

Automatic Cheque Clearance System using MATLAB

B. Janardhana Rao¹, O. Venkata Krishna²

¹CVR College of Engineering, ECE Department, Hyderabad, India
Email: janardhan.bitra@gmail.com

²CVR College of Engineering, EIE Department, Hyderabad, India
Email: venkatakrishna.odugu@gmail.com

Abstract— Nowadays Cheque Truncating System (CTS) compliant cheques are used in banks for saving time and reducing effort in depositing cheques. With the development of image processing it is possible to make machines read the documents instead of humans. Applying the image processing technique in the cheque deposition in banks can reduce the human effort, time and also will be cost effective. In this paper, an automatic cheque clearance system is discussed using Optical Character Recognition (OCR) system by feature extraction and pattern recognition concepts in image processing. Using this one, can reduce time, human effort and money in processing of cheques. The algorithms are executed using MATLAB.

Index Terms— CTS, OCR, cheque, MATLAB.

I. INTRODUCTION

Cheque truncation is the process of stopping to send the physical cheque to the drawee branch issued by the drawer branch. Instead of sending the physical cheque, the electronic image of the cheque along with the relevant information like date of presentation, MICR fields and presenting banks etc. is sent to the drawee branch. This process would help to avoid moving the cheques physically across the branches except in some exceptional cases. This would result in reduction of the time required for payment of cheques thus speeding up the process of realization or collection of the cheques [1].

This paper uses the image processing algorithms like feature extraction and pattern recognition for automation and faster clearance of cheques.

A. Different fields of cheque

Cheque is the main document for payment of money from a payee an account or payment to another bank. So it is necessary to know the different fields in a cheque. In general some fields of the cheque are already filled and some fields are given with dotted lines or boxes that have to be filled and signed by the account holder who issues the cheque. All the cheques contain name of the issuing bank and some cheques contain name of the account holder along with account number printed on each leaf of the cheque book. In case these are not given, one has to fill these details in correctly. Let us discuss different fields in the cheque with reference to the sample cheque, shown in figure 1.

The name of the person to whom the money has to be paid is written and has to be spelt correctly close to the “pay”. Date must be written properly, the current date for the receiver to be able to receive the payment should be written. At the end of the line from “pay” we will find the words “or bearer”. This refers to the person who is carrying the cheque and payment will be made to this person if the cheque does not specify the person whom money has to be paid. It is advisable to cross the “or bearer” printed on the cheque when writing out a cheque to avoid a stolen cheque from being misused.

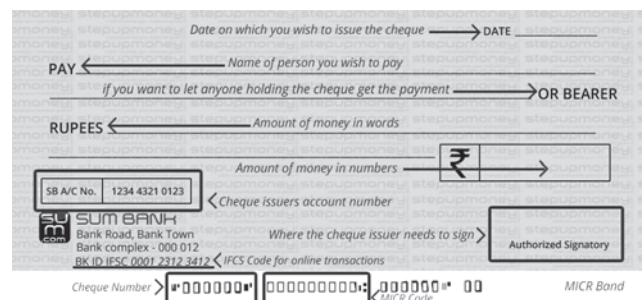


Figure 1: Sample cheque

The payment amount should be written in words next to the printed words on the cheque saying ‘Rupees’ and the same value should be written in the box beside it in figures. The account holders name is printed at the right bottom of the cheque. Signature has to be made above the field where name is printed. For the purpose of identification each cheque has a different number which is different from bank to bank. The cheque number is normally a six digit number enclosed within inverted commas. The cheque bears a nine digit number called Magnetic Ink Character Recognition code (MICR) which is unique for each branch of a bank in India. The cheque also bears the drawee bank, state, branch and sort code which are a series of numbers appearing next to the cheque number. The cheque may also contain the internal code issued by the bank called transaction id.

II. OPTICAL CHARACTER RECOGNITION

Optical Character Recognition Technology (OCR) is the technology used for converting the scanned documents into machine readable format files. It is not possible to search the content in a scanned document that is stored in the memory of a computer. OCR is the technology which

enables recognizing the characters through an optical mechanism automatically [7]. OCR can recognize both printed text and handwritten text [3]. The performance of the system is dependent on quality of input documents and this OCR is designed to process the documents that contain almost text with little non text clutter. OCR is the process of converting electrical and mechanical forms of scanned images of type written and handwritten into machine text. This machine text is digitized for easy search, to store more compactly and for display online.

