

The background is a dark blue gradient with a subtle pattern of white dots. Overlaid on this are several faint, light blue circular elements. On the left side, there is a large circular scale with tick marks and numbers ranging from 150 to 260. To the right of the scale, there are several concentric circles, some of which have arrows indicating a clockwise direction. The overall aesthetic is technical and modern.

# A GENERALIZED APPROACH FOR CHEQUE'S INFORMATION VALIDATION

MANEESH KUMAR MEENA

STATE BANK OF INDIA

# PROBLEM STATEMENT

- Objective of this project is to build a model that can identify hand written information on the cheque such as amount (both in digit and text) and date. It will also validate the information as well as cheque quality. It means it can locate discrepancy in hand written information and also cheque whether cheque is a original one or not.

# SUGGESTED IDEA

- We suggested a image processing model for these objectives.
- This idea works in 3 main steps.
  - Designing a ui that will help us in easy use of model
  - Image processing part, that has both training the model and testing part. Training model mainly focus on how to get our model to its best definition of recognition. And also what type of data will be consider as correct and valid.
  - Image segmentation and localization and further inpating and so on process to get right image for right process

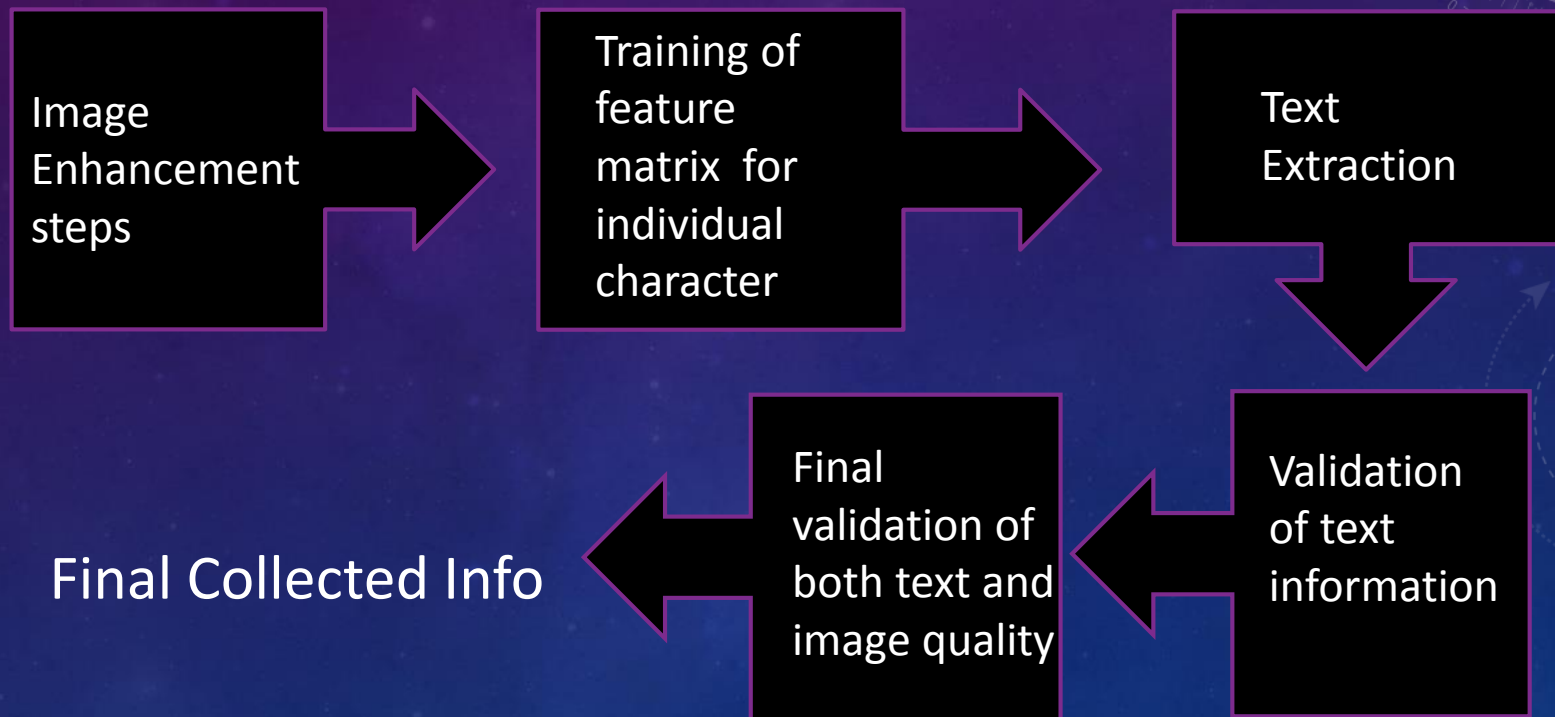
Further we will create quality model as well as text model to satisfy both text validation and checqe quality valiation.

