Locating Destination Address Block On Thai Envelopes

Worapote Chanpongsaeng1,Pinit Kumhom and Kosin Chamnongthai[2]
[1,2,3] Department of Electronic and Telecommunication Engineering, Faculty of Engineering,
King Mongkut's University of Technology Thonburi,
91 Pracha-Uthit Road, Bangkok 10140, Thailand
Tel. +66-2-470-9088, Fax.: +66-2-470-9089
e-mail : worapote_ch@yahoo.com ,ipinmhom @kmutt.ac.th , kosin.cha@kmutt.ac.th

Abstract: About 90% of Thai-style addresses have similar features; e.g. the beginning of each address line is diagonal. In this paper, we propose a method for locating destination address block (DAB) on Thai envelopes based on features of Thai-style addresses. Firstly, we decompose image into smaller blocks and remove all blocks not meeting criteria. Secondly, we search for the DAB candidates. Lastly, heuristic rules and typical features are applied to identify the destination address block. Experimental results using 2,700 envelopes of handwritten and machine printed Thai envelopes show a successful address extraction rate of 91%.

1. Introduction

The Communications Authority of Thailand (CAT) have used automatic mail-sorting system since 1996. This system recognises only zip code so it can sort only standard mails and postcards. If the mails are not in the standard form, the system rejects them, and they will be recognised by human. To decrease the rising labour cost and increase throughput rate, a system can be developed by using both information of zip code and address on envelope to reduce rejection rates. Figure 1 shows the basic block diagram of an automatic mail-sorting system. The output of address block location is an destination address block (DAB) image, which in turn is input to OCR and address interpretation .

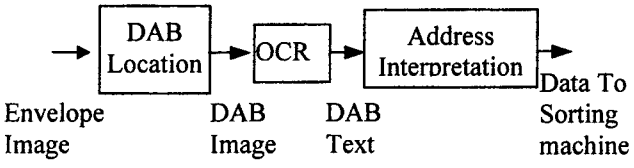

Figure 1. Automatic mail-sorting system.

Destination address block location can be divided into two categories: geometric feature and texture segmentations.

In the first group, selection of the DAB from several candidates is proposed by Downton and Leedham [1]. On the other hand, in [2 – 4], the DAB is located without constructing candidates.

In the second group, the detection is based on a texture segmentation using Gabor filters [5]. This method shows high accuracy rate of detection DAB, but is too slow for usage. Although these previous methods work well for English envelopes, they are not suitable for handwritten Thai envelope because Thai-style addresses possess different characteristics.

From our observations, Thai-style addresses have distinct characteristics as follows:

1) the connecting line between the starting points of the consecutive text lines of the address is incline about 0-110 degree angle,

2) often, the distance between words in address is very far.

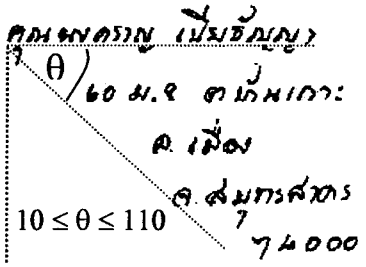

Figure 2. Thai address on envelopes

Based on these two main features, we propose a method for locating the destination address block. The remainder of the paper is organized as follows. In the next section, we explain the basic concept of our proposed method. In Section 3, we introduce the structure of our method, and give the details in Section 3.1-3.3. The experimental results and the discussions are given in Section 4 and 5, respectively. The paper finishes in Section 6 with a conclusion.

