

AUTOMATED CHEQUE PROCESSING SYSTEM

JN SUDARSHAN

Department of Information Science,
RNS Institute Of Technology, Bengaluru, India
sudarshanjn36@gmail.com

KALPAJ PISE

Department of Information Science,
RNS Institute Of Technology, Bengaluru, India
kalpaj12@gmail.com

MANIKANTA YA

Department of Information Science,
RNS Institute Of Technology, Bengaluru, India
manikanta.ya@gmail.com

AKSHATA S BAYYAR

Department of Information Science,
RNS Institute Of Technology, Bengaluru, India
akshata.kittur11@gmail.com

ABSTRACT: Despite of swift advancements happening in digital technology, financial institutions like banks still rely upon conventional medium of processing the bank cheques by humans. The process is cumbersome and takes couple of days for actual transfer of money which involves verification by the intermediaries. This leads to high time and costs. In this paper, we propose an automated system which extracts relevant details of a bank cheque like Payee Name, Amount, Date, Bank Name using Optical Character Recognition and Deep Learning and verifies the signature on the cheque with the existing signature stored in the database using feature extraction and principal component analysis. The signature for a new user is stored using its hash value for security purposes. The proposed system uses modified convolution neural network to extract the handwritten content on cheque leaf where in IAM dataset is used to train the model and get the optimized results. This system will facilitate the process and lead to reduction in time and costs. The efficiency and performance is measured on the self generated data set of bank cheques.

Keywords— Machine Learning, Optical Character Recognition(OCR), Deep Learning, CNN, Image segmentation, Image feature extraction, Support Vector Machine(SVM), Connected component analysis(CCA)

I. INTRODUCTION

With digital innovations happening across the world in all sectors, it becomes crucial for every industry to automate their processes to attain better performance and efficiency for any model. Banking and financial industry plays a crucial role since large number of transactions occur in the form of bank cheques. Automated analysis of bank cheques is an important field to explore since the bank cheques are processed by human intervention. In case manual verification, important parameters like payee name, date, sign and amount of cheque are checked by the employees of the bank. This creates the entire process to be time consuming and more often than not prone to error. This further accumulates the total costs and the resources for executing the process. The automated system of

processing cheque includes extracting and recognition of handwritten and typed information such as payee name, amount, account number, date, cheque number, signature. Since the cheque id and account number are present on a bank cheque in a magnetic format therefore they can be extracted using optical character recognition. In most cases, payee name and date are hand written which makes it an extremely formidable task to extract the information. The signature is extracted and verified with the original signature of the account holder stored in the database of the drawer bank. If the signature is not matched, the system displays an error message. The primary goal of the system is to develop a accurate model which reads the handwritten texts like name of the payee, date and amount in the cheque with the least possible error and then processes the extracted data and performs transaction in encrypted format which maintains security. The amount is hashed using SHA 256 which stores the hashed value of function(Amount) and is eventually stored in the database. And when performing certain transactions the hashed value in the database is decrypted and is added or subtracted with the new value in transaction and the result is hashed again and stored in database. In this way if an intruder manages to break into the database, he still can't modify the contents in the database.

II. EXISTING SYSTEM

There were many models and systems developed and published by many developers and authors in the past decade. Few of the models were successful and others did not attain the necessary criteria to be qualified as successful, one of the models is listed below. This method is a new digit recognizer that can automatically retrieve the account number and courtesy amount that is written on cheque instead of being inputted manually by customers. Some research in the past introduced different solutions to read the information on the cheque, which includes the courtesy amount (numeric), legal amount (textual), signature and particularly written language such as English, French or Korea for cheque processing automation or auxiliary verification. In Malaysia, there is a research conducted to develop the Bank Cheque Recognition System by using the neural network, however, researchers were not satisfied with the performance obtained. To the best of our knowledge, none of the research and

implementation have been done on digital recognition in Malaysia, especially in improving performance in the banking domain. The solution aims to automate the cheque deposit process in a country which will be beneficial to both bank staff and customers. The digit recognizer will incorporate with the cheque deposit machine. Customer is required to insert cheques into the machine and the machine will proceed to scan the image of cheque then read the courtesy amount and bank account number based on the image that has been captured.

