A Prototype System of Courtesy Amount Recognition for Chinese Bank Checks

Dong Liu
Graduate School at Shenzhen
Tsinghua University
Shenzhen, China
liu-d05@mails.tsinghua.edu.cn

Youbin Chen
Graduate School at Shenzhen
Tsinghua University
Shenzhen, China
chenyb@sz.tsinghua.edu.cn

***Abstract*—In this paper, we present a prototype system of courtesy amount recognition (CAR) for Chinese bank checks. The system deals with color check images and consists of five modules: skew correction, binarization, string extraction, segmentation and recognition. The whole system is designed under three principles: information fusion, complementary method combination and multi-hypotheses generation then evaluation. Information from color, gray and binary images is fused and complementary algorithms are applied for binarization and string extraction. Multi-hypotheses are made by keeping all possible candidates when ambiguous solutions exist in extraction and segmentation. Then the most suitable one is selected as the final result by evaluating the probabilities from recognition. Experiments show that our system is very promising based on a large number of real checks collected from different banks.**

Keywords-automatic bank check processing; courtesy amount recognition; document analysis; optical character recognition(OCR)

I. Introduction

Bank checks are widely used all over the world. Despite the rapid growth of credit card and many other electronic means of payment, paper checks are still one of the most popular forms for non-cash payment. Since most of the checks need to be manually processed by human operators, systems for processing bank checks automatically are in great demand. And courtesy amount recognition (CAR) is one of the most important parts in automatic bank check processing.

Generally speaking, a CAR system consists of four modules: binarization, numeral string extraction, segmentation and recognition. Many algorithms have been proposed and applied for constructing these modules in the past years. For example, global thresholding (such as Otsu [1]) and local thresholding (such as Niblack method [2]) are often used for binarization. Stroke models [3, 4] are used for extracting characters in complex background; mathematical morphology [5] or run-length models [6, 7] are usually applied for extraction when guidelines or grids exist. As for segmentation, dissection methods (such as contour difference method [8], drop falling method [9]), recognition based methods, and holistic methods are three kinds of methods mainly used. For recognition, many powerful optical character recognition (OCR) engines have been reported and details from normalization, feature extraction to classifier design in the OCR engines can be found in [10, 11].

Many commercial check recognition systems(such as the A2iA Check Reader [12]) have been reported[13] for processing checks in France, America, Canada and many other countries. Courtesy amount can be recognized automatically with high reliability in these systems. However, the courtesy amount area in Chinese bank checks differs sharply from the other countries' in both layout and image quality. Fig.1 gives a typical image of Chinese bank check, and the courtesy amount area is indicated by white rectangle. As shown in this figure, the courtesy amount is written in pre-printed grids and sometimes covered by red seals. Seals bring a lot of challenges in binarization while the overlap between character strokes and grids makes extraction very difficult. And segmentation under those writing restrictions should be quite different from unconstrained handwriting. So systems mentioned above are not suitable for processing Chinese bank checks and new algorithms are more desirable. Some experimental CAR systems for Chinese bank checks have been reported in [14, 15]. But their performance is not satisfactory enough for practical use.

Figure 1. A typical image of Chinese bank check

In this paper, we design a prototype system of CAR for Chinese checks. The whole system is designed under the guidance of three principles: information fusion, complementary method combination and multi-hypotheses generation then evaluation. Information from color, gray and binary images is fused and complementary algorithms are applied for binarization and string extraction. Multi-hypotheses are made by keeping all possible candidates when ambiguous solutions exist in extraction and segmentation. Then evaluation is carried out by recognition and the one with the highest confidence is selected as the final result.

978-0-7695-4661-2/12 $26.00 © 2012 IEEE
DOI 10.1109/DAS.2012.8

The rest of this paper is organized as follows. Section 2 presents the framework of our system. Implementation details for every step are given in section 3 and experimental results can be found in section 4. Section 5 provides our conclusion and future work.

