Research on Information Recognition of VAT Invoice Based on Computer Vision

Jiaqiao Zhang[1, 2], Fuji Ren[2 *], Hongjun Ni[1 *], Zhenya Zhang[3], Kaixuan Wang[1]

[1]School of Mechanical Engineering, Nantong University, Nantong 226019, China
[2]Graduate School of Advanced Technology and Science, University of Tokushima, Tokushima 770-8506, Japan
[3]School of Economics and Management, Nantong University, Nantong 226019, China
c501937020@tokushima-u.ac.jp, ren@is.tokushima-u.ac.jp, ni.hj@ntu.edu.cn, z.zy@ntu.edu.cn, wkx@ntu.edu.cn

Abstract: With the promotion of the bill exchange system throughout the world, the use of VAT invoices has exploded. In order to solve the problems of low efficiency, high error rate and high labor intensity of manual entry of electronic system of VAT invoice, a method of recognizing VAT invoice information based on computer vision was proposed. Firstly, the invoice image was preprocessed, and the tilt correction was implemented by local adaptive threshold and Hough transform. Then the key information area was segmented and the target object was taken out by projection method. Finally, the characters were recognized by OCR technology. The results show that the method has low error rate and high efficiency, and the average character recognition accuracy can reach 96.21%.

Keywords: VAT invoice; Computer vision; Image processing; Information recognition

1 Introduction

With the development of social economy, the number of invoices used by enterprises is increasing rapidly, and the use of VAT (value-added tax) invoices has shown a leap-forward growth. Usually, some key information on VAT invoices, such as purchaser information, total amount and the name of goods, are generated by manual entry and verification, and then are stored in the information system or tax filing system of the enterprise. This traditional method with low efficiency and high error rate increases the labor of tax staff and prolongs the time for handling business, which is not conducive to the further development of enterprises [1]. Therefore, it is urgent to study relevant VAT invoice automatic recognition algorithm to reduce the pressure on financial staff, and improve the financial management efficiency of enterprises.

At present, scholars have conducted research on the recognition of VAT invoice information based on image processing and computer vision. Aslan Enes [2] proposed an invoice analysis method to achieve the processing of multiple types of invoices. Raymond Ptucha [3] proposed an intelligent character recognition method by using fully convolutional neural networks, and they introduced a probabilistic character error rate to correct errant word blocks. But this method was designed for handwritten fonts and couldn't be applied to complex invoice information recognition. Karthick Kanagarathinam [4] proposed a method for reading text information of electric meter scale image. The meter scale has only one line of numbers, but the invoice image recognition involves multiple areas, which requires area division. Ni Hongjun [5-6] proposed a hardware design method for the invoice recognition devices, including the split type and the air suction type, and the number of invoices was counted. But they did not specifically discuss how to obtain the invoice information.

This paper takes VAT invoice as the research object, and automatically identifies the effective information in the invoice through machine vision and image processing technology, and provides a scheme of invoice information recognition with high efficiency and low cost, which provides strong support for the rapid development of the tax industry. The recognition accuracy rate of the method used in this paper is up to 96%, and it meets the requirements of people.

The rest of the paper is organized as follows: Section II preprocesses the image of VAT invoice; Section III describes method for image segmentation, stamp removal and key information recognition in detail; Section IV illustrates the sample experiment results and discussion; Section V concludes the paper.

2 Image preprocessing

2.1 Binarization of invoice image based on local adaptive threshold

The binary image is a two-dimensional array of logical types with values of only 0 and 1. It is beneficial for reading invoice information correctly and improving recognition efficiency by converting the color invoice image into a binary image. The adaptive threshold is constructed for each seed pixel and it consists of all 8-adjacent pixels that satisfy some similar property to the seed pixel. Taking the gray level as the basic pixel property, the adaptive threshold for preprocessing is based on the additive or multiplicative tolerance interval. All pixels which are 8-adjacent to the seed pixel and whose gray levels are within the tolerance interval belong to the adaptive neighborhood. Let $f(i, j)$ be the seed pixel and $f(k, l)$ be the 8-adjacent pixels, then the pixels' adaptive threshold $f(i, j)$ can be expressed as:

978-1-7281-3863-3/19/$31.00 ©2019 IEEE

$$\begin{aligned} \left| f(k,l) - f(i,j) \right| &\le T_1 \\ \frac{\left| f(k,l) - f(i,j) \right|}{f(i,j)} &\le T_2 \end{aligned} \qquad (1)$$

Where, T_1 and T_2 are parameters of the adaptive threshold, and represent the maximum dissimilarity between adjacent pixels and the seed pixel.

It can be seen that the local adaptive threshold can determine the binarization threshold on the pixel location according to the pixel value distribution of the threshold pixels, thus local image regions of different brightness, contrast and texture will have corresponding local binarization threshold [7-8]. The binarized image of the VAT invoice based on the local adaptive threshold is shown in Figure 1.

Figure 1 Binarized image of VAT invoice

2.2 Tilt correction of invoice image based on Hough transform

The Hough transform is often used to detect straight lines and their inclinations in images. The VAT invoice image has a regular frame border line, so the Hough transform can be used to detect the inclination of the image accurately, and then rotate the corresponding angle to achieve the purpose of image tilt correction [9-10].

The detection of all possible linear pixels in the image can be obtained by applying edge detection, and all pixels whose edge amplitude exceeds the set threshold can be regarded as possible linear pixels. In this study, the Canny operator is used to perform edge detection on the image [11]. First, the image f is convolved with the operator $G_{\mathbf{n}}$:

$$f * G_{\mathrm{n}} = f * \left(\frac{\partial G}{\partial \mathrm{n}} \cdot \nabla G\right) \qquad (2)$$

Where, G is a 2D Gaussian smoothing operator, $\mathbf{n}$ is the normal vector of the local edge, $G_{\mathbf{n}}$ is the first-order directional derivative of G along the $\mathbf{n}$ direction, σ is the standard deviation, * is the convolutional symbol.

Then estimate the normal vector $\mathbf{n}$ of the local edge of each pixel.

$$\mathrm{n} = \frac{\nabla(G*f)}{\left|\nabla(G*f)\right|} \qquad (3)$$

The local maximum position of the convolution of $G_{\mathbf{n}}$ with the image f in the $\mathbf{n}$ direction is the edge:

$$\frac{\partial}{\partial \mathrm{n}} G_{\mathrm{n}} * f = 0 \qquad (4)$$

Further, the calculation formula of the edge position can be obtained by Eq.(5).

$$\frac{\partial^2}{\partial \mathrm{n}^2} G_{\mathrm{n}} * f = 0 \qquad (5)$$

The segmentation result of the Hough transform has strong robustness. Hough transform is not sensitive to data incompleteness and noise sensitivity, so it can solve the problem of thin lines and scattered target objects in the invoice image. Any line can be determined by two points, assuming two points in the image, A=(x_1, y_1), B=(x_2, y_2). According to the relevant knowledge of the linear equation, all the straight lines passing through point A can be represented by $y_1=kx_1+b$. Where, k is the coefficient and b is the intercept. Similarly, the equation can also be interpreted as an equation for the spatial parameters k and b, $b=-kx_1+y_1$, and the linear equation passing point B can be expressed as $b=-kx_2+y_2$. By transforming all the straight lines of these pixels to the corresponding points in the parameter space, the concentrated point (m, n) is detected in the parameter space, which is the line that frequently appears in the original image, $y=mx+n$.

In this process, the linear equation $y=kx+b$ cannot detect vertical lines ($k\to\infty$). In order to overcome this shortcoming, the image can be transformed from the Cartesian coordinate plane to the polar coordinate plane [12]. The specific principle is shown in Figure 2. Let the shortest distance from the origin to the line be ρ, according to the geometric relationship in the triangle, Eq.(6) can obtain:

$$\begin{aligned} \rho &= (x_0 + y_0 \tan\theta)\cos\theta \\ &= x_0\cos\theta + y_0\sin\theta \end{aligned} \qquad (6)$$

$$\rho = x\cos\theta + y\sin\theta \qquad (7)$$

Where, ρ is the vertical distance from the origin to the line, θ is the angle between ρ and the positive direction of the x-axis.

