CS – 7641 PROGRESS REPORT
TOPIC: FORENSIC SIGNATURE VERIFICATION

TEAM 6:
Apurva Sharma (902490301)
Parth Parekh (902566648)
Urjit Singh Bhatia (902671137)

1. Problem description and its evolution:

Our earliest vision for the project involved trying to explore machine learning techniques for breaking CAPTCHA's in an attempt to analyze what are the potential vulnerabilities in the current systems and how they can be improved. The focus shifted when we came across the ICFHR 2010 (12th International Conference on Frontiers in Handwriting Recognition) competition on forensics & signature verification. With a similar underlying theme, participating in this competition seemed like a more feasible option for us for the following reasons:

1. We would get an unambiguous statement of problem, making it easier for us to know exactly what we need to do and giving us a lot of well directed focus.
2. We would be given datasets on which to operate.

The main aim of this project is to develop a system that addresses the issues in the area of "Forensic Signature Verification". The kind of data available for this field of application includes "originals", which are original signatures by writers expressly performing their authentic signatures and "forgeries", which are deliberate imitations of the originals. Thus our efforts are focused on drawing comparative measures between the originals and the forgeries.

Choosing to participate in the competition seems to be a good decision as our problem statement has not undergone major changes since the second submission.

Initially, we assumed that we would get the dataset from ICFHR by mid-March, but later we found out that the dataset would be released in mid-April. This meant that we had to find alternate datasets to start off with our project. A good deal of searching led us to the CEDAR labs at University at Buffalo which is involved in a lot of relevant work in this field. They also had two datasets- of original and forged signatures which are both useful for our study. Worth noting is the fact that this dataset consists of only images; feature engineering thus became a major part of our research goals as explained below.

2. Research goals:

We present our research goals, their treatment and their evolution through the past couple of weeks in this section. Dividing the research goals into 4 logical blocks, we focused on these following major goals:

1. Feature Engineering.
2. Application of learning algorithms on the extracted features.
3. Analysis and optimization of the learning models.
4. Creating a coherent end-to-end application for final submission and the competition.

Our target is to accomplish these goals and push the classification accuracy to as high as possible with the given experimental setup.

Feature Engineering:
As mentioned in the section above, feature engineering was added to our set of research goals after we realized that we would have to extract the features from the raw images since no preprocessed datasets were readily available. Further, feature engineering translated into the following tasks:

1. In-depth knowledge of the domain (the Where?)
2. The features that we should extract based on the domain knowledge (the What?)

3. How to extract these features (the How?)

In an attempt to answers these questions above, we turned to the literature in the field of handwriting recognition, signature forensics & image processing and in due time we started our search for an image processing platform that we could use for the feature extraction algorithms that we would use. While navigating through the domain at hand, a major portion of our focus shifted towards extracting the features from the raw image data.

We decided to use Intel's OpenCV library for image processing for feature extraction. 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 decided to go ahead with using the Macro Features and the Micro Features initially. However, due to our incomplete understanding of the extraction process of the Micro Features and the limited amount of time we had on our hands, we researched solely on the Macro Features. The Macro Features gave us information about the image on the whole and generated 11 features.
These features were: *entropy of gray levels, gray level threshold, number of black pixels, number of interior contours, number of exterior contours, number of vertical slope components, number of horizontal slope components, number of positive slope components, number of negative slope components, slant and height*.
We aimed our research towards implementing the extraction algorithms for these features using the OpenCV library from the given set of scanned signature images. A lot of background research went into analyzing how images are represented using Chain Codes and various other techniques like thresholding using the Otsu algorithm, Sobel operators for edge detection etc.

Application of learning algorithms on the features:
After some deliberation, we realized that using the extracted features directly would not help us solve the proposed problem. Thus we needed to transform our feature-set into a new representative space which would make it compatible to the learning algorithms
We faced the following challenges posed by this goal:

1. How to use the features to build our learning model.
2. What feature transformation would be a helpful representation?
3. Which learning algorithms can be applied?