III. PROPOSED SYSTEM:

The proposed method takes the cheque as input, extracts amount, account number and date from the image, recognizes the digits, constructs the complete numbers and stores it in the excel file. Later this can be entered into database for further transactions. The proposed method has three modules:

1. Pre-processing and digit extraction.
2. Training the classifier and testing of the data set.
3. Post-processing of the recognized digits.

A. Pre-processing and digit extraction from cheque images:

In the first stage of the proposed method, cheque images are scanned and taken as an input to the pre-processing stage. Areas of the amount block, data block and account number are assessed and identified. Then region of interest (ROI) is defined by creating binary mask with pixel. Convolutional Neural Network Approach is used for Extracting remaining portion of the cheque leaf. This mask is applied on cheque image to extract the account number, date and amount fields. The output image is then converted into greyscale, and pre-processing is applied to smooth the image. Further, each individual segmented date, amount and account number images is further segmented into individual digits using connected component analysis (CCA). Projection profile method works well if the spacing between lines is proper. If the spacing is uneven, then segmentation using projection profile method fails to segment. Therefore, CCA which is the number of neurons that connect to the same region of the input. This argument decides the number of feature maps. Padding string and value usually used in the form of key-value pair to add padding to the input feature map. CNN layers are followed by a down-sampling operation that reduces the spatial size of the feature map and removes redundant spatial information. We have used 2×2 max pooling for down sampling. Fully connected layer is having 10 outputs as it corresponds to 10 digits. Next layer in the CNN is soft max which uses activation function to normalize the output of fully connected layer, and output of this layer is positive numbers that sum to one.

B. Training and Testing using Convolutional Neural Network (CNN):

Once the digits are extracted and segmented into individual digits, these extracted digits will be fed to

trained CNN classifier to recognize the digits. To train the CNN classifier, a systematic approach is followed. To recognize the digits extracted from the cheque images, we need to first train the classifier. CNN is used as classifier which takes the sample digit images and learns. Once the classifier is trained, then testing can be done. The image size and channels are to be specified for input layer. The size and channel considered for experimentation are $28 \times 28 \times 1$. First two parameters represent row and columns, and third represents the channel: 1 for greyscale and 3 for colour image. Filter size and number of filters are specified into the convolutional layer; the first parameter is filter size. The second parameter is the number of filters.

C. Post-processing

Post-processing deals with the process of combining the individual digits into complete number for account number, date and amount. Each field contains the collection of digits recognized by classifier that is stored in the form of string. Using simple mathematical formula which is used to reverse the number, the complete number is constructed and used to extract the character/digits properly. However, if the digits are touching each other, then CCA extracts whole connected component as one component.

IV. Literature Review:

In this paper[1] a method for signature verification, which is based on Perception and Probability. It means first the system roughly determines to which class a signature belongs to and then it finally decides whether the signature can be accepted or not. Perception presents the class, which a signature "possibly" belongs and the pattern classification based on state transition determines if at all it belongs to that class. Beside it defines an accurate closeness function. They have proposed such a system where it combine the spatial features of sum graph and HMM and classify them separately by a PNN Knowledge based classifier. In this paper[2], cheque is a payment instrument that requires high-cost processing in banks because it involves significant manual works. The usage of cheque still exists as an important non-cash payment instrument, even though Bank Negara Malaysia has imposed a new processing fee of RM0.50 per cheque since 2015. In this paper, they have proposed a digit recognizer where manual input of payee's account number and cheque amount by the customer will be ceased to simplify the manual process at the cheque deposit machine. In the proposed method[3] verifies a cheque by identifying and examining the account holder's signature. The signature extraction goes through image acquisition, gray scale image translation, binary image extraction, which is localized, segmented. The implementation involves the image extraction and extracted image is divided into characters are going to be localized. The localized data is compared with the collected database which is already collected from the given database. This method is implemented using offline mode, thereby allowing portability. This paper gives

