Robust Character Recognition of Gray-Scaled Images with Graphical Designs and Noise

Minako Sawaki* Norihiro Hagita** Kenichiro Ishii*
NTT Basic Research Laboratories* NTT Communication Science Laboratories**
3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-01, Japan

Abstract

This paper proposes a method for recognizing characters in gray-scale images with graphical designs and noise. Our previous method for binary images is robust against graphical designs, however, it is easily affected by binarization conditions or scanning conditions. Therefore, we improve the method to be tolerant for such conditions by handling gray-scaled images directly. First, a projection profile for text-line extraction and a similarity value for character recognition are expanded for gray-scaled images. Next, learning method of adaptive threshold values against the degree of degradation and variations of intensity levels are introduced to suppress spurious candidates during recognition process. Experimental results for fifty Japanese newspaper headlines show that the proposed method achieves higher recognition rates than our previous method in the wide range of scanning brightness conditions.

1. Introduction

When constructing a large database like a digital library, automatic indexing or structuring methods are needed to handle a lot of incoming data speedily. Characters in data often include important information for indexing or structuring them. Especially characters with graphical designs often indicate important data. However, recognition of such characters suffers from graphical designs.

We proposed a character recognition method for binary images which is robust against heavy noise or graphical designs [1]. This method consists of three technologies: the discriminant function called the complementary similarity measure, the projection value for text-line extraction and an adaptive thresholding for degradation. However, this method is affected by binarization conditions.

In this paper, we extend our previous method by processing gray-scaled images directly in order to make it robust against binarization conditions. To recognize gray-scaled characters directly, several conventional methods are reported [2-4]. They extract character features directly from gray-scale images. These methods work well for images without noise, or with expected noise; however, they are not powerful enough to deal with very noisy images.

In the proposed method, the projection value and the discriminant function called the complementary similarity measure are expanded for gray-scale images using the Pearson's correlation coefficient. When recognizing a gray-scale image, text-line regions in the image are extracted using the proposed projection value. Individual characters in the text-line region are recognized by displacement matching using the complementary similarity measure for gray-scale images. When the maximum similarity value exceeds a threshold, the category of the reference pattern is determined as a recognized category. Adaptive thresholding suppresses spurious candidates yielded by displacement matching, even with graphical designs and different intensity levels to a reference pattern.

We introduce the complementary similarity measure and a projection measure for gray-scale images in Section 2 and 3. Section 4 describes our method of recognizing individual characters by displacement matching and learning adaptive threshold values. Section 5 presents our experimental results. Section 6 concludes the story.

2. Complementary similarity measure for gray-scale images

In this chapter, we describe the expansion of the complementary similarity measure for gray-scale images. Before proceed, we explain the complementary similarity measure for binary images [1]. The normalized pattern ($n = N \times N$ pixels) is expressed as an n-dimensional binary feature vector. Now, let $\boldsymbol{F} = (f_1, f_2, ..., f_i, ..., f_n)$(where $f_i = 0$ or 1) be the feature vector of an input patten and $\boldsymbol{T} = (t_1, t_2, ..., t_i, ..., t_n)$(where $t_i = 0$ or 1) be the feature vector of a binary reference pattern. The complementary similarity measure S_c of $\boldsymbol{F}$ to $\boldsymbol{T}$ is defined as

0-8186-7898-4/97 $10.00

$$S_c(F, T) = \frac{a \times e - b \times c}{\sqrt{T \times (n-T)}}$$

where

$$a = \Sigma_{i=1}^{n} f_i \times t_i, \quad b = \Sigma_{i=1}^{n} (1-f_i) \times t_i, \quad (2)$$

$$c = \Sigma_{i=1}^{n} f_i \times (1-t_i), \quad e = \Sigma_{i=1}^{n} (1-f_i) \times (1-t_i),$$

$$a + e + b + c = n, \; T = \text{a}T \text{ a}. \quad (3)$$

The complementary similarity measure for binary images is a special case of the four-fold point correlation coefficient.

