Translation Camera

Yasuhiko Watanabe
Ryukoku University
Seta, Otsu, Shiga, Japan

Yoshihiro Okada
Ryukoku University
Seta, Otsu, Shiga, Japan

Yeun-Bae Kim
NHK Science and Technical Research Labo.
Kinuta, Setagaya-ku, Tokyo, Japan

Tetsuya Takeda
Ryukoku University
Seta, Otsu, Shiga, Japan

Abstract

In this paper, we propose a camera system which translates Japanese texts in a scene. The system is portable and consists of four components: digital camera, character image extraction process, character recognition process, and translation process. The system extracts character strings from a region which a user specifies, and translates them into English.

1   Introduction

There are many texts in various scenes: signboards, marks, indicators, and so on. Quite a lot of them give us important information: instructions, warnings, explanations, and so on. For example, characters in Figure 1 (a) give us an order, "止まれ *(tomare,* stop)". Natives have no problem to read them. However, it is not easy for foreigners to read them. Especially, it is more difficult when the characters are not their own. To solve this problem, we developed an experimental system which we call *translation camera*. The system translates texts in scene images which are taken with a digital camera. In this paper, we describe the outline of our system and a method for recognizing and translating texts in scene images.

2   Texts in Scene Images

2.1   Related Work

There are many attempts have been done on character recognition in document images [Sakai 93]. On the contrary, there are only a few researches on character recognition in scene images [Choi 90] [Takizawa 95] [Ohya 94][Ueba 96]. It is because there are the following problems on character recognition in scene images:

- it is not easy to obtain the clear contrast between character pattern strings and their background,

- character pattern strings are often disturbed by noises, and

- texts in scene images have a wide variety of the characters and the layouts.

Moreover, the aims of the previous works are generally the automatic detection and extraction of character pattern strings in scene images. However, in the mobile application, a user has no difficulty in specifying a text in a scene image which he wants to translate. As a result, the aim and approach of our research are different from those of the previous works.

2.2   Important Texts in Scene Images

Texts in scene images represent many kinds of information. Especially, the following kinds of information are important:

- instruction (e.g. "止まれ*(tomare,* stop)" in Figure 1 (a)),

- warning (e.g. "路面凍結注意 *(romen touketsu chui,* mind the frozen road)" in Figure 1 (b)),

- explanation (e.g. "三条京阪 *(Sanjo Keihan)"* in Figure 1 (c)), and

- name (e.g. "消火栓 *(shokasen,* fireplug)" in Figure 1 (d)).

Foreigners want to read texts which represent these kinds of information. As a result, in this study, we are concerned with the translation of these texts in scene images.

It should be noted that even foreigners may distinguish texts in scenes which represent these kinds of information although they do not understand what the texts represent. It is because these texts have the following features:

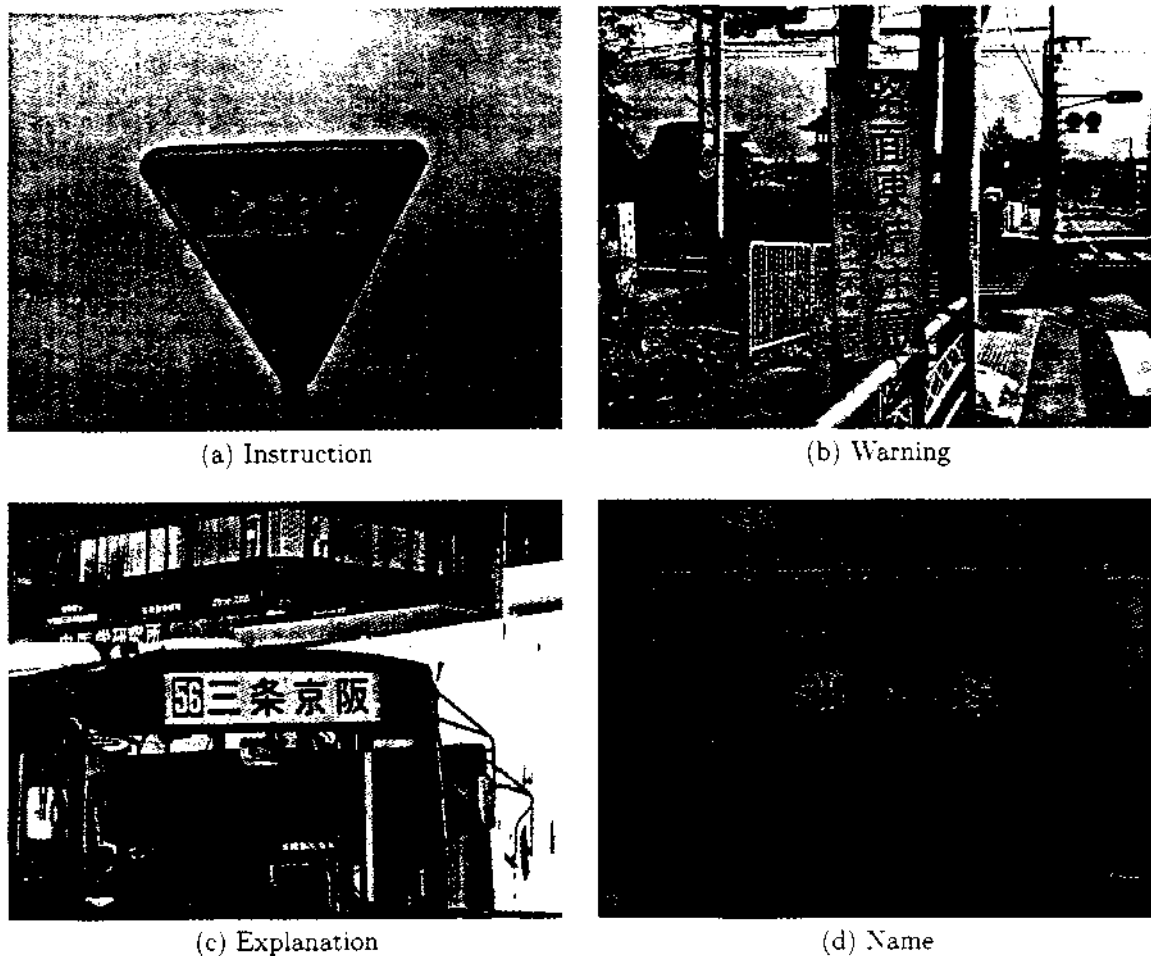(a) Instruction

(b) Warning

(c) Explanation

(d) Name

Figure 1: Texts in scene images which represent important information

- colors of character pattern strings and their background are uniform, respectively.

- characters are written in bold letters or similar letters, and

- texts are written in a straight line (horizontally or vertically).

As a result, we use these features for recognizing characters in scene images.

3 Character Recognition and Translation for Texts in Scene Images

In this chapter, we describe the processes of character recognition and translation for texts in scene images. Figure 2 shows an output of our system. The system shows the result of character recognition and translation when a user specifies a text region and push the button "translate". As shown in Figure 3, the system works in this way:

1. a user specifies a text region,

2. the system extracts character pattern strings from the specified region in the binarization process,

3. the system recognizes the extracted character pattern strings with OCR, and

4. the system translate the results of the character recognition process.

3.1 Manual Operation for Specifying Text Region

As mentioned previously, there are some attempts have been made at the automatic detection of character pattern regions in scene images. However, in our system, the character pattern regions are specified not automatically but by manual operation. The reasons are as follows:

- If an objective text region is not specified explicitly, the system may translate not only the
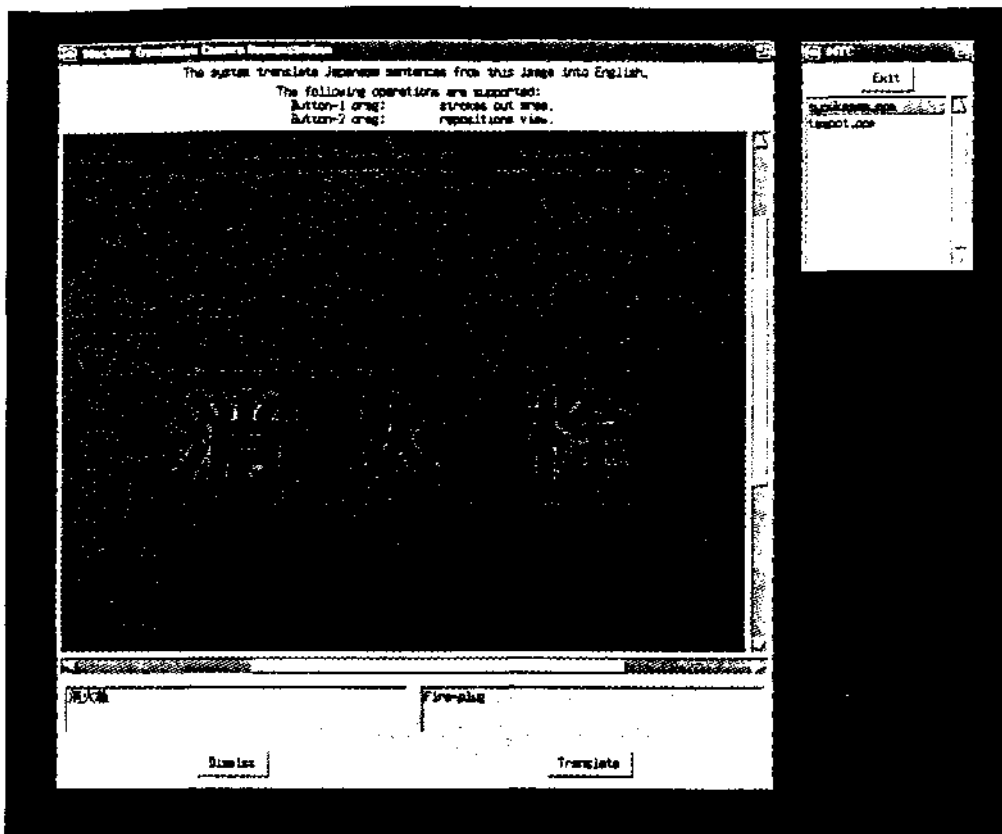
Figure 2: An example translation of a text in a scene image

objective text but also unnecessary texts. For example. Figure 1 (c) contains not only a bus indicator but also many other signboards.

- Texts in scene images should be translated as soon as possible. It is because important texts should be understood, on the spot and at once. But. automatic detection of character regions generally needs much time.

- interactive processing is available in mobile computing.

3.2 Character Pattern String Extraction

As mentioned in Section 2.2. character pattern strings which give us important information have the feature: the colors of the character pattern strings and their background are generally uniform color, respectively. In spite of this, it is difficult to recognize these character pattern strings in scene images. The reasons are as follows:

- a striking contrast between the character pattern strings and their background can not be obtained when the lighting is not uniform.

- character pattern strings are often disturbed by noises.

To solve these problems, we use binarization process [Otsu 79] for the character recognition. [Otsu 79. gives us a good threshold for the binarization of a gray image. From this, we extract the character pattern strings from the specified region in this way:

1. the system converts the color image of the specified region into the gray image.

2. the system obtains the threshold for the biriarization using 'Otsu 79 method.

3. the system converts the gray image of the specified region into the binary image using the obtained threshold.

4. the system extracts the character pattern strings from the binary image.

Figure 4 shows an result of binarization using the threshold which is obtained by using [Otsu 79] method.

3.3 Character Recognition with OCR

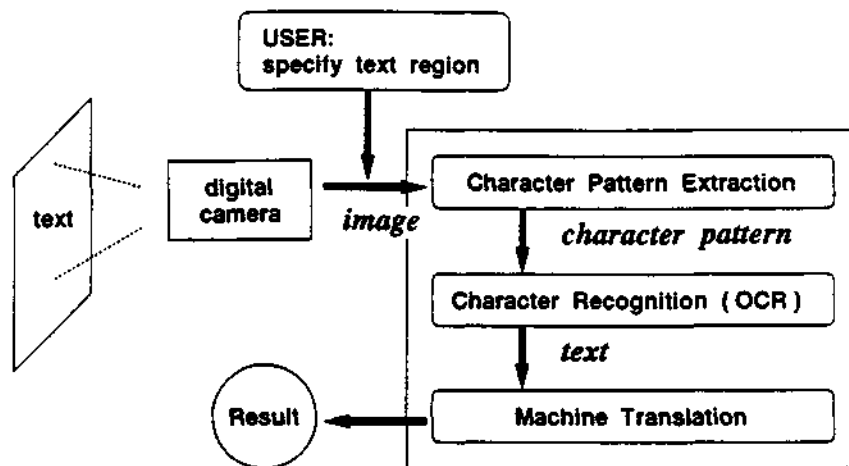As mentioned in Section 2.2. important texts in scene images have these features:

Figure 3: Overview of translation camera

(a) input image (gray image)   (b) output image (binary image)

Figure 4: A result of the binarization using (Otsu 79) method

- texts are written in bold letters or similar letters, and

- texts are written in a straight line (horizontally or vertically).

