Algorithm for Fast Detection and Identification of Characters in Gray-level Images

Zhongliang Fu Fuling Bian Songtao Zhou Qingwu Hu
Wuhan Technical University of Surveying & Mapping,
School of Informatics, Wuhan, 430079, China
Fuzl@public.wuhan.cngb.com

KEY WORDS: Algorithm, Image processing, Neural networks, Segmentation, Character identification

ABSTRACT

This paper discusses methods for character extraction based on statistical and structural features of gray levels, and proposes a dynamic local contrast accommodating line width. Precision locating of character groups is realized by exploiting horizontal projection and character arrangements of binary images along horizontal and vertical directions respectively. Also discussed is the method for segmentation of characters in binary images, which is based on projection taking into account stroke width and character sizes. A new method for character identification is explored, which is based on compound neural networks. A complex neural network consists of two sub-nets, with the first sub-net performing self-induction of patterns via 2-dimentional local-connected 3-order networks, the second sub-set linking up a locally connected BP networks performing classification. Reinforced reliability of the network recognition by introducing conditions for identification denial.
Experiments confirm that the proposed methods possess impressive robustness, rapid processing and high accuracy of identification.

1. INTRODUCTION

Automatic identification of freight car plays an important role in intelligent management of railway transportation. It provides an important evidence for attempering vehicle and freightage. At present, automatic recording system called electronic railway weighing apparatus is adopted in measuring the load of freight car. But the system can only measures gross weight of freight car, not automatically record its number and deadweight. This is operated with manual observation and record. Because freight car runs in high speed, some errors always exist. Therefore, developing automatic identification system of freight car number is valuable.
A feasible way for realization of such an automatic process is real-time acquisition of car numbers image via CCD cameras controlled by infrared sensors, followed by fast detection and identification of characters using computer. Key techniques include fast search, locating, segmentation and identification of characters from image. At last years, some researches about this subject have been done and some systems have been developed. But existing methods and systems are not robust and real-time, and identification rate is not high.
In fact, because images are acquired out of doors, different condition, for example different time, shadow, reflection, strong light in background, stain in bodywork and characters, etc, bring automatic detection and identification difficulty.
Character search includes both characters extraction and locating. Existing algorithms for character extraction have two kinds: a) static threshold method. Although its process speed is fast, but the adaptability to weather and illumination condition is weak. b) line detection method. The kind of method includes line detection operator, Hough transform, contour tracking, line following, etc. All of these operators can't detect wide line and is sensitive to noise.

Character segmentation is the base of character identification. Because a character of number code is not unattached and connective sometimes, this brings character segmentation difficulties. Although the process speed of conventional region-growth method is fast, but they only segment unattached and connective characters. Oscillator neural network (OSNN) goes on visual sense model, but it also only process unattached and connective objects, and the run-time is long.

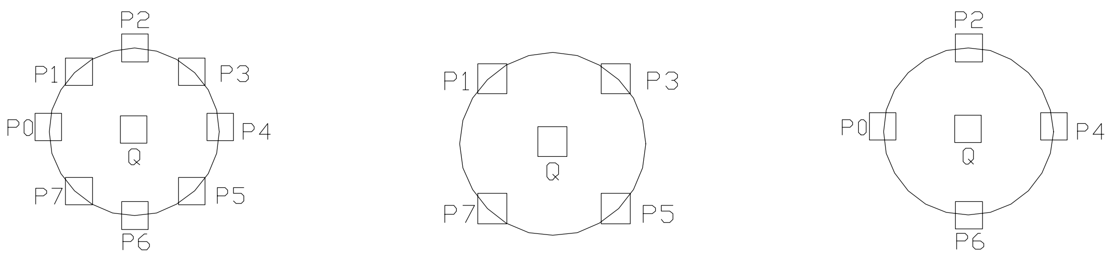
At present, character identification mostly adopts template matching method and feature statistics classing method. Template matching method can't adapt some distortion of characters (for example rotation, scale, local distortion, etc) and is badly affected by noise. Its computing quantity is great. For the feature statistic classing method, choosing feature is difficult. Classing result is affected by statistic distributing rule and noise. Structure information of characters can't be utilized.

Aiming at above problems, the paper first discusses dynamic threshold method based on statistic and structural feature of gray level, and proposes a local contrast method taking into account stroke width feature for fast extraction of character. Then, characters segmentation is realized with an improved projection method. In the end, artificial neural network (ANN) method for character identification is discussed. A new method based on a compound neural network for character identification is proposed.

