Template Matching-Based Method for Intelligent Invoice Information Identification

YINGYI SUN[1], (Student Member, IEEE), XIANFENG MAO[2],
SHENG HONG[1], (Student Member, IEEE), WENHUA XU[1], (Student Member, IEEE),
AND GUAN GUI[1] [iD], (Senior Member, IEEE)
[1]College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[2]Department of Scientific Research and Development, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Corresponding authors: Xianfeng Mao (mxf@njupt.edu.cn) and Guan Gui (guiguan@njupt.edu.cn)

ABSTRACT The increasing use of invoicing has created an unnecessary burden on labor and material resources in the financial sector. This paper proposes a method to intelligently identify invoice information based on template matching, which retrieves the required information by image preprocessing, template matching, an optical character recognizing, and information exporting. The origin invoice image is preprocessed first to remove the useless background information by secondary rotation and edge cutting. Then, the region of the required information in the obtained regular image is extracted by template matching, which is the core of the intelligent invoice information identification. The optical character recognizing is utilized to convert the image information into text so that the extracted information can be directly used. The text information is exported for backup and subsequent use in the last step. The experimental results indicate that the method using normalized correlation coefficient matching is the best choice, demonstrating a high accuracy of 95%, and the average running time of 14 milliseconds.

INDEX TERMS Invoice information identification, template matching, contour extraction, image processing, convolutional neural network.

I. INTRODUCTION

An increasing amount of invoices are being issued as the economy continues to rapidly develop [1]. The process of invoice reimbursement is time consuming and places a burden on financial staff, requiring the signatures of both reimburses and financial staff, a review, record, and backup of the invoices, as well as supervisor approval. The current invoice reimbursement procedure is also an inefficient waste of material resources. A simple and convenient method for invoice reimbursement is required for the sustainable development and efficiency of invoice reimbursement.

Delie *et al*. [2] proposed a solution to increase the operational efficiency of financial staff in which Arabic numerals and Chinese characters were automatically input from the invoice into a computer. Their proposed method preprocessed

The associate editor coordinating the review of this manuscript and approving it for publication was Zhanyu Ma.

the invoices by smooth processing, image inclination angle testing, and image slope adjustment. The information of the processed invoice image could then be entered into the computer by extracting the primary features of the invoice image and identifying fixed field features.

The initial step in the method to improve reimbursement efficiency is to classify various types of invoices. Alippi *et al*. [3] proposed an automatic invoice document classification system to classify invoices. The system analyzed graphical information such as logo and trademarks provided by the document to determine the category, performed well in closed and open world classification tasks.

Hamza *et al*. [4] proposed an initial Incremental Growing Neural Gas (IGNG) [5] model, and subsequently the improved, I2GNG, to solve the incremental classification problem, and applied it to invoice classification. Both models are related to the threshold of the creation of a new neuron. The local thresholds added in I2GNG could determine

to create or delete neurons. Model parameter settings were automatically determined by the model itself in their future research.

The Faster Region-based Convolutional Neural Network method (R-CNN) [6] has experienced a high level of progress in the field of object detection [7]. Zhu *et al.* [8] applied Faster R-CNN to train a detection network which could automatically detect the category and position of books. The model included a region proposal network and an object detection network, which were combined to predict the bounding boxes of books and recognize their categories.

With the help of cloud storage technology [9], Zhang [10] proposed an online invoicing system to greatly improve the efficiency of invoice reimbursement. The system utilized the uniqueness of the QR code on each invoice to identify the invoice to be reimbursed from the cloud storage server. This method dramatically reduced the probability of invoice information errors, speed up invoicing reimbursement, and enhanced the overall efficiency of reimbursement process.

The improvement of invoice reimbursement technology is inseparable from the development of image processing technology. Recently, with the aid of image processing and computer vision [11], methods solving many traditional problems have been developed including object detection [7], [12], water level observing [13], speech language identification [14], [15], intelligent energy networks [16], and image classification [17]–[19].

In addition, deep learning is a possible solutions for realizing intelligent invoice reimbursement technology due to its many success applications in many fields. For example, co-saliency detection [20]–[22] based on convolutional neural network (CNN) is quite related to the feature extraction problem. Zhang *et al.* [23] proposed an effective self-paced multiple-instance learning framework to achieve co-saliency detection. Cheng *et al.* [24] proposed a rotation-invariant layer to the existing CNN architectures to improve the performance of detection. Huang *et al.* proposed deep learning based channel estimation techniques for physical layer wireless communications [25]–[27], especially for millimeter massive multiple antenna systems [28]. Liu *et al.* [29] proposed a deep learning-based message passing algorithm for efficient resource allocation in cognitive radio networks. However, deep learning based invoice reimbursement technology may cause high cost due to high requirements in computation power.

A large number of companies now obtain invoice information using scanners. The use of various scanners leads to inconsistencies in scanning quality. High quality scanners produce scanned images that are clear and regular so that the information can be easily read, while low quality scanned images are blurry and contain a high level of redundant background information and deteriorated images. An example of a low quality scanned invoice image is provided in Fig. 1. In this paper, such problems are solved in the preprocessing phase using image processing technology.
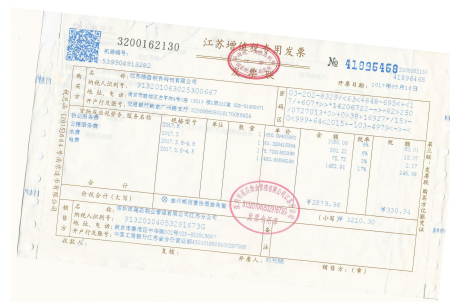

FIGURE 1. Scanned invoice image.