effective sign algorithm and also provides a security by manual checking In the proposed system[4] they can detect handwritten digits from scanned input image by using neural network technique is presented. This methodology of recognition of hand writing is effective and fast compared to earlier molded image pixel comparison methodology, which is comparatively very slow. In the initial phase, hand writing samples are collected from different people and designed a form for handwritten digit input. In this paper[5], They have tackled the problem of general, unconstrained text recognition. They have presented a novel, data and computation efficient, neural network architecture that can be trained end-to-end on variable-sized images using variable-sized line level transcriptions. They have conducted an extensive set of experiments on seven public benchmark datasets covering a wide range of text recognition sub-tasks, and demonstrated state-of-the-art performance on each one of them using the same architecture and with minimal change in hyper-parameters. It discusses[6] the important results reported so far in preprocessing, extraction, recognition and verification of handwritten fields on bank cheques and highlights the positive directions of research till date. The paper has a comprehensive bibliography of many references as a support for researchers working in the field of automatic bank cheque processing In this paper[7], the design and development of the courtesy amount and date of Malaysian bank cheques was reported. The system has successfully implemented the detection and extraction module of the system but the recognition results were not very satisfactory. Possible causes of failure have been discussed to point out improvements that can be made and pitfalls that should be avoided in future work. In this paper[8] they have presented some novel ideas underlying a cheque reading system developed in our institution. They focused on the reading of legal amounts and post-processing of the recognition results. For the legal amount recognition, an approach based on Hidden Markov Models was proposed. HMM (Hidden Markov Model) turned out to be very useful because no segmentation of the legal amount is required, neither into characters nor into literal words.

V. METHODOLOGY

The image acquisition of a bank cheque is crucial for the CTS (Cheque Truncating System). Generally, flatbed scanners are used to acquire such images. Due to orientation and irregularities presented in the scanned image(s), As we are unable to use the acquired image(s) directly for the image processing operations therefore it requires some pre-processing step. Image pre-processing Image [preprocessing is a technique used to scan cheque images. As a scanned image obtained from the scanner cannot be directly used thus it is in need of pre-processing, which involves two primary operations, i.e, rotation and removal of unnecessary background information. In first step, scanned image is rotated with respect to the 'Date Box' (is a common feature presented at the same part of every bank cheque) and then removed the background noise and extra information. Efficiency of

the parameter identification considerably improves with removal of extra background information. Rotation As the scanned images may vary in terms of orientation therefore we used the date box present in all standard cheque leaflets and the relative invariant nature of the position was utilized. In order to perform the rotation of image, It is determined that the point of rotation and degree of rotation. The primary component essential for rotation to work was contour extraction, as we were able to determine the position values of date box by using it which used as our anchor for any set of operation related to length mapping. Also, in order to perform the rotation, we used the midpoint of the image as the rotation point, and we used date box in order to determine the angle necessary for the rotation. Removing background noise There was lot of extra information in an original image which is to be removed. For the task of removing of background information, the date box is used to present in the standard cheque template.

A. Segmentation of Cheque

To use only the desired part of the image for various operations, the good idea is to do image segmentation. For this, image is separated in each pattern locally in the optimum way. Thus, we have performed segmentation to ensure that the tool would use only the required information of image for its processing, while accessing the entire process. As illustrated in figure 1, we have separated each key parameter of the cheque leaflet using segmentation for correct identification and verification. Other than contour extraction in order to determine the position values for the date box as explained in previous section(s), The identified region of interest (ROI) on cheque leaf standard dimensions is done. In order to remove the extra information from the desired region, we have created a standard template using pixel values for bank cheques. After the segmentation of bank cheque image, we have separated each segment of key parameters for OCR to determine the patterns for verification.

