Extraction of Text Areas in Scene Images by Focusing on Local Contour Density and Corner Pixel Ratio

Naoya Taira[1], Fumihiko Saitoh[2]

[1]z4525047@edu.gifu-u.ac.jp
[1.2]Graduate School of Engineering, Gifu University 1-1, Yanagido, Gifu 501-1193, Japan
Keywords: Neural network. Text area extraction, Local contour density, Corner pixel ratio, Scene image

ABSTRACT

The character region extraction required in various fields such as self-driving. Advanced techniques of the character region extraction skills connect improving character recognition system. In this paper, we propose a method for the character region extraction from scene images. This method identify character by neural network with local contour densities and corner pixel ratio. The experimental result show that the proposed method obtains the sufficient accuracy to character region extraction.

1 Introduction

In recent years, character recognition technology has been required in various fields such as web services and automatic driving. Accordingly, research on character recognition has been actively conducted[1]. The current character recognition technology has achieved considerable accuracy for documents with good print quality and images with very good contrast. However, not all landscape images are easy to recognize, and there is a high demand for technology that can maintain a high recognition rate for images lacking in resolution. For character recognition, it is necessary to first extract the character area and then identify the character type. However, while the success rate of character type identification is 97.7%[2], the success rate of character area extraction is as low as 91.3%[3] , so the development of character area extraction technology is expected to have a significant impact on character recognition technology. The extraction of character regions can be divided into "feature extraction" and "identification processing "[4, 5]. In "feature extraction", the parameters for separating the character region from the background region are determined, and in "discrimination", the features obtained in feature extraction are used to discriminate the character region from the background region.

In this paper, we propose a text region extraction method that uses a neural network (NN)[6] to discriminate text regions based on local contour density and corner pixel ratio. Characters are composed of lines, which are complicated in order to represent a large number of characters. For this reason, we propose two feature values: the local contour density, which quantifies the complexity of the lines, and the corner pixel ratio, which quantifies the branching points and tips of the lines. Since the patterns of the features vary depending on the type of characters, we use NNs that can be expected to discriminate complexities such as nonlinearities.

2 Extraction of text area

In the proposed method, closed regions are defined as character candidate regions from the input image, and local contour density and corner pixel ratio are obtained from each character candidate region and identified by NN.

The flow of the proposed system is shown below. The flowchart is shown in Figure 1.

a) Detect the corners of the input image using Harris's corner detection method.

b) Binarize the image using Otsu's Binarization.

c) Extract contour points from the image binarized in b).

d)Extract the closed area from the contour point information obtained in c) and make it a candidate character area.

e) Calculate the local contour density and the corner pixel ratio in the character candidate region obtained in d) using the information in a) and c).

f) Use the values obtained in e) as feature values and identify them by NN.

g) Character candidate regions identified as character regions in f) are judged as characters, and the rectangle of the region is drawn as the output image.