The method proposed in this paper utilizes image processing techniques, and proposes a template matching-based method to automatically identify invoice information. The invoice image is first preprocessed to eliminate the influence of rotation, inclination, and background, then the image is matched to extract the regions of key information. Optical character recognition (OCR) technology [30] is used to identify the characters in the region, and finally, the identification results of invoice information in different regions are exported to the excel form in the same format.

Nowadays, the invoice reimbursement process of the most companies is manually executed, including signatures of both reimburses and financial staff, a review, record, and backup of the invoices, which undoubtedly wastes a lot of human and material resources. Therefore, the proposed novel intelligent method in this paper can effectively solve this problem with high accuracy and fast speed by replacing the staff with computers.

The remainder of this paper is arranged as follows. In Section II, the process of the proposed method is explained. Section III describes each specific step in detail, and experimental results are provided in Section IV.

II. SYSTEM DESIGN
In this section, a template matching-based method is proposed to identify invoice information. The flow chart of the system is provided in Fig. 2, and describes the entire method process including inputting the original image, image preprocessing, template matching, optical character recognition, and information exporting.

A. PREPROCESSING
The original image of the invoice used in this study is provided in Fig. 1. This type of scanned image cannot be directly identified due to its irregularity, and preprocessing is required.

The preprocessing step mainly includes secondary rotation and cutting. The method utilizes the position information of the QR code to rotate the invoice twice, which provides a horizontal image. The horizontal image is then cut to remove useless background information. The preprocessed image is shown in Fig. 3.
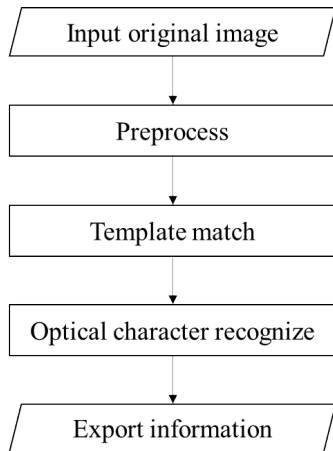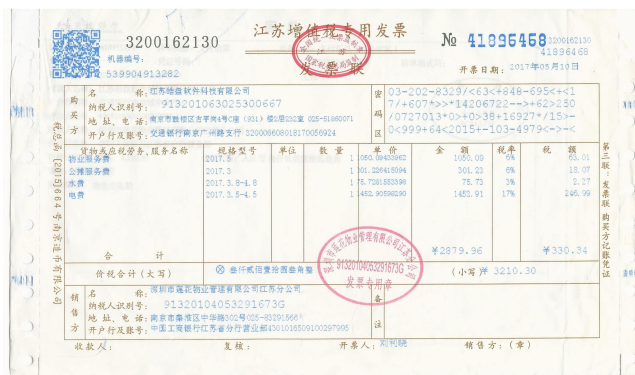
FIGURE 2. Flow chart of the entire process.

FIGURE 3. Preprocessed invoice image.

1) SECONDARY ROTATION

To rotate the image, the point and degree of rotation must be determined. As images scanned by individual scanners are usually different, these two factors may also vary. However, the relative position of the QR code in each kind of invoice image is fixed. As a result, the relative invariance of position can be utilized.

The fundamental principal of secondary rotation is contour extraction, which will be explained further in Section III. Contour extraction can determine the necessary position information of the required contour in the image. In this study, the position information refers to the pixel coordinates of the four vertices of the QR code's edge rectangle in the invoice.

In rotation phase, the image is rotated twice. The first rotation is to ensure that the QR code is in the upper left corner of the image. The second is to rotate the image with a smaller angle to make the image horizontal.

To carry out the secondary rotation, the point and degree must be determined. Because of the relative invariance of position, the rotation angle of each point in the image remains the same while rotating. Thus, to simplify the rotation phase, the rotation point is set to the midpoint of the image. The rotation angle can then be calculated according to the location information of the QR code.

In the first rotation phase, the vertex coordinates of the QR code can be calculated by contour extraction. According to these coordinates and the size of image, the location region of the QR code can then be determined. Four possible regions are illustrated in Fig. 4.
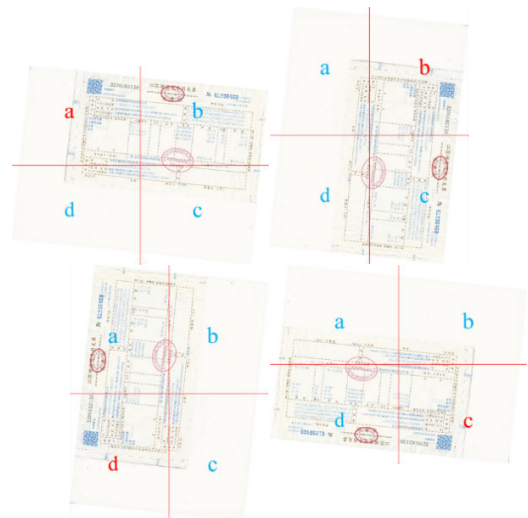

FIGURE 4. Four possibilities for the first rotation.