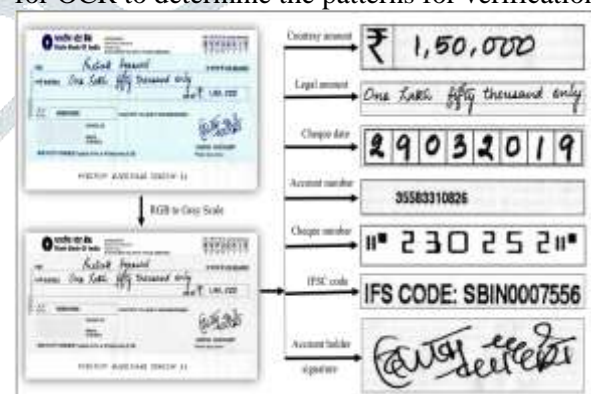


Figure.1 Segmentation of cheque

B. Handwritten Text extraction from Cheque

Handwritten extraction of texts has been a challenging problem in Deep Learning and Natural Language Processing. The handwritten texts can be recognized by making use of various Machine Learning algorithms including the Convolution Neural Networks, Support Vector Machines where we have extract features then work upon them. we have used Neural Networks for this

application. The model extracted the sentences by making use of OpenCv to crop out handwritten texts from certain parts of the cheque and then we have split the sentences into words based on space character between them. Each of the words are then passed into the model which comprises of 5 layers of Neural Network followed by 3 layers of Recurrent Neural Network and eventually retrieving the digital text by making use of Connectionist Temporal Classification.

C. Signature feature extraction and verification

In design of system shown in figure 2, features from signature are extracted and then compared with features which are stored to verify whether the signature belongs to the concerned person. The system is used to avoid the counterfeit incidents taking place in the banking institutions. Features are extracted using the PCA (Principal Component Analysis) method. The extracted features are then matched with those stored in the database. If the signature is verified then the further execution proceeds else execution stops.

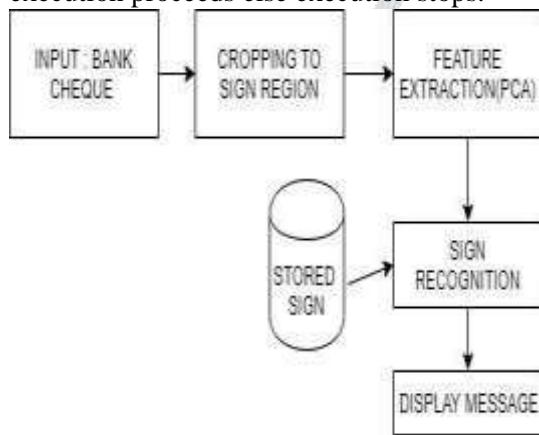


Figure. 2 signature feature extraction

D. CNN model for amount identification

After extracting the image segments, CNN model is used for handwritten numeric digits recognition of courtesy amount as well as to convert the legal amount into string. have used Deep Learning Toolbox (a MATLAB toolbox) for CNN implementation on two convolution layers with six and twelve filters as shown in figure 3. Whereas, in max pooling process, we have used a window of 2x2 with one thousand epochs and one stride. After that, it receives numeric output from the courtesy amount of bank cheques and converted it into string to compare with the legal amount string.