Therefore, we define the complementary similarity measure for gray-scale images as a special case of the Pearson's correlation coefficient, which is a general form of the four-fold point correlation coefficient. Now, let $F_g = (f_{g1}, f_{g2}, ..., f_{gi}, ..., f_{gn})$(where f_{gi} = 0.0 through 1.0) be the feature vector of an input gray-scale pattern and $T_g = (t_{g1}, t_{g2}, ..., t_{gi}, ..., t_{gn})$(where t_{gi} = 0.0 through 1.0) be the feature vector of a gray-scale reference pattern. The complementary similarity measure S_g of F_g to T_g is defined as

$$S_g(F_g, T_g) = \frac{a_g \times e_g - b_g \times c_g}{\sqrt{n \times T_{g2} - T_g^2}} = \frac{n \times a_g - F_g \times T_g}{\sqrt{n \times T_{g2} - T_g^2}} \quad (4)$$

where

$$a_g = \Sigma_{i=1}^{n} f_{gi} \times t_{gi}, \; b_g = \Sigma_{i=1}^{n} (1-f_{gi}) \times t_{gi},$$

$$c_g = \Sigma_{i=1}^{n} f_{gi} \times (1-t_{gi}), \; e_g = \Sigma_{i=1}^{n} (1-f_{gi}) \times (1-t_{gi}),$$

$$F_g = \Sigma_{i=1}^{n} f_{gi}, \; T_g = \Sigma_{i=1}^{n} t_{gi}, \; T_{g2} = \Sigma_{i=1}^{n} t_{gi}^2. \quad (5)$$

The complementary similarity measure S_g has the following characteristics against the reverse contrast pattern $F_g{}^c$ ($f_{gi}{}^c = 1.0 - f_{gi}$) of F_g.

$$S_g(F_g, T_g) = -S_g(F_g{}^c, T_g). \quad (6)$$

3. Text-line extraction from gray-scale images

We describe the text-line extraction from gray-scale noisy images. We first explain the method for binary images. For extracting text-lines from binary images, we proposed a projection feature which emphasizes the difference between text-line regions and background regions, even with graphical designs [1].

Let G_w ($g_w(u,y)$; u = 1, 2, ..., W; y = 1, 2, ..., N_y) be a rectangular window that can include at least one character on an input binary image ($N_x \times N_y$ pixels). Since graphical designs can change gradually along the horizontal or vertical axis, the text-line regions are locally estimated using the projection profile in a local rectangle window G_w, which is shifted pixel-by-pixel in the direction of the text-line. The projection value for binary images is defined below:

(1)

$$p(y) = \frac{a_p \times e_p - b_p \times c_p}{\sqrt{r_T(r-r_T)\, r_X(r-r_X)}} \quad (7)$$

where

$$a_p = \Sigma_{u=1}^{W-1} g_w(u,y)\, g_w(u+1, y)$$

$$b_p = \Sigma_{u=1}^{W-1} (1-g_w(u,y)) g_w(u+1,y),$$

$$c_p = \Sigma_{u=1}^{W-1} g_w(u,y)(1-g_w(u+1,y)),$$

$$e_p = \Sigma_{u=1}^{W-1} (1-g_w(u,y))(1-g_w(u+1,y)), \quad (8)$$

$$r_T = a_p + b_p, \; r_X = a_p + c_p, \quad (9)$$

$$a_p + b_p + c_p + e_p = r. \quad (10)$$

The projection feature for binary images is the same form as the four-fold point correlation coefficient. The generalized form of this feature is the Pearson's correlation coefficient. Therefore, we define the projection feature for gray-scale images as the Pearson's correlation coefficient.

For an input gray-scale image ($N_x \times N_y$ pixels), let G_{gw} ($g_{gw}(u,y)$; u = 1, 2, ..., W; y = 1, 2, ..., N_y) be a rectangular window that can include at least one character. As for binary images, text-line regions are locally estimated using the projection profile in a local rectangle window G_{gw}. The projection value $p_{gray}(y)$ for gray-scale images is defined below:

$$p_{gray}(y) = \frac{r_g \times a_{gp} - r_{gX} \times r_{gT}}{\sqrt{(r_g \times r_{gT2} - r_{gT}^2)(r_g \times r_{gX2} - r_{gX}^2)}} \quad (11)$$

where

$$a_{gp} = \Sigma_{u=1}^{W-1} g_g(u, y)\, g_g(u+1, y), \quad (12)$$

$$r_{gT} = \Sigma_{u=2}^{W} g_{gw}(u, y), \; r_{gX} = \Sigma_{u=1}^{W-1} g_{gw}(u, y),$$

