Digit Classification on Signboards for Telephone Number Recognition

Takuma Yamaguchi†, Yasuaki Nakano‡, Minoru Maruyama†,
Hidetoshi Miyao† and Toshihiro Hananoi‡

†Dept. of Information Engineering, Shinshu University
4–17–1 Wakasato, Nagano–shi, Nagano, 380–8553, Japan
{zyama, maruyama, miyao}@cs.shinshu-u.ac.jp

‡Kyushu Sangyo University, Fukuoka–shi, 813–8503, Japan
{nakano, hananoi}@is.kyusan-u.ac.jp

Abstract

This paper presents a digits classification system to recognize telephone numbers written on signboards. Candidate regions of digits are extracted from an image through edge extraction, enhancement and labeling. Since the digits in the images often have skew and slant, the digits are recognized after the skew and slant correction. To correct the skew, Hough transform is used, and the slant is corrected using the method of circumscribing digits with tilted rectangles. In experiments, we tested a total of 1,332 images of signboards with 11,939 digits. We obtained a digit extraction rate of 99.2% and a correct digit recognition rate of 98.8%.

1. Introduction

There are many characters and indicators on signboards around us. Sometimes we see telephone numbers on boards in natural scenes. The phone numbers can give us some information if we have a database to check them. For example, we can get the shop or company name and address from the phone number. If the numbers are recognized, we may even find out the place where we are now.

In this paper, we propose a digit recognition system to recognize phone numbers. Characters are printed in various languages in the world, however digits usually are written in Arabic numerals. So this system can be used all over the world.

Due to the auto-focusing and zooming capabilities of recent cameras, it is possible to obtain high-quality images which are sufficient for use. In this research, we assume that the telephone numbers area can be detected and captured with good resolution. The target of the investigation is recognizing phone numbers on signboards.

Previously number recognition in natural scenes, for example license plates and road traffic sign for speed limit classification[1], have been investigated. As character recognition systems in natural scenes, Fujisawa et al. proposed a camera system called ICC (Information Capturing Camera)[2]. We also presented a system of place-name recognition in natural scenes[3]. The system extracts and recognizes place-names in natural scenes. It may be almost impossible for many foreign visitors in Japan to understand Japanese words including Kanji characters. As a result place-names on signboards in natural scenes cannot be understood and some visitors may get lost. Therefore it is useful to translate place-names into the visitor's mother tongue.

Although these systems can be very useful, they cannot be used all over the world, because the characters vary depending on countries. However telephone number usually are printed in Arabic numerals, so telephone number recognition can be used anywhere. There has been much research on number recognition in natural scene. Typical examples are recognition of license plates and speed limit signs. However, in these cases, the degree of skew and slant of digits is very limited. On the other hand, location of phone numbers are not known in natural scene images, range of inclination and deflection of digits changes is also not known and not limited. Characters in the images of signboards in natural scenes have larger skew and slant than ones in the document images scanned using scanner. Thus skew and slant correction is needed to recognize phone numbers in natural scenes. A lot of research for skew and slant correction has been presented. Chaudhuri et al. proposed skew angle correction method of Indian scripts[4]. In this method skew angle is corrected using the features of Indian scripts head line. Sun et al. proposed skew and slant correction method

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 **IEEE**

Figure 1. Sample Input Images

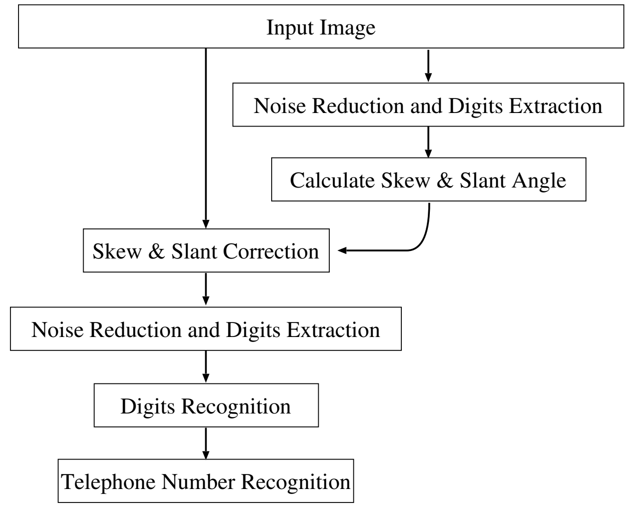

Figure 2. Flowchart of this system

using gradient direction[5]. In the research a simple and fast algorithm for determining the skew angle of an image and the slant angle of text characters only using the gradient orientation histogram was proposed.