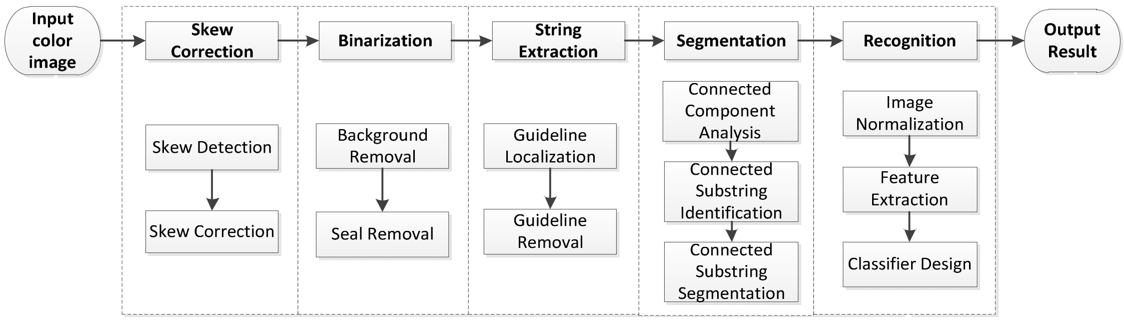

Figure 2. Flow chart of proposed system

II. FRAMEWORK

The whole system is constructed mainly by the four generally used modules: binarization, numeral string extraction, segmentation and recognition. And skew correction module is added as shown in Fig.2.

The images are sometimes skewed because they are got from different scanners by different operators. So in the first module, skew angle is detected and the image is deskewed. The layout of all Chinese checks from different banks is nearly the same because they are printed following the same standard. So courtesy amount area can be roughly extracted from a pre-defined region and sent to the next module. Uniform background and red seals which may cover the characters are removed in the binarization module. In string extraction module, constrained grids should be removed. Guidelines in the grids are localized exactly and then removed to get the numeral string. Multi-hypotheses are made and more than one extraction results are kept when it is not sure whether one part of guideline should be removed or not. In segmentation module, all the connected components (CCs) in an extracted string are analyzed first and the ones which should be segmented further are identified by their size and relative position to the guidelines. Then cut points for these CCs are determined and multi-hypotheses are also made when uncertainties exist. All possible candidates generated previously are sent to the last module and a recognition result as well as its probability is given for each input. The one with the highest probability is selected as the final result. Implementation details for each module will be given in the next section.

III. IMPLEMENTATION DETAILS

A. Image Skew Correction

In this module, operations are carried out following the detection then correction procedure. The original image is converted from RGB to gray space first and edges are detected using the canny method. Then standard Hough transformation is applied for detecting edges corresponding to the horizontal long lines in the original image. Skew angle is got by averaging the angles of top few long lines and the original image is rotated according to the skew angle. In our system, skew angle of checks never exceeds 15 degrees. So angles within 15 degrees are calculated only in the Hough transformation for acceleration. Finally, a pre-defined region which locates at the upper right side of a check image is extracted as courtesy amount area and sent to the next module for further processing.

B. Image Binarization

The binarization module consists of two steps: background removal and seal removal. Uniform background and red seals which cover numeral strings sometimes are removed in the two steps respectively. And numeral digits as well as the pre-printed grids in which digits are written should be preserved.

In the first step, operations are based on gray images and two complementary thresholding methods are combined. Otsu method is first used on the whole gray image. It is a global thresholding method and removes uniform background by calculating a single threshold adaptively. But the single threshold is not precise enough and some background pixels touching with characters are usually left. So a local thresholding algorithm called Niblack method is employed. Threshold is calculated for every pixel according to its neighborhood and characters can be extracted exactly with sharp edges from background. However noise appears in the background area without text. So in our system Niblack method is only applied with the non-background pixels got by Otsu method and characters are extracted cleanly while avoiding noises in background area and saving a lot of computational efforts.