2. CHARACTER EXTRACTION

The purpose of character extraction is to separate the pixels in character stroke and other pixels.

Suppose the width of character stroke is W, gray value of pixel (x,y) is $f(x,y)$, the checked pixel is Q, Point P_i (i=0,1, …,7) is in neighborhood of point Q. They distributes as Fig 1. The average gray value in neighborhood of P_i is A_i. The size of neighborhood is $(2W+1) \times (2W+1)$.



(a) (b) (c)

Fig. 1 neighborhood distribution

Suppose $N=(2W+1)\times(2W+1)$, then:

$$A_i = \sum_{i=-w}^{w}\sum_{j=-w}^{w} f(x-i,y-j)/N \tag{1}$$

If the gray value of the pixel in character stroke is great than one of other pixel, then

$$L(P_i) = \begin{cases} 1, \text{ if } f(x,y) - A_i > T_1; \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

Where, T_1 is a threshold. $f(x,y)$ may be replaced with the average gray value in neighborhood of point Q.

If the pixel Q is in character stroke can be confirmed by following formula.

$$B(x,y) = \begin{cases} 1, \text{ if } \bigvee_{i=0,1,2,3} [L(P_i) \wedge L(P_{i+4})] = 1; \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

Where $B(x,y)$ is equal to 1 means that the pixel Q belongs to character stroke and otherwise to background.

Shown from Fig. 1 and formulation (2), (3), P_1, P_3 and P_5, P_7 are used to detect vertical and horizontal stroke.

The distance of P_1 and P_3 is equal to $2W$. P_0, P_2 and P_4, P_6 is used to detect diagonal stroke.
The algorithm features dynamic of stroke width without limitation, preserving character shapes and resisting noise.

3. LOCATING CHARACTER GROUP

Due to the complexity of background and the effect of noise, no all of 1 pixels detected in binary process belong to character stroke. These pixels must be separated with the pixels in character. The role of character location is to determine the location and range of characters. The algorithm is as follows:

1) Scan image row-by-row from left to right, and from up to down, record the numbers N_1 of 1 pixel in each row, the numbers N_2 of 0 pixel to 1 pixel and the numbers of 1 pixels appearing continuously N_3.
For ith row,

$$N_1 = \sum_{j=0}^{N} B(i,j), \qquad N_2 = \sum_{j=0}^{N} B(i,j)\cdot(1-B(i,j)), \qquad N_3 = {N_1}/{N_2} \tag{4}$$

Where, N is the width of image.

2) If following inequations yield,

$$\begin{aligned} &LC * W < N_1 < 4 * LC * W \\ &LC < N_2 < 3 * LC \\ &W/2 < N_3 < 2 * W \end{aligned} \tag{5}$$

then character pixels may exist in ith row. LC is the length of character group. For the number code of freight car, LC=7.

3) If inequations (5) yields in continuous five rows starting from ith row, then row number i is recorded, $r(k)=i$.

4) Set $R_1 = \min\,(r(k))$,

$$R_2 = \max\,(r(k)), \quad k = 1,2,..... \tag{6}$$

Where, R_1, R_2 are minimum and maximum number of row where character pixels are in.
Similarly, characters between R_1 and R_2 are projected in horizontal direction. With the project value, the range of characters in vertical direction can be determined.
For the jth column, suppose project value is N_4.

$$N_4 = \sum_{i=R_1}^{R_2} B(i,j) \tag{7}$$

If the following inequation is satisfied in continuous five column,

$$W < N_4 < HC \tag{8}$$

then, the column number $c(k)=j$ is recorded. Where, HC is the height of character. Finally, minimum and maximum column number is computed.

$$\begin{aligned} &C_1 = \min(c(k)), \\ &C_2 = \max(c(k)), \ \mathrm{k} = 1,2,... \end{aligned} \tag{9}$$

C_1, C_2 is the range of character group in horizontal direction.

4. CHARACTER SEGMENTATION

The role of character segmentation is to segment each character in character group. It may utilize information as the space between characters, character height and width. The algorithm is as follows:

1) The character group image is projected in vertical direction, have $G(i,j) = \sum B(i,j)$.

2) If $G(i) < T_2$, then set $G(i)=0$. Character group is segmented into n sub-blocks in $G(i)=0$.

