MAT JOURNALS

Journal of Image Processing and Artificial Intelligence
e-ISSN: 2581-3803
Volume 5 Issue 1

Automatic Extraction of Attributes from Printed Indian Cheque Images by Template Matching Technique

Priya M.P
MscTech by Research, Department of Studies in Computer Science, University of Mysore, Mysuru, Karnataka, India
Email: *22priyamp@gmail.com*
DOI: *http://doi.org/10.5281/zenodo.2597063*

Abstract

Reserve Bank of India (RBI) has introduced Cheque Truncation System (CTS) for Indian banks in order to reduce the time required for physical movement of cheques between the clearance departments. However, other processes including database entry and verification are carried out manually. The proposal here is to eliminate the manual intervention by extracting the attributes from the input cheque image and updating the database automatically which significantly would reduce the time lapse on filling up the data into the database. Automatic database updating also contributes to provide secure data retrieval through querying system for verification of attributes by concerned banks. In this paper, a novel approach to extract printed attributes from Indian Bank cheque images based on their template structures is proposed. Template structures are determined by extracting the MICR code from the input cheque image. Important attributes region is segmented, and the printed data is recognized. Extensive experiments demonstrate the efficacy of the proposed method.

Keywords: *Indian bank cheque, Cheque image processing, attributes extraction, template matching.*

INTRODUCTION

The banking and financial sectors in India use Cheques as mode of payment till date to attain cashless financial transactions as well as accountability. Initially, cheques presented at the drawer banks were physically moved to the drawee banks for clearance which involved manual labour costs and delay in clearance. As a solution, Reserve Bank of India (RBI) introduced Cheque Truncation system (CTS) which reduced the time required for physical movement of cheques between banking institutions. Here, Magnetic Ink character recognition (MICR) code is acquired by the drawer bank to determine the drawee bank and branch and the cheque images are transmitted electronically to the corresponding bank branches.

Various methods are available in the literature for extraction of MICR code from cheque images. The method [2] makes use of the preliminary information of spatial location of MICR code, extraction is performed by applying horizontal and vertical projection profile [18] [19]. However, database entry and verification of data from the cheque is still manually undertaken by the transaction-clearance departments which is again a tedious process and involves manual errors. This urged for research to automatically retrieve data from cheque images. Though different banks possess their own cheque formats/standards, most of the attributes such as Indian Financial System Code (IFSC), Magnetic Ink Character Recognition (MICR) code, account number and cheque number are printed. However, it is challenging to extract account number when it is ink-stamp printed. A sample cheque image marked with key attributes that are updated as part of transaction into the database by the banks is shown in Figure.1.