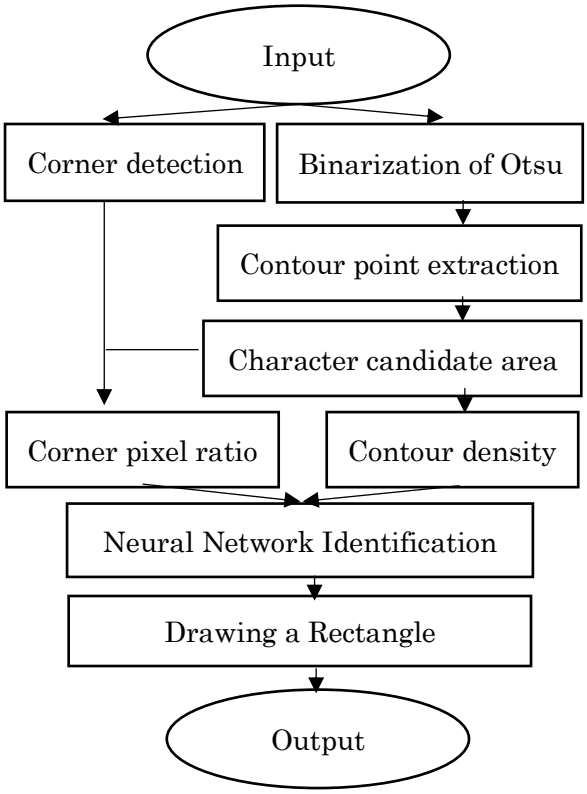

Fig. 1 Flow of the system

2.1 Local contour density

We propose a feature that quantifies the complexity of character lines, defined as the local contour density, which is the percentage of outermost contour point coordinates included in a character candidate region. It indicates how much of the candidate object's contour points are occupied in the candidate region. The formula of the local contour density is Equation (1).

$$\text{Local contour density}(\%) = \frac{Outermost\ contour\ point}{character\ candidate\ area} \times 100 \quad \cdots(1)$$

2.2 Corner density

We propose a new feature called corner density, which is defined as the ratio of corner pixels in a candidate character region, to quantify the complexity of the tips and branch points of candidate objects. It is defined as the ratio of corner pixels in the candidate region. We also used Harris's corner detection method to detect corner pixels. The formula for the corner pixel ratio is given as Equation (2).

$$corner\ pixel\ ratio(\%) = \frac{Number\ of\ corner\ pixels}{character\ candidate\ area} \times 100 \ldots(2)$$

2.3 Identification by Neural network

NN is a mathematical model of neurons in the human brain nervous system. NN consists of input layer, hidden layer and output layer, and neurons are connected to each other by weights.

2.3.1 Models of neural network

We use MLP (Multi-Layer Perceptron) as the NN, and the input, hidden, and output layers are 1566, 450, and 1566, respectively. Learning is done by the error back propagation method. The training data consists of 1566 candidate character regions including character regions and background regions in 50 scene images.

2.3.2 Hyperparameter settings

In a machine learning model, if the hyperparameters that need to be set by a human among the parameters are not set efficiently, the prediction accuracy will decrease because the performance of the model will change. Therefore, we optimized the hyperparameters by using grid search, which is a method to find the optimal values by trying all combinations of the specified parameters in a brute force fashion.

(1) Hidden layer

In general, increasing the number of layers improves the accuracy, but it is not enough to simply increase the number of layers. There is a risk of accuracy degradation due to overlearning. In addition, the cost also increases, so it is necessary to specify the optimal value. In this paper, we adopt a single hidden layer of 450. Although there was almost no change in accuracy even with multiple layers, we judged it to be inefficient because of the increase in cost.

(2) active function

It is a function that represents the activation of a neuron with respect to the sum of its inputs. In this paper, we adopt the ReLU (Rectified Liner Unit) function, which can solve the gradient vanishing problem because the derivative is always 1 since the input value never becomes negative.

$$\mathrm{f(x)} = \max(0, x) \quad \ldots(3)$$

(3) L2 regularization

To suppress overlearning, we specify the regularization term of L2 regularization, a method that introduces a regularization term into the loss function. Instead of the loss function E(ω), we use the function in Equation (4). In this paper, we set λ to 0.001.

$$\mathrm{E}(\omega) + \frac{\lambda}{2}\|\omega\|^2 \quad \ldots(4)$$

3 Experiment and Results

3.1 Experimental and evaluation conditions

In order to evaluate the proposed method, we conducted a character region extraction experiment using 452 characters in 50 scene images of 1188×890 pixels in 256 shades. The target characters are alphabets. A part of the experiment is shown in Figure 2. The area enclosed by the rectangle in the output image is the area detected as the character region. It can be said that the extraction of the character area was performed well. As an evaluation index, we define the extraction rate as the number of extracted character areas out of the total character areas, as Equation (5).

$$\text{extraction rate}(\%) = \frac{Number\ of\ correct\ answers\ in\ character\ area}{total\ number\ of\ characters} \times 100 \quad \ldots(5)$$

However, it can be said that the extraction rate alone is insufficient for evaluation.In the case where the number of background regions mis-extracted as text regions is large, i.e., all candidate regions are considered as text regions, the extraction rate is high, although it cannot be said to be highly accurate.Therefore, we defined the evaluation index of the part of the background region that was mistakenly extracted as the text region as the false extraction rate in Equation (6).