OCR technology enables the scanned images to become documents that can be fully searchable with text content that is recognized by the computer. OCR is the process to extract the document information and enter automatically instead of manually entering the information into the electronic database which results in efficient information processing in less time and is more accurate. The block diagram of the OCR is shown in figure 2 [4] [5].

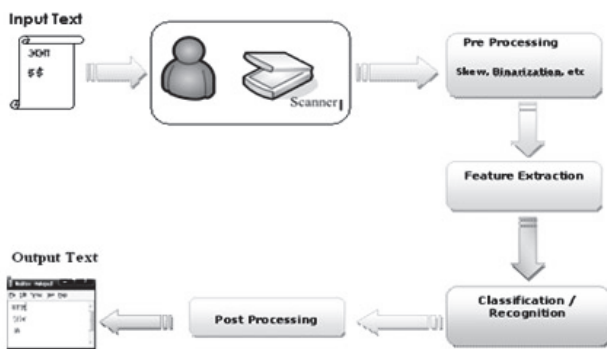


Figure 2: Block Diagram of OCR

The main functional blocks in OCR are Image Acquisition, Pre-processing, Feature Extraction, Pattern Recognition, Post-processing and display. Image acquisition is the process of obtaining text image from the scanner or pre stored files.

A. Scanning:

The scanner contains the camera which captures images and converts original document into digital image. In general original documents are made up of black colored text printed on the white background. Scanning comprises, thresholding which converts digital image as gray scale image. Thresholding is the process of converting multi-level image into bi level image called black and white image. The fixed value is defined in thresholding, if the gray levels are below the threshold level then it identified as black, whereas, if the threshold is above it is identified as white which results in saving memory and computational efforts. The scanner scans the pixels of the text and not the characters. The data may carry some noise while patterns are scanned and digitized. The scanner with low resolution may produce smeared images and touching line segments.

B. Preprocessing:

The Preprocessing of image contains the operations like RGB to GRAY conversion, Segmentation, Edge

detection and Image enhancement etc. The noise occurring in scanning process results in poor recognition of the characters and this problem is overcome by preprocessing the image. The preprocessing comprises normalization and smoothing. The normalization handle the uniform size, slant and rotation of characters and in smoothing some rules are applied to the contents of the image with the help of thinning and filling techniques.

C. Feature Extraction:

The feature extraction is the process of obtaining information about the object or group of objects in order to facilitate classification. This is the important module in the OCR system. The characteristics of the image are called features. It extracts the features of the symbols. In this process the symbols are characterized and unimportant attributes are left out. The feature extraction identifies the features like intersections, open spaces, lines etc. but does not match concrete character patterns. An algorithm is used to implement feature extraction which is concerned with the representation of symbols.

The document applied as input contains several lines of text that needs to be categorized into single character for recognition. The following steps are applied for this purpose: The initial darker pixel is named as the top of the row in the scanned document. The next blank line is detected for the bottom. The area between this top and bottom is the row of characters in the image.

Now each character is identified for the rows. This is obtained by scanning the row vertically from top to bottom. The leftmost pixel of the character is the first darker pixel detected. Now if all the pixels are found to be blank, then this is right of character. The character obtained from the scanned image is normalized to 15 X 15 pixel from any pixel size. This is done by cropping the image using top, left, right, and bottom boundaries. Now the 15 X 15 size cropped image can be converted into array of 15 X 15 binary image using thresholding. Here logic 1 represent black and logic 0 represent white.

D. Pattern Recognition:

The pattern recognition is done by matching the obtained binary format with the existing templates. In order to obtain this, the binary format is divided into 5 tracks and each track is again subdivided into 8 sectors. The matrix is generated corresponding to each track- sector for identifying number of pixels in each region. The obtained track-sector matrix is to be matched with the existing templates [2]. The template which is already existing contain each track-sector intersection value i.e. each track and sector value. The character is identified if all the parameters are matched with the template values. Then each and every font consists unique values in the matrix, this makes it easy to identify each font separately.

E. Post Processing:

The post processing consists of the steps that improve the quality of obtained data like filtering.

III. DESIGN AND IMPLEMENTATION

The implementation of the cheque clearance system is explained with the flow chart shown in figure 3.