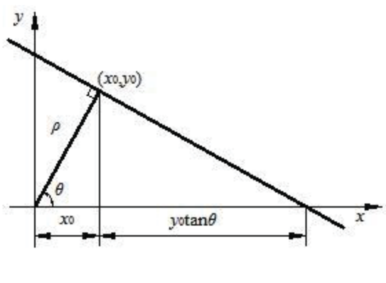


Figure 2 Hough transform principle diagram

Further, the peak value of the point in the polar coordinate plane is calculated, and the coordinate corresponding to the peak position is transformed into a straight line equation in the Cartesian coordinate plane, thereby obtaining the tilt angle θ of the image. In this embodiment, the tilt angle $\theta = -7.99$ °.

After determining the tilt angle of the invoice image, we perform the tilt correction by geometric transformation. The coordinates of the points in the output image after geometric transformation are usually approximated by a polynomial formula:

$$
\begin{aligned}
x' &= \sum_{r=0}^{m}\sum_{k=0}^{m-r} a_{kr} x^r y^k \\
y' &= \sum_{r=0}^{m}\sum_{k=0}^{m-r} b_{kr} x^r y^k
\end{aligned}
\qquad (8)
$$

Where, (x', y') is the coordinate of the point in the output image, (x, y) is the coordinate of the point in the original image, a_{rk} and b_{rk} are the coefficients.

This transformation is linear for the coefficients a_{rk} and b_{rk}, so if the corresponding pairs of points (x, y) and (x', y') in the two graphs are known, the coefficient a_{rk} and b_{rk} can be determined by the way of solving the of linear equations.

$$
\begin{aligned}
x' &= a_0 + a_1 x + a_2 y \\
y' &= b_0 + b_1 x + b_2 y
\end{aligned}
\qquad (9)
$$

Substituting the coordinates of the relevant points and the tilt angle obtained above, the polynomial equations of this embodiment are calculated as:

$$
\begin{aligned}
x' &= x\cos\theta + y\sin\theta = 0.99x - 0.14y \\
y' &= -x\sin\theta + y\cos\theta = 0.14x + 0.99y
\end{aligned}
\qquad (10)
$$

The invoice image is reversely rotated according to the tilt angle, and the result is shown in Figure 3.

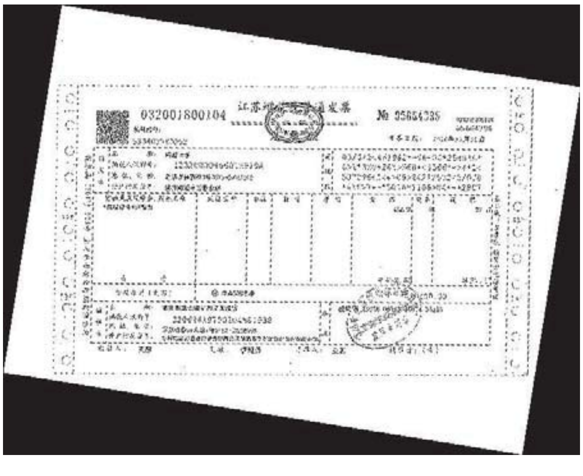


Figure 3 VAT invoice image after tilt correction

3 Image segmentation and information recognition

3.1 Image segmentation and information location based on projection method

The main body of the invoice image is divided by a plurality of horizontal and vertical lines, and the key information to be recognized is surrounded by a rectangular frame, and the pixel gray values of these rectangular frame edges are all zero. Therefore, by counting the number of pixels whose gray value is 0 in each row (column) of the invoice image, the peak of the number is corresponding to the edge of the different rectangular frame [13]. The horizontal projection of the image $f(x, y)$ along the x-axis is:

$$
I_x = \sum_{y=0}^{W-1} f(x, y), \quad (0 \le x \le H) \qquad (11)
$$

And the vertical projection along the y-axis is:

$$
I_y = \sum_{x=0}^{H-1} f(x, y), \quad (0 \le y \le W) \qquad (12)
$$

Where, $I(x)$ is a vertical projection, $I(y)$ is a horizontal projection, $f(x, y)$ is the original image, W is the width of the image, H is the height of the image.