We transformed the original feature space into a distance space by taking 'Inter-class' distances between the known forgeries of the given original signature samples of a given writer. Similarly, we took the 'Intra-class' distances of the original samples of a given writer. This allowed us to see how forgeries are distant from the originals in this new space.
The learning algorithms we are applying to the final dataset include Neural Networks, Decision Trees & SVM and SVM with Adaptive Boosting with 10 fold cross validation. We have built some preliminary models using these algorithms and will present the results we have obtained till now in the later sections.

Analysis and optimization of the learning models
After running our preliminary tests on the extracted dataset, we are focusing on the alternatives available to us at this point:

1. Extract more features from other feature categories mentioned previously.
2. Exploit the existing feature set by improving and tweaking the learning algorithms.

Since we had decided to use only the Macro features for constructing the dataset our aim is to determine if adding more features and more feature types will be helpful to improve the performance of the models we have obtained and the other option is to solely focus on the dataset we have obtained. There are pros and cons for both the approaches here, adding new features will be time consuming and might not guarantee significant improvements. It will also make us revisit and add more feature extraction algorithms on the OpenCV platform but it might empower us with more highly correlated features. On the other hand, trying to optimize the learning algorithms may prove unworthy since some important features might be missing but at the same time tweaking the algorithms and combining them might give us good results.
Creating a coherent end-to-end application for final submission and the competition
Currently, our application is fragmented into various modules. The major module sits on the OpenCV platform while another component written in Python used this OpenCV module to extract the image features and then

generate the datasets. The third part is the set of learning algorithms for which are currently using Weka. However, our aim to provide a single point of entry (a set of original and forgery samples) and a single point of exit which would be a classification – deciding whether the sample in question is an original or a forgery.

3. Project plans and accomplishments

We almost followed our plan that we had anticipated. The description about our bi-weekly plan is explained below:

1. Preprocessing (Week 1 & 2)
   - During this week, as planned in our initial proposal, we read most of the papers, close to 30, published by CEDAR (Center of Excellence for Document Analysis and Recognition) and finalized the approach
   - In our initial proposal we presumed that we will consolidate different datasets into one and train different algorithms on the resultant dataset. Finally we used the CEDAR dataset which is collection of 24 X 55 original signatures and their forgeries – a total of 1320 samples of each class.
   - As the data we obtained consisted only of raw images, we had to spend a lot of time in identifying and extracting the relevant features.
2. Implementation (Week 3 & 4)
   - Our main aim during this week was to implement the approach we finalized in the earlier weeks.
   - We spent most of the time this week learning the OpenCV library for image manipulation.
   - As described in above we initially spent some time extracting GSC (Gradient Structure Concavity) micro features. The main problem we faced here was that we could not find any description of converting the Gradient feature bit vector to Structure bit vector. We also tried to contact Dr. Sargur Srihari, the author of most of the papers in this domain, without any success. Moreover this method (GSC) would have yielded only one feature vector of 1024 bits which would not give us much information about the data. Thus we discarded this method and concentrated on extracting the 11 Macro features mentioned in above sections. All the features were extracted using the OpenCV library.
   - As our dataset was still not ready, we could not run any learning algorithms during this week as otherwise planned.
3. Optimization (Week 5 & 6)
   - In the fifth week, we migrated from feature space to distance space thereby transforming the n-class problem in d-dimensional feature space to a 2-class problem of same or different writers in multi-dimensional distance space. We were able to do so because of the key finding that the within-writer distance (the distance between two samples written by the same writer) will be less than the between-writer distance (the distance between two samples written by two different writers).
   - After the dataset was ready we ran ANN, SVM, SVM with ADABoost and Decision tree (J48) algorithms over this dataset. We have compared all the results in below section.
4. Final testing and Application development (Week 7, 8, 9 & 10)
   - We intend to stick to our original plan for next week.
   - We might try to extract more features and try to improve the accuracy. We also intend to implement few other approaches like DTW and compare the results with our current implementation.