In the second step, information from both RGB and HSI color spaces is used for removing red seals. All the foreground pixels got in the first step are transformed into HSI color space first. Then all the red-like pixels whose hue values are around 0 degree are extracted if $\cos(H)>0.85$ is satisfied. The condition is so loose that all the pixels belonging to seals and strokes covered by seals are extracted no matter what saturation or intensity they have. Then the R value of extracted pixels in RGB color space is gathered as a new data set. A threshold T is calculated according to this data set and all the pixels with a lower R value less than T are regarded as strokes and brought back to the original image. The threshold is defined as

$$T = \min(T_{Otsu}, T'), (T' = \text{minimum} + k*\sigma) \quad (1)$$

σ is standard deviation and k is a constant parameter (3 is empirically selected in our experiment).

Both the Otsu threshold and standard deviation are considered for the data set. When many stroke pixels exist in the extracted data set, T' is always very high as a result of high standard deviation. Threshold calculated by Otsu is suitable for this typical two-class problem and is always selected as the final T by the min operator. If only a few or even no stroke pixels exist in the data set, standard deviation is very low. Threshold got by Otsu is always too high to separate the stroke pixels while T' is relatively low. So T' is more suitable and is naturally used as the final T by the definition.

Finally, all the remaining pixels are regarded as foreground and binarization result is got. An example of the proposed binarization algorithm is given in Fig.3. And more details about the binarization module can be found in our previous work [16].

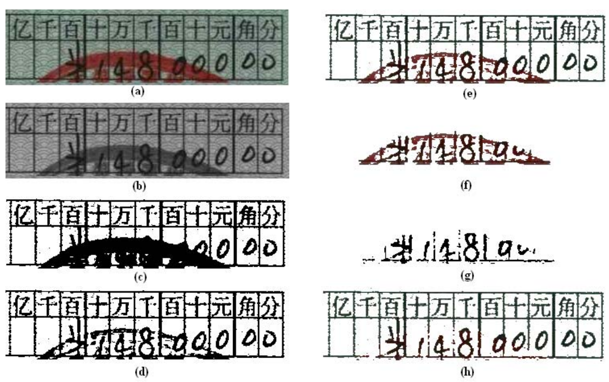

Figure 3. (a)Original color image (b)Gray image (c)Result by Otsu method (d)Result after employing Niblack method (e)Color image with background removed (f)Seal-like pixels extracted (g)Stroke pixels got by threshold T (h)Final result after stroke recovery

C. Numeral String Extraction

In this module, guidelines in restricted grids are localized then removed in order to extract the courtesy amount string.

In the first step, guidelines are roughly localized by projection in binary image. With the interference of characters and uncertain quality of binary image, peaks corresponding to guidelines are not always the most significant ones in projection histograms. So a lot of prior knowledge got from the standard layout of Chinese checks is used in order to make localization correct and robust.

Positions of three horizontal lines are searched first in horizontal projection histogram of the whole courtesy amount area. All the peaks whose values are greater than a pre-defined threshold are labeled and all the combinations of three labeled peaks are regarded as candidate groups of positions for three horizontal lines. Because all the checks in China are printed following a single standard and scanned with a pre-determined resolution, distances between the three lines are almost fixed and can be approximately calculated using the scanner DPI. Distance restriction is used and most of candidates are eliminated. If more than one position group left, transition times between foreground and background in the binary image are calculated at the upper line position of each group. The one with fewest transition times is more likely to be a line and its corresponding group is selected as positions for three horizontal lines.

Vertical projection is then carried out within the area between the first and last horizontal lines. The most significant peak in projection histogram is selected as the position of one vertical line. According to the standard layout of Chinese checks, width between adjacent vertical lines is nearly the same and can be approximated by the distance between the upper two horizontal lines. So a reference width for finding other vertical lines can be obtained by using positions of horizontal lines. And other vertical lines can be localized sequentially by finding peaks on both sides of existing lines within specific regions, whose centers are just away from the nearest vertical line already found by the reference width. When positions of all eleven vertical lines are got, the first step of localization is finished.

Positions for both upper and lower edges of each horizontal line (and left and right edges for vertical lines) are to be determined exactly in the second step. Gradient is computed within a small neighborhood of guidelines in the original gray image by employing Sobel operator. And maximum and minimum can be found precisely in the projection histogram, which correspond to the double edges of each guideline. Some localization results got by our proposed method are shown in Fig.4.