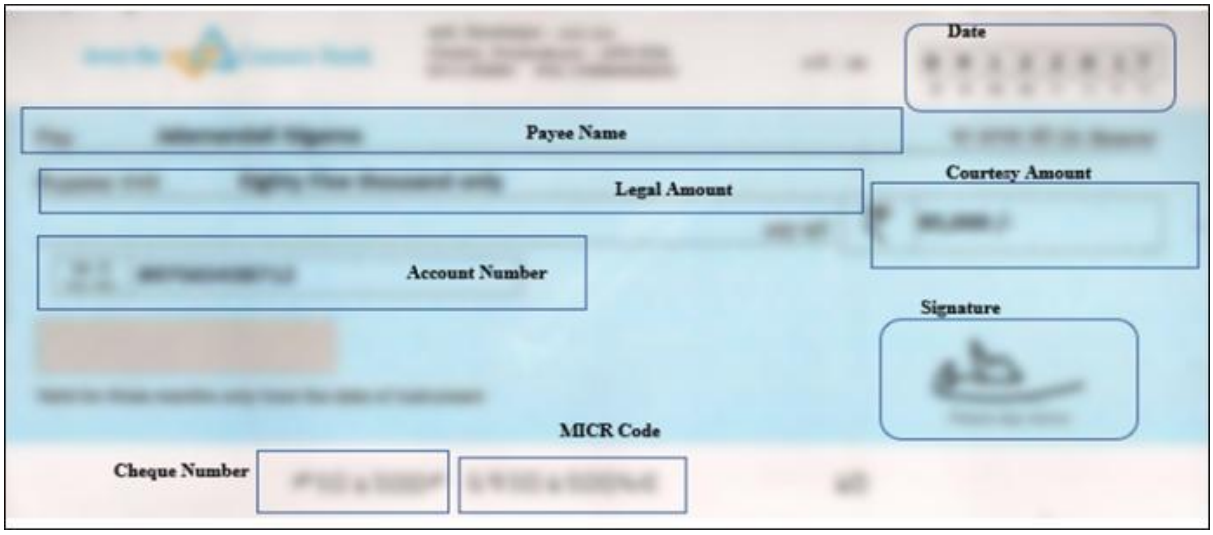

Figure 1: *Key Attributes Recorded from Cheque to the Database by Banks.*

Literature Review

In the literature one can find various proposals to extract important regions [13] in a cheque image as discussed below.

Latest work found in the literature to extract important regions in a cheque image is proposed by Prabhat et al. [1] based on identification of lines in cheque images to locate the data regions.

Ashar et al. [3] have proposed research work on detection and extraction of rectangular pantograph region from a cheque image. Techniques like Maximally Stable Extremal Regions (MSER) are used to isolate a specific region on the cheque image. Agarwal et al. [5] proposed method to distinguish between handwritten and printed text data in the cheque images.

Histogram encoded set of heuristic rules are applied to courtesy amount region for extraction. Research works have also emphasized approaches to extract courtesy amount using surrounding bounding box, connected component analysis [6] [7]. Payee name data field can be extracted using graph-based algorithm which also uses connected components [8]. V. K. Madasu [9] have proposed sliding window approach to estimate entropy-based pixel density in the window to binarize [15] the cheque image using this entropy.

Connected components are labelled and these labelled regions are categorized using four fuzzy texture features and hence fuzzy neural network is trained to extract these labelled regions in the cheque images. Lal Chandra et.al [6] proposed an algorithm to automatically detect the courtesy amount region, pre-process it and segmenting the courtesy amount into individual characters before recognition. Sankari. M [12] et.al proposed techniques to categorize cheque image regions based on Cartesian coordinate space, converting cheque image to gray scale and segmenting contours leads to localization of account number and verification of details pertaining to it.

Gray scale conversion of images reduces dimension of the cheque size, segmentation and localization helps to extract account number and signature attributes. Using Hamming Distance Measures (HDM), extraction of account number and signature can be done with ease. Extensive research works, and experimentations are carried out on automating the process [11] of clearance transaction by automatically extracting the cheque fields [14] to enhance the speed, performance, and efficiency and interoperability features of cheque processing systems.

From the above discussion, it can be concluded that prior knowledge of standard structure of the cheque images is important for extraction of key attributes. Accurate segmentation of key attributes is very essential for the automation of Indian cheques. Once segmented, the data in the segmented region must be efficiently recognized to be updated to the database. This paper presents a systematic model to extract key attributes from Indian cheque images using three stages. Firstly, for any given bank, a cheque image template is created by marking the locations of all the attributes necessary for a transaction. Further, for a given input cheque image, MICR code band is extracted using horizontal projection profiling from the bottom of the image. Sliding window-based profiling [20] is performed to segment the Cheque number and MICR code from the MICR code band.

The cheque template for the drawee bank branch is fetched using the extracted MICR code. The template image is scaled to the size of the input cheque image. The important regions marked in the template image is matched with the input image using Scale-Invariant Feature Transform (SIFT) features [3] and is segmented and Optimal Character Recognition (OCR) [16] is performed. However, it is beyond the scope of this paper to perform signature verification [21]. Also, the proposed work is limited to printed cheque images. As the bank cheque images contain account number, signature and amount details, which are inherently confidential in nature and sensitive which requires maximum accuracy while extracting the attributes, the data samples available in the literature are limited. Hence, the proposed method is evaluated using 96 Indian cheque images of five different banks.

The major contributions of this paper can be summarised as follows:

1. Extraction of key attributes in MICR band- The paper improves on the work presented in [2], efficiently extracting MICR code as well as cheque number by adapting sliding window method.
2. A novel method for locating the attributes in the input cheque image and recognizing the printed text.