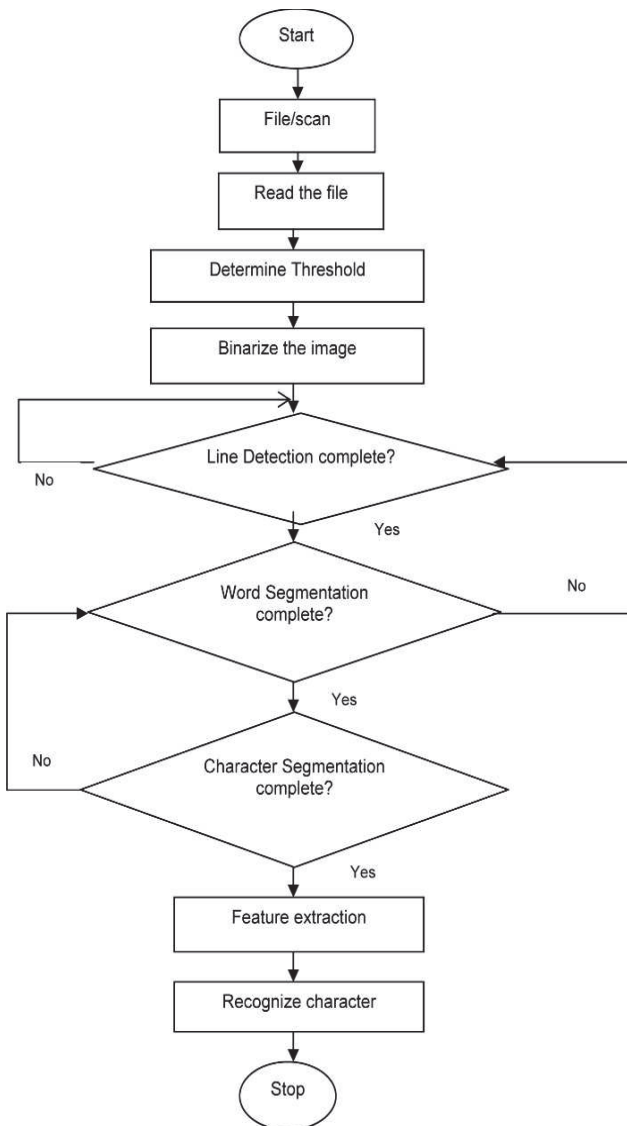


Figure 3: Flow chart for the Cheque Clearance system

Here for the operations to take place, we consider the sample CTS cheque of SBI and fill the necessary details in the cheque and then by using a normal scanner we scanned the cheque and converted it into an image of .jpg format. This process is called Image acquisition in Image processing. Then the image file is read for doing necessary operations on it. The color image is converted from RGB to grayscale, then the image is binarized using thresholding so as to store the image as matrix in digital format. Then the size of the image or matrix is obtained. It is called line or margin detection. Then we determine the necessary fields of the cheque to be extracted and crop it to the new image. This is called feature extraction. The noise obtained from the cropped image is removed by using filtering and

segment the cropped image is segmented as individual characters.

We initially have to create the template of characters for both alphabets and numbers for matching with the segmented characters of the cropped image. This is called template formation. Then we have to compare the individual characters with templates created to find the best match. This is called template matching. After the best match is found then we have to store that character. Then we have to calculate cheque number, using those characters as

$$ChequeNum = num(1) \times 10^5 + num(2) \times 10^4 + num(3) \times 10^3 + num(4) \times 10^2 + num(5) \times 10 + num(6)$$

Then we will extract remaining features of the cheque like amount, account number, digital signature and MICR number by using image preprocessing operations followed by optical character recognition algorithm implemented in Matlab.

IV. SIMULATION RESULTS

The scanned copy of State Bank of India cheque with necessary details filled in cheque is taken as input file to extract features as shown in figure 4.



Figure 4: Input file

The binary image obtained after the RGB to gray scale conversion and threshold operation is shown in figure 5.

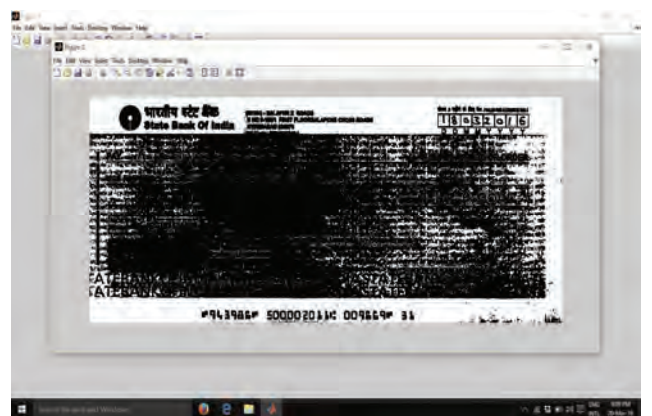


Figure 5: Binary image

Now we go with edge detection to find the areas where key information is located. So we can resize the image and find the number of rows and columns and filter the image for removal of any noise. Now we will crop the cheque number to another matrix. The cropped cheque number stored in another image is filtered for removal of noise. Then we will crop the individual numbers in cheque number for template matching. Templates are either characters or alphabets that are created using a Matlab program as shown in figure 6.

