AUTOMATIC RECOGNITION OF FREIGHT CAR NUMBER

FU Zhongliang
BIAN Fuling
ZOU Songtao
HU Qingwu

KEY WORDS  image processing; neural networks; segmentation; character identification

ABSTRACT  This paper discusses methods for character extraction on the basis of statistical and structural features of gray-level images, and proposes a dynamic local contrast threshold method accommodating to line width. Precise locating of character string is realized by exploiting horizontal projection and character arrangements of binary images in horizontal and vertical directions respectively. Also discussed is the method for segmentation of characters in binary images, which is based on projection taking stroke width and character sizes into account. A new method for character identification is explored, which is based on compound neural networks. A complex neural network consists of two sub-nets, the first sub-net performs self-association of patterns via 2-dimensional local-connected 3-order networks, the second sub-net, linking with a locally connected BP networks, performs classification. The reliability of the network recognition is reinforced by introducing conditions for identification denial. Experiments confirm that the proposed methods possess the advantages of 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. However the system can only measure gross weight of freight car, not automatically record its number and deadweight which is operated with manual observation and record. Because freight car runs in high speed, some kinds of 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 image of all cars 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. Recently, some researches about this subject have been done and some systems have been developed. But existent 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 conditions, for example, time, shadow, reflection, strong light in background, stain in bodywork and characters, etc, bring difficulty to automatic detection and identification.

Character search includes both characters extraction and locating. There are two kinds of existent algorithms for character extraction. One is static threshold method, although its processing speed is
fast, the adaptability to weather and illumination condition is weak. The other is line detection method, which includes line detection operator, Hough transform, contour tracking, line following, etc. All of these operators can not detect wide line and are sensitive to noise.

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

At present, for character identification template matching method and feature statistics classing method are mostly adopted. Template matching method can not adapt some distortion of characters (for example, rotation, scale, local distortion, etc) and is badly affected by noise and the computation load is heavy. 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 not be utilized.

Aiming at above problems, this paper first discusses dynamic threshold method based on statistic and structural feature of gray-level images, and proposes a local contrast method taking into account stroke width feature for fast extraction of character. Then, characters segmentation is realized by 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 from other pixels. Suppose the width of character stroke is \( W \), gray value of pixel \((x, y)\) is \( f(x, y)\), the checked pixel is \( Q \), pixel \( P_i(i = 0, 1, \cdots, 7)\) is in neighborhood of pixel \( Q \), their distribution is shown in Fig. 1. The average gray value in neighborhood of \( P_i \) is \( A_i \).

\[
A_i = \frac{\sum_{i=-w}^{w} \sum_{j=-w}^{w} f(x - i, y - j)}{N} \quad (1)
\]

If the gray value of the pixel in character stroke is larger than other pixels, then:

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

where \( T_1 \) is a threshold. \( f(x, y) \) may be replaced by the average gray value of pixels in neighborhood of pixel \( Q \). The pixel \( Q \) in character stroke can be confirmed by the following formula.

\[
B(x, y) = \begin{cases} 1, & \text{if } \forall_{i=0,1,2,3} [L(P_i) \land L(P_{i+4})] = 1; \\ 0, & \text{otherwise} \end{cases} \quad (3)
\]

where, \( B(x, y) \) is equal to 1, which means that the pixel \( Q \) belongs to character stroke and otherwise to background. As shown in Fig. 1, Eqs. (2) and (3), \( P_1, P_3 \) and \( P_5, P_7 \) are used to detect vertical and horizontal stroke. The distance between \( P_1 \) and \( P_3 \) is equal to \( 2W \). \( P_0, P_2 \) and \( P_4, P_6 \) are used to detect diagonal stroke.
The algorithm features dynamic of stroke width without limitation, preserving character shapes and resisting noise.

3 Locating character string

Due to the complexity of background and the effect of noise, all the 1-pixels detected in binarizing do not belong to character stroke. These pixels must be separated from 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 times \( N_2 \) of 0-pixel to 1-pixel and the average numbers of 1-pixel appearing continuously \( N_3 \).