3) From left to right, check the width D of each sub-block. If D is approximate to the width of character, then the sub-block is corresponding to a character. If D is approximate to the half width of character, then check next sub-block. If the width of next sub-block is approximate to the width of character, then current sub-block is corresponding to a character. If the width of next sub-block is half the width of character, then the current sub-block is merge into a character with next sub-block.

The algorithm can process linked character and the character composed of multi-sub-blocks.

5. CHARACTER IDENTIFICATION

The method for character identification must possess the ability to adapt shift, rotation, scaling of character. The ANN scheme implementing the invariance to shift, rotation and scaling have:

1) The invariant feature of extracted pattern is used as network input.

2) The transformed pattern that is shift-, rotation- and scale- invariant is used as network input.

3) Constructing a network model that is shift- and rotation- invariant.

These schemes have respective merit and shortcoming. But the third scheme can more incarnate the mechanism of human brain identifying pattern and is easy to be realized in computer. Usually adopted network model is full connective three-order NN. But the network model has some obvious limitation. a) It doesn't utilize the structure information of pattern. b) It can only process regular distortion. The tolerance to errors that are brought by the complex distortion of pattern and noise is low. Its rule realizing shift- and rotation- invariance is that if pattern is shifted and rotated, the shape of triangle structured with the random three points in this pattern is steady. In fact, when irregular distortion of pattern and noise exists, the rule isn't hold. The research in physiology shows that human brain cell isn't full connected. ANN used for simulating human brain must possess the characteristic of human brain.

By researching the identification method of handwritten character with BP network, Keiji Yamade proves that locally connected BP network can resist a noise and the distortion of pattern.

In the paper, 3D neural network introduces 2D local connection and composes of a compound NN and BP network for character identification.

A compound network composes of three sub-nets as Fig. 2. First sub-net is a 2D locally connected three-order network. It realizes self-associations of pattern. Second sub-net is a classing network. It finishes classing patterns.

The model and learning algorithm of two sub-nets is as follows.

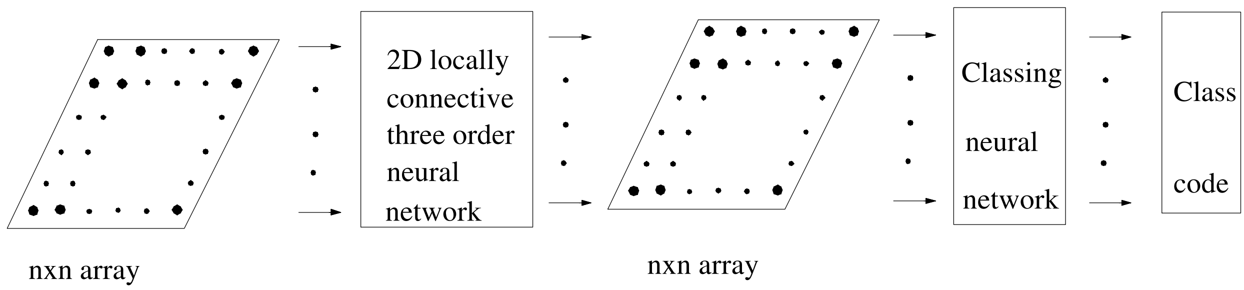


Fig. 2 A compound neural network architecture

5.1 2D Locally Connected Three-order Network

2D locally connective three-order network is substantively a monolayer feedback network. Neurone is ordered by $n \times n$ 2D array. Every neurone is connected with adjacent neurone by weight. As Fig 3.

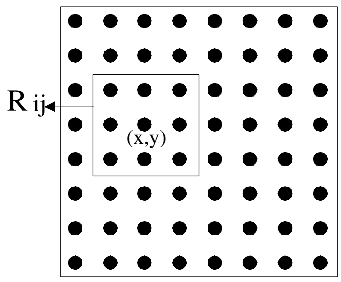


Fig.3a 2D locally connection

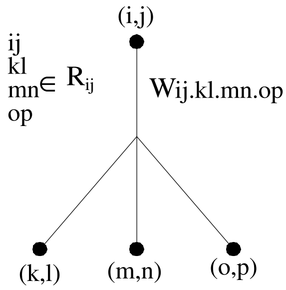


Fig.3b 2D locally connective three order neural network

Suppose the input of neurone (i,j) is X_{ij}, the output is Y_{ij}, then:

$$y_{ij} = f[\sum_{\substack{kl \\ mn \in R_{ij} \\ op}} W_{ij.kl.mn.op} \cdot X_{kl} \cdot X_{mn} \cdot X_{op}] \tag{10}$$

Where, R_{ij} is a neighborhood $n_1 \times n_2$ of neurone (i,j), usually the size is taken as 5×5. $f[.]$ is the output function of neurone. Here $f[x]=sgn(x)$.

