Design and Implementation of English Text Recognition System Under Robot Vision

Gang Wang
Design and Implementation of English Text Recognition System Under Robot Vision

Gang Wang
Department of intelligent science and technology, JiMei University
Xiamen, Fujian Province
wangweb@sina.com

Abstract: the traditional methods of English letter visual recognition are machine learning and contour feature extraction. However, limited by the hardware resources such as camera and processor, it is difficult for the robot to realize fast and accurate text extraction and recognition. In this paper, median filtering method is used for image preprocessing, corner detection method for text extraction, overall framework of text recognition based on improved SOFM recognition algorithm of concentric rectangles with equal area, and based on the NAO robot platform to achieve system inspection.

Key words: robot, character recognition, SOFM, concentric matrix

I. THE INTRODUCTION

Robots are increasingly widely used in warehousing and logistics, building distribution, emergency safety, home services and other fields. Various laws and regulations have been promulgated to promote the development of the robot industry. Computer vision image processing technology is used for target recognition, image characteristics analysis[1], in particular, the optimization of optical character recognition technology is the basic guarantee to promote the robot to expand its application field and improve its application capability, the robot is used in bill recognition, license plate tracking, sign processing and other scenarios, due to the robot hardware resources are limited and there are many different kinds of fonts, how to realize the rapid and accurate robot based on the text recognition is becoming more and more attention and research.

There are many methods to recognize English letters, the most popular are machine learning and contour feature extraction. However, the former has a large amount of data processing and a high demand for system resources, while the latter has a high demand for camera hardware and low efficiency due to deformation, so it is not suitable for mobile robot platform.

In this paper, median filtering method is used for image preprocessing, corner detection method for text extraction, overall framework of text recognition based on improved SOFM recognition algorithm of concentric rectangles with equal area, and based on the NAO robot platform to achieve system inspection. Because business card has the physical characteristics of license plates, bills, signage and other common scenarios, this paper takes robot business card recognition as an example to discuss in detail.

II. SYSTEM DESIGN AND IMPLEMENTATION

The system is developed based on the V4 version of NAO robot platform. The effective image acquisition pixel is 1288*968, the pixel size is 1.9 million * 1.9 million, the resolution is 1.22 Mp, and the processor is ATOM E3845 1.91GHz.

The system is divided into three parts: image preprocessing, text region extraction and text recognition.

A. Image preprocessing

Image preprocessing includes graying, smoothing, sharpening, grey scale segmentation and tilt correction. Among them, the smoothing operation can remove the noise in the image, and it has a large influence factor on the extraction and recognition of the later text, which is the key
part of the system in the pre-processing stage. Experimental comparison methods: neighborhood average method, median filter method and KNNF filter method.

1) Neighborhood average method: All pixels of the image are traversed. When noise is determined in the neighborhood, the pixel is set as the average of the neighborhood. Otherwise, the pixel value remains unchanged. The process can be described as:

\[
g(x,y) = \begin{cases} \frac{1}{M} \sum_{(i,j) \in S} f(x,y), & \left| f(x,y) - \sum_{(i,j) \in S} f(x,y) \right| > T \\ f(x,y), & \text{otherwise} \end{cases}
\]

Where, \( S \) is the selected neighborhood, \( M \) is the number of pixels in the neighborhood, and \( T \) is the threshold value.

Since most of the noise is concentrated in the high frequency band and the main energy of the image is concentrated in the low frequency stage, the method of attenuation of high frequency can effectively smooth the noise. Suppose the impact response of the filter is \( H(r,s) \), then the output result of the filter \( g(x,y) \) can be expressed as the convolution form, namely

\[
g(x,y) = H(x,y) * f(x,y) = \sum_{r=-K}^{K} \sum_{l=-L}^{L} f(x-r,y-l)H(r,s)
\]

2) Median filtering method: Select an odd-point sliding window \( W \), traverse the image with the window, sort the selected pixel points in the window, take out the gray value in the middle to replace the gray value of the pixel points, can be described as:

\[
g(x,y) = Med \{ f(x-k, y-l), (k,l) \in W \}
\]

The neighborhood average method uses a low-pass filter to suppress the noise, but the image edge usually contains a large amount of high-frequency information. Therefore, while denoising, the smoothing operation also blurs the image boundary, which is the most effective for eliminating the isolated points, line segment pulse interference and image scanning noise [2]. As shown in Figure 3 and Figure 4:

3) KNNF class filtering method: The grayscale values of pixels of the same type are highly correlated. Therefore, an \( m \times m \) window can be selected, and the gray value of the center pixel in the window is replaced by the average gray value of
the closest K adjacent pixels. As shown in Figure 5 and Figure 6:

\[ M_I = \nabla I \nabla I^T = \begin{bmatrix} I_x & I_y \\ I_y & I_x \end{bmatrix} \begin{bmatrix} I_x^2 & I_x I_y \\ I_y I_x & I_y^2 \end{bmatrix} \]

(4)

Where \( \nabla I \) is the image gradient containing the derivatives \( I_x \) and \( I_y \). Because of this definition, the rank of \( M_I \) is 1, and the eigenvalues are \( \lambda_1 = |\nabla I|^2 \) and \( \lambda_2 = 0 \). The convolution can be calculated by the weight matrix

\[ \tilde{M}_I = W \ast M_I \]

(5)

Get the local average of \( M_I \) in the surrounding pixels, the calculated matrix \( \tilde{M}_I \) is also called Harris matrix[2]. The \( W \) value determines the size of the area of interest around pixel \( x \).

The text region can be extracted effectively by the region threshold of the number of corner points. See Figure 7 and Figure 8:

C. Improvement of recognition oriented method based on SOFM

SOFM neural network algorithm[3] is an unsupervised learning vector quantization method, which does not need to know the attributes of training samples, but conducts pattern clustering through self-organization[4].

1) Feature extraction [5]: Feature extraction is a key issue in English character recognition. It should be: ① Differentiability, the obvious difference of features is the basis for rapid recognition; ② Independence, characteristics should be independent of each other[3]; ③ With a small
number, the complexity of pattern recognition systems increases rapidly with the increase of dimensions.

To ensure that in the limited hardware conditions to achieve a high accuracy of recognition. In this paper, single character image feature is used as the input data of SOFM network. 15 * 15 single characters can be converted to templates, calculate the gray average value of 255 feature blocks $F_i = \frac{1}{n}\sum_{j=1}^{n} f_j$, of which $i = 1, 2, \ldots, 225$, $n$ the number of pixels in this region. Then 255 $F_i$ accumulative average:

$$F_0 = \frac{1}{72}\sum_{i=1}^{72} f_i.$$ Then by $X_i = |F_i - F_0|$, the eigenvalue $X_i$ of 255 feature areas in the image is obtained. The eigenvalue $X_i$ of 255 eigenblocks is used as the 255 dimensional eigenvector of the input layer of SOFM network.

Each letter has a 255 dimensional feature vector to reflect the image distribution, which is different from each other. Affected by the robot viewing Angle may exist four directions. After self-organization learning, the output layer must correspond to $4*2*26$ character classification marks. After 2380 experiments on text recognition of multi-angle and multi-type business cards and signage, the NAO robot can achieve the English letter recognition rate of 98.4%. But the recognition speed is around 12ms/ unit.

2) Improvement of SOFM network based on equal-difference area concentric rectangle slices: There are 255 input nodes and 208 output nodes, which need to match 208 weight vectors. Therefore, the computation speed is slow, and it brings great challenges to the robot system resources, making it impossible for the robot to realize parallel task processing. This paper presents a method to reduce dimension of SOFM based on concentric rectangle slice.

![Concentric rectangle processing diagram](image)

As shown in Figure 9 and Table 1, after slicing the concentric rectangle according to equal area, case classification can be quickly carried out according to the gray value of the outer slice, and the gray value of the concentric rectangle slice remains unchanged no matter whether the letter is in any of the four directions. According to the above analysis, the algorithm flow is as follows:

① Perform equal-difference area concentric rectangle cutting on the letter image, and quickly realize uppercase or lowercase classification according to the results of the outer slice.

② In this paper, for each letter input layer reduced to 50 layers of slices of the feature vector, the output node is 26, can be more accurate text processing, and for different case, design a model with 2 sub-sofm networks.

③ Each identification is based on the result of the number of ○1, the corresponding sub-sofm network is selected for identification. Only 26 types of weight vectors need to be matched for each recognition. Compared with 208 types of matching, the computation of network weight has been greatly reduced and the recognition speed has been effectively improved. The specific algorithm is as follows:

① Initialize the learning coefficient $\alpha$ and the link weight of the input and output nodes;

② Input sample data of the $m$ type image $X_m = [x_1, x_2, \ldots, x_m]$ $m = 1, 2, \ldots, 26$;

③ Calculate the distance between the input $X_m$ and the connection weight $W_j$ of all output nodes:

$$L_i = \sum_{i=1}^{n}(x_i - w_{ji})^2$$ $i = 1, 2, \ldots, 26$.

