Recognition on User-Entered Data from Myanmar Bank Cheque

Nang Aye Aye Htwe, San San Mon, Myint Myint Sein
Mandalay Technological University, University of computer Studies,Yangon
htwe.aye@gmail.com, mlm@wwlmail.com

Abstract

In this paper, we propose a model to recognize the payee's name and legal amount, courtesy amount on bank cheque. These three items are written in Myanmar format. An innovative approach for extracting from bank cheque images and other documents is proposed based on the integration of the crop method with Hidden Markov Model(HMM)The capable of extracting the user filled information and character segmenting have been done automatically from the scanned image of Myanmar bank cheque image. Depend on nature of Myanmar handwriting style, feature extraction and classification is done. Recognition on extracted parameter writing in Myanmar character from check image demonstrates by the effectiveness of Hidden Markov Model (HMM). The system not only to recognize handwritten words but also to display the output in machine printed fonts. It will be providing to easy and convenient to staffs in bank cheque system. It saves time and money on the bank cheque system because it is automated and saved labor consuming.

Key words: Offline Handwriting, Preprocessing, Normalization, Feature extraction, Classification, Hidden Markov Model (HMM)
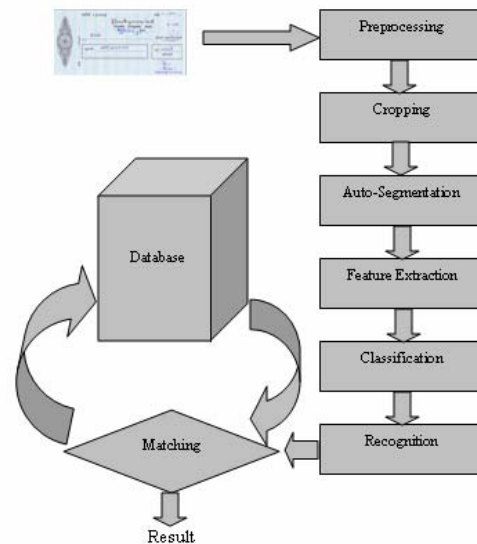
1. Introduction

Millions of handwritten or machine printed bank checks have to be processed everyday. Since the bank check processing is merely a repetition task it is desirable to realize it in an automatic fashion.

Bank cheques and financial documents in paper format are still in enormous demand in spite of the overall rapid emergence of e-commerce and online banking.

The extraction and recognition of handwritten information from a bank cheque propose a formidable task which involves several subtasks such as extraction and recognition of signatures, courtesy amount, legal amount, payee and date.

Currently, thresholding and image subtraction techniques are being used for extracting the user-entered data. The techniques based on image subtraction have shown more robustness to segment the user-entered data.

Figure. 1. Overall system of bank cheque recognition

Okada -Shridhar have suggested an approach based on a subtraction operation between one filled-in bank check image and the same bank check image without the filled -in information. This approach might not be feasible for real-life

applications

Recognition on Myanmar handwritten date on bank checks using Zoning methods is presented by H.H. Thaung. Cheriet et al. have been proposed to eliminate complex background of checks. Baselines extraction is used to localize the information of the principal items such as the amount and the payee. Djeziri et al. also developed a technique to eliminate the background picture of checks. This technique can extract the line even from the low contrast images. Two fusion strategies are employed on the feature extracted using filiformity and Gabor filter techniques. In order to produce a successful cheque processing system, many subproblems have to be solved such as background and noise removal, recognition of the immense styles of handwriting and signatures, touching and overlapping data in various field of information and errors in the recognition techniques. Madasu et al. implemented a system to read courtesy amount, legal amount and date fields on cheques. Oliveira el al. defined verification as the post processing of the results produced by recognizers. Method to recognize handwritten words are well known and widely used for many different languages. Hidden Markov Models (HMM) have been very successfully implemented for recognizing cursive written words. Several type of decision method including statistical methods, neural networks, structural matching (on trees, chains, etc) and stochastic processing (Markov chains, etc) will use along with different types of features.
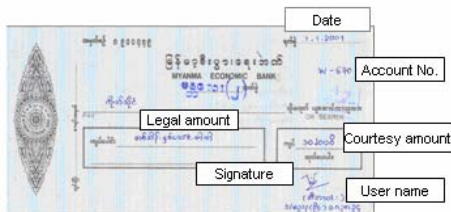
2. Nature of the bank cheque

Figure. 2. Nature of Myanmar Bank cheque

A cheque document normally comprises machine printed character and written texts as well as graphic images.

The nature of bank cheques is varied and complex and this makes the problem of automatic bank cheque processing very difficult. The only way to extract signatures from bank cheques and other forms is to have some sort of prior information about the layout of the document. In this paper, we have proposed a similar technique to approximate the area of segmentation before extracting the user-filled in formations from the region of interest.