The rest of the paper is organized as follows. The methods presented in this paper are discussed in methodology and the corresponding experimental results are discussed further followed by conclusions.

METHODOLOGY

The method for automatic extraction of attributes from Indian cheque images using template matching is presented in the below algorithm in two stages.

Algorithm:

Phase 1: MICR code extraction
Input: An Indian cheque image having transaction details
Output: Cheque number and MICR code

1. Convert the given input image to gray scale image
2. Compute the horizontal projection profile for the image.
3. Starting from the bottom of the image, mark the second entropy to segment the MICR code band.
4. Perform morphological operations such as dilation and erosion on the MICR code band and compute rectangular bounding boxes.
5. Segment the cheque number and MICR code regions from the input image using the coordinates of the rectangular bounding boxes computed in step 4.
6. Perform OCR using the fonts for MICR code band specified in the CTS by RBI, for training.
7. Update cheque number and MICR code for the transaction.

Phase 2: Segmentation of important regions and recognition of printed text data
Input: Grayscale image obtained from step 1 in phase 1, corresponding template cheque image.
Output: Printed attributes recognized

1. Scale the template image to the size of the input cheque image.
2. Compute SIFT features for both the input cheque image and the template image
3. Locate the important attributes regions in the input cheque image by matching with the template image using SIFT features computed in the previous step.
4. Perform OCR on the printed attributes such as Account number and IFSC code.
5. Update the recognized attributes into the database.

MICR code extraction

To determine the template of the cheque image to be used for further cheque processing, MICR code from cheque image must be extracted. The MICR band in Indian bank cheques is present at the bottom of the cheque image and it is on the white background separated from the cheque watermark region [17]. Horizontal projection profile [22] is plotted for the input cheque image converted to grayscale.

Using the horizontal projection profile [22] marked in Figure.2.b, MICR region present at the bottom of the cheque image can be determined by segmenting the image from second entropy as shown in the Figure.2.c.