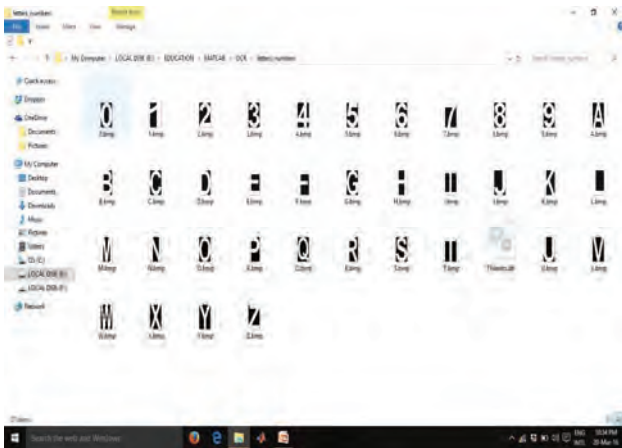


Figure 6: Templates

The extracted individual cheque numbers are then matched with existing templates to find correct match. This is called template matching. Then the cheque number is calculated and displayed from the characters extracted from template matching as shown in figure 7.

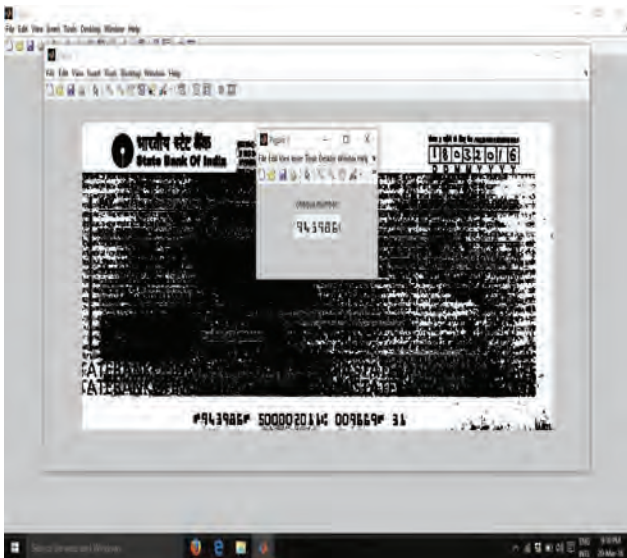


Figure 7: Cheque Number Extraction

In a similar fashion the MICR number extracted is shown in figure 8. Similarly the Account number extracted is shown in figure 9.

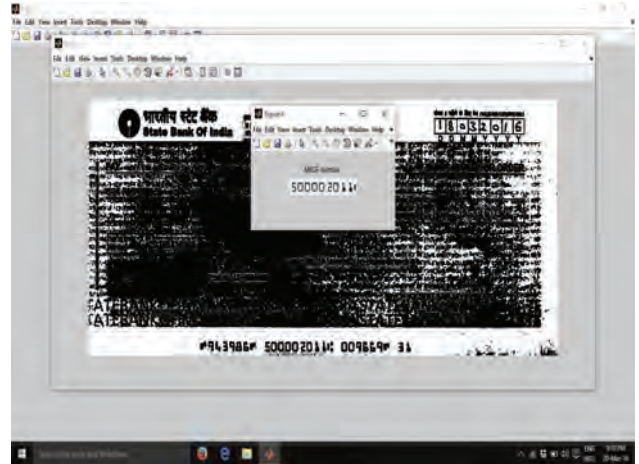


Figure 8: MICR number Extraction

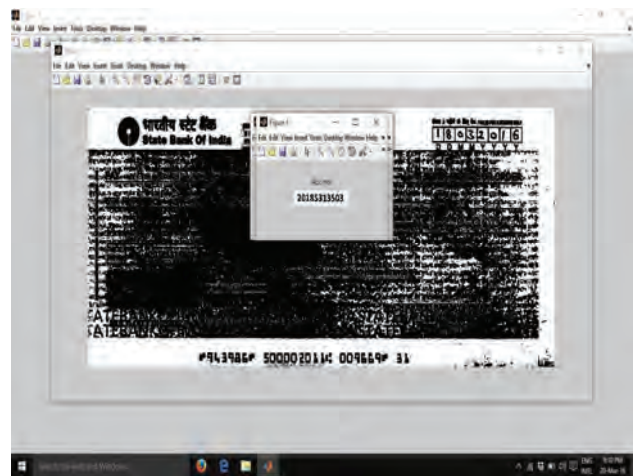


Figure 9: Account number Extraction

Similarly the Amount extracted is shown in figure 10.

