# *Signature Detection and Verification on Bank Cheques*

Chintan Gandhi
School of Engineering & Applied Science, Ahmedabad University
Gujarat, India 380009
chintan.g.btechi15@ahduni.edu.in

Anshul Jethvani
School of Engineering & Applied Science, Ahmedabad University
Gujarat, India 380009
anshul.j.btechi15@ahduni.edu.in

Jay Modi
School of Engineering & Applied Science, Ahmedabad University
Gujarat, India 380009
jay.m.btechi15@ahduni.edu.in

Darshan Shah
School of Engineering & Applied Science, Ahmedabad University
Gujarat, India 380009
darshan.s.btechi15@ahduni.edu.in

Harsh Shah
School of Engineering & Applied Science, Ahmedabad University
Gujarat, India 380009
harsh.s.btechi15@ahduni.edu.in

Hardik Udeshi
School of Engineering & Applied Science, Ahmedabad University
Gujarat, India 380009
hardik.u.btechi15@ahduni.edu.in

***Abstract*— Signatures are widely used as a means of personal authentication for different financial documents such as bank cheques. Frauds due to signatures can be calamitous because a high amount of information can be made available by signature authentication. This means that signatures need to be verified from financial documents to guarantee security of sensitive data. Offline signature verification systems work on the scanned images of dataset and identifying whether a signature is genuine or fake. These systems are particularly helpful for processing cheques in banks. This system inherently involves detection of signatures from cheque images, localizing the signature and verifying its authenticity. We demonstrate the end-to-end pipeline required to build such a system. The possible approaches for constituent problems, implementation details and discuss our results in the paper.**

***Keywords*— *Signature detection, Bank cheques, SVM, Connected Component, Optical Character Recognition***

## I. Introduction

Bank cheques and legal documents are integral components in modern economy and are important documents which are authenticated by the account holder's signature. Signatures on these documents can be forged and this entails a high-risk factor while approving these documents' veracity.

Approaches to signature verification fall into two kind of categories. Online systems record location, velocity, pen pressure, etc. to classify the signature as genuine or forged. Offline systems, on the other hand, work on scanned images of signatures. These systems prove favourable for working on cheques or legal documents' images.

From scanned images of cheques, we first need to extract the signature and localize it. To achieve this goal, we have used approaches based on edge-detection, contours formation, OCR, connected components labelling and line-sweep method. In part II of this paper we have first discussed the pre-processing steps and then we have discussed the approaches used by us. We also have discussed merits and demerits of using all the approaches along with the outcomes of each approach

For signature verification, we use different features for characterizing signatures such as convex hull area, contour area, aspect ratio, bounding rectangle area, etc. These features are then used in SVM and ANN classifiers to detect genuineness of a signature. Part III deals with pre-processing steps and feature extraction for signature verification followed by SVM & ANN construction procedures. Part IV talks about results of our signature verification approaches. Finally, conclusions are drawn in Part V.

## II. Signature Detection

### A. *Pre-processing Steps*

First, the RGB input image is converted to grayscale and then binarization is performed on the image to remove background noise and emphasize the signature to be verified.

### B. *Approaches used for signature detection*

To detect and localize the signatures from the documents, we used following three approaches.

#### 1) Contour based detection

In this approach, we first carried out edge-detection and for that we used Canny Edge Detector and for finding thresholds for Canny Edge Detection, we use Otsu Thresholding approach. After that, we formed contours based on edge-detection. After that, for each individual contours, we checked the number of corners present in that particular contour. Then, the contour with maximum number of corners was selected as the signature and based on that, we could fit rectangle on that contour.


*Fig 1: Correct signature detection using contours*

This approach could not work well in the scenarios where the signatures were not fully connected or when the document contained some printed/handwritten letters which were very close to each other. In the latter case, the canny edge detector could not perform accurately and hence contour formation was incorrect.

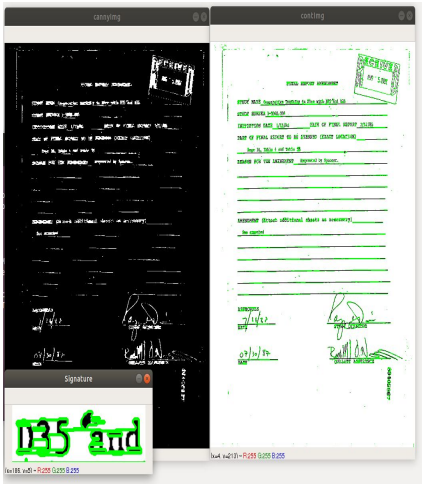


*Fig 2: Wrong detection using contours*

*2)* OCR based detection with connected components labelling

For OCR based detection, we used some heuristics. We assumed all the documents to be having some phrase written (mostly printed) under the signatures. This phrases could be anything like “Please Sign Above”, “Sign Here” etc. We gave this phrase as an input to OCR and based on the length of the found phrase, we fit a box above it assuming the signature will be inside that box.