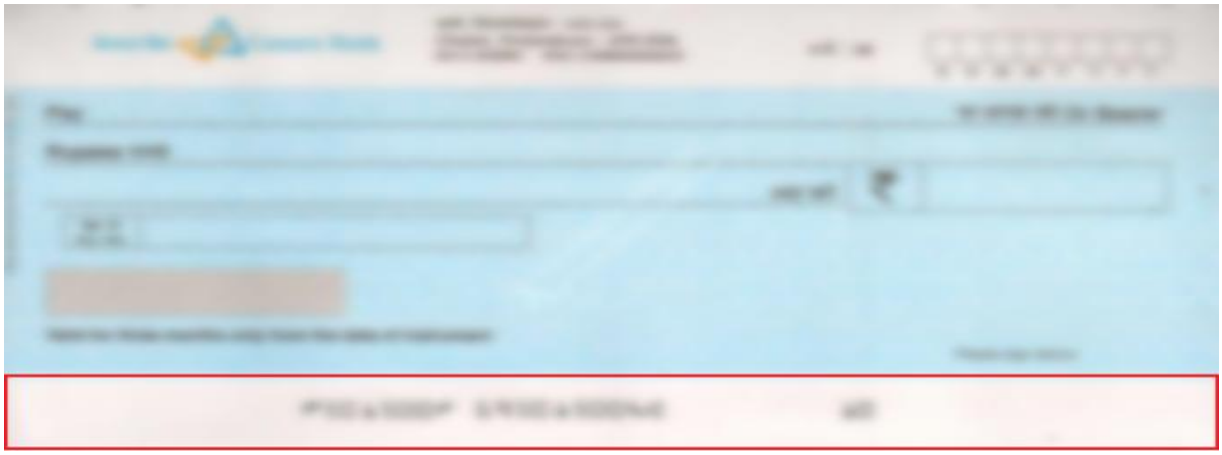

Figure 2a: *MICR Code Region of a Cheque Image.*

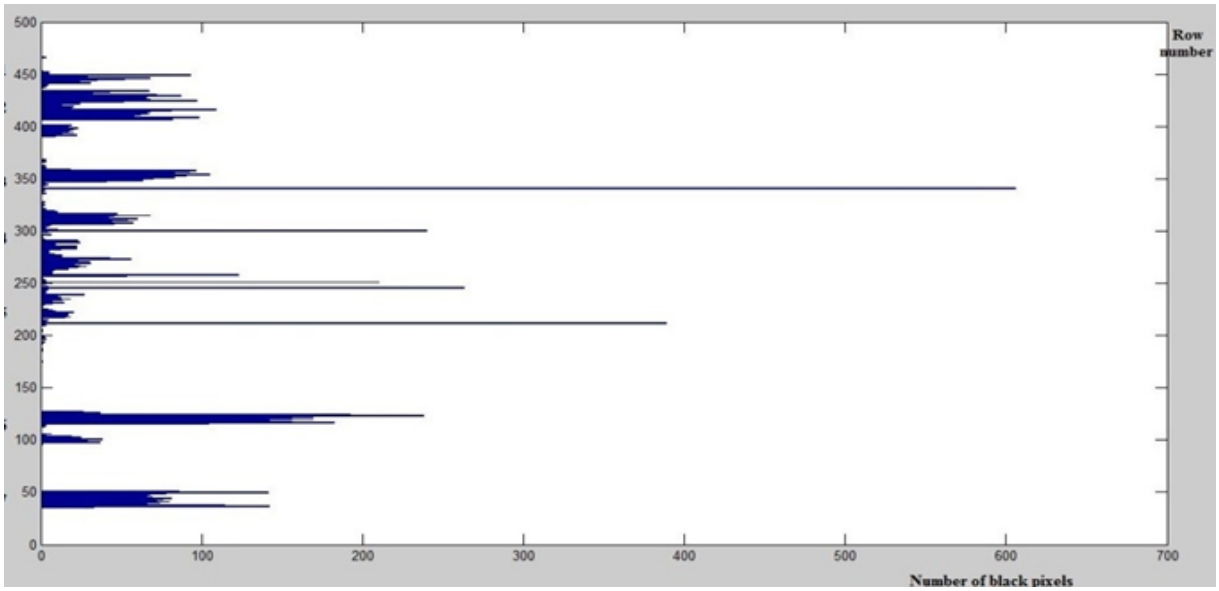

Figure 2b: *Horizontal Projection Profile of MICR Code Region.*

Figure 2c: *MICR Code Region Extracted.*

Extraction of Key attributes

With the MICR code extracted, the bank name is determined, and the corresponding cheque template is fetched from the database as shown in Figure.3. With the co-ordinates available for each attribute region from the template, data is extracted using connected components in that region from the input cheque image and the data is retrieved using OCR technique as discussed in the next section.

Figure 3: *Cheque Template Available for a Sample Bank.*

The key attributes are extracted based on their properties in the cheque images.

Data extraction for each attribute is as follows.

Date

Date is extracted by detecting rectangles in the H2 region as shown in Figure.4a. is obtained by marking the coordinates from the template by performing vertical projection profile. Value in each rectangle is fetched and is stored in the order DD-MM-YYYY.

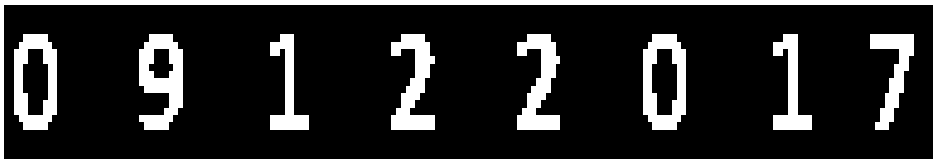

Figure 4a: *Date Extracted from Rectangle Block of Date Region.*

Payee name

Payee name is extracted by extending the H3 region as in Figure.4b from template to the entire row and first connected component obtained by applying word threshold.

Figure 4b: *Payee Name Extracted.*

Legal Amount

Legal amount is extracted by extending the region from template to the entire row and to the next row i.e., at H4 and H5 region and thus capturing two rows and subtracting the courtesy amount region from it.