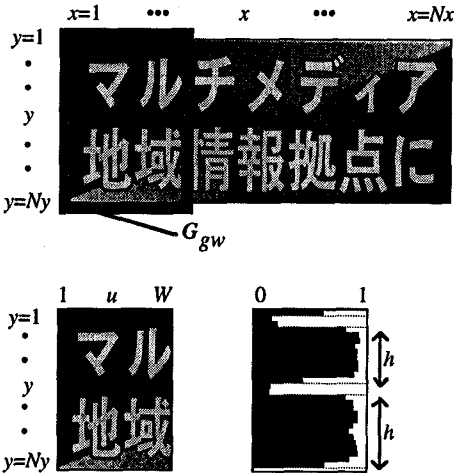
$$r_{gT2} = \Sigma_{u=2}^{W} \{g_{gw}(u, y)\}^2, \; r_{gX2} = \Sigma_{u=1}^{W-1} \{g_{gw}(u, y)\}^2,$$

$$r_g = W - 1. \quad (13)$$

To avoid notches, the projection axis is divided into N sections and projection values $p_{gray}(y)$ are averaged in each section. A range h of high projection values in the projection profile is then selected as a candidate of a text-line region at a G_{gw} location. More specifically, when the averaged projection value exceeds 30% of the maximum of the projection profile, the section is determined to be a text-line region. This threshold was defined empirically. In practice, the text-line regions fluctuate somewhat over all G_{gw} due to different character heights at each location and graphical designs. To avoid these notches, the text-line regions are averaged over all G_{gw}.

Figure 2 shows an example of a rectangular window G_{gw} and the averaged projection profile of $p_{gray}(y)$ for G_{gw} at x = 1. Two groups of high values indicated with h

in Fig. 2 (b) show the existence of two text-line regions.



(a) Observation window G_{gw} (b)Averaged projection profile of $p_{gray}(y)$

Fig. 2 Observation window for text-line region extraction and projection profile of $p_{gray}(y)$.

4. Character recognition using displacement matching

4.1 Displacement matching

Since it is difficult to select cut positions of individual characters even with $p_{gray}(y)$, displacement matching is applied for character recognition. The extracted text-line region with height h is normalized to yield a normalized text-line region L with a height of N. An N×N-pixel square window F_g is selected for matching since reference patterns consist of N×N pixels. F_g is shifted pixel-by-pixel along the direction of the text-line and is compared to the reference patterns using the absolute value of the complementary similarity measure $|S_g|$ at each position (Fig. 3).

4.2 Adaptive Thresholding

When the maximum $|S_g|$ at each position exceeds the threshold, the recognized category and its position are determined. To recognize characters with high precision, a relevant threshold should be determined according to the degradation state of F_g and the reference pattern T_{gmax} with the maximum $|S_g|$. In the case of binary image recognition, when a window F is located at the correct character position during displacement matching, $|S_c|$ of F to the reference pattern of the correct category decreases according to the degree of degradation while $|S_c|$ of F to the reference patterns of the other categories decreases more rapidly. Therefore, thresholds for F are determined using the degree of degradation and T_{max}.

Besides the degradation, when recognizing a gray-scale image, the similarity value also decreases as the intensity values of all pixels decrease. Therefore, thresholds for gray-scale images should be determined using the intensity value, the degree of degradation and T_{gmax}.

We estimate thresholds for gray-images from those for binary images. When learning thresholds for binary images, the similarity of the nearest pattern F to T_{max} is determined as a threshold of T_{max} at $\|F\|$ [1]. In the case of gray-scale images, when an input pattern F_g is not degraded, but the intensity values of all pixels are decreased by ω, the similarity of F_g decreases by ω. Therefore, thresholds should be decreased by ω.

In recognition, ω is estimated from the ratio of the standard deviation of F_g and T_{gmax}. The modified threshold table for F_g is obtained using ω. The $\|F_g\|$ of the maximum threshold is obtained as $\|F\|\times\omega$ and the maximum threshold value is obtained by multiplying ω by the maximum threshold for F. The modified thresholds for other $\|F_g\|$ are interpolated so that thresholds for $\|F_g\|$=0 or n are 0. The threshold tables for the reference pattern of "報" is shown in Fig. 4.