In our research, when the skew is corrected, Hough transform is used. The skew angle is obtained from center points of the extracted candidate digits regions using Hough transform. Then to correct slant, a method of circumscribing digits with tilted rectangle is used. The rectangle is tilted to 45 degree. The slant angle is calculated using the method, and the slanted image are corrected with the angle. The details are shown later.

2. Data

Due to high-performance cameras, it's not difficult to take images of objects which high precision. Input images such as shown in Fig.1 were used. In the input images there is at least one phone number. We collected a total of 1,332 images with 11,939 digits. These input images were taken in fine and cloudy daylight hours. Each image consists of 640×480 pixels with 8 bit gray scale.

3 Overview of Our Method

A block diagram of this system is shown in Fig.2.

Since there are a lot of noise in natural scene images, noise is removed from the input image in the first stage using a labeling method, and the candidate regions of digits are extracted from the labeled image. Digits and characters in scene images often have skew and slant. To reduce them, skew and slant angles are calculated in the next phase. Then the digit regions in the image are corrected using the calculated angles. After skew and slant of input image were corrected, noise is deleted, and the candidate areas of numbers are obtained using the labeling method again. Then, the areas are classified by pattern matching based on directional feature patterns[6]. Finally, telephone numbers are recognized using the results of digits classification. The details of each method are shown later.

4. Preprocessing

In the preprocessing, we try to extract candidate regions of digits without skew and slant. Each method of preprocessing is described in this section.

4.1 Noise Reduction and Digits Extraction

Since the natural scene contains a lot of noise unlike the scanned document, noise reduction is very important. The first step is to extract the edges using a Roberts filter. The operator is not sensitive to weak edges, so the amount of noise is decreased. However when the digits have only weak edges, digits may not be extracted. Therefore the contrast of the input image should be enhanced before input images are filtered. At this point the following transformation is used:

$$OUT = IN + OFFSET \times (IN - 127.5),$$

where IN denotes the level(the intensity) of a input pixel, OUT denotes the level of a output pixel and $OFFSET$ denotes the correction value. Value range of $OFFSET$ is between 0 to 1. When the value is 0, the contrast is not enhanced. On the other hand when the value is 1, a strong contrast is obtained. In this research, we set $OFFSET = 1$,

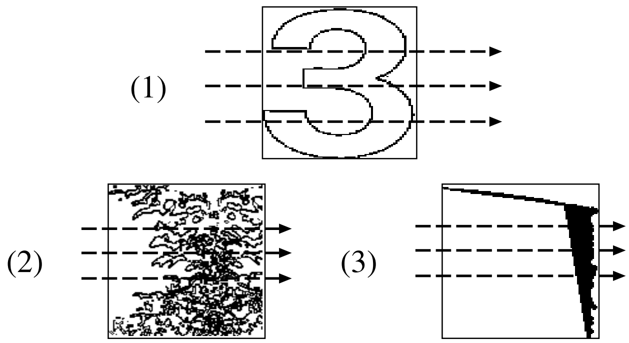

Figure 3. Noise Reduction

because edges of digits are very important. The second step is to binarize the filtered image to get the labeled image. And then both the pixels with many connected neighbors and few connected pixels are excluded from the labeled image. Next, if the areas are too large or too small or the height is 10 (or more) times the width, then the regions are removed. If the width is 1.5 (or more) times the height, the regions are also removed. This is because heights of digits are longer than width in most cases. Even after the processing, there may be noise remaining in the image. Using the property of the edges of digits, we can reduce the noise remarkably. When the scan lines go across the candidate area, the lines hit the black edges(Fig.3). If the number of edges hit by a line too many(Fig.3-(2)) or too few(Fig.3-(3)), the area is not a digit region. Finally, the remaining regions are chosen as the candidate areas of the digits.

