Table Recognition and Recovery in the Camera-Captured Image with Complex Background

Guiyang Xie[1, a], Yirong Mao[1,b] , Zifan Nan[1,c], Shizhe Wang[1,d],

Yue Li*[1,e], Wei Wang[1,f]

[1]College of Software, Nankai University, Tianjin 300071, P. R. China

[a]keen.guiyang.xie@gmail.com, [b]yirong_mao@hotmail.com, [c]nanzifan@mail.nankai.edu.cn

[d]shizhe_wang@163.com, [e]liyue80@nankai.edu.cn, [f]kevinwangwei@nankai.edu.cn

Keywords: Form Recognition; OCR; Form Representation; Table Retrieval

Abstract. With the development of OCR techniques, people start to use image processing methods to analyze and process the image-type documents (typically scanned images) for obtaining the information of characters and the layout. On the other hand, table, draw relatively less attention in OCR even it is an important writing structure containing plenty of classified massages in special format. This paper comes up with an innovative method aiming to detect, recognize and recover the table on a camera-capture image with complicated background. This method mainly includes the preprocessing, binarization and edge detection of the image, together with the detection of cells, as well as the analysis and recovery of table structure. In practical experiments, Emgu CV is employed to achieve more accurate result with better efficiency.

Introduction

At current stage, OCR technology for recognition of characters has reached a rather satisfactory accuracy. The research direction of OCR has changed from text OCR to characters detection and recognition with complicated background, or even further, real scene in natural environment.

However, table, as an important part of the text, has drawn less attention. The current OCR technologies generally regarding tables as usual lines or text boxes, whereas, table is an important writing structure containing abundant information about format, and is widely used in various writing situations.

Another emerging issue worthy discussion is table retrieval in the camera-captured image. Currently, professional business OCR systems are mostly designed for scanned images, and are not feasible for camera-captured images. Such technologies usually use threshold-binarization to identify the pixels as text or not, which do not have well adaptability on captured images. The different luminance distribution causes the failure of the binarization and thus causes the discontinuity of the border of the table, leading the border mis-organized as the wrong characters. Moreover, some of table text has complicated background, such as watermark, which increases the difficulty of table detection and recognition.

To solve the aforementioned problems, this paper comes up with a complete table recognition and recovery system based on the computer vision technology, able to process the camera-captured images with complicated background. Specially, we extract the table structure by the visual features, morphology as well as binarization on text's images. Our experiment shows that our methods have achieved a satisfactory result under the complex background.

Related Works

It is a challenge to design a complete system for the table recognition and recovery on the camera-captured pictures with complex background, due to the inconsistent luminance distribution through the image and the existence of the background. Fortunately, many publications have addressed some subtasks.

Image Rectification. The general study of image rectification is lens distortions, such as radial distortion. This paper research the trapezoid rectification result from the imaging principle of camera itself. [4], [5] contain a detailed statement on camera imaging principle.

Document Image Binarization. Document image binarization is a traditional and challenge problem in Document Image Analysis and Retrieval (DIAR), [1] proposed an OTSU-based adaptive and parameterless method for document image binarization. Inspired by Niblack, Sauvola's method is one of the most famous adaptive algorithms for image binarization. Wolf et al. [7] proposed a binarization method for multimedia documents based on Niblack's and Ostu's algorithms.

Table structure Analysis and Represent. Daniel P. Lopresti [3] did abundant research on form recognition and analysis. [5] paid attention to extract and represent form from various kinds of documents, such as PDF, postscript and image.

System Design

A scheme of our proposed system is presented in Figure 1. We will describe the algorithms of the detection and recovery of form in this section.

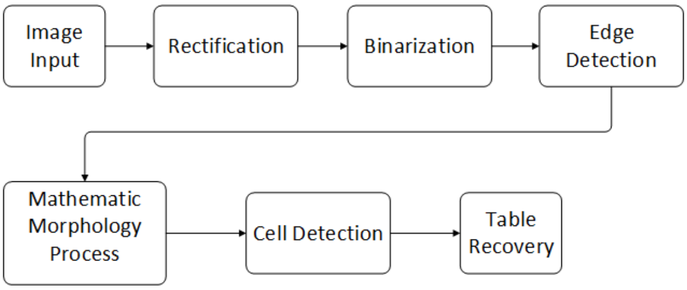


Figure 1: the Flowchart of the Table Recognition and Recovery system

Image trapezoid rectification. In those pictures, taken by pinhole camera, realistic parallel lines always intersect at infinity [4]. Furthermore, when the shooting angle is a sort of slant, the paper and the table are likely to get distorted to a trapezoid. Fortunately, the distortion can be rectified by a projective warp.