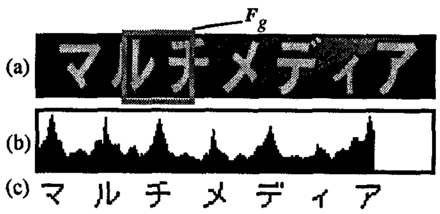


Fig. 3 Displacement Matching.
(a) normalized text-line region L
(b) maximum value of $|S_g|$
(c) recognized categories

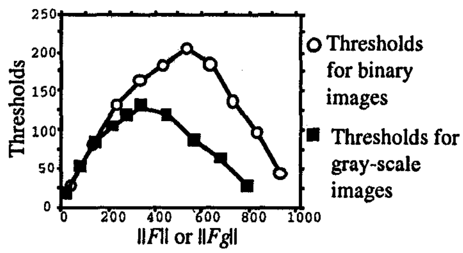


Fig. 4 Thresholds for gray-scale images. (examples for reference pattern of "報")

5. Experimental results

5.1 Headline data

We used 50 headlines with graphical designs in Japanese newspapers (25 horizontal text- and 25 vertical text-lines) including 529 characters as test data. They were gathered using three brightness levels: level A (dark), level B (middle), and level C (bright). The number of text-lines in one headline was either one or two. The character font was gothic. Binary reference patterns without graphical designs were extracted manually from 121 binarized headline images at level B. As the aspect ratio of characters differs with the direction of text-lines, the reference patterns were stored in either a horizontal or a vertical text-line dictionary according to the direction of the text-line. When recognizing characters, the horizontal (vertical) text-line dictionary was used when the width of the input headline image was larger (smaller) than its height. The number of reference patterns were 913 (500 categories) for the horizontal text-line dictionary and 988 (525 categories) for the vertical text-line dictionary. The size of the reference patterns was $n=N \times N=32 \times 32$.

5.2 Character recognition

The recognition experiments for 50 test headline images were made. For comparison, two conventional methods were applied to the test data. The first one employed our previous method [1]. The test data was binarized at 50% of the intensity level and the resulting images were recognized by [1]. With the second conventional method, the test data was binarized at 50% of the intensity level. Graphical designs in the resulting images were removed with a morphological approach [5] and then recognized with commercial OCR software. The recognition rates are shown in Table 1. Table 1 shows that the proposed method achieves recognition rates of over 79% in the three binarization conditions, while the others are affected by the conditions.

At level B, the recognition rate for method [1] is slightly higher than that for the proposed one. This is because binary reference patterns obtained from images at level B were used.

Table 1 Recognition rates(%).

Method	Brightness level		
	level A (dark)	level B (middle)	levelC (bright)
(1)proposed method	83.6 (208)	79.6 (221)	85.3 (150)
(2)binarization + our previous method	62.0	80.5	48.8
(3)binarization + a conventional method	25.0	28.4	15.7

(): number of extra candidates

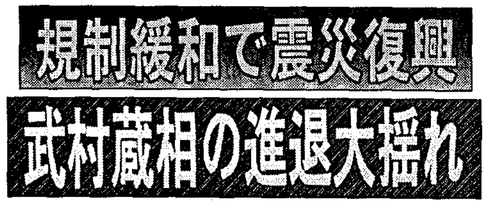


Fig. 5 Correctly recognized headlines.

6. Conclusion

In this paper, we proposed a method for recognizing characters with graphical designs directly from gray-scale images. Our previous method, which is robust for binary images with graphical designs, was expanded for the gray-scale images. We extend the complementary similarity measure for binary images and the projection value for text-line extraction for binary images to those for gray-scale images. The new similarity measure and the projection value are based on the Pearson's correlation coefficient. Also, adaptive thresholding for gray-scale images are developed for precise recognition under noisy conditions. The experimental results for Japanese newspaper headlines with graphical designs show that the method achieves higher recognition rates than our previous method in the wide range of brightness conditions.

In the near future, our method will be applied to the recognition of characters in color images.

Acknowledgments

We would like to acknowledge Dr. Kazumi Odaka for his stimulating discussions.

References

[1] M. Sawaki and N. Hagita, "Text-line Extraction and Character Recognition of Japanese Newspaper Headlines with Graphical Designs", *Proc. of ICPR*, Vienna, Austria, Aug., Vol.III Track C, pp.73 - 78, 1996.

[2] L. Wang and T. Pavlidis, "Direct Gray-Scale Extraction of Features for Character Recognition", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.15, pp.1053-1067, 1993.

[3] S-W Lee, D-J Lee and H-S Park, "A New Methodology for Gray-Scale Character Segmentation and Recognition", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.18, pp.1045-1050, 1996.

[4] D. Liu, Y. Yamashita and H. Ogawa, "Recognition of handwritten characters in the presence of noise", *IEICE Technical Report PRU*92-168, pp.95-101, March, 1993.

[5] M. Okamoto and H. Hayashi, "Character Extraction from Headlines with Background Patterns by Using Shrinking/ Expanding Methods", *IEICE Technical Report PRU*90-151, pp.47-54, 1991 (in Japanese).