TRANSLATION CAMERA ON MOBILE PHONE

Yasuhiko Watanabe  Kazuya Sono  Kazuya Yokomizo  Yoshihiro Okada
Dept. of Electronics and Informatics, Ryukoku University, Seta, Otsu, Shiga, Japan
watanabe@rins.ryukoku.ac.jp

ABSTRACT

In this paper, we report a mobile application which translates Japanese texts in a scene into English. The application is designed to run on a mobile phone with a camera. It recognizes Japanese characters sensed by an onboard camera of a mobile phone and translates them into English.

1. INTRODUCTION

The functions of mobile phone are rapidly progressed: telephone access, Internet access, the ability to download and execute applications from Web sites, and onboard sensors (e.g., camera and GPS receiver). Onboard sensors can change the fundamentals of mobile phone styles. This is because an onboard sensor can sense data in real world and generate the interaction between real world and cyber world. This paper reports an experimental application for a mobile phone with a camera, translation camera on mobile phone. It translates Japanese texts sensed by an onboard camera of a mobile phone into English.

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, “なるめ (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 mobile application, translation camera on mobile phone. The application translates texts in scene images which are taken with an onboard camera of a mobile phone. In this paper, we describe the outline of the application and a method of 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]. This 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. Taking account of this point, we developed translation camera which translated Japanese texts in a scene into English [Watanabe 98]. Also, Yang developed a similar system which detected Chinese signs in video images and translated them into English [Yang 01]. However, both of them have the same problem, portability. To solve the problem, we refined translation camera to run on a mobile phone.

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. “なるめ (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. This is because these texts have the following features:

- 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).

We use these features for recognizing characters in scene images.
3. APPLICATION ENVIRONMENT ON MOBILE PHONE

Translation camera on mobile phone was implemented on Java-enabled i-mode mobile phone. Java-enabled i-mode mobile phone combines telephone access, Internet access, the ability to download and execute Java applications from any i-mode Web site, and onboard sensors (e.g. camera and GPS receiver). Figure 2 shows the i-mode Java application Environment [NTT 00]. It is comprised of:

- Operating system,
- the mobile phone’s native applications (e.g. telephone and browser),
- the Java Application Manager (JAM),
- the K virtual machine (KVM),
- the CLDC class libraries.

Fig. 2. i-mode Java application environment

The i-mode Java extension APIs are comprised of User Interface API, Networking API, Low-level graphics API, text processing API, and ScratchPad manager.

4. 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. As shown in Figure 3, the system works in this way:

1. a user takes a picture by an onboard camera of a mobile phone and specifies a text region,
2. the mobile phone sends the picture and the specified region to a remote host via Internet,
3. the remote host extracts character pattern strings from the specified region in the binarization process, then, recognizes and translates them.
4. the remote host sends the translation result to the mobile phone.

Fig. 3. Overview of translation camera on mobile phone

- the i-mode Java Extension APIs,
- JAR Storage and Data Storage (ScratchPad).

The i-mode Java extension APIs are comprised of User Interface API, Networking API, Low-level graphics API, text processing API, and ScratchPad manager.

1. i-mode is services offering wireless web browsing and e-mail from mobile phones in Japan. It was introduced in February 1999 by NTT DoCoMo.
2. The JAM manages JAVA applications and communicates with the K virtual machine.
3. The KVM is a version of a Java virtual machine that was re-designed for small embedded devices. It is designed to run on memory and CPU power constrained devices.
4.1. 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 objective text but also unnecessary texts.
- Texts in scene images should be translated as soon as possible. This 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.

4.2. 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 binarization 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.

4.3. Character Recognition

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

- 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

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

4.4. Machine Translation

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

5. EXPERIMENTAL RESULTS

For evaluating our method, we used 141 texts in 119 scene images in our experiment. The results are shown in Table 1. The performance was 66%. Table 2 shows the causes of the incorrect translation. We will discuss each process of our system.

5.1. 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 4 (a), “安全第一 (anzen daiichi, safety first)”, could not be extracted because the background was not uniform. Figure 5 (a) shows the result of the binarization on the specified text region in Figure 4 (a). Character pattern strings in Figure 4 (b) and (c) could not be extracted because the contrasts were not enough. In Figure 4 (b), “報” could not be extracted correctly because the background near the character was brighter than the rest. On the contrary, in Figure 4 (c), the characters “注意事項 (chui jikou, directions)” could not be extracted because the characters got discolored. Figure 5 (b) and (c) shows the results of the binarization on the specified region in Figure 4 (b) and (c). Character pattern strings in Figure 4 (d) could not be extracted because they were damaged. Figure 5 (d) shows the result of the binarization on the specified region in Figure 4 (d).

<table>
<thead>
<tr>
<th>Table 1. Results of translation camera</th>
</tr>
</thead>
<tbody>
<tr>
<td>translation</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>translation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Causes of incorrect translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causes of incorrect translation</td>
</tr>
<tr>
<td>------------------------------------</td>
</tr>
<tr>
<td>character pattern extraction</td>
</tr>
<tr>
<td>character recognition</td>
</tr>
<tr>
<td>mistranslation</td>
</tr>
<tr>
<td>total</td>
</tr>
</tbody>
</table>
5.2. Character Recognition Process

The correct recognition score of character recognition was 85%. 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.

In Figure 6, the first letter “つ” of “つねたいお飲物” could not be recognized correctly because the first letter “つ” is similar to katakana letter “コ”.

5.3. Machine Translation Process

The system often failed to translate texts in case of long sentences. Also, the system failed when texts 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].