As is shown in Fig. 4, the rotation angle changes as the region of the QR code changes. The invoice image is divided into the four regions of a, b, c, and d, and the region that the QR code belongs to is expressed by the red letter in each image. If the QR code is judged at region a, the image will remain still. In other three cases of b, c, and d, the image must be rotated counterclockwise at 90°, 180°, and 270°, respectively.

After the first rotation phase, the QR code is in the upper left corner of the image. However, the first rotation is a rough adjustment to produce a relatively regular image rather than a horizontal image.

In the second rotation phase, due to the changes in coordinates caused by the first rotation, it is necessary to recalculate the vertex coordinates of the QR code. Next, to obtain the second rotation angle, the coordinates of the four vertices must be reordered according to the value of the x-coordinate. According to the value of the x-coordinate, the four vertices are named A, B, C, and D in order. Here, only the first two vertex are required to calculate the rotation angle. The angle can be calculated as in Eq. (1):

$$\alpha = \arcsin(\frac{x_B - x_A}{\sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}}) \qquad (1)$$

In which, α denotes the rotation angle, and x_A, y_A, x_B and y_B represent the horizontal and vertical coordinates of A and B respectively.

As seen Fig. 5, determining only the rotation angle cannot produce a horizontal image. The angles in these two cases are the same, so additional restrictions are required. In comparing these two cases, it can be seen that the direction of rotation depends only on the value of the y-coordinate of A and B.
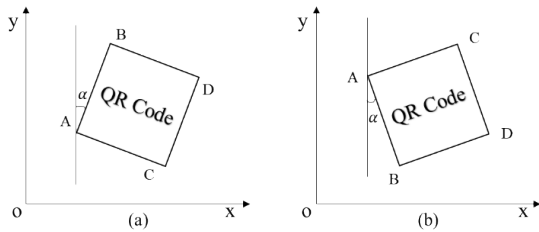
FIGURE 5. Schematic diagram of the second rotation with two possible directions of the second rotation. (a) The rotation that will be counter clockwise. (b) The rotation that will be clockwise.

If $y_B > y_A$, the image should be rotated counter clockwise. In other cases, the image should be rotated clockwise. After the second rotation, the image will be horizontal, as shown as Fig. 6.

FIGURE 6. Image after secondary rotation.

2) CUTTING EDGE

The image with secondary rotation contains a lot of useless background information. Efficiency of identification can be improved by removing these useless regions. Considering the same template for each type of invoice, the QR code is utilized in this step.

It is easy to obtain the position information of the QR code for an invoice image after secondary rotation. The range of coordinates in the image that should be preserved can then be calculated using length mapping between the QR code and the whole invoice image. The schematic diagram of the cutting edge is presented in Fig. 7.
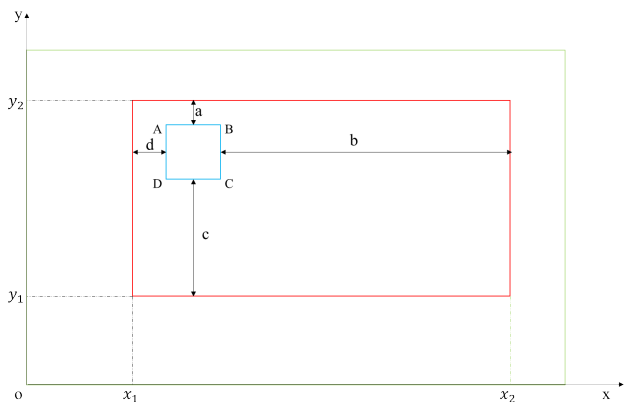

FIGURE 7. Schematic diagram of the cutting edge.

The schematic diagram of the cutting edge is illustrated in Fig. 7. The green rectangle represents the scanned image with secondary rotation, the red rectangle denotes the edge of invoice, and blue represents the QR code. The purpose of cutting edge is to remove the middle part between the red and the green rectangles, so that only the region within the red rectangle remains.

The mapping between QR code and the whole image is described by Eq. (2).

$$\begin{cases} \dfrac{AD}{AD+a} = \dfrac{y_A - y_D}{y_2 - y_D} \\ \dfrac{AD}{AD+c} = \dfrac{y_A - y_D}{y_A - y_1} \\ \dfrac{AB}{AB+d} = \dfrac{x_B - x_A}{x_B - x_1} \\ \dfrac{AB}{AB+b} = \dfrac{x_B - x_A}{x_2 - x_A} \end{cases} \tag{2}$$

Eq. (2) is simplified as

$$\begin{cases} y_2 = \dfrac{(y_A - y_D)*(AD+a)}{AD} + y_D \\ y_1 = y_A - \dfrac{(y_A - y_D)*(AD+c)}{AD} \\ x_1 = x_B - \dfrac{(x_B - x_A)*(AB+d)}{AB} \\ x_2 = \dfrac{(x_B - x_A)*(AB+b)}{AB} + x_A \end{cases} \tag{3}$$

Here, the parameters involved are described in Fig. 7. The AB and AD represent the length of the QR code in the actual invoice, and a, b, c, and d denote the distance from each edge of the QR code to the edge of the invoice, respectively. The horizontal and vertical coordinates of the vertices of the QR code are x and y, and the subscripts represent the corresponding vertices. The four parameters to be solved are the boundary of the invoice in the entire scanned image, that is, the coordinate range of cutting.