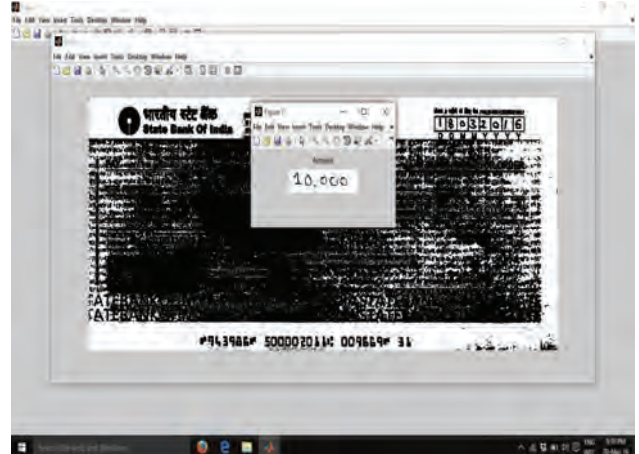


Figure 10: Amount Extraction

Similarly the signature extracted is shown in figure 11.

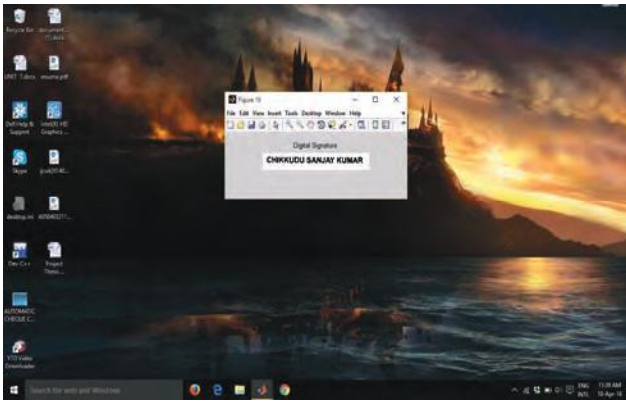


Figure 11: Signature Extraction

V. CONCLUSIONS

In this paper an automatic cheque clearance system is implemented in MATLAB using image processing applications. The different fields in the Cheque truncating System (CTS) compliant cheque like cheque number, MICR number, amount, account number and signature are extracted by using Optical Character Recognition (OCR) technology by image processing concepts which are feature extraction and pattern recognition. Using this one can reduce the human effort, time and money for automatic processing of cheques. The results are obtained for CTS, SBI cheque by extracting cheque number, MICR number, Account number, amount and signature.

REFERENCES

- [1] Mohit Mehta, Rupesh Sanchati and Ajay Marchya, "Automatic Cheque Processing System", International Journal of Computer and Electrical Engineering, Vol. 2, No. 4, August, 2010.
- [2] S. T. Kahan, T. Pavlidis, and W. Baird, "On recognition of printed characters of any font and size", IEEE Transactions of Pattern Recognition and Machine Intelligence, PAMI-91987, pp.274-285.
- [3] R. Smith. "An overview of the Tesseract OCR Engine" Proc 9th Int. Conf. on Document Analysis and Recognition, IEEE, Curitiba, Brazil, Sep 2007, pp629-633.
- [4] Vamsi Krishna Madasu, Brian Charles Lovell, "Automatic Segmentation and Recognition of Bank Cheque Fields", Proceedings of the International Conference on Digital Image Computing: Techniques and Applications.
- [5] Richa Dhaundiyal, Surbhi Mathur, Rupesh Bhirud and AjayVideka, "Automated Cheque Processing System", Proceedings of the NCNTE-2012, Third Biennial National Conference on Nascent Technologies.
- [6] CTS Clearing House Interface – Specifications NCR Corporation & RBI/NPCI, 24th August, 2010.
- [7] Nallasamy Mani and Bala Srinivasan, "Application of Artificial Neural Network Model for Optical, Character Recognition", IEEE Computer 0-7803-4053-1/97