Since the lines of the original image are thin and not continuous, the error of direct projection processing is large. Therefore, the image is first processed in morphology. The purpose of this study is to enlarge the black areas in the image, so the 5×5 cross-shaped structural elements are selected to inflate the image. The two sets are merged by the aggregate addition to enlarge the area of the scattered distribution points in the invoice binarized image, increasing the connectivity of the lines. The inflated VAT invoice image is shown in Figure 4.

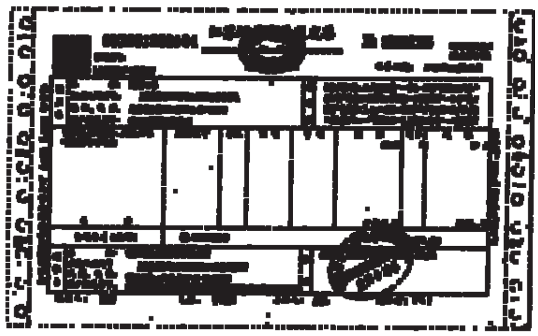
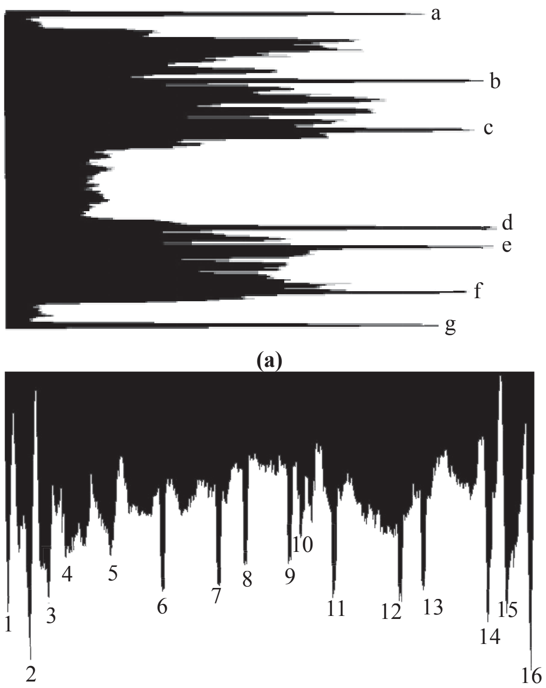


Figure 4 Inflated VAT invoice image

The invoice image is horizontally projected and vertically projected, and the result is shown in Figure 5. As can be seen from the figure, there are 7 distinct peaks in the horizontal projection direction and 16 distinct peaks in the vertical projection direction, which respectively correspond to the rectangular frame edges, so the positions of the target regions can be determined by these peak coordinates. For example, the area where the total amount is surrounded by the horizontal line d-e and the vertical line 5-10, and contains the key information of the par value. The other areas can be divided in the same way.



(a)

(b)

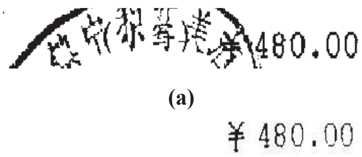
Figure 5 (a) Horizontal projection **(b)** Vertical projection

3.2 Stamp removal based on color channel

Most of the characters in the VAT invoice are printed characters, but the sellers often stamp the invoice manually. Sometimes the stamp is stuck in the target area where the information needs to be recognized, which affects the accuracy of character recognition. The main affected information is the total amount.