Handwritten text on the bank cheque is composed of not only Myanmar format but also English format. Bank cheque image contained the following six items.

- Account Number
- Legal amount
- Courtesy amount
- Signature
- Date
- Payee name

In this paper, we used to recognize courtesy amount, legal amount and user name. These three items is used for recognition and interpretation into machine printed character.

3. Word Segmentation

The bank cheque is scanned by using 300 DPI (Dots per Inch) and 256 gray levels scanner. Firstly, preprocessing for bank cheque image is performed. It removes the noise form the bank cheque image. The image should be noise free image.

Scanned documents often obtained noises that arise from the printing process, scanned process and age of print quality, etc. The crop method is applied to extract required user-filled information from the bank cheque image. If necessary, our system need to remove some noise and unwanted background effect. Skew detection is also performed on the preprocessing step.

Figure. 3. Cropping the area of the interested

Cropping require user-filled items from the bank cheque is performed. Thinning on the sub image to get one-pixel size for the user filled information of extracted images.Then, the system will convert RGB to BW (black and white) form.

Normalization is used to convert random spacing to the equal spacing. The original input is a binary image. It contains the character of different sizes and spaces. The purpose of normalization is to improve recognition rates in this system.
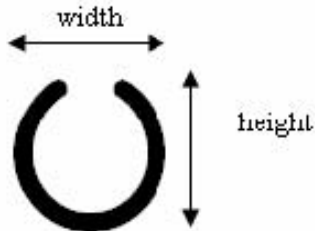
In the segmentation step, we have to perform row segmentation and then column segmentation. Row segmentation is intended to extract the characters line by line. Column segmentation is used to segment each item into sub-image. Segmentation is used to get easy and effective way in the remaining steps. By using segmentation, the space between line can be easily found ant the text can be separated in the easy way.

Upper and lower boundary values of the minimum bounding boxes are sending to the next step for feature extraction and classification.
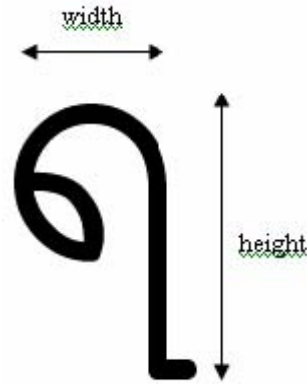
4. Feature Extraction

In the feature extraction step using MWR algorithm, the height and width of the sub image is the most important things. Feature points are extracted from the whole word of the sub image.

Firstly, the feature points are extracted from the sub images. For Myanmar character "pasouth" "ပ", we make rectangular box around this character. Feature points are start extracted from the left top corner of the sub image.
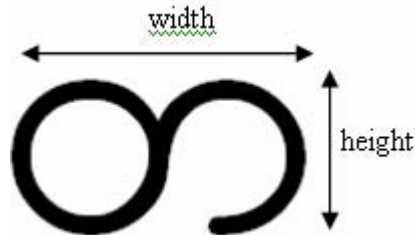
Figure. 4. Character with the same size of width and height

Figure. 5. Character with height is longer than width

Depend on the nature of the result from one of these three tests; we classified them into many groups. We find white pixels on the top of the character and black pixel for other three directions.

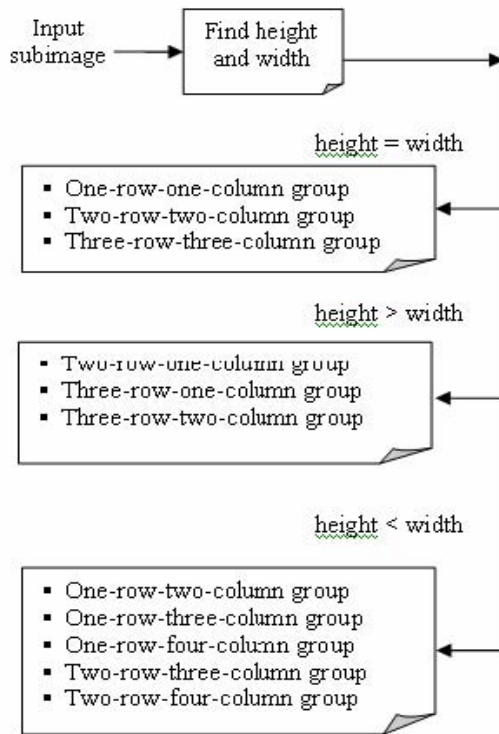**Figure. 6. Character with width is longer than height**

The system need the standard sub image for one-row-one –column.

Width= Height=n

Where n= standard length from normalization
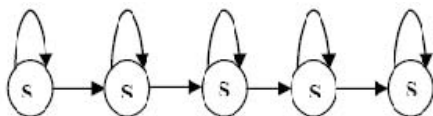
5. Classification and Recognition

After the feature extraction on the sub images have done, we classified them into many group depend on the nature of its writing style. By using many classified group on this system, it improves the recognition rate and time on the bank cheque processing system.