Figure 4. Examples of localization results

After all the edge positions are got, run-length of each guideline is analyzed in order to remove the restricted grids. For both horizontal and vertical lines, procedures of run-length analysis are nearly the same. So only vertical lines are taken into account in the following and operations for removing horizontal lines can be carried out in the same way. For vertical lines, run-length is defined as a set of connected black pixels in a single row. And a line can be represented by a group of run-lengths from different rows. If one run-length exceeds the area defined by double edges of the guideline, it is likely that stokes are connected with the current vertical line at this point. So guidelines at connected positions should be preserved while others are removed. If more than one connected position exists on the same side of a vertical line, it is not sure whether the gaps between them are pure line pixels or stroke pixels covered by guideline. So multi-hypotheses are generated in this situation by keeping two possible candidates: one candidate is got by keeping all the pixels in the gaps while in the other one all of them are removed. Fig.5 illustrates the run-length analysis for a vertical line.

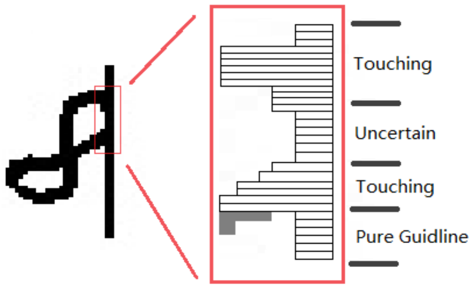

Figure 5. Example of run-length analysis for vertical line

All the guidelines in restricted grids are analyzed then removed one by one. Finally, elements out of the exact courtesy amount region are removed and candidates of extracted numeral string are left for further processing.

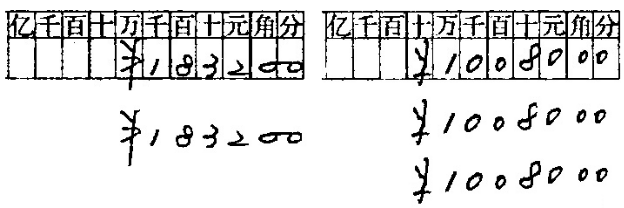

Figure 6. Examples of extracted candidates

D. *Numeral String Segmentation*

In the segmentation module, numeral string is segmented into isolated digits. Because digits in courtesy amount string are written inside preprinted grids, whose positions have been got in previous module, it is easy to tell if one part of the numeral string needs to be segmented according to its size and location. In this module, isolated digits are extracted first and dissection method is employed for finding out cut points of connected characters. Multiple candidates are kept if more than one possible cut point can be found in order to make sure that correct results can be obtained.

For each courtesy amount string, all connected components (CCs) are found first by contour tracing. Then CCs in the same grid are grouped together for further analysis. If the main body of one group lies in a single grid, it is extracted as an isolated digit. After extraction of all isolated digits, CCs containing connected digits are left for segmentation. The number of connected digits can be estimated by how many grids the CCs cover and vertical boundary of the grids can be used for reference in finding out cut points. With the help of these information, segmentation is carried out using the contour difference method proposed by Fujisawa [8]. Outer contour of the CCs to be segmented is separated into upper and lower part at the leftmost and rightmost point, respectively. And contour difference is calculated by formation:

$$C_{Diff}(x) = H_U(x) - H_L(x) \qquad (2)$$

$H_U(x)$ stands for the y value of the lowest point in the upper contour at x while $H_L(x)$ stands for the highest one in lower contour. We search the contour difference in a pre-defined interval around the referenced cutting point and the point with minimum absolute value is set as an original cutting position.

With the constraint of grids, there is little overlap between main bodies of connected digits. But they are connected by ligatures in most cases. So positions of original cutting points should be modified further in order to detect the ligature. If contour difference in the original cutting point is nearly the same as width of stroke, it is probably ligature at this point. Search is continued along both sides of the original cutting point until contour difference expands abruptly or contour direction changes significantly. And two candidate cutting points with ligature between them can be found for every two connected digits. Sometimes, the ligature is part of the digits and cannot be removed directly. In order to make sure the correct segmentation result is got, we keep two candidates for each isolated digit from segmentation: one without ligature and the other with part of ligature.