④ Find the winning node: $L_k = \min\{L_i\}$

| A | a |
|---|---|
| b | B |
| C | c |
| d | D |
| e | e |
| F | f |
| G | g |
| h | H |
| i | i |
| j | j |
| k | k |
| L | l |
| m | M |
| N | n |
| o | O |
| p | P |
| Q | q |
| R | r |
| s | S |
| t | t |
| u | u |
| v | V |
| w | w |
| x | X |
| y | Y |
| Z | z |

**TABLE I. GRAY SCALE SEQUENCE OF EQUAL-DIFFERENCE AREA CONCENTRIC RECTANGLE SLICES**
⑤ Adjust the connection weight of the output node k and
the input node and learning coefficient α:

\[ w_k(p + 1) = w_k(p) + \alpha (p) [X_m(p) - (p)] \]  

(7)

\[ \alpha (p + 1) = \alpha (p) - \frac{\alpha}{n} \]  

(8)

Where n is the number of sample data of the m class
currency;

The system first slices 50 images of concentric rectangles
of equidistant area by extracting the letter images to be
recognized, and extracts 50-dimensional feature vectors, and
then based on the concentric rectangle circumscribed
characteristic value of the image to select the corresponding
SOFM network, and will have to extract 50-dimensional
feature vector data as a child of the selected SOFM network
input data, and according to the \[ L_j = \sum_{i=1}^{n} (x_i - w_j)^2 \], calculate the value of \( L_j \). Finally, the letter recognition
is realized according to the mark of the output node.

III. EXPERIMENT AND CONCLUSION

In this paper, the system was tested on NAO (V4) robot
platform for 2380 business cards and signage, and compared
with other methods. Table 2 is the experimental results.

| Methods | Contour feature matching [6] | Global feature [7] | SOFM | After the improvement of this paper |
|---------|-----------------------------|------------------|------|-------------------------------------|
| Recognition rate | 96.1% | 96.3% | 98.4% | 97.8% |
| Mistakenly identified | 2.3% | 1.2% | 1.1% | 1.2% |
| Test sample | 2000 | 2000 | 2380 | 2380 |
| Speed (MS/PCS) | 14 | 15 | 12 | 19 |

Although the recognition rate of the improved method
decreases slightly, the recognition speed is improved
obviously.

IV. CONCLUSION

In view of the limited hardware conditions, this paper
proposes a system architecture and optimization algorithm
design from image processing to text extraction and
recognition, and tests it on the NAO robot platform to verify
the feasibility of the system and method, and realizes
the application of the robot in recognition scenes such as signage
and business card.

ACKNOWLEDGEMENTS

This work was supported by the Natural ScienceFoundation of Fujian Province of China (2019H0021), and (2019J01712). The establishment of
experimental environment and experimental data acquisition
is form the staff of laboratory. Special thanks again: Wei Chen, Nan Chen, Zhen Chen etc. Thanks for their wholehearted
support. Special thanks to Pro Yunming Pu’s guidance. And
thanks for a lot of scholar’s achievements in others
organization.

REFERENCES

[1] Yuxia Li. Design and implementation of English letter recognition
software system under computer aided vision Modern electronic
technology 2017.02(40):44-48.

[2] Xiaoping Yuan, Minming Tong, Jinlin Xu. Research on application of
image processing in TV image alarm in coal mine industry Computer engineering and design 2008.09(29):4833-
4835.

[3] Bin Duo, Xufeng Wang. Design and implementation of banknote
identification system Journal of Harbin university of science and
technology 2008(4):13-16.

[4] Ruiqing Qie, Linfu Xue, Jing Sun, Lin Zhang. Comprehensive
reservoir evaluation of chengzihe formation in jixi basin by
SOFM Science of surveying and mapping 2010.05(3):91-94.

[5] Fei Ye, Feng Li. Fast handwritten numeric character recognition
based on the whole feature Computer engineering and design
2006.10(27):4347-4349.

[6] Kunping Xiao, Jiwei Li, Hong Huang. License plate character
recognition based on combinatorial feature and support vector machine
Computer knowledge and skills 2010.21(36):51-54.

[7] Jia Luo, Dagang Liu. A fast English letter recognition method based
on the whole feature Computer and digital engineering
2012.9(275):117-119.