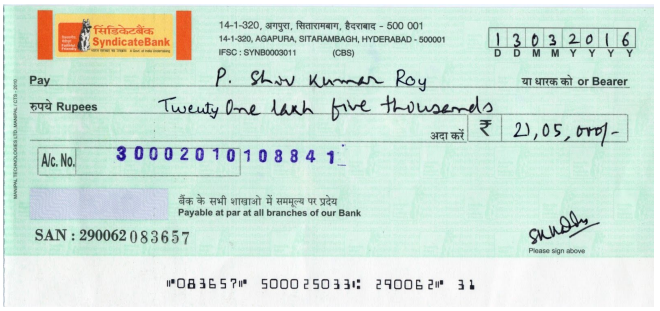



*Fig 3: Cheque with a phrase “Please Sign Above”*

This could not work in cases where the actual signature overlapped the phrase given to OCR. Obviously, when the document did not have any such phrases, this would fail, but it is very commonly seen that all the documents contain some phrases like these.

We cropped the box that we had formed. But in some cases, when there were some other text above the input phrase, the box contained that text too. For e.g. in our case, some cheques contained the name of the signatory above the phrase “Please Sign Here”.

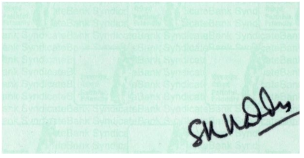


*Fig 4: Cropped signature using OCR approach*

To remove that non-useful text, we used connected components labelling algorithm to fit a tight box around the signature. Two passes of connected component labelling was applied to get the connected components in an image. Then rectangle fitting was applied to the detected connected components. Components are selected on the basis of maximum area of the rectangle.

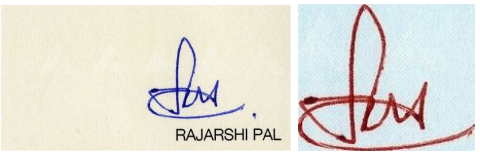



*Fig 5: Removing non-useful text using connected components labelling*

But, this algorithms fails when the components would be unconnected. At that time, it will not detect the complete signature.

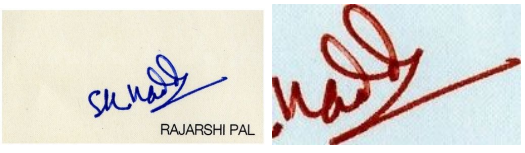



*Fig 6: Failure scenario of the approach*

*3)* OCR based detection with lineSweep method

We have used OCR for the same context as mentioned in approach-2 above. But, we have used different algorithm for rectangle fitting across the signature.

Line-Sweeping algorithm was run in both ways: vertically and horizontally. On the basis of this, the exact signature is detected and after that this image is passed as an input to the verification algorithm.

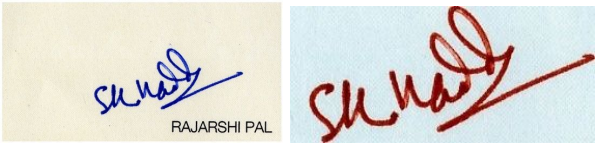



*Fig 7: Successful localization of signature*

## III. Signature Verification

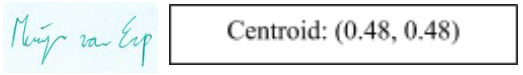
### A. Pre-processing Steps

First, the RGB input image is converted to grayscale and then binarization is performed on the image to remove background noise and emphasize the signature to be verified.

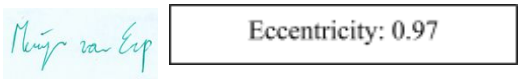
### B. Feature Extraction

Features are extracted from signature images to feed into the classifier. We use SIFT features along with geometrical features, based on shape and dimensions of a signature, to accurately represent the image. The description of features is as follows:

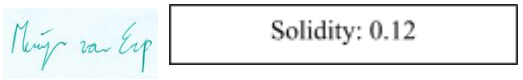
- **Centroid**: Horizontal and vertical centres of signature region.



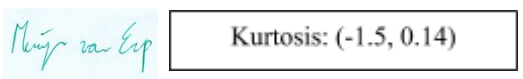

- **Eccentricity**: The ratio of the distance between the foci of the ellipse and its major axis length.



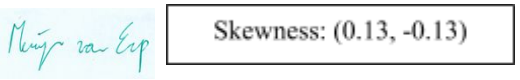

- **Solidity**: The Ratio of pixels in the region to pixels of the convex hull image.