Through analysis, it can be found that the character information to be recognized is blue in the original color image, and the stamp color is red, so the interference information can be filtered by setting the color channel [14]. This study chooses to determine whether the stamp is in the target area after the completion of the segmentation. If there is no stamp in the target area, the subsequent processing will continue. Since the positional parameter of the target area is known, if there is a stamp in the target area, we can only process the target area in the original image, which can reduce the amount of calculation and improve the efficiency of the algorithm.

The original invoice image is in RGB mode, and there are three color channels. The color value of each pixel can be represented by (b, g, r), and the value range of b, g, r is [0-255]. Read the length and width of the image and traverse all the pixels. When $G_b(x, y)$ (the color component of b-channel) exceeds the set threshold (in this embodiment, $G_b(x, y)$=150), the point is retained, otherwise, forcing the point color value to be converted to (255, 255, 255). The comparison of the experimental results is shown in Figure 6. It can be seen that the method can effectively remove the bad influence of the stamp on the information recognition.



(a)

(b)

Figure 6 (a) Image with stamp **(b)** Image without stamp

3.3 Character information recognition based on OCR

After the target area is divided, the area to be recognized is projected horizontally, then the characters are divided into rows. The specific principle is the same as the invoice image segmentation and information recognition. The invoice information is further divided in detail, and each type of information is labeled accordingly. In this study, optical character recognition (OCR) technology is used to read the character information of the divided pictures [15]. At the same time, the Tesseract engine is used, which can be compatible with character recognition of more languages, and greatly improve the recognition accuracy. The recognized character information is written into the Excel report according to the corresponding label, thereby realizing the automatic recognition and storage of the key information of the VAT invoice.

4 Results and discussion

The final information recognition results of the embodiment used in this paper is as shown in Table I. A total of 66 characters were recognized, two of which were recognized incorrectly. The author processes 300 invoices and compares the results with other methods. It can be seen that the accuracy rate of the invoice information recognition method proposed in this paper can reach 96.21%, which is higher than other methods.

Table I Comparison of information recognition results

Information name	Correct results	Recognition results
Purchaser name	南通大学	南通大学
Taxpayer identification number	12320000466012919A	12320000466012919R
Address and telephone	南通市啬园路 9 号 0513-85012108	南通市凋园路 9 号 0513-85012108
Bank	建行南通分行营业部	建行南通分行营业部
Name of goods and services	住宿服务+住宿费	住宿服务+住宿费
Total amount	480.00	480.00

Table II Comparison of experimental results

Methods	Invoice type	Accuracy rate/ %
Xia Ting [16]	Used car invoice	85.40
Xie Zhigang [17]	VAT invoice	90.00
Li Zhiyi [18]	Jingdong invoice	93.49
Wu Junliang [19]	VAT invoice	95.10
Our Paper	VAT invoice	96.21

The analysis results show that the characters recognized by mistake are the English letter "A" and the Chinese character "啬". The author enlarged the size of the recognized image and successfully recognized the English letter, but the Chinese character was still not recognized successfully. This shows that the image size has an influence on the recognition results. If possible, the image size should be enlarged. Secondly, some seldom used Chinese characters were not successfully recognized. This indicates that the number of Chinese characters in the training library is not enough. Later, we can further improve the recognition effects of invoice information by expanding the character training library or by adding deep learning and neural network theory.

5 Conclusions

In this paper, the image of VAT invoice is preprocessed by local adaptive threshold and Hough transform to realize the tilt correction. The projection method is used to locate and segment the key information area of the invoice. The OCR is used as the classifier to recognize the characters of target area. Experiments show that the method studied in this paper has higher efficiency and lower error rate, which effectively reduces the labor intensity of tax practitioners. The next step of this paper is to continue to optimize the algorithm by adding deep learning, neural network and other theories, and further expand the training library for improving the accuracy rate of character recognition.

Acknowledgements

This research was supported by A Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), Key Research and Development Program of Jiangsu (Industry Prospects and Common Key Technologies) (BE2018093), Jiangsu Province University Students Innovation Training Program Key Project (201910304055Z).