As a result, there is no need to give consideration to these problems:

- wide variety of characters (letter forms), and

- wide variety of layouts

Consequently, OCR module in our system recognizes character pattern strings just as they are extracted in the binarization process. We built OCR module by using OCR library [RICOH 95].

3.4 Machine Translation

MT module translate the results of the character recognition process. We used MT system "PENSÉE" [Oki 98] as MT module in our system.

4 Experimental Results

For evaluating our method, we used 190 texts in 100 scene images in our experiment. The results are shown in Table 1. The performance was 61.1 %. Table 2 shows the causes of the incorrect translation. We will discuss each process of our system.

Table 1: Results of translation camera

	success	failure	total
translation	116	74	190

Table 2: Causes of incorrect translation

| Causes of incorrect translation	text
character pattern extraction	8
character recognition	42
mistranslation	24
total	74

4.1 Character Pattern Extraction Process

There are three causes of the incorrect character pattern extraction:

- the color of the character pattern strings or their background were not uniform.

- the contrast between the character pattern strings and their background was not clear.

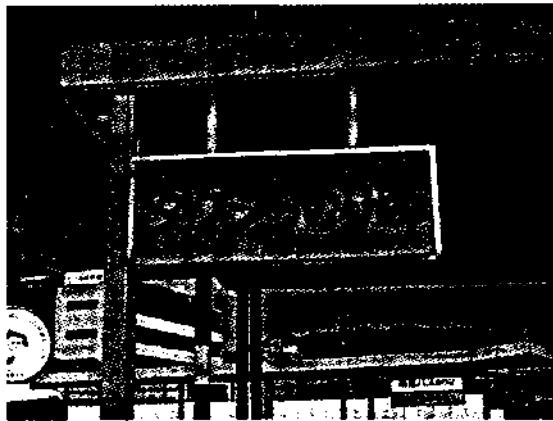- character pattern strings were damaged or dirty.

Characters in Figure 5 (a), "メニュー (menyu, menu)", could not be extracted because the background was not uniform. To put it more precisely, there are a grain
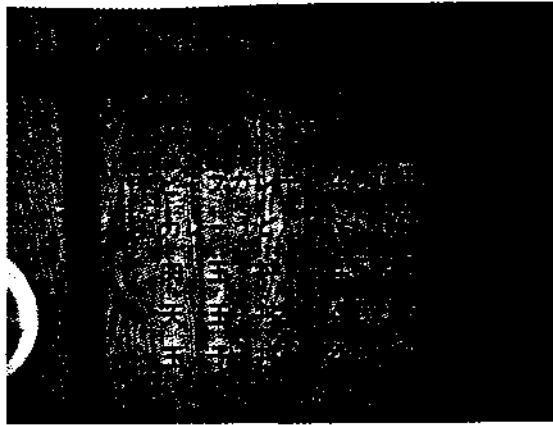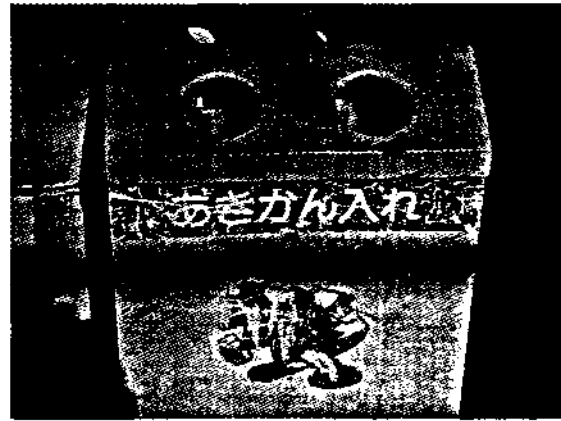
(a) cause: the background was not uniform
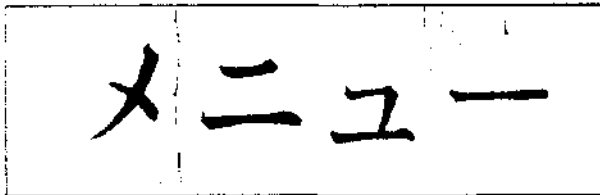