Let $X = [\mathrm{x}, \mathrm{y}]^T$ represent the points in distorted image, $X' = [\mathrm{x}', \mathrm{y}']^T$ represent the points in warp image. Thus $H_i = [\mathrm{h}_{i0}, \mathrm{h}_{i1}, \mathrm{h}_{i2}]'$ is the homograph vector and hence $H = [\mathrm{H}_0; \mathrm{H}_{1;} \mathrm{H}_2]$ represents the homograph matrix, which satisfies

$$\tilde{X}' = H\tilde{X} \tag{1}$$

where $\tilde{X}$ is the homogeneous form X .

The task is to find a proper H with following derivation that

$$x_i' = \frac{h_{00}x_i + h_{01}y_i + h_{02}}{h_{20}x_i + h_{21}y_i + h_{22}}, y_i' = \frac{h_{10}x_i + h_{11}y_i + h_{12}}{h_{20}x_i + h_{21}y_i + h_{22}} \tag{2}$$

$$\Rightarrow \begin{bmatrix} x_i & y_i & 1 & 0 & 0 & 0 & -x_i'x_i & -x_i'y_i & -x_i' \\ 0 & 0 & 0 & x & y & 1 & -y_i'x_i & -y_i'y_i & -y_i' \end{bmatrix} h = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{3}$$

And $h = [H_1', H_2', H_3']$.

h , the parameter, can be figured out with at least four matching points and the constraint $\|h\| = 1$.

In our system, we point out four corners of the paper in distorted image X , then find a minimal rectangle that covers the trapezoid corresponding to X . If there is any distortion, we calculate H to rectify the distortion by a projective warp. The camera-captured image and rectified image is shown in Figure 2 (a), (b).

Image binarization. Table detection relies on image binarization, aiming to abstract text and table pixels and remove the background pixels. Owing to uneven lighting, complex background, binarying the document images in a conventional way is a hard job need extra luminance balancing and background remove. In our experiments the table lines prone to breakage, even disappearance due to the uneven lighting. Therefore, Wolf's method [7], thresholding surface by sliding a rectangular window across the image, is employed in the proposed method. The threshold T for the center pixel of the window is calculated according to the formula (4), where M, S, and Min_I stand for the mean, variance and minimum of the gray values in the window. In addition, Max_S is the maximum of all the variance and k is a constant.

$$T_{x,y} = M_{x,y} + k*(S_{x,y} / \text{Max\_S} - 1)*(M_{x,y} - Min\_I) \quad (4)$$

$$B_{x,y} = \begin{cases} 255 & if\,(I_{x,y} > T_{x,y}) \\ 0 & else \end{cases} \quad (5)$$

The result is presented in Figure 2 (c).

Mathematical Morphology. The proposed method extracts the table from images using the operators of Mathematical morphology (MM).

The table lines may be broken after binarization, so firstly we enhance the images by detecting the edges with Canny Algorithm and then performing the erode operation. The result is presented in Figure 2 (d).

Though the binarization has filtered some background pixels, some other non-text and non-table pixels still remain. According to the feature that table is mainly composed of vertical and horizontal lines, two special structure elements, seH and seV, which are like a horizontal line and a vertical line with specific length L, have been designed to cleaning the non-straight-line component.

After performing the erosion operation in the image with seH, the horizontal lines are extracted whose length is longer than the specific number. Then we carry out the dilation operation to connect the broken lines. Likely, using the structure element seV, we extract the vertical lines.

Rectangle Detection. There are non-table lines exist in image after former step. We then remove them based on the feature that cells of tables are close rectangles. As seen in Figure 2 (f), imagine that the image is covered by lattices, and all lines in the image locate on the lattices' lines. Traverse all lattices and validate whether the rectangle is closed, then we can find out all rectangles which belong to the table.

Table Structure Analysis. According to the statistic of the detected rectangles, a table can be generated with the same number of rows and columns as original table. Then, we traverse each rectangle and compare it with the information about rows and columns to confirm the position and other information of the cell.

Table Represent. The table information, including word size, word orientation, word space and line space, are all useful and valuable.