$W_{ij.kl.mn.op}$ $(kl,mn,op \in R_{ij})$ is the self-adaptive weight of network. Suppose the weight is relative with the distance of relative neurones, then equal weight class is constructed as follows:

$$\{W_{ij.kl.mn.op} = W_{ij.d_1.d_2.d_3.d_4} \,|\, d_1 = m-k, d_2 = n-l, d_3 = o-k, d_4 = p-l, d_1, d_2, d_3, d_4 > 0\}$$

$W_{ij.kl.mn.op}$ may be found with correct error learning algorithm. Namely:

$$W_{ijd_1d_2d_3d_4} = W_{ijd_1d_2d_3d_4} + \eta \cdot (T_{ij} - y_{ij}) \cdot (\sum_{kl \in Rij} X_{kl} \cdot X_{k+d_1, l+d_2} \cdot X_{k+d_3, l+d_4}) \tag{11}$$

In the equation (11), T_{ij}, y_{ij} is respectively desired and actual output of the neurone (i,j), η is learning rate.
In equation (11), the dimension of weight $W_{ijd_1d_2d_3d_4}$ increases than ones in 1-D connection. But because R_{ij} is a small neighborhood, d_1, d_2, d_3, d_4 which only change in R_{ij} is very small. So the size of network and computing quantity don't remarkably increase.

5.2 Classing Network

Substantively, classing network is locally connected BP network. It implements different associative memory. First layer of network is an input layer. Its input is the output of first sub-net. Second layer is a hidden layer. It locally connects with the first layer. Third layer is an output layer. The connective mode between output layer and hidden layer is full. Every neurone in output layer is corresponding to a preconcerted class code of pattern.
If the output value of ith neurone in mth layer, then

$$y_i^m = f[\sum_j W_{ij}^m y_i^{m-1} + \theta_i^m] \tag{12}$$

In above equation, y_j^{m-1} is the output value of jth neurone in m-1th layer. W^m_{ij} is the connective weight between y^m_i and y^{m-1}_j, θ_i^m is bias, $f[.]$ is a Sigmoid function. It yields following equation:

$$f(x) = \frac{1}{1+e^{-x}} \tag{13}$$

Weight W^{m}_{ij} may be computed with δ-rule learning algorithm:

$$W_{ij}^{m} = W_{ij}^{m} + \eta \cdot \delta_i^{m} \cdot y_j^{m-1} \tag{14}$$

η in the equation is study speed. Commonly -0.01< η <0.3, δ_i^{m} is an epoch error.
For output layer,

$$\delta_i^{m} = y_i^{m} \cdot (1 - y_i^{m}) \cdot (T_i - y_i^{m}) \tag{15}$$

T_i, y^{m}_{i} is respectively the desired and actual output of ith neurone.
For hidden layer,

$$\delta_i^{m} = y_i^{m} \cdot (1 - y_i^{m}) \cdot \sum_j \delta_j^{m} W_{ij}^{m} \tag{16}$$

To avoid that correction value waves, an addition value is added to each correction value.

$$\Delta W_{ij}^{m}(n+1) = \Delta W_{ij}^{m}(n+1) + \alpha \cdot \Delta W_{ij}^{m}(n) \tag{17}$$

n in the equation (17) is iterative numbers. α is a positive impulse coefficient, α =0.9.
The equation (16) shows, when y^{m}_{i}=1 or 0, even if $T_I \neq y^{m}_{i}$, δ_i^{m} =0 make ΔW_{ij}^{m} equal to 0. For avoiding the case, when y_i=0 or 1, set y_i=0.1 or 0.9.
For improving the identification reliability, Reject identification is introduced. The rule is:

a. All of network output value is less than V_1=0.75;
b. Hypo-maximum output value is great than a threshold V_2=0.4;
c. The difference between maximum and hypo-maximum is less than threshold V_3=0.35.

Output result of first sub-net in a compound network is less different with ideal pattern. But this doesn't affect right identification to pattern. It is for second sub-net may also tolerate error.

6. EXPERIMENT AND CONCLUSION

To make sure the validity and reliability of all of above algorithms, 28 images are acquired and processed. Fig. 4 shows an original image and its process results.