# MOTIVATION

- It is a really efficient model in the sense of effectiveness and usages. It is an automatic approach and can be modified in future if some changes needed.
- Use of training model for hand recognition gives freedom to detect any regional language , for now we are doing it for English but it can be implement for any other language as long as we have available training data of hand written characters .
- Quality validation is done based on the color properties of the cheque image, We did it in 4-different color spaces (total 12 spaces), which will give us a effective model of cheque quality detection and thus prevention of cheque fraud.

# METHODOLOGY

- Basic idea





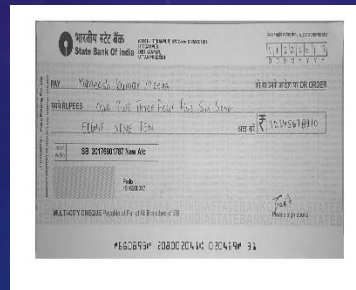
# IMAGE ENHANCEMENT STEPS

- This step includes correction of information like correction in level values and property enhancement steps.
- Since scanned image will be taken as input to this model, further color image processing and image space conversion will be done.

original



Gray



HSV



LAB



# TRAINING OF CHARACTER MODELS

- Collecting data related to hand written character
  - This steps involve the collection of hand written digits, small and capital latter characters. We will use this data to train our detection model. Here are some sample images

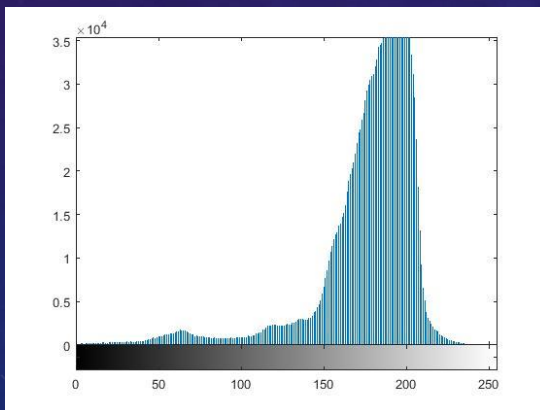


- Saving features of text image into templates to use for detection
  - This collected information related to each and individual character and save it in a template . Ex PCA, correlation matrix etc.
  - These templates contains data related to each character so while predicting we will call these properties to identify text.

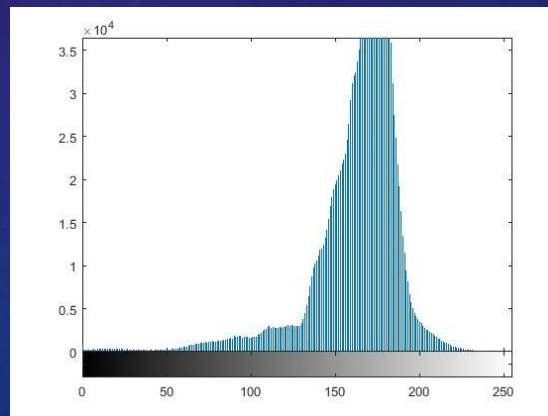
# CONTINUE...

- Now for the prediction of cheque quality. We will train threshold model/multi-classification model. It will create histogram of an image and look for the basic properties of the cheque.
- We will use total 12 color spaces to do so. Below histograms are shown for three spaces.

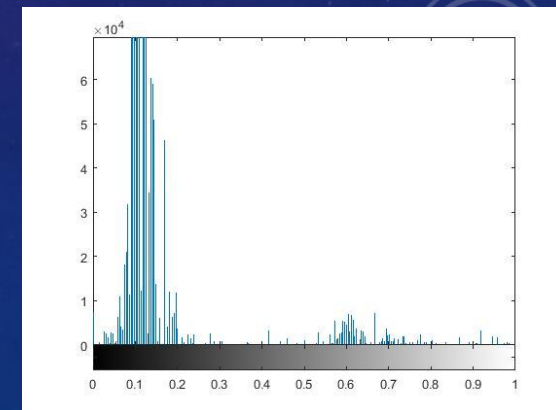
Gray histogram



B histogram



H histogram





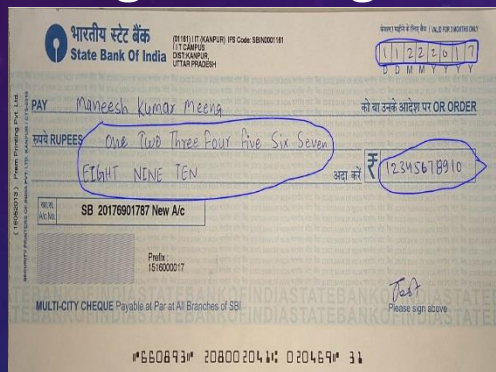
# TEXT EXTRACTION

- This is an important step, this includes multiple processes
- 1<sup>st</sup> collected the image that we need to extract.
- 2<sup>nd</sup> divide this image into sub-divisions or image segmentation.
- 3<sup>rd</sup> , for each segmented region, run text detector
- For digit region like amount in digit and date, we have implemented digit detector. And for text part we have separate text detector