The invoice image can then be cut out using the four solved boundary ranges, and the cut image is illustrated in Fig. 3.

B. TEMPLATE MATCHING

After preprocessing the original scanned invoice image, the next step is to extract the region where the required information is located. This paper uses template matching to achieve this.

The process of template matching is to find a similar image of the template in a larger image, and determine its position information.

Here, the information before locating what needs to be identified in the image can be seen as a template. The similar image block can then be found in the invoice image. Using the geometric positional relationship between templates and information, it is easy to determine the position of the required information.

To obtain as much information as possible, twenty templates are constructed in this paper. These twenty templates are divided into six categories according to positional relationship, which are billing information,

purchaser information, goods information, amount information, seller information and biller information.

Billing information mainly includes the invoice number and billing date. Purchaser information contains the purchaser's company name, taxpayer identification number, contact information and bank card number. Goods information includes name, unit, quantity, unit price, amount, tariff, and tax. The amount information is comprised of uppercase and lowercase total-price-tax. Seller information includes the seller's company name, taxpayer identification number, contact information, and bank card number, and biller information is the name of biller.

As the image in the preprocess phase is gray, the template image should also be gray. The twenty template images used in this paper are provided in Fig.8. For convenience of display, the twenty template images are stacked together.

FIGURE 8. Twenty stacked template images.

After constructing the template images, the corresponding blocks in the invoice image can be detected using the template matching method. The details of template matching are explained in Section III.

For the location of goods information, the required details are located under the template. Thus, vertical search is used, and performs well in this situation. As a result, the number of goods under one template is often more than one, and constant detection is required while searching vertically, until the lower boundary is detected.

For other information, the required details are located on the right side of the template. In this case, a horizontal search is useful, and it only must detect the right region.

A selection of matching results are illustrated in Fig. 9, and contain five pieces of information with Chinese characters and numbers. As seen in Fig.9, the main aspects of the required information are completely extracted.
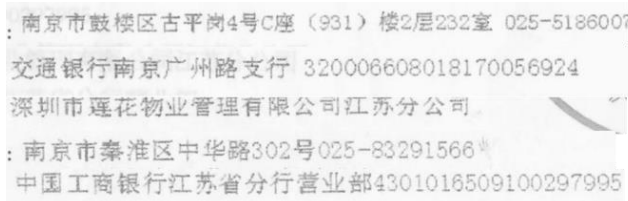

FIGURE 9. A selection of matching results.

C. OPTICAL CHARACTER RECOGNIZING

The above operations can recover the images of the required information, however, the obtained results are images instead of text. A method to convert image information into text is therefore required.

Optical character recognition is a method used to carry out image to text conversion as fast speed and with high accuracy. It detects the characters in an image according to brightness, and converts the characters to text.

Invoice images tend to contain too much information however, so the identification effect will be poor if the entire image is input. Thus, the image is separated into small parts for inputs, which is the undertaken in the first two operations.

After preprocessing and template matching, about twenty lines of the image is input. Every line of image contains only one line of characters, so that these lines can be easily and accurately recognized. Figure 10 displays some identification results for Fig. 3.

A selection of the OCR results are provided in Fig. 10. It can be seen that the identification effect is very successful compared to the scanned invoice image, and the Chinese characters, letters, and numbers are all identified accurately.

FIGURE 10. Some OCR results.

D. EXPORTING INFORMATION

Optical character recognition output is in the form of strings which can't be used directly and must be converted into saved texts.

For information other than goods, each matched image block contains only one piece of information and each piece of information is distributed in one line of image block. In this case, only exporting the recognized strings to their corresponding location in the excel table is required without any format modification.

On the other hand, for goods information, each matched image block generally cannot contain a complete piece of goods information, for example, when the goods name is too long to be contained in the same line. In this case, a strategy to judge if the information in a line is complete is necessary.

In this paper, the scanned invoice image was analyzed, and it was determined that this situation occurs only in the region of the goods names. Each line in this region usually contains Chinese, English letters, and numbers, and one line

can hold up to around 30 characters. Among them, one Chinese character occupies two characters, and one English letter or number occupies only one character.

Thus, when beginning to export the information in this region, it is necessary to judge whether the total characters of each line is equal to thirty. If so, the current goods name will be incomplete and the remaining information will be in the next line, meaning that, the identified information from the current and following line must be connected together. If the total characters of the current line is less than thirty, the length of the current goods name is complete, and the identified information can be exported directly. The exported excel table results are shown in Fig. 11.
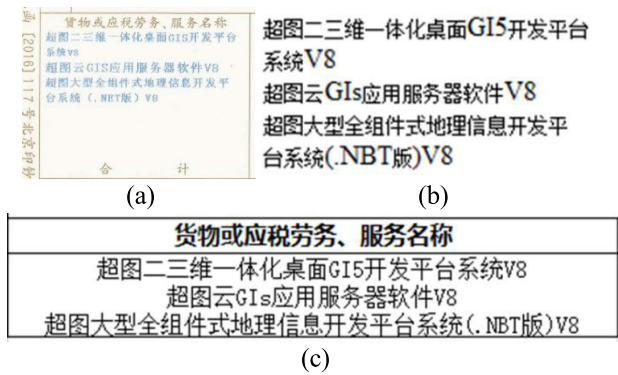

FIGURE 11. A selection of the exported excel results. (a) The original region of the identified part containing three pieces of information distributed in five lines. (b) The corresponding results of optical character recognition with the same format. (c) Results of exporting with the above processing which contains three pieces of information, each of which is distributed on the same line.

