Recognition of Chinese Character in Snapshot Translation System

Fai Wong, Sam Chao, Wai Kit Chan and Yi Ping Li
Faculty of Science and Technology of University of Macau
{derekfw, lidiasc, ma96562, ypli}@umac.mo

Abstract

In this paper, we presents a mobile based snapshot translation system, Cyclops. The system takes an image of Chinese text as its input and translates the textual content of the image into either Portuguese or English, or both. The underneath design idea is to provide user a comprehensive user interface in using language translation tool to access the meaning of non-native text. The system has been developed based on different technologies including image processing, Chinese optical character recognition and machine translation. In this article, we focus on describing the character recognition module which uses Peripheral Direction Contributivity (PDC) for representing the features of Chinese character. Most importantly, it has been tailored to run on common use mobile devices which have memory and storage limitations.

1. Introduction

As the evolvement of language technologies and the growth of mobile devices, it brings new opportunities and platforms for the translation tool to further support the cross-language communication [1]. This seems to be an ideal environment and platform for potential use of machine translation and related technologies. There are number of translation systems developed for personal hand-held devices such as mobile phones and PDAs. However, the user interfaces of these systems are not friendly enough. The entry of text is very time-consuming and is intrinsically limited by the keypad. They are usually provided with letters assigned to different buttons, miniaturized hard or soft QWERTY keyboard, or handwriting recognition using a stylus on touch screen [2]. Furthermore, many of the Chinese related translation systems are designed for Chinese people. English-speaker cannot make use of systems to translate Chinese text. This is because Chinese and English languages are quite different in computation aspect. Chinese is a non-alphabetic language. It relies on particular input system and is infeasible for non-Chinese users. This can be a problem if users do not have any knowledge about the Chinese language. Therefore, we propose a novel mobile translation framework, Cyclops – snapshot translation system. It takes the advantage of the digital camera of hand-held devices and optical character recognition (OCR) to address the bottleneck of text entry. That allows both native and non-native users to easily use mobile translation systems. In this paper, we primarily focus on describing the recognition module of Cyclops for the translation from Chinese to Portuguese and Chinese to English, since the processing of Chinese language in terms of character recognition and machine translation technologies is more challenge.

2. Design model of Cyclops

Since Cyclops is designed to be run in tactical environment and should be convenient to use, we developed an intuitive user interface. To translate a text, the user simply presses the *Recognize* button on the device, the system then switches to the camera shooting mode, where a user is allowed to focus and shoot a snapshot of any interested Chinese text. After the confirmation, the system will process the image and

Figure 1. User interface of Cyclops.

display the recognized text, as well as its translation in target languages on the screen. Figure 1 shows a screenshot of the graphical user interface (GUI) of Cyclops. The GUI window is basically divided into four parts, the upper one shows the recognized Chinese text output by the OCR component. The central box displays the translation results, and the content is dominated by the *language* options on the right side, where the user can select to show the output translation in both Portuguese and English, or either one of them. The lower part of the screen shows the *Recognize* button to trigger the translation.

An overview of the proposed snapshot translation model is given in Figure 2. The translation process begins with a preliminary analysis of the given image captured by the digital camera of mobile device, including the conversion of image from color to binary format, the detection and segmentation of characters from the image, as well as the normalization of each sub-image of character. Based on the segmented images, features of Chinese character derived from peripheral directions are constructed and further analyzed by using the K-L transform. In the classification process, a 3-level matching strategy is adopted to minimize the searching space and optimize the computation due to the constraint of money and processing power of hand-held devices. The output of *k*-best candidates for each character is then taken as input to the transfer-based machine translation engine and let the subsequent language analytical modules determine the correct recognition based on different linguistic information and context, and finally translate the text into target languages of Portuguese and English.

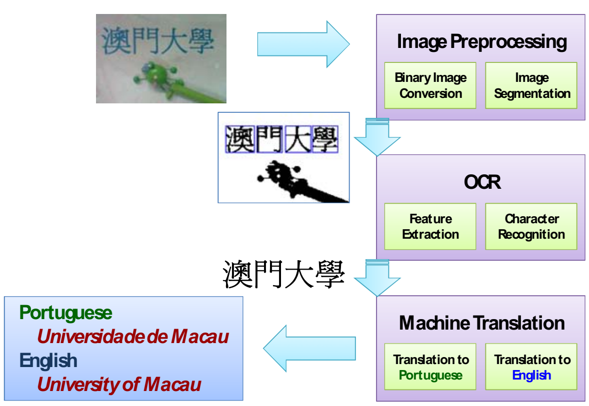

Figure 2. Design model of Cyclops.