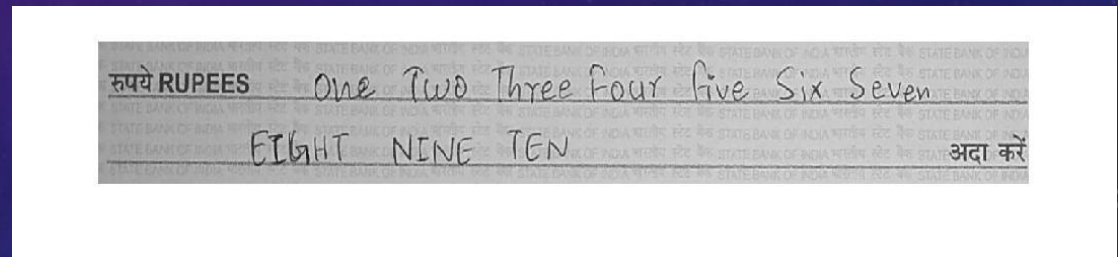
# CONTINUE...

- Working mechanism of both detector are same, both find region of interest and segment ROI in lines and lines into digits or latters.

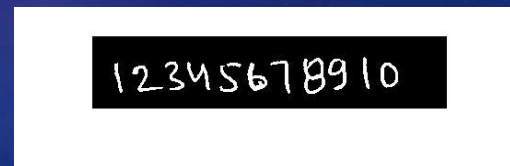
original image



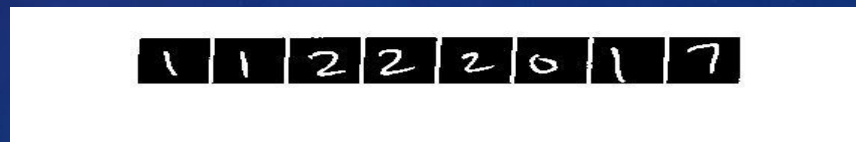
segmented text amount



segmented digit amount



segmented date



# GUI

- Simplest GUI to interact with system.

GUI

Load

Extract

Validate

Cheque Date

Amount in Digits

Amount in Text

Quality Warning

Account details

Account Prefix

Branch Details

Account details goes here

Show prefix if any

Branch details goes here

# EXTRACTION OF TEXT AND DIGITS

- Matching of patterns from training modules saved in templates . It includes following steps.
- It first extract the feature out of segmented image and correlate with template features. Closet matching will be the final detected digit.
- correlation function for images

```
global templates
comp=[ ];
for n=1:num_letras
    sem=corr2(templates{1,n},imagn);
    comp=[comp sem];
end
vd=find(comp==max(comp));
```

- Loading data from saved template and matching features and finding out the most correlated one

# CHEQUE QUALITY VALIDATION

- Quality model based on color spaces. Fortunately SBI bank have amazing color space properties.
- Used 12 different color spaces to identify quality of cheque
- In future, energy of color space will be added as a feature which will increase its effectiveness .
- This is a light and color sensitive information so all scanned cheque should be in same DPI the rules and threshold can be regulated (**REALLY IMPORTANT**).
- For now, I'm considering my camera quality image, In full model it will be applicable to bank standard too.



# FINAL VALIDATION

- Predicting final results. It includes modification of extracted results.
- For example, modification of text amount based on extracted digit amount and extracted info, since it is too difficult to get all characters correct.
- Raising warning level if cheque quality looks defective and also if extracted info does not matches mentioned info.

# RESULTS

- This includes following extraction and validation results
  - Extraction of hand written information (these are further validated with original entity)
  - Extraction of cheque's OCs , that includes account no, branch details, prefix etc.
  - Check of image quality, level of alarm, predict for fake cheque.

# CONCLUSIONS

- Matching is done based on available training data on hand written text. So , it can be applied to any language as long as training data is available.
- Different detector are built to detect digit and character so if we about to change language digit identifier will not need modifications.
- Extracted information is being validate so conflicts can be removed.
- Image quality has several layer of warning so we can act accordingly.
- Image quality cheque can prevent frauds in cheque banking but it can be applied once standard cheque image are available.
- Further upgrades and modification can be done to enhance the model.

# REFERENCES

- ML references from my own project.  
Maneesh, Varun and others, “Data Mining and Knowledge Discovery”, IME 672 (course project).
- Matlab, [www.mathworks.com](http://www.mathworks.com)
- Maneesh, Arvind, “OCR for camera based handheld devices”, Image Processing.
- A little bit of IOT.

# QUESTIONS







THANK YOU