Fig.4a Original image

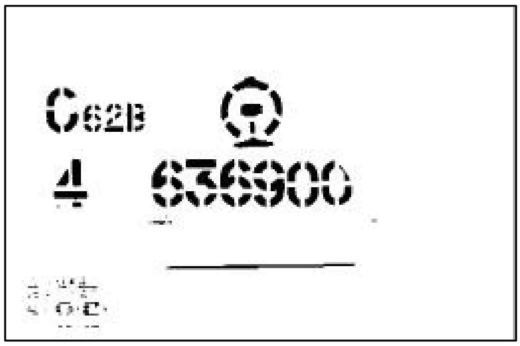


Fig.4b Binary image

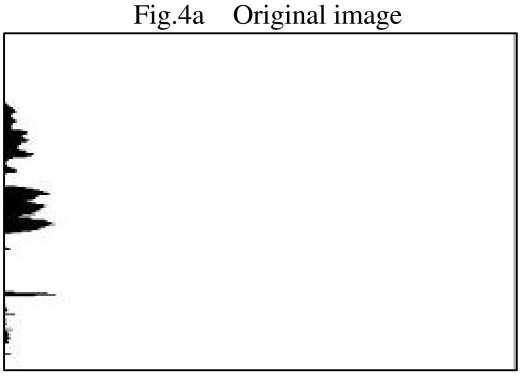


Fig.4c Horizontal project

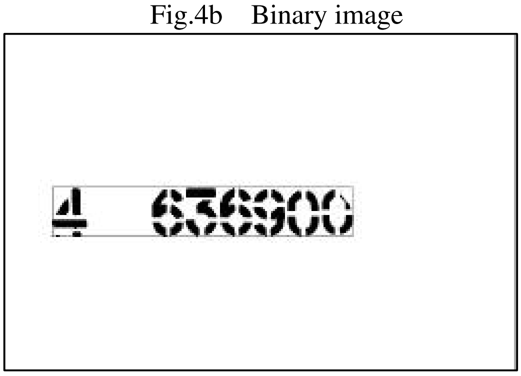


Fig 4d Locating character group

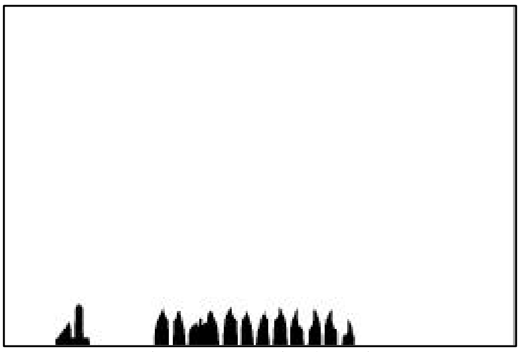


Fig.4e Vertical project

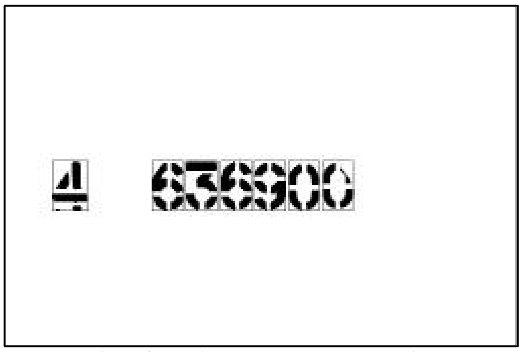


Fig.4f Character segmentation

Experiment results shows that all characters in 24 images is accurately identified, respective only one character in two images is falsely identified because these images are partly sheltered, and extraction of characters in other two images is difficult because reflex is too strong. Right identification rate reaches 92.8%. If the number of images increases, the rate will be improved. By combining automatic identification with manual back-check and modification, identification rate may reach 100% to satisfy the demand in applications. For every image, the time spent is less than 0.2 second.

REFERENCES

1. Andreas Uert and Jung Yang, Extraction of Text and Symbol from Unstructured Vector Data, Hannover, 1987,3
2. Keiji Yamada, Handwritten Numeral Recognition by Multi-layered Neural Network with Improved Learning Algorithm. International Joint Conference on Neural Networks, Washington, D.C. Jun,18-22,1980
3. M. Karnel and A. Zhao, Extraction of Binary Character/Graphics Images from Grayscale Document Images. Graphic Models and Image Processing, 55, 1993, 203-217