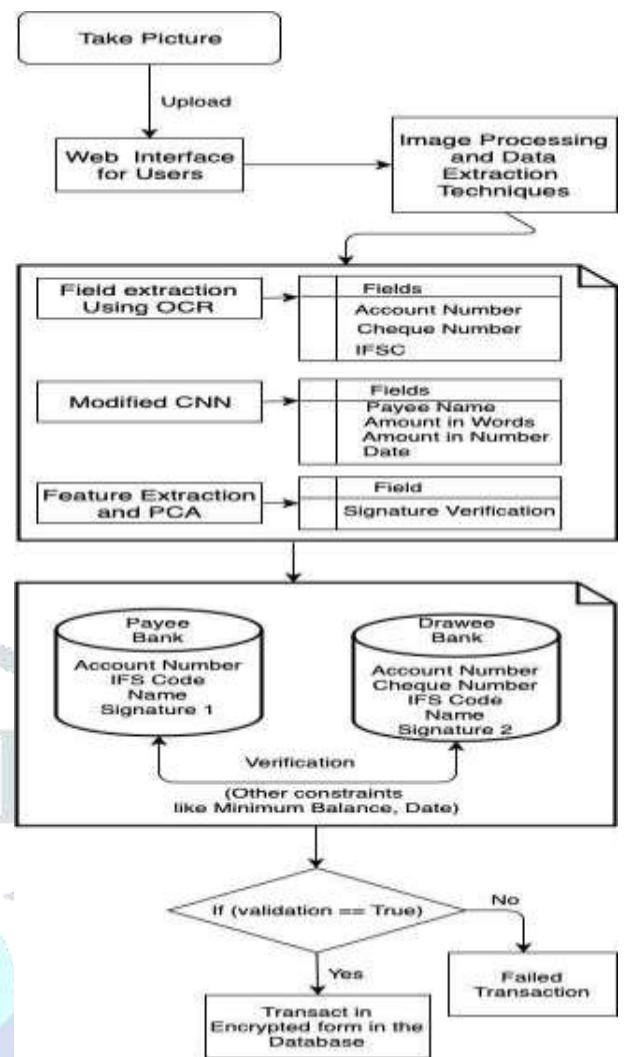


Figure 3 CNN architecture for handwritten digit recognition of courtesy amount

VI. SYSTEM DESIGN

As illustrated in figure 4 ,the cheque is fed to the system which performs OCR at first stage on whole data as a pre-processing step. Currently we have made templates for each bank like HDFC, SBI, PNB, OBC, BOI so that data can be extracted using Opencv to crop out the required parts effectively and efficiently. After getting the results we perform string matching and manipulation methods on the result extracted to get the specific bank template. System design is the process of the defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development. Object-oriented analysis and methods are becoming the most widely used methods for computer systems design. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user. The UML has become the standard language in object oriented analysis and

the CNN used for digit recognition. Similarly, for character recognition using CNN, we were able to achieve the accuracy up to 99.94%. For signature recognition, SIFT and SVM classifier models are used to determine the relevancy for verification of signature and achieved 98.1% accuracy.

IX. REFERENCES

- [1] Sook Chin Chiew , Xin Yuan Law , Ren Zhang Tan , XinYing Chew, Khai Wah Khaw “Digital Recognition by Deep Learning Techniques: A Proposed Digit Recognizer to Automate Cheque Deposition”, In Amity Journal of Computational Sciences (AJCS) 2019
- [2] Ganga Holi and Divya K. Jain, “Convolutional Neural Network Approach for Extraction and Recognition of Digits from Bank Cheque Images”, In International Conference on Intelligent Sustainable Systems, 2019
- [3] Mukesh Jha, Madhur Kabra, Sahil Jobanputra, and Prof. Rupali Sawant, “Automation of Cheque Transaction using Deep Learning and Optical Character Recognition”, In Second International Conference on Smart Systems and Inventive Technology (ICSSIT 2019)
- [4] Saleem Ulla Shariff, Maheboob Hussain, Mohammed Farhaan Shariff, “Automated bank cheque verification using image processing and deep learning methods”, Springer Science+Business Media, LLC, part of Springer Nature 2020
- [5] Victor Carbune, Pedro Gonnet, Thomas Deselaers, Henry A. Rowley, Alexander Daryin, Marcos Calvo, Li-Lun Wang, Daniel Keysers, Sandro Feuz, Philippe Gervais, “Fast multi-language LSTM-based online handwriting recognition”, International Journal on Document Analysis and Recognition (IJ DAR) (2020) 23:89–102
- [6] Girish C. J, Mrs. Geetha G. P “Design of Bank Cheque Validation System”, International Journal of Engineering Research & Technology (IJERT) 05, May-2015
- [7] Mohit Mehta, Member, IACSIT, Rupesh Sanchati and Ajay Marchya, “Automatic Cheque Processing System”, International Journal of Computer and Electrical Engineering, 2018
- [8] Sebastian Salazar-Colores, Eduardo Cabal-Yepez, “A Fast Image Dehazing Algorithm Using Morphological Reconstruction” IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 28, NO. 5, MAY 2019
- [9] Beresneva A, Epishkina A, Shingalova D (2018) Handwritten signature attributes for its verification. In: IEEE conference of Russian young researchers in electrical and electronic engineering (EIconRus), pp 1477–1480