4.2 Skew and Slant Correction

Characters obtained from natural scenes often have skew and slant. Because most of the images are not taken squarely. Skew and slant are created by the angle of pan, tilt and inclination of the camera. Skew interrupts digit string extraction, and slant disturbs numbers recognition. Therefore they should be removed. For that purpose, Hough transform and a method of circumscribing digits with tilted rectangles are used.

4.2.1 Skew and Slant Angle Calculation

When the skew was corrected, Hough transform was used to determine the angle of skew. The equation is $\rho = x\cos\theta + y\sin\theta$. First, the center of the candidate regions of the digits are calculated. Next, their centers are transformed with Hough transform. Because thetransform is only applied to the centers, the computation cost is quite low. An angle with maximum number of votes is the skew angle α (Fig.4). Therefore skew correction angle ϕ is equal to $-\alpha$

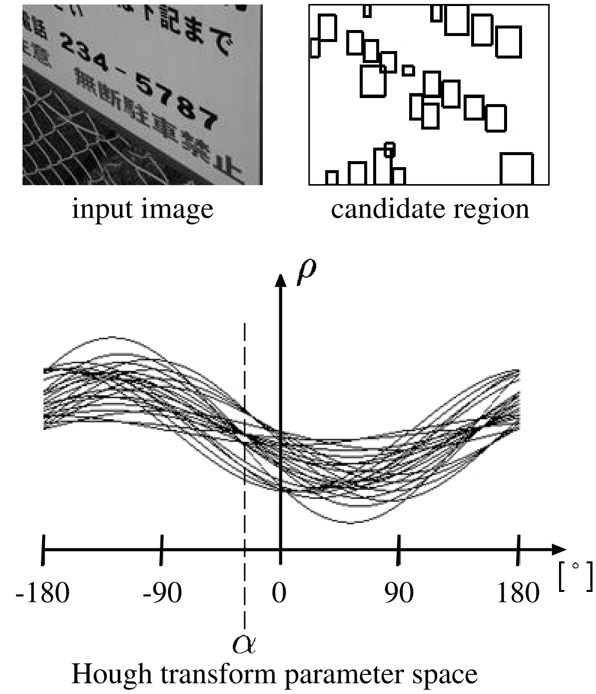

Figure 4. Skew Angle Calculation

(in this research, $-45 \le \phi, \alpha \le 45$).

After the skew correction, we extract slant angle by circumscribing digits with tilted rectangles. Using the resultant slant angle, digit regions are properly transformed so that number recognition method is applicable.

First, skew corrected numbers are circumscribed with rectangles. The rectangle is slanted to 45 degree as shown in Fig.5. In other words, candidate digits are surrounded by the following four lines:

$$y = x + b_1, \qquad y = x + b_2,$$
$$y = -x + b_3, \qquad y = -x + b_4,$$

where $b_3 > b_4$ and $b_1 > b_2$ (Fig.5-(1)).

Next, the surrounded area is transformed so that dot Q is just above dot P (Fig.5-(2)). Let P, Q be the top and bottom intersections. The correction angle (Fig.5-(3)) is given as $\theta = \arctan(B/A)$, where A denotes height of slanted character and $B = (b_2 + b_3 - b_1 - b_4)/2$.

The same operation is applied to all candidate areas. And then, the average of the θs is given as the skew angle of the image.

4.2.2 Skew and Slant Correction

The skew and slant of the input images is corrected using the equation as follows:

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 **IEEE**

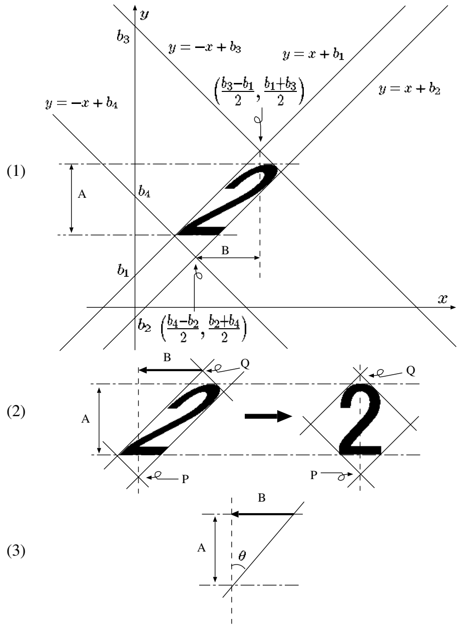

Figure 5. Slant Angle Calculation

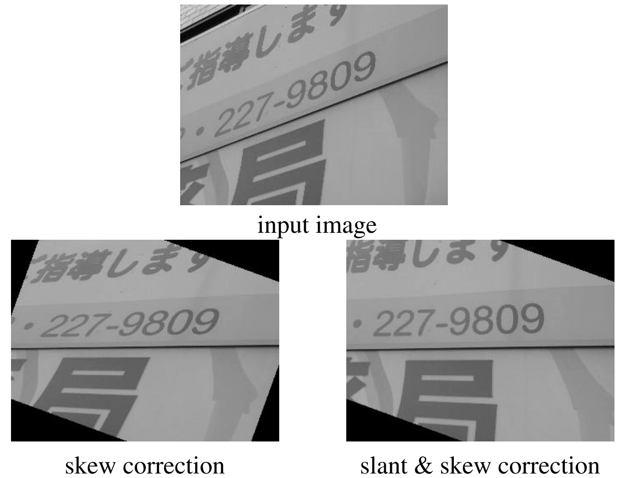

Figure 6. Results of Skew and Slant Correction

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi + \tan\theta \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} H\tan\theta \\ 0 \end{bmatrix}$$

where ϕ and θ denote the correction angle of skew and slant, and H denotes height of the input image.

The results of skew and slant correction are shown in Fig.6. In the results, we can see the skew and slant were corrected. The size of each digit is not compensated for, but that is not a problem, because digits are normalized to the same size in the recognition stage. To recognize number strings, it does not cause a problem either, because the centers of digits areas are lined up in nearly horizontal position.

Finally, the candidate areas of digits without noise, skew and slant are extracted using the noise exclusion and digits extraction method which was proposed in the section 4.1. Some regions may not be digit, but this is removed when they are recognized.

The results of the noise exclusion and digits extraction processes are shown Fig.7.

5 Digits Recognition

5.1 Features

Extracted digits are normalized to 64×64 pixels, and the normalized patterns areas are divided equally into 8 × 8 blocks. For each block, directional feature extraction method is applied, and elements which are vertical, horizontal, upwards slanting to the right and upwards slanting to the left are extracted.

5.2 Recognition Method

For the digit recognition, pattern matching based on directional patterns was used in this research.

Let $\boldsymbol{X}$ represent a feature vector of the unknown pattern and $\boldsymbol{S}_i$ be that of a template of class i.

The similarity is defined by :

$$\frac{\boldsymbol{X} \cdot \boldsymbol{S}_i}{\sqrt{\boldsymbol{X} \cdot \boldsymbol{X}}\sqrt{\boldsymbol{S}_i \cdot \boldsymbol{S}_i}},$$

where "·" denotes the inner product of two vectors.

For each pattern the similarities are calculated with all templates. The recognition result is the class with maximum similarity. For each class, we use 8 templates, which are made from 8 different fonts. That is, total number of the template pattern is 80. To make our method robust and reliable perturbed correlation method[7] is also used. In our method, the templates are moved 1 pixel to vertical and horizontal direction to find an optimum match.