3. Recognition of Chinese scripts

As the main input modality, we implemented the character recognition module using the integrated digital camera of the mobile device. In Cyclops, Chinese is chosen as the source language to be analyzed due to that Chinese is widely used in many Asian communities worldwide. Similar to Korean and Japanese, these languages are related by shapes, or syntax and semantics [3], and consist of thousands of characters. Moreover, because of the sophisticated formation of Chinese characters, their computerized input and automated recognition are much more difficult than that of western languages. Thus, the development of Chinese OCR, in this stage, may help in ease the development for other languages to the extension of Cyclops system in future.

3.1. Feature representation using PDC

Chinese characters are made up of strokes in four main directions, and previous works show that the peripherals of character strokes contain a lot of information regarding the shape of characters, hence are useful features for character recognition [4]. In this work, we adopt the features-based recognition approach by using Peripheral Direction Contributivity (PDC) information to represent the features of character [5-6]. As illustrated in Figure 3, the *direction contributivity density* (DC), referring to stroke pixel *P*, representing the distances along the eight directions from *P* to the boundary pixels of character strokes are used as features, and is described as an 8-dimentsion vector $< d_0, d_1, ..., d_7 >$, calculated according to

$$d_i = \frac{l_i}{\sqrt{\sum_{k=0}^{7} l_k^2}}. \quad (1)$$

where $i = 0, \ldots, 7$, and l_i is the number of black pixels in the i^{th} direction to the boundary of stroke. *Peripheral Direction Contributivity* (PDC) corresponds to the distance *A* from a point on the character frame to the first background to foreground transition (as front edge point), while *B* corresponds to the distance between the edge of the stroke to the closest second background to foreground transition of stroke in the 2^{nd} layer (known as 2^{nd} order peripheral feature) along the given scan direction (row or column), and similar for the distance *D* in 3^{rd} layer, as shown in Figure 3.

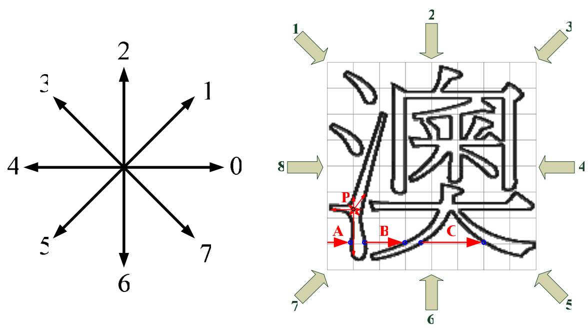

Figure 3. 8-direction DC and PDC scan.

For each of the visited layers, front edge point of stroke is served as the reference point and a corresponding DC feature is constructed. All such distances are effective in representing the external shape as well as the internal structure of a Chinese character [7]. Under the described representation, the total number of feature elements of a character can be formulated as

$$\textit{no. of feataures} = \textit{rows} \times \textit{scan directions} \times \textit{layers} \times \textit{PDC dimensions} . \quad (2)$$

In our system, the size of each character image is normalized to 32×32 pixels to compensate the limited resources of the mobile devices. The number of feature elements becomes: 32 (*rows*) × 8 (*scan directions*) × 3 (*layers*) × 8 (*PDC dimensions*) = 6144. This number is large and is infeasible for a hand-held device to process. Therefore, we reduced the feature dimension by dividing the 32 rows or columns of a given scan direction into 4 groups and use the average of the corresponding PDC components in each group as the features. As a result, the dimensionality is effectively reduced to 768. In practice, the feature vector is further reduced via Karhunen-Loeve transform to remove the redundancy of feature elements. More importantly, it can help to retain only the most relevant features against noise and differentiate one character from others to achieve a good recognition performance. In our design, 128 feature elements are transformed from the original 768 features, and is used for the subsequent matching of an input feature vector.