(b) cause: the contrast was not clear

(c) cause: the contrast was not clear
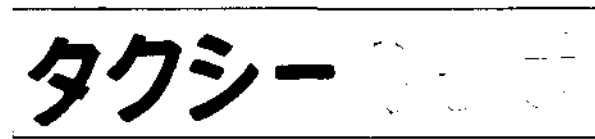

(d) cause: the text region was damaged

Figure 5: Example images from which the system could not extract character pattern strings correctly

(a) binary image of the text region in Figure 5(a)

(b) binary image of the text region in Figure 5(b)
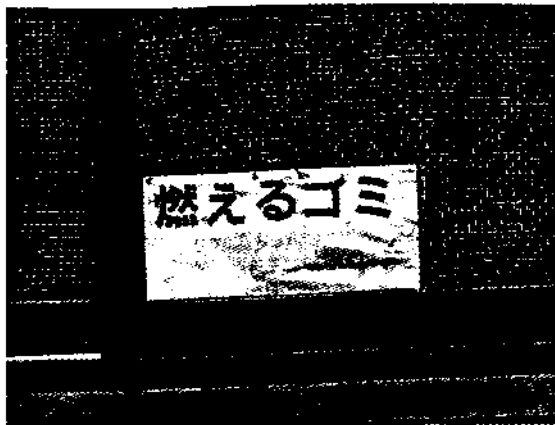
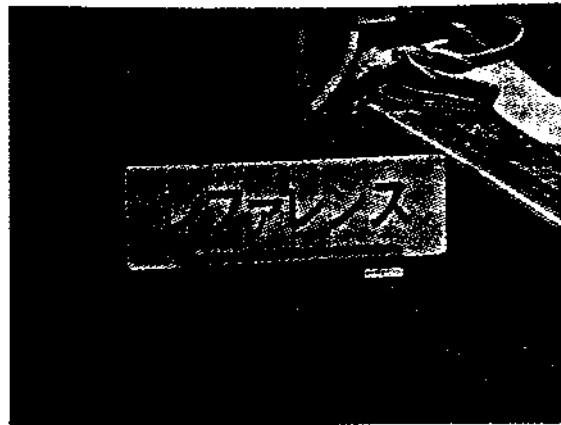

(c) binary image of the text region in Figure 5(c)

(d) binary image of the text region in Figure 5(d)

Figure 6: Binary images of the specified text regions of the images in Figure 5

(a) cause: distortion of character pattern strings   (b) cause: similar character

Figure 7: Characters which the system could not recognize correctly

of wood between katakana letters "メ" and "ニ". As a result, the system extracted this grain of wood in mistake for a character pattern. Figure 6 (a) shows the result of the binarization on the specified text region in Figure 5 (a). Character pattern strings in Figure 5 (b) and (c) could not be extracted because the contrasts were not enough. In Figure 5 (b), the background near the last character "所 (place)" was darker than the rest. Consequently, "所" could not be extracted correctly. On the contrary, the background of the signboard in Figure 5 (c) was tinged with red because the signboard was exposed to the setting sun. As a result, the black characters "タクシー *(takushi,* taxi)" could be extracted while the red characters "の り ば *(noriba.* stand)" could not be extracted. Figure 6 (b) and (c) shows the results of the binarization on the specified region in Figure 5 (b) and (c). Character pattern strings in Figure 5 (d) could not be extracted because they were damaged. Figure 6 (d) shows the result of the binarization on the specified region in Figure 5 (d).