First, we find out the relationship between the sizes in MS-Word's setting and the sizes of printed paper. Our input image contains a standard card which can be used to calculate the realistic size of the paper, word space and line space.

Let w represent the width of the cell, h represent the height of the cell.

1. Judge whether the cell is empty

   If there isn't any word in a specific area, the cell is empty. The center of the area is also the center of the cell. The width of the area is *f(x)*, and the height is *g(h)*.Simply we can let *f(w)=w/2*, *g(h)=h/2*.

   So only if there isn't any word in area *(f(w), g(h))*, the cell is empty and need to configure further more settings, such as the word font, orientation, etc.

2. Set the word font, orientation and space

   Case 1: words in the cell

   We set the orientation, space, size and font of the word the same as the existing words.

   Case 2: none words in the cell

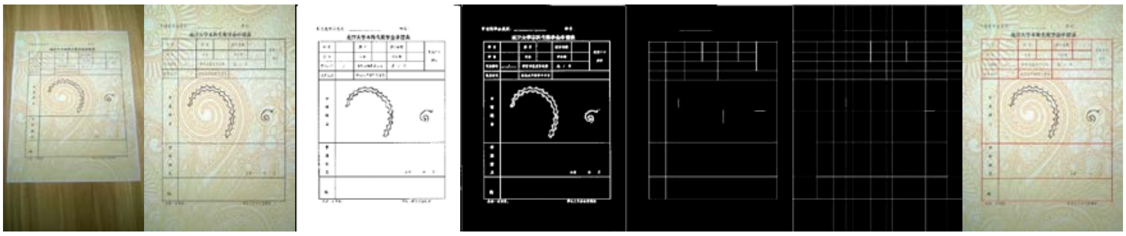
In this case, it is so hard to determine the setting, if only considered the current cell. Instead, we can refer to the surrounding cells just like case 1.
Eventually, we reconstruct the table and configure the settings in those empty cells.

Results

This test uses the algorithm we designed to perform the table represent. Putting the table and the original picture in the same image, and calculate the accurate rate of each cell's coincident area and original area. The test data is shown in Table 1.

Table 1. The accuracy rate of each cell's coincides area and original area.

No.	Accurate rate	No.	Accurate rate	No.	Accurate rate
1	0.9241	6	0.4340	11	0.9864
2	0.9718	7	0.9067	12	0.8315
3	0.9793	8	0.9219	13	0.8820
4	0.8911	9	0.8708	14	0.8119
5	0.9341	10	0.8504	15	0.9629



(a) (b) (c) (d) (e) (f) (g)

Figure 2: The intermediate results during the table detection process: Camera-captured image (a), the rectified image (b), the binarized image (c), the image after MM process (d), the detected table (e), the lattices image (f), the reconstructed table (g)

Conclusion

We have described a complete solution about table recognition and recovery in the pictures with complex background. Our system is based on the image processing and analysis technique and we have adapted it to the requirements of specific scenes. Acquiring a satisfying result, our system has a vast prospect of application.

Acknowledgement

This work is supported by the "National University Student Innovation Program".

References

[1] R. Farrahi Moghaddam and M. Cheriet, "AdOtsu: An adaptive and parameterless generalization of Otsu's method for document image binarization," Pattern Recognit., vol. 45, no. 6, pp. 2419–2431, Jun. 2012.
[2] J. Sauvola and M. Pietika, "Adaptive document image binarization," vol. 33, pp. 225–236, 2000.
[3] J. Chen and D. Lopresti, "Table Detection in Noisy Off-line Handwritten Documents," 2011 Int. Conf. Doc. Anal. Recognit., pp. 399–403, Sep. 2011.
[4] D. W. Embley, D. Lopresti, and G. Nagy, "Notes on Contemporary Table Recognition," no. April 2006.
[5] J. Ramel, M. Crucianu, N. Vincent, C. Faure, G. E. T. Enst, and P. France, "Detection , Extraction and Representation of Tables," no. Icdar, 2003.

[6] Trier, O.D.; Jain, A.K., "Goal-directed evaluation of binarization methods," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.17, no.12, pp.1191,1201, Dec 1995
[7] Wolf, C., Jolion, J.-M., & Chassaing, F. (2002). Text localization, enhancement and binarization in multimedia documents. Object Recognition Supported by User Interaction for Service Robots, 2, 2–5. doi:10.1109/ICPR.2002.1048482