3.2. Hierarchical classification

In the recognition module, a 3-level hierarchical matching strategy from coarse to fine is adopted to speed up the classification process, and this is closely related to how the feature templates are organized. In Karhunen-Loeve transform, the covariance matrix of features derived from training samples is computed, and then its Eigen values are derived and ranked in descending order which indicate the information magnitude. The bigger the Eigen value is, the more the information it has. The Eigen vectors of the first m largest Eigen values are taken to form a transform matrix T, where $m < n$, and is used to transform the original n-dimension PDC feature to an m-dimension one. In the first level coarse classification, we utilize a small number of transformed feature elements that associated with the biggest Eigen values and yielding a set of candidates. During the classification process, we use 24 feature elements to construct a coarse classifier, and keep c_1 candidates (where c_1 is determined at the design time, in our prototyping system, 25 candidates is used for this coarse level of classification). The matching is based on the Euclidean distance

$$L_2(v, \hat{v}_j) = \sqrt{\sum_{i=0}^{m-1} (v_i - \hat{v}_{ij})^2} . \quad (3)$$

where v is the transformed feature vector for the character to be recognized, $\hat{v}_j$ is the transformed feature template of j^{th} character in lookup table, v_i is the i^{th} element of the feature vector for the character to be recognized, $\hat{v}_{ij}$ is i^{th} element of the transformed feature vector of j^{th} character in lookup table, and m is the dimensionality of transformed vector. Smaller distance implies the two characters are more similar.

For the finer classifier in the second level, we use 48 feature elements to narrow the candidates to c_2 (where $c_2 < c_1$). The finest classifier in the third level uses all of the 128 feature elements to select the closest character among the c_2 candidates as the final recognition result.

$$\min_{1 \le j \le t} L_2(v, \hat{v}_j) . \quad (4)$$

where t is the total number of characters in the lookup table. In our algorithm, five candidates with the best scores are retained and used to feed into the translation module to further determine the best recognition pattern based on context.

4. Text post-processing and translation

This module takes the task to further validate the recognition results and translate the Chinese text into the corresponding languages of Portuguese and English according to the selection preference of user. As described in section 3, the recognition module yields the best five candidates for each character as the output results. The idea is tried to maintain a high recognition *recall* rate (where the correct character should always be included in the short list) and let the subsequent tasks to determine the final recognition text based on possible surrounding context. The correction algorithm in Cyclops is based on lexicon lookup approach, where a dictionary is used to verify if a possible combination of characters that forms a word (or phrase) is a registered lexical item or not, then classifies it as the ultimate result, and meanwhile, the syntactic-semantical information is also retrieved from the dictionary for later translation.

4.1. Text translation

In the translation module, we use a light version of the transfer-based MT system [8] as the component for translation from Chinese to Portuguese and Chinese to English. Where the translation module is mainly composed by two components: 1) dictionaries for lexical translations of constituents analyzed by the shallow parser; and 2) syntactic parser developed based on *constraint synchronous grammar* (CSG), where the rules of source and target languages are formulated in parallel, and the translation for the analyzed sentence can be immediately inferred after parsing. Since the analysis of Chinese and the generation of translation in Portuguese and English share the same analytical mechanism and translation algorithm, two sets of lexical and syntactic data, one for Chinese-Portuguese and another for Chinese-English, are used for guiding the algorithm to produce the translation in corresponding target language according to user preference. The hand-written (transfer structural) rules are declarative and defined according to the format of CSG formalism as described in [8]. It uses a pair of *context free grammar* (CFG) productions to describe the syntactical pattern of source and target languages, and is possible to associate with extra feature constraints for guiding the parsing process and selects the corresponding translation pattern.

5. System evaluation

In order to evaluate the prototyping system of Cyclops that proposes the *snapshot translation* based on hand-held device, two experiments are carried out. Independent from the whole system, the first experiment is setup to investigate the performance of the Chinese OCR module based on the designed algorithm, and is carried out offline on laptop computer instead of mobile device. Like many other OCR system, the construction of our module has two steps, namely, training and testing. The training phase is to create the recognition lookup table using the training samples. Where the samples are produced with four fonts (*Ming*, *Kai*, *Shu*, *Song*), three styles (*normal*, *boldface*, *italic*) and five different sizes from laser printer and scanned with default setting. Totally, 20,000 samples are used to construct the model. In the evaluation, a test suit of 200 samples, that are not included in the training set for creating the lookup table, are used to evaluate the model, and the recognition rate is 95.2%.

In the second part of the evaluation, according to the nature of the snapshot translation system, the experiment is designed and setup as follows, 1) 100 Chinese words or phrases with different fonts and styles are produced from laser printer with sufficient larger size, in order that the digital camera of mobile device is able to capture for recognition. The distribution of the samples with different fonts and styles is given in Table 1; 2) two users are invited to participate in using the system to translate the prepared context, among them, User 2 is one of the developers himself and is quite familiar with the characteristic of the system. We would like to see how different the system is used by people with different background knowledge regarding the system; and 3) the data of image captured by the device, image after binarization, segmented sub-images of characters, as well as the translation result of each operation is collected for us to analyze and evaluate the overall performance of Cyclops under different operational environment and conditions.