Figure 4c: *Legal Amount Extracted*

Courtesy amount

Courtesy amount is extracted by capturing the rectangular block in the region marked at H4 and H5 region using the template.

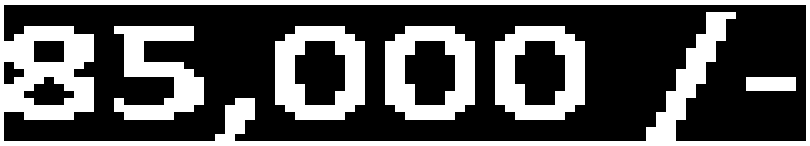

Figure 4d: *Courtesy Amount Extracted.*

Account number

Account number is fetched from the rectangle coordinates obtained from the template image. Prior knowledge of placement of account number region in the cheque is used in extraction. Seal-printed account numbers are fetched using connected components applied on H6 region.

Figure 4e: *Account Number Extracted.*

Signature

The signature is fetched by computing the connected components in the region marked from the template by applying vertical projection profile to the H6 region in Figure.4f. The second part of the H6 region provides the signature region.

Figure 4f: *Signature Extracted.*

Experimentation and Results

The proposed model is applied on 86 Indian bank cheque images. The model is accurate in extracting all the key attributes from each input image. However, the account number field that are seal printed pose challenge for segmentation and number recognition and the model performs satisfactorily for this attribute. MICR code extraction and Figure.5b.shows attributes extraction from a sample of input cheque image.

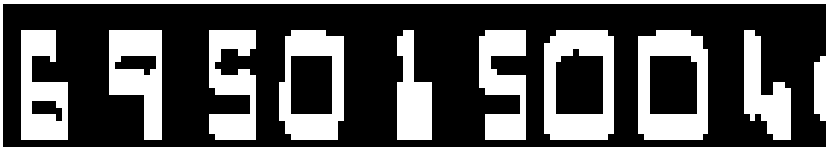

Figure 5a: *MICR Code Extracted.*

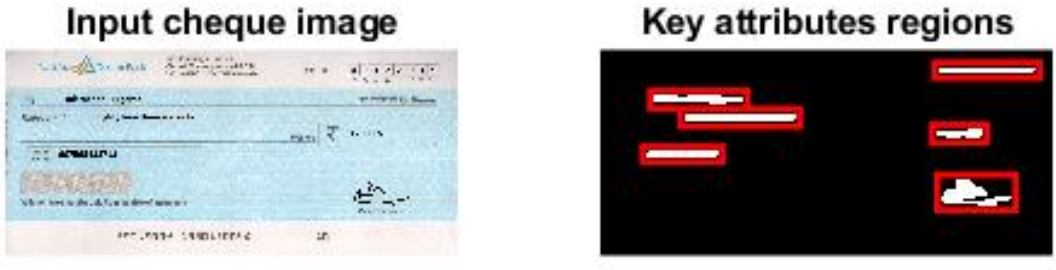

Figure 5b: *Key Attributes Extraction from Input Cheque Image by Template Matching.*

CONCLUSION

This paper proposes a conceptual model for extraction of key attributes from Indian bank cheque images.

The spatial locations of data regions in the cheque image is determined using template matching technique and fields like date, account number, payee name, courtesy amount, legal amount and signature are detected using surrounding bounding box and connected component approach.

Optical Character Recognition methods are applied to these detected regions to extract the attributes from the input cheque image.

Verification process is performed using the extracted cheque number and account number details and hence the extracted attributes are stored in the database.

Various works are available in the literature addressing the problem. However, the proposed model attained better computation time as the processing is required once during template computation and thus for every input image data is extracted using soft computing techniques