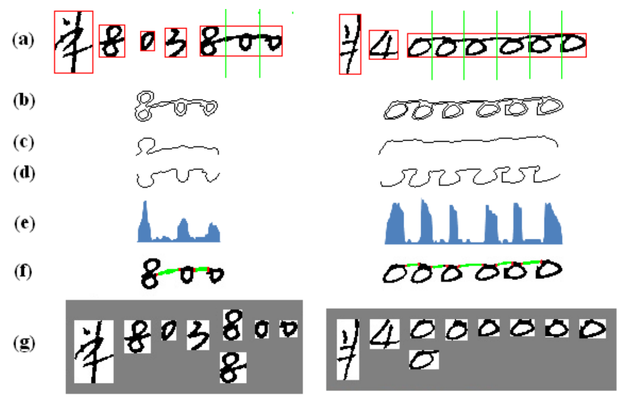

Figure 7. (a)Connected substring identification by analyzing CCs, positions of vertical lines are taken as referenced cutting points (b)Contour of connected substring (c)Upper contour (d)Lower contour (e)Contour difference (f)Candidate cutting points and ligatures (g)Segmentation results with candidates

In Chinese checks, there are often several zeroes connected together at the end of a courtesy amount string. In this situation, ligatures should be taken out directly in order to save computation time. So in our system, if a CC with connected digits lies at the end of the whole string and inner loops can be detected, two candidates are kept only for the first digit in order to avoid mistake when the substring does not start with zero. And single candidate is kept for the following digits by removing the ligatures for acceleration. Illustration of the segmentation module is given in Fig.7.

Finally, every courtesy amount string is segmented into groups of isolated digits. All the groups are sent to the next module for further recognition and evaluation.

E. *Recognition*

The OCR engine in this module is constructed using state-of-the-art techniques mentioned in [10, 11]. Moment normalization is applied and input images are transformed into a standard size. Gradient feature is extracted and support vector machine is utilized for classification.

We extract all digits from 2,000 real check images. And 1,000 samples for each digit are selected as training set, while the others are left for testing. The LIBSVM [17] toolbox is used for training and testing our OCR engine and a 99.5% accuracy rate is achieved on the testing set.

For each group of segmentation candidates, all the isolated digits except the first RMB sign are recognized by our OCR engine. Probabilities of all the recognition results in one group, which are calculated by algorithms in [18], are multiplied together as the confidence for this group.

All the groups are processed in the same way and results with corresponding confidence are obtained. Evaluation is carried out by comparing confidence of each result. The one with highest confidence is selected as the final output.

IV. EXPERIMENTS

In China, banks are nearly paranoiac about the confidentiality of the checks they provide for research. It is almost impossible to get a public check database for testing our system. So all the experiments are carried out with the data collected by ourselves. In our experiments, 10,346 real checks are collected from different banks and scanned into color images with the resolution of 200DPI for testing.

Our courtesy amount recognition system is just a prototype at present but very promising. A list of result candidates with corresponding confidence is given for each input check image and the one with highest confidence is regarded as the final result. The recognition rate is defined as the ratio of checks with correct results in the total number of checks. Experimental results are shown in Table 1 compared with the other two reported courtesy amount recognition systems for Chinese bank checks.

TABLE I. PERFORMANCE OF DIFFERENT SYSTEMS

System	No. of test samples	Recognition rate		
		1st candidate	*Top 2 candidates*	*Top 3 candidates*
System in [14]	84	89.3%	/	/
System in [15]	1053	86.5%	/	/
Our system	10346	90.3%	93.2%	94.1%

The system is tested on a PC with Intel E7200 @ 2.53GHz CPU, 2GB RAM, and running Windows XP SP3. All the 10,346 check images can be processed successfully within one hour in our experiments. As shown in table 1, our system can get high recognition accuracy on a large data set. And higher recognition rate can be achieved by considering more candidates. Although the three systems are tested with different data sets, we are still able to say that performance of our system is very promising according to the experimental results.