Table 1. Distribution of testing data with different fonts and styles.

Font	No. of Phrases
Ming (normal)	25
Kai (normal)	25
Song (normal)	25
Ming (italic)	25

Table 2 and Table 3 show the translation results conducted by User 1 and User 2 respectively, and we found that they have produced significantly different results. The successful (correction) rate that User 1 obtained from using the system is 80%, while User 2 achieved the recognition and translation rate up to 96%. After reviewing the logged information, we found that User 1 may not be familiar with system, the pictures he shot are distorted with the shooting angles and the poor lighting (with shadows) also has a bad affect on the images, as illustrated in Figure 5 (a) and (b), which make the system hard to recognize the text successfully. Moreover, we found that the text printed with *italic* style is also the main cause to the failure of recognition. Figure 5 (c) shows the captured image of the text being affected by the font style, together with uneven lighting condition in the picture. However, the overall performance of the system achieves 88% of recognition and translation rate.

Table 2. Recognition results conducted by User 1.

	Ming normal	Kai normal	Song normal	Ming italic	%
Correct	22	23	24	11	80%
Incorrect	3	2	1	14	20%

Table 3. Recognition results conducted by User 2

	Ming normal	Kai normal	Song normal	Ming italic	%
Correct	25	25	25	21	96%
Incorrect	0	0	0	4	4%

(a) (b) (c)

Figure 5. The captured images failure at recognition and translation - (a) shooting with perspective angle, (b) uneven lighting with shadow, and (c) text with italic font style.

6. Conclusion

In this paper, we proposed a new translation framework, Cyclops, that realizes the translation of the snapshot of text by using the integrated digital camera of mobile device, targeted at providing a comprehensive GUI for user to access the meaning of any non-native text in the most natural and efficient way. The challenge of the development of this system is due to the severe constraints on computation and memory of mobile device, and these issues have to be considered in the design of each component of Chinese OCR, image processing and machine translation. The preliminary empirical results show that the proposed system is feasible and achieves an average translation result of 88% in selected domain.

Acknowledgement

This work was partially supported by the Research Committee of University of Macau under grant CATIVO: 5868, and also supported by Science and Technology Development Fund of Macau under grant 057/2009/A2.

References

[1] N. Ward, "Machine Translation in the Mobile and Wearable Age," in *The 9th International Conference on Theoretical and Methodological Issues in Machine Translation Workshop: Machine Translation Roadmap*, Keihanna, Japan, 2002.

[2] B. O'Riordan, K. Curran, and D. Woods, "Investigating Text Input Methods for Mobile Phones," *Journal of Computer Science,* vol. 1, pp. 189-199, 2005.

[3] C. Y. Suen, S. Mori, H. C. Rim, and P. S. P. Wang, "Intriguing aspects of oriental languages," *Int. J. of Pattern Recognition & Artificial Intelligence,* vol. 12, pp. 5-29, 1998.

[4] M. Cheriet, N. Kharma, C.-L. Liu, and C. Y. Suen, *Character Recognition Systems - A Guide for Students and Practioners*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2007.

[5] T. Wakahara, Y. Kimura, and M. Sano, "Handwritten Japanese Character Recognition Using Adaptive Normalization by Global Affine Transformation," in *2001 Sixth International Conference on Document Analysis and Recognition*, 2001, pp. 424 - 428.

[6] C. Y. Suen, S. Mori, S. H. Kim, and C. H. Leung, "Analysis and Recognition of Asian Scripts - the State of the Art," in *Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)*, Edinburgh, Scotland, 2003, pp. 866- 878.

[7] T. Akiyama and N. Hagita, "Automated entry system for printed documents," *Pattern Recognition,* vol. 23, pp. 1141-1154, 1990.

[8] F. Wong, D. C. Hu, Y. H. Mao, M. C. Dong, and Y. P. Li, "Machine Translation Based on Constraint-Based Synchronous Grammar," in *Natural Language Processing (IJCNLP 2005)*. vol. 3651 Jeju Island, Republic of Korea: Springer Berlin / Heidelberg, 2005, pp. 612-623.