III. DETAILED EXPLANATION

A discussion of the detailed methodology mentioned in previous sections is provided in this section, including contour extraction and template matching.

A. CONTOUR EXTRACTION

Contour extraction is a method that can extract position information of the required rectangle. In this study, the position introduction mainly refers to the four vertex coordinate values of the required rectangle.

1) GRAYSCALE OPERATION

The input is a colorful image with three channels, which are red, green, and blue, respectively. To reduce computational complexity, the three-channel image is converted to a single-channel image. The single-channel image has the same overall brightness distribution characteristics as the corresponding three-channel image. The conversion for these two images is described as Eq. (4).

$$Gray = 0.072169 \cdot B + 0.71516 \cdot G + 0.212671 \cdot R \quad (4)$$

Here, Gray denotes the pixel value of the grayscale image, and B, G, and R represent the blue, green, and red component of the corresponding colorful image, respectively.

2) GAUSSIAN DENOISING

The grayscale image from the original scanned invoice image still contains a lot of noise information. To improve the quality of identification, the noise must be eliminated in advance.

Gaussian filtering is a linear smoothing filter used to smooth the signal, which can be seen as a convolution operation between the image and normal distribution. The operation of two-dimensional Gaussian function is described as Eq. (5).

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5)$$

Here, $G(x, y)$ denotes the two-dimensional normal distribution, and σ represents the standard deviation of distribution.

The convolution operation for images can be illustrated as Eq. (6).

$$g(i, j) = \sum f(i - k, j - l)h(k, l) \quad (6)$$

In which, $g(i, k)$ denotes the pixel value of each point of the convolved image, $f(i - k, j - l)$ is the pixel value of the point of the flipped image, and $h(k, l)$ represents the value of each point of the Gaussian filter.

3) EDGE DETECTION

Edge detection can provide image edge information, which is the key step in obtaining the position information of the QR code. This process can recover all edge information for one image so the required details are not omitted. The canny edge detection algorithm is used in this paper.

Edge information is detected according to the first derivative information of image in Canny. Two thresholds are then set in order to limit the range of edge pixel values. A high threshold is used to locate the point of the edge, and the background is clearly distinguished in the obtained edge. However, it is precisely because of the strict distinction that it is often difficult to form the obtained edges into a closed graph rather than intermittent lines. In this case, using the low threshold is essential.

Following this step, the regions near the edges with high threshold are detected again to find the missing edges using low threshold. The new edges are made up of the first and second detected edges.

4) EROSION AND DILATABILITY

The scanned invoice image usually contains some black dots which are larger than the noise. These dots are too large to be eliminated using the above process.

In this case, two morphological operations, erosion and dilatability, are introduced. Erosion is a process of eliminating boundary points and shrinking the boundary to the inside. It is essentially an operation to locate the local minimum of an image. This process can reduce the highlights and increase the darkness in the image. The second opposite operation is dilatability, which is the process of merging all background points that come into contact with an object into the object, and expanding the boundary to the outside. Similarly, it is

essentially an operation to find the local maximum of an image, and can increase highlights and reduce the darkness in the image.

The two operations are often used in combination to improve performance. In this paper, dilatability is used first, following by erosion. This is called the closing operation in morphological operations. Closing operation can fill small cracks, leaving the overall position and shape unchanged. The effects of this operation is illustrated in Fig. 12.
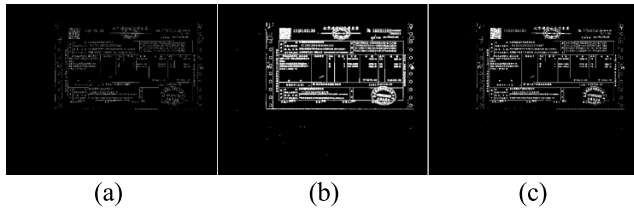

(a)   (b)   (c)

FIGURE 12. Effects of the closing operation. (a) Image with Canny edge detection, it is obvious that the edge is too thin to identify with human eyes. (b) Image obtained after dilatability, its edges are enhanced enough to be recognized by human eyes. (c) Image obtained after erosion, its edges are still clear, with some elimination of highlights.

5) EXTRACTING CONTOUR

Some complete contours are still unable to be found using edges detected with the above process. Thus, the next step is to combine these edge pixels into a contour. The image with Canny edge detection is a binary image which only contains points with a pixel value of 1 or 0. Among them, 0 denotes the black region, while 1 represents the white.