V. CONCLUTION AND FUTURE WORK

In this paper, we propose a prototype system of courtesy amount recognition for Chinese bank checks. Skew correction, binarization, string extraction, segmentation and recognition are sequentially carried out with the input check image in order to get the recognition result. Information from different sources is used and complementary methods are combined in the system. Multiple candidates are kept, and then verified in order to deal with uncertainty. The promising experimental results on a large real check image set show that: 1) the binarization and extraction modules, which fuse information from different sources and combine complementary methods, can get satisfactory results for further processing; 2) keeping multiple candidates is a powerful way of dealing with uncertainties in string extraction and segmentation, and the most suitable result can be obtained successfully by evaluating the recognition confidence; 3) our proposed system is able to process real bank checks from different banks within an acceptable time cost and it is very promising for practical use.

In the future, we aim at improving our system in the following ways: 1) constructing more complementary OCR engines; 2) designing a verification module for further analysis of recognition results and rejection to the uncertain ones.

REFERENCES

[1] N. Otsu, A threshold selection method from graylevel histograms, IEEE Trans. Syst., Madn. & Cybern., 9 (1979) 62-66.

[2] W. Nilback, An introduction to digital image processing, Prentice-Hall International N.J., 1985.

[3] Y. Xiangyun, M. Cheriet, C.Y. Suen, Stroke-model-based character extraction from gray-level document images, Image Processing, IEEE Transactions on, 10 (2001) 1152-1161.

[4] Chongyang Zhang, "Studies on prepocess in automatic financial document processing system," Ph.D. Dissertation, Nanjing University of Science&Technology, 2004

[5] X. Ye, M. Cheriet, C.Y. Suen, K. Liu, Extraction of bankcheck items by mathematical morphology, International Journal on Document Analysis and Recognition, 2 (1999) 53-66.

[6] Yan Zhang, "Research on Key Technologies of Automatic Financial Document Processing System," Ph.D. Dissertation, Nanjing University of Science&Technology, 2008

[7] Y. Bin, A.K. Jain, A generic system for form dropout, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 18 (1996) 1127-1134.

[8] H. Fujisawa, Y. Nakano, K. Kurino, Segmentation methods for character recognition: from segmentation to document structure analysis, Proceedings of the IEEE, 80 (1992) 1079-1092.

[9] G. Congedo, Segmentation of numeric strings, in: G. Dimauro, S. Impedovo, G. Pirlo (Eds.), 1995, pp. 1038-1038.

[10] C.-L. Liu, Handwritten digit recognition: benchmarking of state-of-the-art techniques, Pattern Recognition, 36 (2003) 2271-2285.

[11] C.-L. Liu, K. Nakashima, H. Sako, H. Fujisawa, Handwritten digit recognition: investigation of normalization and feature extraction techniques, Pattern Recognition, 37 (2004) 265-279.

[12] N. Gorski, V. Anisimov, E. Augustin, O. Baret, S. Maximov, Industrial bank check processing: the A2iA CheckReader; TM, International Journal on Document Analysis and Recognition, 3 (2001) 196-206.

[13] G.DIMAURO, S.IMPEDOVO, G.PIRLO, A.SALZO, Automatic bankcheck processing : A new engineered system, World Scientific, Singapore, SINGAPOUR, 1997.

[14] Hongdong Song, Yali Wang, Shaowei Xia, Recognition System of Handwritten Numerals on Checks Using Support Vector Machine, Computer Engineering and Applications, 3 (2003), 58-60

[15] Junxia Gu, Xiaoqing Ding, Multi-model Segmentation and Recognition Algorithm of Courtesy Amount, Journal of Image and Graphics, 4 (2008), 696-701

[16] Dong Liu, Youbin Chen, Binarizing the Courtesy Amount Field on Color Chinese Bank Check Images, the 11th International Conference on Document Analysis and Recognition (ICDAR2011), Beijing, China, pp.774-778

[17] C.-C. Chang and C.-J. Lin. LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.

[18] T.-F. Wu, C.-J. Lin, R.C. Weng, Probability Estimates for Multi-class Classification by Pairwise Coupling, J. Mach. Learn. Res., 5 (2004) 975-1005.