Fig. 2 Part of the experiment

Table.1 Experimental results

	Teacher	Test	Comparison method
Extraction Rate	94.3	91.9	91.3
Error Rate	1.98	2.77	10.5

Table.2 Results without "I" and "O"

	Teacher	Test
Extraction rate other than "I" and "O"	97.10	93.94
Extraction rate of "I"	44.44	72.72
Extraction rate of "O"	73.68	88.89

$$\text{Error Extraction Rate}(\%) = \frac{Number\ of\ extracted\ background\ regions}{total\ number\ of\ characters} \times 100 \quad \ldots(6)$$

The proposed method is evaluated with the aim of improving the extraction rate and suppressing the false extraction rate.

3.2 Experimental results and discussion

The experimental results are shown in Table 1. The extraction rate and false extraction rate of the proposed method are 91.9% and 2.77%, respectively. The characters that were not extracted were "I" and "O" in large numbers, and the background regions that were incorrectly extracted were mostly stick-shaped artifacts and round marks. It is thought that the similarity of the feature values led to the decrease in accuracy. Since these two characters are very similar to the outermost contour lines of the artifacts in the background region, it is thought that similar artifacts existed in the background region of the training image and were not identified well due to the fact that they were trained. Therefore, it is necessary to devise and investigate contour features that are not limited to the outermost contours. The extraction rates excluding the two characters are shown in Table 2. If we exclude these two characters, we can obtain a very high extraction rate. If the extraction rate of these two characters is improved, the overall extraction rate can be further improved.

3.3 Comparison experiment and discussion

In the comparison experiment, we compared the proposed method with the method using line density[5] .The experimental results show that the extraction rate of the proposed method is 0.6% higher than that of the comparison method. In the comparison experiment, the extraction rate of the proposed method was 0.6% higher than that of the comparison method. In addition, the false extraction rate was 7.6% higher than that of the comparison method. In addition, the false extraction rate was 7.73% lower than that of the comparison method. It can be said that we succeeded in differentiating objects other than artificial objects similar to "I" and "O".

4 Conclusions

In this paper, we propose a method for identifying text in a scene image using NN, focusing on the local contour density and corner pixel ratio. In order to cope with various surrounding conditions of the scene image, we used the features of the character candidate object itself for identification.

In the extraction experiments, the extraction rate and the false extraction rate were 91.9% and 2.77%, respectively. It can be said that the proposed method is effective in detecting text regions in scene images, which is the purpose of this study. However, one of the problems is the low recognition rate for specific characters. It is necessary to develop features using not only the outermost contour but also the contour lines, to select training images without bias, and to increase the number of target characters.

References

[1] Minoh Michihiko: Document Image Analysis,IEICE,Vol.76,No.5,pp.502~509

[2] Hori Keitaro, Sugawara Hiroki, Itoh Akiyoshi: An Unified Neural Network for Handwritten Similar Characters Recognition by Using Peripheral Local Outline Vector, Information Processing Society of Japan,Vol.40,No.12,pp.4239~4247(1999)

[3] Shiku Osamu, Anekawa Masanori, Nakamura Akira, Kuroda Hideo: A Method for Character String Extraction from Maps Using String Dictionary,IEICE(D-11),Vol.J78-D- Ⅱ ,No.12,pp.1919-1922(1995).

[4] Y. Xu and G. Nagy: "Prototype Extraction and Adaptive OCR" ,

IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.21, No.12, pp.1280-1296(1999-12)

[5] Shiku Osamu, Nakamura Akira, Kuroda Hideo:A Method for Character String Extraction Using Local and Global Segment Densities: Information Processing Society of Japan: Vol.40,No.8,pp.3230~3238(1999)

[6] Hagiwara Katsuyuki: Basics and Some Recent Advances in Neural Networks: IIEEJ,Vol.39,No.3<pp.255~263(2010)