2. Basic concept

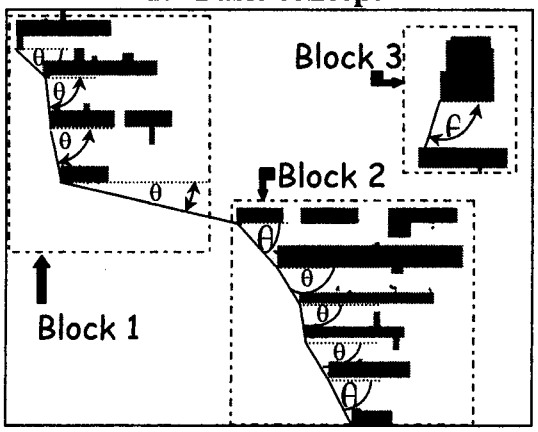

Figure 3. Basic concept

From the two main features in Section 1, we measure the distance and angle between two consecutive address lines as shown in Fig. 3. The DAB candidates are found based on these two parameters; i.e. the angles between two consecutive address lines are between 10-110 degree and the distances are not too far. For example, in Fig. 3, Block 1 and Block 2 are the DAB candidates, but Block 3 is not because its angle does not fall into the 0-110 degree range. Notice that Block 1 and Block 2 are not the same block because the distance between the last line of Block 1 and the first line of Block 2 is too far. Finally, we select the DAB based on the location and area of the candidates; i.e. the candidate at the middle of the envelop is the DAB.

3. Proposed method

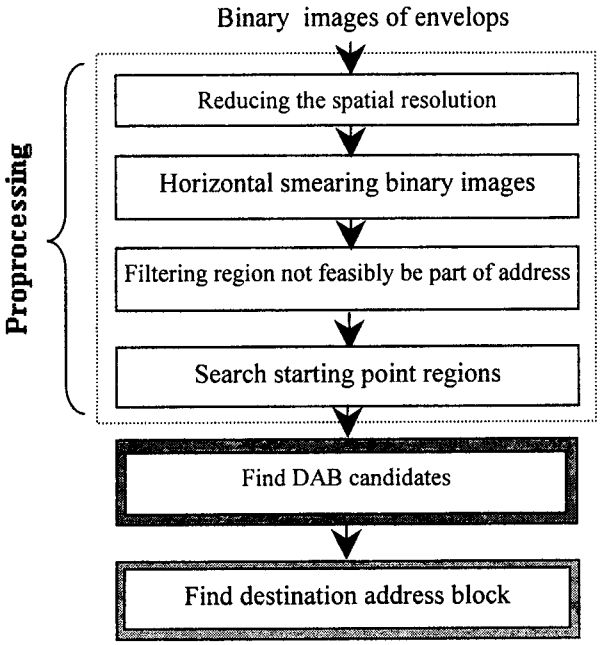

Figure 4. Flowchart of our proposed system

An overview of the proposed method which uses geometric feature for detection DAB show in figure 4. First step to fourth step are preprocessing step. The bottom-up method is applied. First step, in order to get a faster processing time, the spatial resolution of the binary image is reduced. Then, the resulting image is decomposed into regions by horizontal smearing technique. Smearing binary image has many regions and some regions are not feasible to be an address. So these regions are eliminated. Next process is to find the starting points of regions possibly be the DAB candidates by locating regions without other regions on their left and upper sides. Then, the DAB candidates are found following the rules based on the angles and the distances between two consecutive addresses lines as mentioned in the previous section. Finally, the DAB is located based on the location of the candidates. The following subsections explain each step in details.

3.1 Preprocessing

The input to this process is a binary image of envelope with 50 % resolution of the original image so that computation time is reduced. We connect components with windows of size 6x1 pixels. The window is horizontally and vertically moved, without overlapping, across the reduced envelope images. For each position, If number of black pixel above the threshold, the pixel inside the window are set to zero.Next, the filtering regions,The aim of process is to eliminate regions in the smearing image that are not feasible to be part of the address.These regions will be eliminated based on the following rules: i.e.regions with heights less than 2 pixels. Last process is finding starting point region, upper-left corner of address, from remained regions.