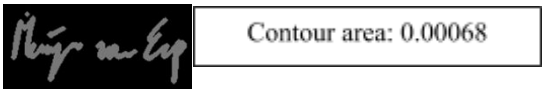

- **Kurtosis**: It is a measure of flatness of distribution. High kurtosis denotes low noise.



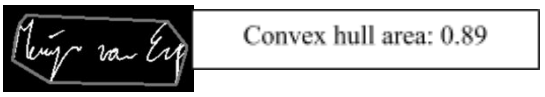

- **Skewness**: Measure of asymmetry of distribution. Hazy and smooth surfaces are more positively skewed.



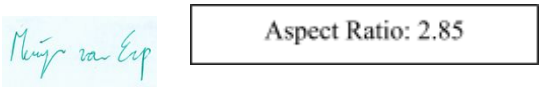

- **Contour area:** Area of the contour formed around the signature region.




- **Convex hull area:** Area of tightly bounded region of the signature.




- **Aspect ratio:** Ratio of width and height of signature images




- **Bounding rectangle area:** Area of the rectangle formed around the signature region; product of width and height in calculating aspect ratio.

- **SIFT Features:** To make up for the rotation and translation of signature present in image region

These features describe a whole image and are fed to the SVM and ANN classifiers which will be described in the following section.

### C. Dataset in [1]

This dataset consists of a set of genuine and forged signatures with 5 signatures for each of the 12 individuals. The signatures are clear images without any noisy background. The forgery signature in this dataset is the mixture of random, simple and skilled forgeries.

(a) Genuine signature (b) Forged signature

### D. IDRBT Cheque Image Dataset [3]

This dataset was the result of running our signature detection algorithm on the cheque dataset provided by IDRBT. For signature detection, we input the cheque images and use OCR to detect texts ‘Please’, ‘sign’ and ‘above’. First, we approximate a bounding rectangle based on the

location of these texts. Second, we run the connected component labelling algorithm or a line sweeping algorithm to isolate the signatures from the cropped image. These final output images from detection module contain background noise and hence, they are subject to pre-processing steps, mentioned in II-A, before the signature verification module. For signature verification, the model needs genuine and forged signatures. We manually generate forged signatures found on cheques and feed the true data along with this fabricated data to the classifier.

(a) Genuine signature

(b) Forged signature

*E. Approaches*

*1) SVM*

A Support Vector Machine is a discriminative classifier defined by a separating hyperplane. Given a labeled training data, the algorithm outputs an optimal hyperplane which forms the decision boundary and categorizes new examples.

The geometrical features and SIFT features that are extracted in II-A are concatenated into a long vector and this forms the representation of a signature image. For each image, this vector is appended to the training data matrix. The labels are also provided based on whether the signature image is drawn from genuine training or genuine forged set.

This feature vector is passed to a SVM with a linear kernel. For the test dataset, genuine and forged signatures are input to the SVM and the predicted and actual labels are matched to compute the accuracy.

*2) ANN*

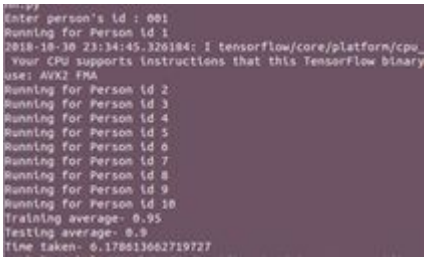
An Artificial Neural Network is a brain-inspired system deployed to find complex patterns in data. Neural networks for classification, consist of input and output layers, as well as a hidden layer consisting of units that transform the input into something that the output layer can use. These networks and their deep layer counterparts have proven to be extremely useful when dealing with image data.

The geometrical features and SIFT features that are extracted in II-A are concatenated into a long vector and this forms the representation of a signature image. For each image, this vector is appended to the training data matrix. The labels are also provided based on whether the signature image is drawn from genuine training or genuine forged set.

The neural network architecture used consists of an input layer, a hidden layer with 7 neurons and an output layer with 1 neuron which outputs 1 if the signature is genuine or forged. For the test dataset, genuine and forged signatures are input to the ANN and the predicted and actual labels are matched to compute the accuracy.

## IV. Results

- ANN Accuracy with 5 features:




Training Accuracy: 95%, Test Accuracy: 90%

- ANN Genuine:

Genuine Image is the original image that is cropped after applying signature detection algorithm. We applied an ANN classifier to classify whether the image is Genuine or Forge.