For \( i \)th row,

\[
N_1 = \sum_{j=1}^{N} B(i,j)
\]

\[
N_2 = \sum_{j=1}^{N-1} (1 - B(i,j)) \cdot B(i,j+1)
\]

\[
N_3 = N_1/N_2
\]

where \( N \) is the width of image.

2) If the following inequalities yield,

\[
\frac{1}{4} L C \times W < N_1 < 4 \times \frac{1}{4} L C \times W
\]

\[
L C < N_2 < 3 \times L C
\]

\[
W/2 < N_3 < 2 \times W
\]

then character pixels may exist in \( i \)th row. \( L C \) is the length of character string. For the number code of freight car, \( L C = 7 \).

3) If inequalities (5) yield in continuous five rows starting from \( i \)th row, then row number \( i \) is recorded, \( r(k) = i \).

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

\[
R_2 = \max(r(k)), k = 1, 2, \ldots
\]

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 \( j \)th column, suppose project value is \( N_4 \).

\[
N_4 = \sum_{i=I_{R_1}}^{R_1} B(i,j)
\]

If the following inequality is satisfied in continuous five column

\[
W < N_4 < H C
\]

then, the column number \( c(k) = j \) is recorded. Where, \( H C \) is the height of character. Finally, minimum and maximum column numbers are computed.

\[
C_1 = \min(c(k)), k = 1, 2, \ldots
\]

\[
C_2 = \max(c(k)), k = 1, 2, \ldots
\]

\( C_1 \) and \( C_2 \) describe the range of character string in horizontal direction.

4 Character segmentation

The role of character segmentation is to segment each character in character string. Information such as the space between characters, character height and width may be utilized. The algorithm is described as follows:

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

2) If \( G(i) < T_2 \), then set \( G(i) = 0 \). Character string 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 a character, then the sub-block corresponds 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 corresponds to a character. If the width of next sub-block is half the width of character, then the current sub-block is merged into a character with next sub-block.

The algorithm can process attached characters and characters composed of multi-sub-blocks.

5 Character identification

The method for character identification must possess the ability to shift, rotation, scaling of character. There are three kinds of ANN schemes as follows:

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

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

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

These schemes have their respective merit and shortcoming, but the third scheme can better incarnate the mechanism identifying pattern of human brain and is easy to be realized in computer. Usually adopted network model is fully connective three-order NN. However the network model has some obvious limitations: 1) It does not utilize the structure information of pattern. 2) 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 random three points in this pattern is steady. In fact, when irregular distortion of pattern and noise exists, the rule is not hold. The research in physiology shows that human brain cell is not fully 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 Yamada[2] proved that locally connected BP network can resist a noise and the distortion of pattern. In this paper, a 2D local connection is introduced into the three-order neural network which comprises a compound NN and BP network for character identification as well. A compound network is composed of three sub-nets as shown in Fig. 2. First sub-net is a 2D locally connected three-order network which realizes self-associations of pattern. Second sub-net is a classing network which implements classing patterns.

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

Fig. 2 A compound neural network architecture

5.1 2D locally connected three-order network

2D locally connected three-order network is a monolayer feedback network in substance. Neurone is ordered by $n \times n$ 2D array. Every neurone is connected with adjacent neurone by weight. (see Fig. 3 and Fig. 4).

Suppose that the input of neurone $(i, j)$ is $x_{ij}$, the output is $y_{ij}$, then:

$$y_{ij} = f\left[ \sum_{kl, mn, op} W_{ij, kl, mn, op} \cdot x_{kl} \cdot x_{mn} \cdot x_{op} \right]$$

$$kl, mn, op \in R_{ij}$$

(10)
where $R_{ij}$ is a neighborhood $n_1 \times n_2$ of neurone $(i,j)$, usually the size is taken as $5 \times 5$. The $f[.]$ is the output function of neurone. Here $f[x] = \text{sgn}(x)$. $W_{ij, kl, mn, op} \in R_{ij}$ is the self-adaptive weight of network. Suppose the weight is relative to the distance of relative neurones, then equal weight class is constructed as follows:

$$W_{ij, kl, mn, op} = \begin{cases} 
1 & \text{if } d_1 = m - k, d_2 = n - l, d_3 = o - k, d_4 = p - l, d_1, d_2, d_3, d_4 > 0 \\
0 & \text{otherwise}
\end{cases}$$