4. Individual tasks

- In the initial week after reading some of the papers all three of us brainstormed on different approaches and finalized the roadmap for rest of the project
- The tasks were mainly divided into gathering domain expertise for feature set identification and how to extract them, and then implementing those algorithms.
- Apurva was responsible for the first task and Urjit and Parth did the latter one.
- In next couple of weeks Parth and Urjit concentrated on learning OpenCV library and getting familiarized with its image processing operations. During that time Apurva was gathering all the

domain knowledge and identifying various features and their extraction procedures.

- Later on all three of us collaborated in converting thosee images to feature dataset.
- Once the dataset was prepared each one of us ran several ML algorithms and analyzed the preliminary results described below.

5. Preliminary results

Being new to the domain, one of the first things we were curious to see was the relative importance of the features chosen by us- which features seem to have a greater impact on what the label of a particular instance is. One of the more direct ways of getting a feel of this included running a standard decision tree learning algorithm on the data set and seeing which attributes are utilized higher up in the decision tree formed.

The decision tree formed seemed to suggest that *entropy of gray levels* was one of the key features which the *average slant* and *number of vertical slope components* were at the bottom of the decision tree, indicating that they may be not as information rich. With this knowledge we were curious to see the distributions of samples based on these features. The following are these findings in the feature space as well as the distance space:

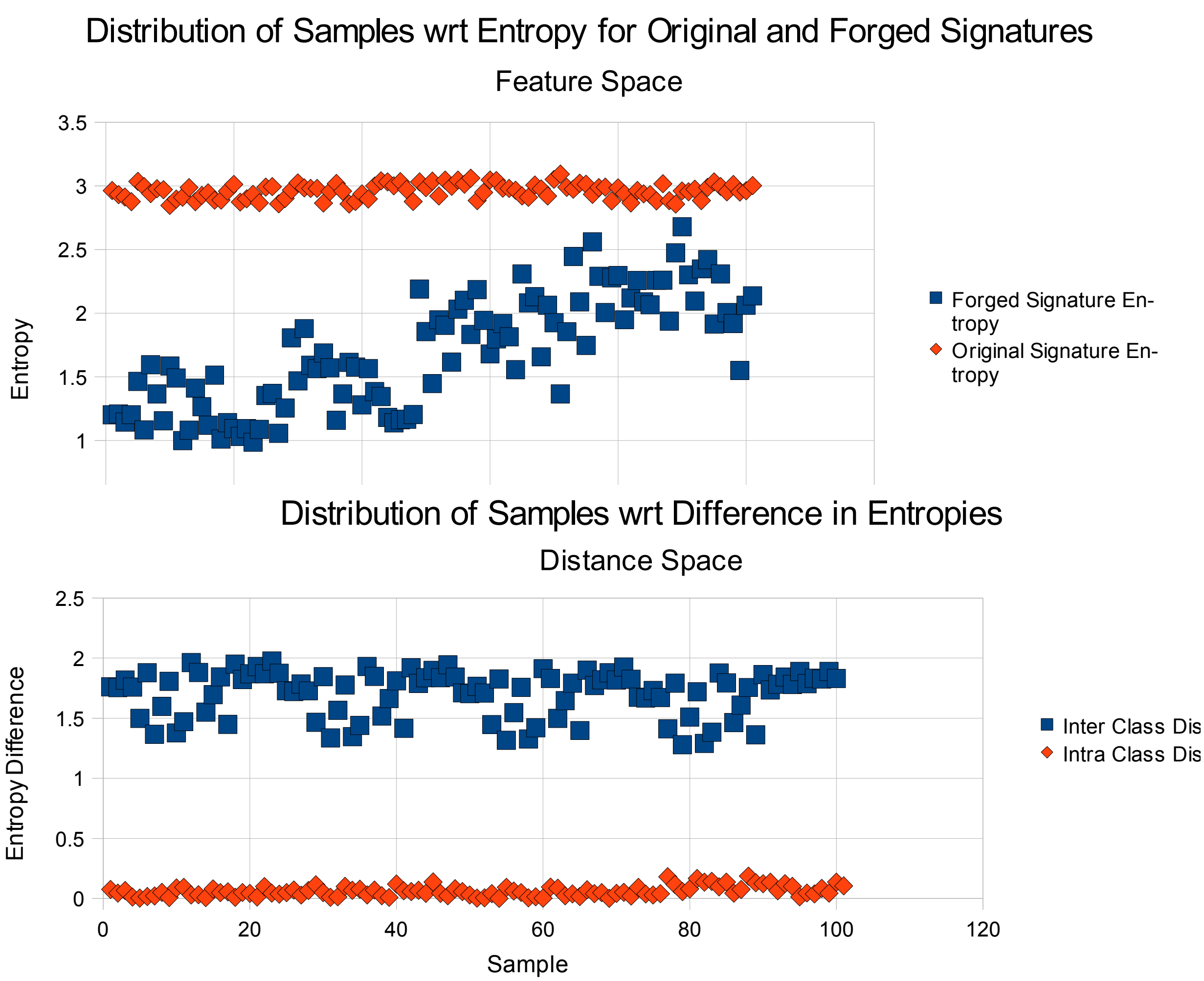
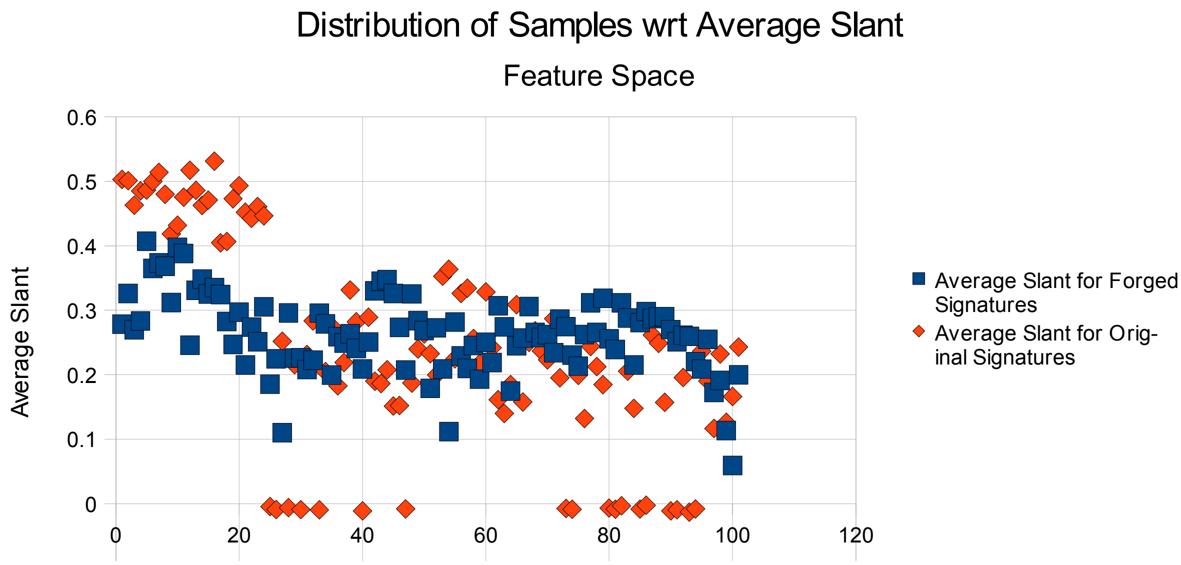


Fig 1 (a)-(b)Distributions of selected samples w.r.t. Entropy of Gray Levels in Feature and Distance Space