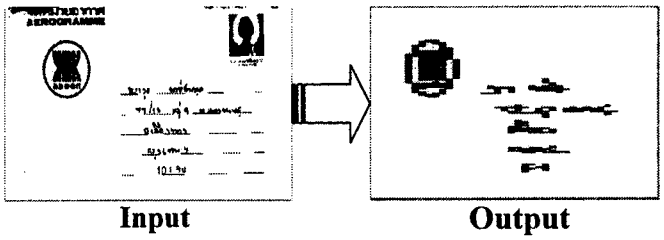

Figure 5. Input and Output of preprocessing

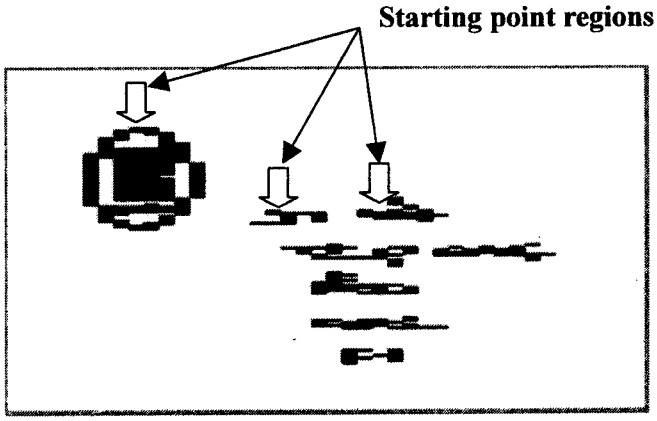

Figure 6. Staring point regions

3.2 Finding DAB candidates

Each starting point region is checked whether it is a part of address using rules based on the distinct features of Thai address styles. The algorithm for checking the conditions can be illustrated in eight steps as follows:

Step 1: Select one of starting point regions. Evaluate maximum and minimum vertical positions of starting point region, denoted by x^s_{max} and x^s_{min} and those of horizontal positions, denoted by y^s_{max} and y^s_{min}. Position of x^s_{max} , x^s_{min} , y^s_{max} and y^s_{min} show in figure 7.

Step 2: Select regions that have $y^s_{min} < y_{max}$ or $x^s_{min} > x_{min}$

Step 3: Select a region from step 2 with starting point region from step 1 and find θ and D defined as follows.

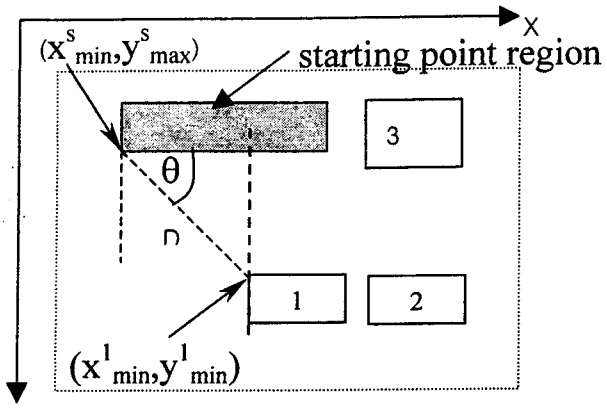

Figure 7. The angle θ and distance D between starting point region from step 1 and region from step 2

$$D = \sqrt{(x^s_{min} - x_{min})^2 + (y^s_{max} - y_{min})^2}$$

$$\theta = \cos^{-1}\left(\frac{(x^s_{min} - x_{min})}{D}\right)$$

Step 4: Check θand D from step 3 for correct condition as follows:

$$0 \leq \theta \leq \frac{\pi}{1.8}$$

and

$$D \leq Ts$$

where Ts is predefined threshold

If θand D do not agree with the conditions, the region from step 2 is not part of address block with the starting point region; otherwise the region is a part of address block with starting point region, and y_{max} , y_{min} , x_{max}, and x_{min} are kept.

Step 5: Repeat step 2-4 until no region with correct condition of step 2 and count number of regions considered to be part of address block with the starting point region from step 1. If the number of regions is more than 3, we keep them as a DAB candidate.

Step 6: Repeat 1-5 until no starting point region.

Step 7: The candidate with the location at the middle of an envelope is selected to be the part of the DAB.

Step 8: If the regions selected in step 7 have a region on right-hand with distance less than a threshold, merge them to region form step 7.

3.3 Finding DAB

We select DAB form DAB candidates based on position and area of DAB candidate. By using 200 envelopes images for finding appropriate weight . If more

$$P_{DAB} = W_{area} * A + W_{loc} * loc$$

Where A is area of DAB candidate per area of envelopes and loc is distance between starting point region and reference line

4. Experiment and Results

The proposed system has been evaluated on 2,700 Thai envelopes consisting of 1,800 machine-printed and 900 handwritten in a simple background from The Communications Authority of Thailand (CAT). The results are compared with Formbase-mehtod [4]. The system has been implemented on AMD 400 MHz PC using Microsoft Windows system and MATLAB version 5.3.1. The results are shown in Table 1.