In a binary image, the contour consists of some points with a pixel value of 1, and the pixel value of its neighboring point is 0. The boundary points of the contours are determined by Eq. (7) and (8).

$$f(i, j-1) = 0, \quad f(i, j) = 1 \tag{7}$$
$$f(i, j) \geq 1, \quad f(i, j+1) = 0 \tag{8}$$

Here, $f(i, j)$ denotes the pixel value of the point in the binary image. Both $f(i, j-1)$ and $f(i, j+1)$ represent the neighboring points of $f(i, j)$. It is a simple process to locate the contours using Eq. (7) and (8). However, the contour of the QR code is specifically required, and this is determined in the following step.

This paper uses mapping relations between the QR code and the whole image. The proportion of area occupied by the QR code in the image is the limitation used to find the required contour. The detected contour of the QR code is provided in Fig. 13.

B. TEMPLATE MATCHING

Template matching is a method utilized to find the part most similar to the template in a bigger image. It is carried out by sliding the template block in the image to find the most alike part.

The three principles for template matching, are square error matching, correlation matching, and correlation coefficient matching, respectively.
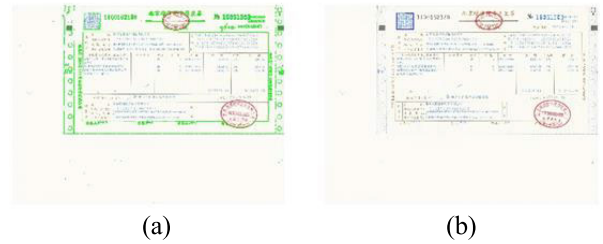

(a)   (b)

FIGURE 13. Detected contours of QR code.(a) All the contours detected are green, including the QR code. (b) The selected contour of the QR code which is colored blue.

1) SQUARE ERROR MATCHING

Square error matching is a principle based on the square error between the template and image. Its value is 0 for best matching, and 1 for the worst. This principle is described by Eq. (9).

$$R(x, y) = \sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2 \tag{9}$$

Its corresponding normalized form is described by Eq. (10).

$$R(x, y) = \frac{\sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2}} \tag{10}$$

Here, T denotes the template, and I is the image, and R represents the result of matching. As described in Eq. (9), the more similar the two images are, the closer the value is to 0. Therefore, the best matching is the image with the smallest $R(x, y)$.

2) CORRELATION MATCHING

Correlation matching is a principle based on multiplication operation between the template and the image. Its value is 0 for the least matching, and a larger value indicates better matching. This principle is described as Eq. (11).

$$R(x, y) = \sum_{x', y'} (T(x', y') \cdot I(x + x', y + y')) \tag{11}$$

Its corresponding normalized form is described as Eq. (12).

$$R(x, y) = \frac{\sum_{x', y'} (T(x', y') \cdot I(x + x', y + y'))}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2}} \tag{12}$$

Here, T denotes the template, I is the image, and R represents the result of matching. Eq. (11) indicates that the more similar the two images are, the larger the value. Therefore, the best matching is the image with the largest $R(x, y)$.

3) CORRELATION COEFFICIENT MATCHING

Correlation coefficient matching is an improvement to correlation matching. Both principals are based on multiplication operation, but the multiplications are replaced by correlation

with their mean. Thus, as above, a larger value indicates better matching. This principle is described by Eq. (13)-(15).

$$R(x, y) = \sum_{x', y'} (T'(x', y') \cdot I'(x + x', y + y')) \quad (13)$$

$$T'(x', y') = T(x', y') - 1/(w \cdot h) \cdot \sum_{x'', y''} T(x'', y'') \quad (14)$$

$$I'(x + x', y + y') = I(x + x', y + y') - 1/(w \cdot h) \cdot \sum_{x'', y''} I(x + x'', y + y'') \quad (15)$$

Its corresponding normalized form is described as formula (16).

$$R(x, y) = \frac{\sum_{x', y'} (T'(x', y') \cdot I'(x + x', y + y'))}{\sqrt{\sum_{x', y'} T'(x', y')^2 \cdot \sum_{x', y'} I'(x + x', y + y')^2}} \quad (16)$$

Here, T and I denote the template and image, and T' and I' are the correlation with their mean, respectively. R represents the result of matching, and w and h denote the width and height of the image, respectively. As above, the image with largest R(x, y) is the best match.

The mathematical expressions of the three principles indicate their computational complexity. It is clear that square error matching has the simplest calculation and correlation coefficient matching is the most complicated. Therefore, in the same situation, first principle square error matching may be the most time efficient, but third principle correlation coefficient matching may have the highest accuracy.

IV. EXPERIMENTAL RESULTS

This paper proposes a template matching based method to identify invoice information. In Section III, three principles for template matching are presented and explained. A selection of experimental results concerning the accuracy and running speed of invoice information identification are compared in this section.

A. IDENTIFICATION ACCURACY

To compare the identification accuracy of the three principles and their normalized forms, experiments are carried out on thirteen items in the invoice image, including billing information, purchaser information, amount information, seller information and biller information. The specific classifications are provided in Section II.

The identification results of the same invoice image are provided in Fig. 14. Subgraph a, c, and e in Fig. 14, represent the results with square error matching, correlation matching, and correlation coefficient matching, respectively, and b, d, and f are their normalized forms, respectively.

Using these principals, each method is tested on the same image, as shown in Fig. 3. Results clearly show that correlation matching and its normalized form have the poorest performance, and the correlation coefficient matching method and its normalized form show the best performance.