6 Experimental Results and Telephone numbers Recognition

The results of the digits extraction and recognition are shown in Table.1. 11,843 digits were extracted from the input images with 11,939 digits (i.e. number extraction rate is 99.2%), and 11,695 digits were correctly recognized from

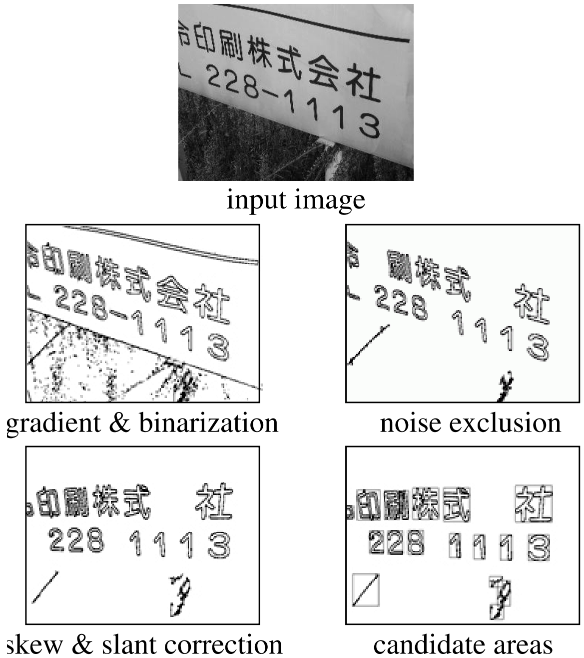

Figure 7. Results of Noise Exclusion & Digits Extraction

those extracted numbers (i.e. correct number recognition rate is 98.8%). However there are some digits recognized incorrectly. Slant and skew correction failed when there was too much noise in the candidate regions of the digits. If the appropriate angles are not found, the characters may be recognized incorrectly.

When skew and slant correction were used, the digit extraction rate was almost the same but the digit recognition rate was better as shown in Table.1. If digits have large skew and slant, some digits were recognized incorrectly without the correction.

Table 1. Results of digits recognition

skew & slant correction	without	with
digits extraction rate	99.2%	99.2%
digits recognition rate	97.4%	98.8%

There may be still some digits which aren't telephone numbers in the input images. But they can be removed when telephone numbers are recognized. Phone numbers should always consist of digit strings. An isolated number cannot be an element of telephone numbers. Furthermore the strings should line up in a horizontal position, because the skew of the image was already corrected. When there are some candidate phone numbers in a image, a telephone numbers lexicon is needed. The digit strings can be distinguished from phone number and non-phone number using the lexicon. Local digits features are also available, for example, in Nagano-City long-distance code is "026" and local office number always includes "2". In Japan 3–3–4, 4–2–4, 5–1–4 and 6–0–4 telephone numbering is also used according to the size of cities, towns and villages.

Nagano's long-distance code was "0262" until several years ago. Some samples are written so. But the problem can be solved easily.

7 Conclusions

In this paper, we propose a digit classification system to recognize telephone numbers on signboards in natural scenes. We can get some information (which is the shop or company name and address etc) using a telephone number database. Skew and slant correction method for recognizing the digits in natural scenes was also proposed in this paper. In experiments we tested total of 1,332 images of signboards with 11,939 digits, and obtained a digit extraction rate of 99.2% and a correct digit recognition rate of 98.8%

References

[1] K. I. Kim, K. Jung and J. Kim; Color Texture-based Object Detection: an Application to License Plate Localization, Proc. of SVM, pp. 293–309 (2002.)

[2] H. Fujisawa, H. Sako, Y. Okada and S.-W. Lee; Information Capturing Camera and Developmental Issues, Proc. of ICDAR, pp. 205–208 (1999.)

[3] T. Yamaguchi and Y. Nakano; Extraction of Place-name from Natural Scenes, Proc. of IWFHR, pp. 239–243 (2002.)

[4] B. B. Chaudhuri and U. Pal; Skew Angle Detection of Digitized Indian Script Documents, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 2, pp. 182–186 (1997).

[5] C. Sun and D. Si; Skew and Slant Correction for Document Images Using Gradient Direction, Proc. of ICDAR, pp. 142–146 (1997.)

[6] S. Mori, C. Y. Suen and K. Yamamoto; Historical Review of OCR Research and Development, Proc. of IEEE, 87(7), pp. 1029–1058 (1992.)

[7] M. Yasuda, K. Yamamoto and H. Yamada; Effect of the Perturbed Correlation Method for Optical Character Recognition, Pattern Recognition, Vol. 30, No.8, pp.1315-1320 (1997.)