REFERENCES

1. Dansena P. Pramod Kumar K, Pal R.: 'Line based extraction of important regions from a cheque image',Eighth International Conference on Contemporary Computing (IC3). 2015, Volume: 1, pp: 183-189.
2. Karthik V: 'Detection of MICR Number and Verification of Water Mark in a Cheque'IRDBT, Hyderabad 2015, pp.5-13.
3. Ali, A., Pal, R.: 'Detection and extraction of pantograph region from bank cheque images'3rd International Conference on Signal Processing and Integrated Networks (SPIN) IEEE 2016.pp. 498-501.
4. Ke Liu, Ching Y. Suen, Nadal, C.: 'Automatic extraction of items from cheque images for payment recognition', Proceedings of 13th International Conference on Pattern Recognition. IEEE 1996, pp.798-802
5. Hussein, K.: 'Detection of courtesy amount block on bank checks'. Journal of Electronic Imaging, Vol 5, 1996, pp.214.
6. Chandra, L., Gupta, R., Kumar, P., Ganotra, D.: 'Automatic courtesy amount recognition for Indian banks cheques', IEEE Region 10 Annual International Conference, Proceedings/TENCON, 2008.
7. Palacios, R., Gupta, A.: 'A System for Processing Handwritten Bank Checks Automatically', Social Science Research Network Electronic Paper Collection,2002, pp.2-35
8. Madhvanath, S., McCauliff, S., Mohinddin, K.M.: 'Extracting patron data from check images.'Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR IEEE 1999,pp. 523-526.
9. V. K. Madasu, B. C. Lovell, "Automatic Segmentation and Recognition of Bank Cheque Fields", Digital Imaging Computing:

Techniques and Applications (DICTA'05)., 2005, pp. 223-228.
10. Dhanalakshmi, S., Kaviya, J.: 'Cheque Clearance System using various validation techniques'. IInternational Journal of Computer Science & Engineering Technology (IJCSET),2013.pp.1412-1416
11. Sankari, M., Benazir, M., Bremananth, R.: 'Verification of bank cheque images using Hamming measures.' 11th International Conference on Control Automation Robotics & Vision. ICARCV 2010,pp. 2531-2536.
12. N. Otsu, 'A threshold selection method from gray-level histograms', IEEE Trans. Sys. Man. Cyber. vol. 9, no. 1, 1979, pp. 62-66.
13. Jayadevan, R., Kolhe, S.R., Patil, P.M., Pal, U.: 'Automatic processing of handwritten bank cheque images: a survey'. International Journal on Document Analysis and Recognition IJDAR- 2011,pp.267–296
14. Liu, K., Suen, C.Y., Cheriet, M., Said, J.N., Nadal, C., Tang, Y.Y.: 'Automatic Extraction of Baselines and Data From Check Images'. International Journal of Pattern Recognition and Artificial Intelligence.1997, pp.675–697
15. R., J., Pal, U., Kimura, F.: 'Recognition of Words from Legal Amounts of Indian Bank Cheques',Proceedings - 12th International Conference on Frontiers in Handwriting Recognition, ICFHR 2010,pp. 166-171.
16. Rajender, M., Pal, R.: 'Detection of manipulated cheque images in Cheque Truncation System using mismatch in pixels',2nd International Conference on Business and Information Management, ICBIM 2014,pp.30-35.
17. Li, S., Shen, Q., Sun, J.: 'Skew detection using wavelet decomposition and projection profile analysis.', Elsevier,Pattern Recognition LettersVolume 28, Issue 5, 1 April 2007, pp. 555-562
18. Shivakumara, P., Hemantha Kumar, G.: 'A novel boundary growing approach for accurate skew estimation of binary document images',Elsevier, Pattern Recognition Letters Volume 27, Issue 7, May 2006, pp. 791-801
19. Canny, J.: 'A Computational Approach to Edge Detection', IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8 (6), 1986, pp. 679-698.
20. Djeziri, S., Nouboud, F., Plamondon, R.: 'Extraction of signatures from check background based on a filiformity criterion'. IEEE Transactions on Image Processing.7 (10), 1998. Pp.1425–1438
21. Kasturi, R., O'Gorman, L., Govindaraju, V.: 'Document image analysis: A primer'. Sadhana.Vol.27, [Online].Feb2002, pp.3–22 (2002).

Cite this article as: *Priya M.P. (2019). Automatic Extraction of Attributes from Printed Indian Cheque Images by Template Matching Technique. Journal of Image Processing and Artificial Intelligence, 5(1), 30–37. http://doi.org/10.5281/zenodo.2597063*