FIGURE 14. Comparison of identification results on a scanned invoice.

A comparison of the results on an electronic invoice is provided in Fig. 15. The method with correlation coefficient matching and its normalized form again display superior performance. In this case, after careful comparison of e and f, it is obvious that f normalized correlation coefficient matching is the most accurate. The superior region is circled in green.

FIGURE 15. Comparison of identification results on an electronic invoice.

Therefore, experimental results indicates that the method with the principle of normalized correlation coefficient matching has the best performance, which is the highest identification accuracy.

B. IDENTIFICATION SPEED

In practical applications, speed is essential as well as accuracy. The test method speed is evaluated with six principles on seven invoice images in this study, including scanned and electronic invoices. A comparison of the individual speeds is displayed in Table 1.

TABLE 1. Comparison of method speed.

	img1	img2	img3	img4	img5	img6	img7
Square Error Matching	4	6	5	5	4	5	4
Normalized Square Error Matching	11	13	12	13	13	15	14
Correlation Matching	3	5	4	4	5	4	4
Normalized Correlation Matching	13	13	9	12	13	15	14
Correlation Coefficient Matching	10	10	7	8	8	9	8
Normalized Correlation Coefficient Matching	14	14	15	14	15	14	14

Table 1 is a comparison of the identification method speed when utilizing various principles. The numbers in the table represent the average running time (millisecond) for template matching on seven different invoice images.

By carefully analyzing the data, it is obvious that correlation matching demonstrates the fastest identification speed, and square error matching is slightly slower. Correlation coefficient matching takes the most time of all methods. The normalized forms are all slower than the methods themselves.

C. FURTHER EXPERIMENTS

While normalized correlation coefficient matching is not the fastest, the average time for template matching is so small that a few milliseconds of slowness will have little effect when compared to the entire invoice identification process. Additionally, the accuracy of normalized correlation coefficient matching is the best. Therefore, it is clear that normalized correlation coefficient matching is the best choice.

More experimental results are provided in Fig. 16. The total number of identified information is 98, with four misidentifications which are circled in red. Thus, the overall identification accuracy is 95.92 %. After careful analysis of the wrong identification information, it can be determined that most unimportant in the reimbursement process. Information like money, goods, and purchaser are identified accurately, which are the foundation of invoice reimbursement.

FIGURE 16. A selection of final results for invoice identification.

V. CONCLUSION

This paper proposed a template matching based method to intelligently identify invoice information. The method includes four steps, which are preprocessing, template matching, optical character recognizing, and information exporting. Identification issues including secondary rotation, contour extraction, and branch information were then discussed. Experimental results determining accuracy and speed indicated that the normalized correlation coefficient matching method is the best choice for template matching. The accuracy of this method can be reach 95.92 %, and the most important information including money, goods, and purchaser were identified accurately. This proposed invoice information identification method can be applied to invoice reimbursement to greatly improve efficiency.

REFERENCES

[1] S. Bayar, "Performance analysis of e-archive invoice processing on different embedded platforms," in *Proc. IEEE 10th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Baku, Azerbaijan, Oct. 2016, pp. 1–4.

[2] M. Delie, L. Jian, and T. Jinwen, "The design and implementation of a Chinese financial invoice recognition system," in *Proc. Int. Symp. VIPromCom Video/Image Process. Multimedia Commun.*, Zadar, Croatia, Jun. 2002, pp. 79–82.

[3] C. Alippi, F. Pessina, and M. Roveri, "An adaptive system for automatic invoice-documents classification," in *Proc. Int. Conf. Image Process.*, Genova, Italy, Sep. 2005, p. II-526.

[4] H. Hamza, Y. Belaid, A. Belaid, and B. B. Chaudhuri, "Incremental classification of invoice documents," in *Proc. 19th Int. Conf. Pattern Recognit.*, Tampa, FL, USA, Dec. 2008, pp. 1–4.

[5] F. Alilat and S. Loumi, "Modelling of suspended matter by hybrid RBF-IGNG network," in *Proc. 8th Int. Conf. Intell. Syst., Theories Appl. (SITA)*, Rabat, Morocco, May 2013, pp. 1–7.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[7] K. Kuan, G. Manek, J. Lin, Y. Fang, and V. Chandrasekhar, "Region average pooling for context-aware object detection," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 1347–1351.

[8] B. Zhu, X. Wu, L. Yang, Y. Shen, and L. Wu, "Automatic detection of books based on faster R-CNN," in *Proc. Int. Conf. Digit. Inf. Process.*, Moscow, Russia, Jul. 2016, pp. 8–12.

[9] L. A. Tawalbeh and S. Habeeb, "An integrated cloud based healthcare system," in *Proc. 5th Int. Conf. Internet Things, Syst., Manage. Secur.*, Valencia, Spain, 2018, pp. 268–273.

[10] W. Zhang, "Online invoicing system based on QR code recognition and cloud storage," in *Proc. 2nd IEEE Adv. Inf. Manage., Commun., Electron. Autom. Control Conf. (IMCEC)*, Xi'an, China, May 2018, pp. 2576–2579.

