0.97 on validation data set(created by splitting dataset4) when combined together.

**Explanation of the method used for building signature verification system**

**“Before proceeding further please run the final\_siganture.py file once and see how it works as it will help in understanding the model.”**

Offline handwritten signature verification system composed of two part:

* Data preprocessing (Image Processing, Feature Extraction)
* Model training (Training Classifier, Feature Matching)

**Data preprocessing:**

In this each image is processed such that bounding box is created around the signature and that part of image is only taken which consists of signature (lets name it as cut\_piece image), and many important features are extracted.

List of features extracted from images are:

* Hu Moments
* Haralick
* Eccentricity Solidity
* Skew Kurtosis
* Aspect ratio,
* Bounding\_rect\_area,
* Contour area
* Centroid
* Center of mass
* Baseline shift
* Signature actual Height
* Signature actual Width
* Cut\_piece image is broken into 16 parts and for each part center of mass is calculated and then angle is calculated between center of mass of each part with the right bottommost part, these angles is then used as features along with above mentioned one, these angles turn out to be the most important features for signature verification system.

Once these features (total of 52 features for each images) are extracted from images a data set is created such that training and validation data is obtained after splitting the image database (containing the features of each image in the chosen directory) into 80:20 ratios.

**Model training**

Bagging technique (type of Ensemble technique) has been applied as this technique gives the best accuracy when datasets is small.

Two Random Forest Classifier model is trained one for identity check of signatures and another one for status check (i.e. 1 for genuine and 0 for forged signature)

And each model is trained independent of each other and finally they are combined once they are fully trained.

Each model is tuned and best parameters are obtained for which model gives maximum accuracy score on validation set.

Training time depends upon the size of database used for training for example it took approx. 10 min to train both the model on dataset4(folder present inside dataset folder containing signature of 18 persons 5 for each person) and approx. 16-17 min to train both the model on sample\_signatures database (containing signature of 30 persons 5 for each person).

This model performs better than many existing models, its accuracy is better than LVQ, VQ, SVM approach.

**RESULT:**