References

[1] Ming Delie, Liu Jian, Tian Jinwen. Research on Chinese Financial Invoice Recognition Technology. PATTERN RECOGNITION LETTERS, 2003, 24(1-3): 489-497.

[2] Aslan Enes, Karakaya Tugrul, Unver Ethem, et al. An Optimization Approach For Invoice Image Analysis. IEEE Signal Processing and Communications Applications Conference (SIU), 2015: 1130-1133.

[3] Raymond Ptucha, Felipe Petroski Such, Suhas Pillai, et al. Intelligent Character Recognition using Fully Convolutional Neural Networks. PATTERN RECOGNITION, 2019, 88: 604-613.

[4] Karthick Kanagarathinam, Kavaskar Sekar. Text Detection and Recognition in Raw Image Dataset of Seven Segment Digital Energy Meter Display. ENERGY REPORTS, 2019, 5: 842-852.

[5] Ni Hongjun, Zhang Jiaqiao, Lv Shuaishuai, et al. Invoice Recognition Device [P]. China, CN108961536A, 2018.

[6] Ni Hongjun, Zhang Jiaqiao, Lv Shuaishuai, et al. Invoice Automatic Counting Device and Invoice Counting Method [P]. China, CN109117936A, 2019.

[7] Guo Jiebin, He Chuanjiang, Zhang Xiaoting. Nonlinear Edge-Preserving Diffusion with Adaptive Source for Document Images Binarization. APPLIED MATHEMATICS AND COMPUTATION, 2019, 351: 8-22.

[8] Shi Ji-chuan, Ray, Nilanjan, Zhang Hong. Shape Based Local Thresholding for Binarization of Document Images . PATTERN RECOGNITION LETTERS, 2012, 33(1): 24-32.

[9] Wang Jinjiang, Fu Peilun, Gao RobertX. Machine Vision Intelligence for Product Defect Inspection Based on Deep Learning and Hough Transform. JOURNAL OF MANUFACTURING SYSTEMS, 2019, 51: 52-60.

[10] El Hajjouji, Salah Mars, Zakariae Asrih, et al. A Novel FPGA Implementation of Hough Transform for Straight Lane Detection. ENGINEERING SCIENCE AND TECHNOLOGY, AN INTERNATIONAL JOURNAL, 2019.

[11] Kumar Gaurav, Umesh Ghanekar. Image Steganography Based on Canny Edge Detection, Dilation Operator and Hybrid Coding. JOURNAL OF INFORMATION SECURITY AND APPLICATIONS, 2018: 41-51.

[12] Filippus S.Roux. Coordinate Transformations and the Hough Transform. ENCYCLOPEDIA OF MODERN OPTICS (SECOND EDITION), 2018, 4:182-192.

[13] Wang Lei, Zhang Libo, Yang Xiao-dong, et al. Level Set Based Segmentation using Local Fitted Images and Inhomogeneity Entropy. SIGNAL PROCESSING, 2020, 167.

[14] Ma Ruiqiang, Zhang Shanjun. An Improved Color Image Defogging Algorithm using Dark Channel Model and Enhancing Saturation. OPTIK, 2019: 180. 997-1000.

[15] G. Abdul Robby, Antonia Tandra, Imelda Susanto, et al. Implementation of Optical Character Recognition using Tesseract with the Javanese Script Target in Android Application. PROCEDIA COMPUTER SCIENCE, 2019, 157: 499-505.

[16] Xia Ting. The Research and Application of the Character Recognition Algorithm for the Invoice Date [D]. Xi'an University of Technology, 2016.

[17] Xie Zhigang. Researches on Unsupervised Image Processing of VAT Invoices [D]. Shanghai Jiao Tong University, 2015.

[18] Li Zhiyi. Research on Automatic Identification Method of Invoice Content Based on Probability Graph Model [D]. Harbin Engineering University, 2016.

[19] Wu Junliang. Research and Application of Testing Technology Based On the Graph Theory [D]. Qingdao University of Science and Technology, 2018.