[11] K. Azhar, F. Murtaza, M. H. Yousaf, and H. A. Habib, "Computer vision based detection and localization of potholes in asphalt pavement images," in *Proc. IEEE Can. Conf. Elect. Comput. Eng. (CCECE)*, Vancouver, BC, Canada, May 2016, pp. 1–5.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 580–587.

[13] J. Pan, Y. Yin, J. Xiong, W. Luo, G. Gui, and H. Sari, "Deep learning-based unmanned surveillance systems for observing water levels," *IEEE Access*, vol. 6, pp. 73561–73571, Nov. 2018.

[14] Z. Ma, H. Yu, W. Chen, and J. Guo, "Short utterance based speech language identification in intelligent vehicles with time-scale modifications and deep bottleneck features," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 121–128, Jan. 2019.

[15] Z. Ma, H. Yu, Z.-H. Tan, and J. Guo, "Text-independent speaker identification using the histogram transform model," *IEEE Access*, vol. 4, pp. 9733–9739, 2017.

[16] Z. Ma *et al.*, "The role of data analysis in the development of intelligent energy networks," *IEEE Network*, vol. 31, no. 5, pp. 88–95, May 2017.

[17] X. Zhang, K. Jiang, Y. Zheng, J. An, Y. Hu, and L. Jiao, "Spatially constrained bag-of-visual-words for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Beijing, China, Jul. 2016, pp. 501–504.

[18] Y. Yamamoto, S. Tsuruta, S. Kobashi, Y. Sakurai, and R. Knauf, "An efficient classification method for knee MR image segmentation," in *Proc. 12th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Naples, Italy, Nov./Dec. 2016, pp. 36–41.

[19] Z. Ma, Z.-H. Tan, and J. Guo, "Feature selection for neutral vector in EEG signal classification," *Neurocomputing*, vol. 174, pp. 937–945, Jan. 2016.

[20] J. Han *et al.*, "Representing and retrieving video shots in human-centric brain imaging space," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2723–2736, Jul. 2013.

[21] D. Zhang, J. Han, C. Li, and J. Wang, "Co-saliency detection via looking deep and wide," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 2994–3002.

[22] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2015.

[23] D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 865–878, May 2017.

[24] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[25] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.

[26] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, to be published. doi: 10.1109/TVT.2019.2893928.

[27] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[28] Y. Lu, C. Cheng, J. Yang, and G. Gui, "Improved hybrid precoding scheme for mmWave large-scale MIMO systems," *IEEE Access*, vol. 7, no. 1, pp. 12027–12034, 2019.

[29] M. Liu, T. Song, J. Hu, J. Yang, and G. Gui, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641–653, Jan. 2019.

[30] N. Prameela, P. Anjusha, and R. Karthik, "Off-line Telugu handwritten characters recognition using optical character recognition," in *Proc. Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Coimbatore, India, vol. 2, Apr. 2017, pp. 223–226.

YINGYI SUN (S'18) received the B.S. degree in mathmatics from the Nanjing University of Posts and Telecommunications, in 2017, where he is currently pursuing the master's degree. His research interests include deep learning, and optimization and its application in image processing.

XIANFENG MAO received the B.S. and master's degrees in electronic information engineering from Guangxi Normal University, in 2005 and 2009, respectively. He is currently with the Department of Scientific Research and Development, Nanjing University of Posts and Telecommunications.

SHENG HONG (S'18) received the B.S. degree in electronic information engineering from the Nanjing University of Posts and Telecommunications, in 2016, where he is currently pursuing the master's degree. His research interests include deep learning, and optimization and its application in image processing.

WENHUA XU (S'18) received the B.S. degree in electronic engineering from Shangdong Normal University, in 2017. She is currently pursuing the master's degree with the Nanjing University of Posts and Telecommunications. Her research interests include deep learning, and optimization and its applications in image processing.

GUAN GUI (M'11–SM'17) received the Dr.Eng. degree in information and communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2012.

From 2009 to 2012, he was supported by the China Scholarship Council, and the Global Center of Education, Tohoku University, where he joined the Wireless Signal Processing and Network Laboratory (Prof. Adachi laboratory), Department of Communications Engineering, Graduate School of Engineering, as a Research Assistant and as a Postdoctoral Research Fellow. From 2012 to 2014, he was supported by the Japan society for the promotion of science fellowship as a Postdoctoral Research Fellow of the Wireless Signal Processing and Network Laboratory. From 2014 to 2015, he was an Assistant Professor with the Department of Electronics and Information System, Akita Prefectural University. Since 2015, he has been a Professor with the Nanjing University of Posts and Telecommunications, Nanjing, China. His current research interests include on deep learning, compressive sensing, and advanced wireless techniques. He has received the Member and Global Activities Contributions Award and seven best paper awards from ICC 2014 and VTC 2014-Spring, ICC 2017, ICNC 2018, ADHIP 2018, CSPS 2018, and ICEICT 2019. He was also selected as a Jiangsu Specially Appointed Professor and a Jiangsu High-Level Innovation and Entrepreneurial Talent. He has received the Nanjing Youth Award. He was an Editor of the *Security and Communication Networks* (2012–2016). He has been an Editor of the IEEE Transactions on Vehicular Technology, since 2017, the *KSII Transactions on Internet and Information System*, since 2017, and the IEEE Access, since 2018.

• • •