*(a) Binary Image of the Genuine Signature*

*(b) Boundary Box of the Genuine Signature*

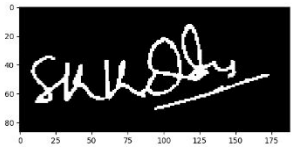
*(c) Convex Hull Area of the Genuine Signature*

*(d) Contour formation around the Genuine Signature*


*(e) Result: Clearly classifying Genuine Image*

- ANN Forged:

Forged Image is the forged signature that we made manually and cropped the signature part of the image.



*(a) Binary Image of the Forged Signature*

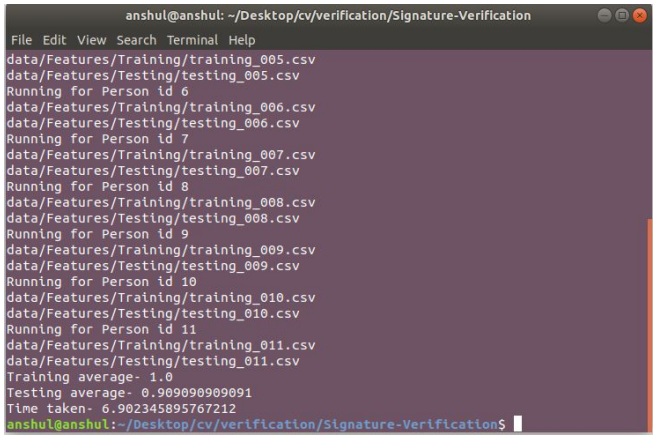

*(b) Boundary Box of the Forged Signature*


*(c) Convex Hull Area of the Forged Signature*


*(d) Contour formation around the Forged Signature*

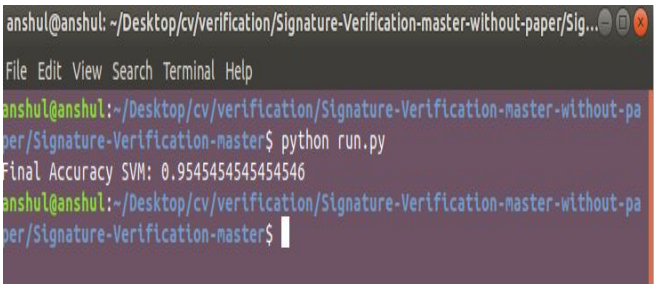

*(e) Result: Clearly classifying Forged Image*

- ANN Final Accuracy




*ANN Final Accuracy: 90.90%*

- SVM Final Accuracy




*SVM Final Accuracy: 95.45%*

The features computed before passing to the SVM Classifier were the geometrical features and the SIFT features while to the ANN only geometrical features were passed.

We added contour features and SIFT Features to the set of geometric features in both approaches. Because of the SIFT features, accuracy in SVM based approach turns out to be 95% while it is 91% in the ANN based approach. To stress on the significance of SIFT features and compare their effect on classifiers, we run our approach after augmenting dataset and removing the SIFT features from the feature set in ANN classification method and noticed an 11% decrease in the accuracy. So the final accuracy of ANN without SIFT features decreased to 80% from 90.90%.

## V. Conclusion

OCR based approach followed by line-sweeping method (which actually is heuristics based) gave the best results as it has very reasonable assumptions and most of the cheques/signatures follow those assumptions. Other approaches have some major assumptions which may not always be feasible in real scenario.

For signature verification, we have used some geometric features like centroid, solidity, contour area, convex hull area and others. We arrive on the conclusion that SIFT features along with geometric features provide the required characterization of signatures for them to be distinguished as genuine and forged.

## VI. Future Work

We can apply our signature detection approach on all financial documents instead of just cheques. Financial documents have different structures than cheques so our approach will have to modified to incorporate those changes.

We have increased the size of data set on our own but it is still not very large. We can create a much larger dataset so that we can use deep learning techniques on that and compare/improve the quality of the results.

## VII. References

[1] Karouni, Ali, Bassam Daya, and Samia Bahlak. "Offline signature recognition using neural networks approach." *Procedia Computer Science* 3 (2011): 155-161.

[2] Chandra, Subhash, and Sushila Maheskar. "Offline signature verification based on geometric feature

extraction using artificial neural network." In *Recent Advances in Information Technology (RAIT), 2016 3rd International Conference on*, pp. 410-414. IEEE, 2016.

[3] IDRBT Cheque Image Dataset [http://www.idrbt.ac.in/icid.html ]

[4] Liwicki, M.: ICDAR 2009 Signature Verification Competition. http://www.iaprtc11.org/mediawiki/index.php/ICDAR_2009_Signature_Verification_Competition_(SigComp2009). Accessed 24 Feb 2015 (2009)

[5] Cüceloğlu, İ., Oğul, H.: Detecting handwritten signatures in scanned documents. Proceedings of the 19th Computer Vision Winter Workshop, pp. 89–94 (2014)