Figure. 7. Classification of groups

There are several established methods of estimating a sequence of probabilities form a sequence of data which have been applied in the fields of both speed and handwriting recognition. Two methods are used for hidden Markov Model: Continuous and Discrete Hidden Markov Models (HMMs). Hidden Markov Models (HMMs) also have been very successful in the field of Optical Character Recognition (OCR) due to their ability to model the passage of time.

The Hidden Markov Models is a doubly stochastic variant of Markov Model, with the underlying stochastic process that is not observable (hidden), but cannot be observed through another set of stochastic process that produces the sequence of observation sequence.

Figure. 8. Simple left right hidden Markov Model

It contains state, state transitions, and transition probabilities. Initial state probabilities exist for each state which defines the chances of the model being found in the particular state at the beginning of an observation sequence.

Each state in the model has a number of parameters associated with it which describe the probability of making a transition from that state to another state in the model, the probability of particular observation being produced while in that state, as well as the probability that a particular state was the start state for the sequence under observation.

The following parameters are used to define in Hidden Markov Model. And the training model is shown in Figure 6.
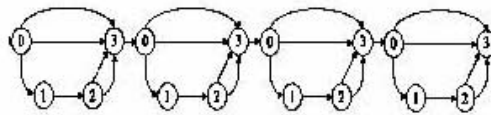
• T: Length of the observation sequence

$$o = (o_1, o_2, ..., o_3)$$

• N: Number of states in the model;
• M: Number of possible observation symbols;
• S : Set of possible states of the model;
• q t : State of the process at time t

$$v = \{v_1, v_2, .., v_m\}$$

: Discrete set of possible observation symbols;

Figure. 9. Training Model

• A $= \{a_{ij}\}$: State transition probability distribution, in which a_{ij} denotes the Probability of going from state s_i at time t to state s_j at time t + 1;

• B $= \{b_j(k)\}$: Observation symbol probability distribution, in which bj(k) denotes the output symbol probability in state sj of producing a real observation

• Symbol $o_t = v_k$;

• $\prod = \{\pi_j\}$: The initial state distribution, in which π_i denotes the probability of being in state i at time t = 1.
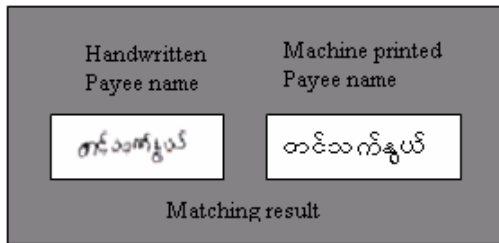
Given a model, to be represented by the

compact notation $\lambda = \{A, B, \pi\}$ three basic problems of interest must be solved for the model.

Maximum likelihood parameter estimation for HMM is obtained by the iterative procedure, with multiple observation sequences. Two-Hidden Markov Model is used for recognition. The first one is to recognize the handwriting word and the second one is for handwritten lexicon recognition. The two log probabilities for recognition are using two hidden markov model. The log probabilities are together added to get the correct final result.

6. Experimental Results

We create a large database fro recognition which contains the lexicon format and many of the sub word in each group. By using these information, we have conducted several experiments to access the performance under known variation of lighting, scale and orientation.

Figure. 10. Transforming of Handwriting Payee name

7. Conclusion

Handwriting recognition and printed character translation techniques are presented in this paper. The recognition process is based on the Hidden Markov Model. The output of the system is machine printed character in

Myanmar Language. This approach can be improved into many languages. In next step, we aim to recognize the Myanmar handwritten characters and transform to the printed characters in multiple languages and interpret by voice. For authentication, signature verification is used for bank cheque processing system.

References

[1] N.A.A.H, "Automatic Extraction of Payee name and Legal Amount from Bank cheque", *In Proceeding of Fifth International Conference on Computer Application*, February 2007.

[2] H.H.Thaung, "Recognition on Handwritten Date on Bank Checks using Zoning Methods," *PhD Dissertation University of Computer Studies*, Yangon, 2004.

[3] K.L.Thaw, "Recognition for Myanmar Handwriting digits and Characters", I*n Proceeding of Fifth International Conference on Computer Application,* pp.415-419 February 2007.

[4] K.Sandar, "Off-line Myanmar Handwriting recognition using Hidden Markov Models". *In Proceedings of Third International Conference on Computer Application*, pp.463-469 March 2005.

[5] M.M. Nge, " Automatic Segmentation of Handwriting Myanmar Digit and Script on the Boucher Sheet", *In Proceedings of Fourth International Conference on Computer Application* , pp.499-505 March 2005.

[6] S.H.Mar, "Myanmar Character Identification of Handwriting between Exhibit and Specimen", *In Proceedings of Third International Conference on Computer Application,* pp.471-475 March 2005