4.2   Character Recognition Process

The correct recognition score of character recognition was 91.7 %. The causes of incorrect recognition were as follows:

- failures in the character pattern extraction process,

- distortion of the character pattern strings, and

- existence of the similar character patterns

Texts in Figure 7 (a), "燃えるゴミ *(moeru gomi.* burnable garbage)". was not straight because the signboard was bent.   To be precise. "燃える" and "ゴ

ミ" sloped forward and backward, respectively. Consequently, "燃えるゴミ" could not be recognized correctly although "燃える" and "ゴミ" could be recognized correctly. Texts in Figure 7 (b). "レファレン ニ *(refarense,* reference)", could not be recognized correctly because the first letter "レ" is similar to hiragana letter "し"

4.3   Machine Translation Process

The system often failed to translate texts in case c: long sentences. Also, the system failed when text-contained a word which is not in the translation dictionary. However, it is good for MT system to translate a text in a scene image because the sentence structure tends to be simple. In particular, it is easy to translate when it consists of only one word. Also, texts in signboard have many typical expressions. As a result, EBMT (Example Based Machine Translation) systems are suitable for this task [Sato 91].

There are one problem which needs to consider visual context. Usual MT systems are concerned with only textual context. On the contrary, translation camera should be concerned with visual context. For example, the text in Figure 8. "普通 *(futsuu)*", implies a door of a "local train".   From this, "普通" should be translated as "local train" But. in this experiment, the system translated it as "usual" because the system did not use visual context into account. As a result, for realizing precise translation on translation camera, we must investigate a method for extracting visual context from scene images.

References

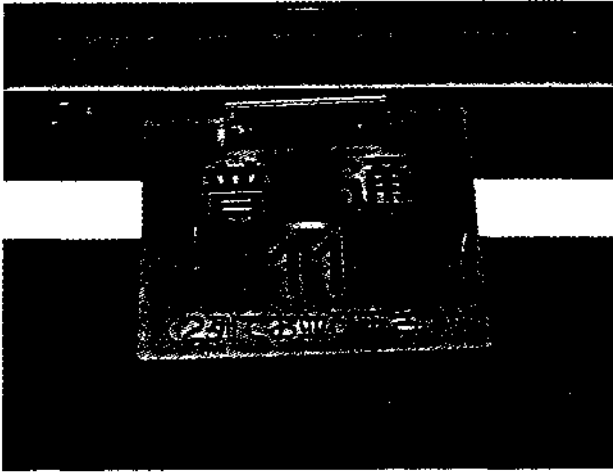[Choi 90] Choi, H.. Agui. T.. Nakajima. M.. Yokomae. T.: Extraction of Passing Car and Its Number Plate Us-

Figure 8: A signboard implies a door of a "local train"

ing a Series of Images, (in Japanese), Transaction of IEICE, Vol.J63-D-II No.3, (1990).

[Liu 96] Liu. Y., Yamamura, T., Ohnishi, N., Sugie, N.: Extraction of Character String Regions From a Scene Image, (in Japanese), IEICE-WGPRU, 95-222. (1996).

[Ohya 94] Ohya, J., Shio, A., Akamatsu, S.: Recognizing Character in Scene Image, IEEE transactions on pattern analysis and machine intelligence, Vol.16 No.2, (1994).

[Oki 98] Oki Electric Industry Co., Ltd.: Machine Translation System PENSÉE entirely written in Java http://www.oki.co.jp/OKI/RDG/English/java/pensee. (1998).

[Otsu 79] Otsu, X.: A threshold selection method from gray-level histograms. IEEE Trans. Syst., Man.& Cybern.. SMC-9, (1979).

[RICOH 95] Ricoh Ltd.: Ricoh IMAZONE Version 2.0 install manual. Ricoh Ltd., (1995).

[Sato 91] Sato, S.: Example-Based Machine Translation. Doctoral dissertation. Kyoto University, (1991).

[Sakai 93] Sakai, T.: A History and Evolution of Document Information Processing, Proc. of 2nd ICDAR, (1993).

[Takizawa 95] Takizawa, K., Senda, S., Minoh, M., Ikeda, K.: Extraction of Character Pattern Strings from Signboard. (in Japanese), IEICE-WGPRU. 94-133, (1995).

[Ueba 96] Ueba, A., Takeda, T., Okada, Y.: Extraction of Character String Images from Color Images based on Isochromatic Lines Processing., (in Japanese), journal of IIEEJ. Vol.25 No.4, (1996).