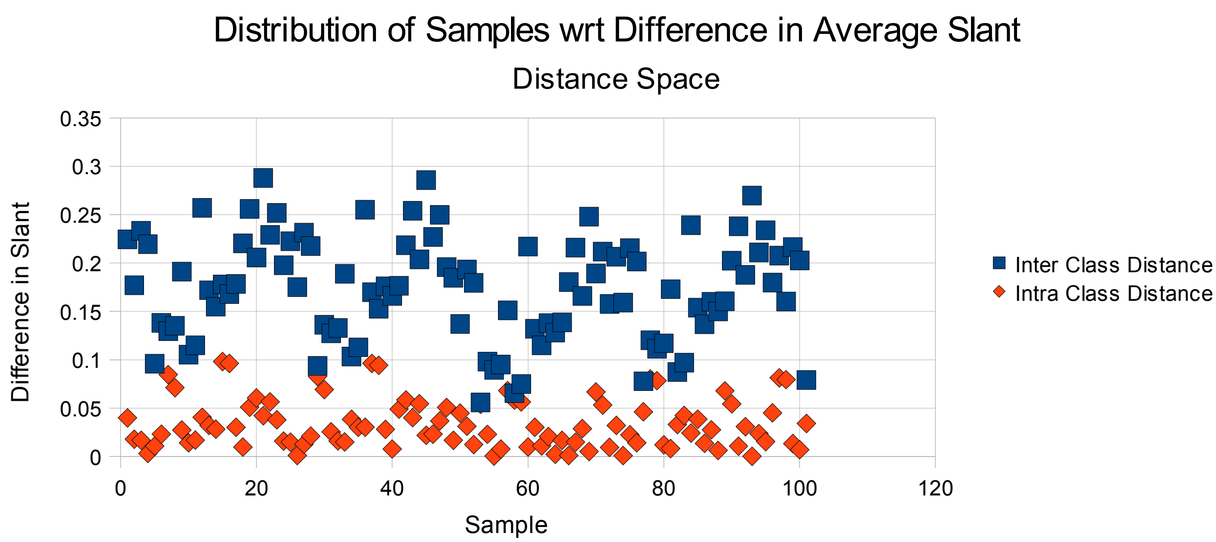


Figure 1 (c)-(d) Distributions of selected samples w.r.t. Average Slant in Feature and Distance Space

Figure 1 (a)-(d) actually convey a lot of information. We can clearly see that if we just consider the feature space, then Entropy tends to separate out the two classes fairly nicely. In comparison, for *Average Slant* such separation is not straightforward. Even more interesting is the case of transformation to the distance space which indicates a far more clear separation, even for attributes like average slant, thus leading us to believe that this might be a reasonably good approach to take.

Once convinced about our feature selections, we ran the following algorithms (so far) on the extracted data set. Then, we tried to apply Ada-boost to some of these algorithms to see if the results improve. The following table captures our results so far :

Algorithm Used	Classification Accuracy	Area under the ROC curve (same for both labels)
Decision Tree (J48)	97.92%	0.978
Decision Tree (J48) + Ada Boost	98.58%	0.997
SVM	68.36%	0.513

SVM + Ada Boost	97.38%	0.985
ANN	98.29%	0.997
kNN (k=1)	98.14%	0.98

Table 1: Comparison of performances of different learning methods on the handwritten signatures Dataset. All runs were done using 10 fold cross validation

These results are much better than our expectations and almost make us wonder if we did something wrong! A discussion with Dr. Isbell regarding this leads us to believe that there may be a chance of leakage of information from the training set to the test set because of using the entire dataset while applying 10 fold cross validation. Top on our list of things to do next is a detailed exploration of this very possibility. We plan to take the following approaches :

1. Instead of exposing the entire dataset, we plan to split it separately into Training set and Test set with the test set not being exposed at all while creating the model. We currently have 24 versions for each original and forged signature. We can use half of these for training and the other half for test.
2. We will be receiving an entirely new dataset from ICFHR for the competition pretty soon. In that case use that entirely new data set for testing on the models that we have currently built.

Once clear about the above possibility, we can shift focus on trying to explore more features as well as fine tune the learning algorithms (as of now, most models were prepared using default settings and we all know by now how much of a difference tweaking those parameters makes).