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

$$W_{ij, dl, dz, d3, d4} = W_{ij, dl, dz, d3, d4} + \eta \cdot (T_i - y_{ij}) \cdot \left( \sum_{kl \in R_{ij}} x_{kl} \cdot x_{k+l, d1} \cdot x_{k+dz, l+d2} \cdot x_{k+d3, l+d4} \right)$$

(11)

In Eq. (11), $T_i$, $y_{ij}$ is desired and actual output of the neurone $(i,j)$ respectively, $\eta$ is learning rate, and the dimension of weight $W_{ij, dl, dz, d3, d4}$ becomes more than that in 1D connection. However because $R_{ij}$ is a small neighborhood, $d_1, d_2, d_3, d_4$ which change only in $R_{ij}$ is very small. So the size of network and calculating load do not remarkably increase.

5.2 Classing network

In substance, classing network is locally connected with BP network, it implements heter-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 which 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 corresponds to a preconcerted class code of pattern.

If $y_{ij}^m$ is the output value of $i$th neurone in $m$th layer, then

$$y_{ij}^m = f[ \sum_{kl \in R_{ij}} W_{ij, kl, mn, op} y_{ij}^{m-1} + \theta_{ij}^m ]$$

(12)

In above equation, $y_{ij}^{m-1}$ is the output value of $j$th neurone in $m - 1$ layer. $W_{ij}^m$ is the connective weight between $y_{ij}^m$ and $y_{ij}^{m-1}$, $\theta_{ij}^m$ is bias, $f[.]$ is a Sigmoid function, hence we have the following equation:

$$f(x) = 1/(1 + e^{-x})$$

(13)

Weight $W_{ij}^m$ may be computed with $\delta$-rule learning algorithm:

$$W_{ij}^m = W_{ij}^m + \eta \cdot \delta_{ij}^m \cdot y_{ij}^{m-1}$$

(14)

where $\eta$ is study speed. Commonly $-0.01 < \eta < 0.3$. $\delta_{ij}^m$ is an epoch error.

For output layer,

$$\delta_{ij}^m = y_{ij}^m \cdot (1 - y_{ij}^m) \cdot (T_i - y_{ij}^m)$$

(15)

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

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

(16)

$n$ in the Eq. (17) is iterative numbers. $\alpha$ is a positive impulse coefficient, $\alpha = 0.9$.

Eq. (16) shows, when $y_{ij} = 1$ or 0, even if $T_i \neq y_{ij}^m$, $\delta_{ij} = 0$. $\delta_{ij} = 0$ makes $\Delta W_{ij}^m$ equal to 0. For avoiding this case, when $y_i = 0$ or 1, set $y_{ij} = 0.1$ or 0.9.

For improving the identification reliability, a reject identification is introduced. The rule is:

1) All of network output value is less than $V_1 = 0.75$;
2) Hypo-maximum output value is greater than a threshold $V_2 = 0.4$;
3) The difference between maximum and hypo-maximum is less than a threshold $V_3 = 0.35$.

Output result of first sub-net in a compound network is less different from ideal pattern, but this does not affect right identification of pattern, for second sub-net may also tolerate error.

6 Experiment and conclusion

To confirm the validity and reliability of all the above algorithms, standard templates for 0~9 are generated, and 28 images are acquired and processed. Fig. 5 shows an original image and its processed result.

Experimental results show that all characters in 24 images are accurately identified, 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 identifi-
cation with manual back-check and modification, identification rate may reach 100% and satisfy the demand in applications. The time spent in processing each image is less than 0.2 second.

References

1 Keiji Yamada. Handwritten numeral recognition by multi-layered neural network with improved learning algorithm. International Joint Conference on Neural Networks, Washington, D.C., 1980

2 Karnel M, Zhao A. Extraction of binary character/graphics images from grayscale document images. Graphic Models and Image Processing, 1993, 55, 203-217