Table 1. DAB location

Type	Machine-printed Thai envelope	Handwritten Thai envelope
Proposed methods	94%	84%
Form-Based Method (4)	93%	70%

Table 2. Average processing time per Envelope image

Type	Proposed methods	Form-Based Method (4)
Average Time(s)	6.3	6.2

5. Discussion

The experimental results of the proposed method show that our method can locate destination address block on machine-printed and handwritten Thai envelope. In machine-printed the detection rate is 94%. The incorrect detection is because the following reasons. First,form preprocessing some addresses on Thai envelope have very short space between text lines and they are incline such as the one shown in Figure 8(a). As a result, when they are passed to the smearing process, every text line becomes the same region as shown in the right picture of Figure 8(a). Secondly, there exists some edge resulting from a transparent plastic wrap or plastic windows such as the image in Figure 8(b). As a result, it increases x_{min} of a text line so that the θ (Step 4, section 3.2) does not correct. For the case of the handwritten Thai envelope, the detection rate is 84%. One cause of the incorrect detection seems to be some parts in address is too far from the main address parts as shown in Figure 8 (c).

6. Conclusions

This paper proposed algorithm for locating destination address block for Thai style of addresses. The experimental results show that the method can locate DAB with detection

rate of 94 % for 1,800 machine-printed and 84 % for 900 handwritten Thai envelope.

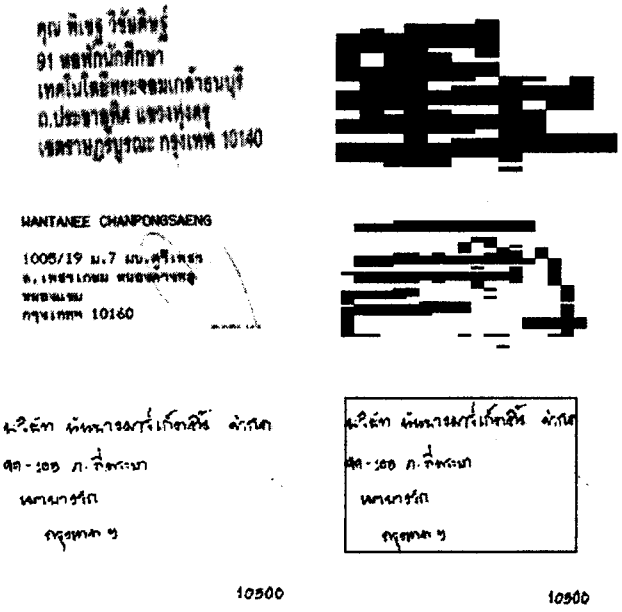

Figure 8. (a) address with text lines being very close and incline (b) address with edge of transparent plastic wrap or plastic windows (c) address with some parts far (left).

References

[1] A.C. Downton and C.G. Leedham " Pre-Processing of Envelope Images For Optical Character Recognition ", *Pattern Recognition. IEEE*, May 1988,vol. 23, no. 3/4, : 27-28.

[2] Bin Yu, Anil K. Jain and Moidin Mohiuddin, " Address Block Location on Complex background ", *IEEE Document Analysis and Recognition*, 1997 : 897-901

[3] M. Wolf, H. Niemann, W.Schmidt, "*Fast address block location on handwritten and machine printed mail-piece images* ", *IEEE Document Analysis and Recognition*, 1997,Vol. 2 : 753-757, .

[4] M. Wolf, H. Niemann, " Form – based Locatlization of Destination address block on Complex Envelops ", *IEEE Document Analysis and Recognition* ,1997, vol. 2 :908-913,

[5] Anil K. Jain, Sushil K. Bhattachrjee, "Address block location on envelopes using Gabor filters: supervisedmethod", *Pattern Recognition, 1992. IAPR '92.* , 1992. ,Vol 2 : 264-267

[6] Junliang Xue, Xiaoqing Ding, Changsong Liu, Shiyan Pan, Hongwei Kong , "Destination address block location on handwritten Chinese envelope", *Document Analysis and Recognition, 1999. ICDAR '99.* :737-740.

[7] D. Beymer, "Locating destination address block on handwritten Korean envelopes", *IEEE* Pattern Recognition,1994,Vol. 2 : 619-621 .