Cyclops – Snapshot Translation System Based on Mobile Device

Fai Wong
Department of Computer and Information Science, University of Macau, Macao S.A.R., China
Email: derekfw@umac.mo

Sam Chao and Wai Kit Chan
Department of Computer and Information Science, University of Macau, Macao S.A.R., China
Email: {lidiasc, ma96562}@umac.mo

***Abstract*—This paper presents a snapshot translation system, Cyclops. It realizes the translation of textural information being captured as an image via a digital camera of mobile device. The design framework of Cyclops is targeted at providing user the most comprehensive interface in using language translation tool to access the meaning of non-native text in a more natural and efficient way. It supports the translation of languages from Chinese to Portuguese and English. In the course of system realization, several technical challenges were encountered. The system has been developed based on the technologies of image processing, optical character recognition (OCR), and machine translation (MT). Most importantly, the system is designed to run on the common usage mobile devices which have memory and storage limitations.**

***Index Terms*—Snapshot Translation; Natural Language Processing; Machine Translation, Constraint Syncrhonous Grammar; Mobile Device; Chinese-Portuguese; Chinese-English; OCR; Peripheral Direction Contributivity**

I. INTRODUCTION

In the era of fast development and business globalization, automatic translation of languages from one to another through the use of computers is becoming more and more attractive. This is not only to the system developers and language translators, but also to the public and tourists. Nowadays, the use of language translation systems has become popularization due to the continue emerging of various translation systems in the market [1]. 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 [2]. This seems to be an ideal environment and platform for the potential use of machine translation and related technologies. Over the past years, there are number of translation systems (including the electronic dictionaries [3]) developed on personal hand-held devices. Some of them are even available for mobile phone, Pocket PCs and PDAs [4]. For example, the translation system, likes MobileTran [5], provides the translation service through the use of mobile phone and the WAP (Wireless Application Protocol) communication protocol as the media. Where the translation program itself is hosted at the remote server. IdiomaX [6] implements the translation system based on Windows Mobile environment. It is able to install and run on different hand-held devices such as Pocket PCs, PDAs or Smart phones, as long as these devices support the proper versions of operating system. On the other hand, the work of Waibel et la. [7] and Hsiao et al. [8] focus on the development of mobile translation system by using speech as the input of source language and outputs the translated speech as the target language. That makes use of the components of automatic speech recognition (ASR), statistical machine translation (SMT) and text-to-speech (TTS) to realize the speech to speech translation.

However, the user interfaces of the described systems in some way are not friendly enough for people to access the meaning of non-native text. Take the first two mobile translation systems for example, the input of text of these translation systems is very time-consuming and is intrinsically limited by the design of the text input methods. The most common methods for text entry on these hand-held devices are the keypad. And which are usually provided with letters assigned to different buttons, miniaturized *hard* or *soft* QWERTY keyboard, or handwriting recognition using a stylus on touch screen [9]. This is more difficult for users to enter text than that of using computer keyboard. This, as a result, will definitely waste most of the valuable time of users due to the high overhead involved in typing words if the user needs to frequently referencing documents in different languages [10]. For the speech translation systems, the use of speech as the input source does not rely on the conventional text input methods. It is quite natural for native speaker to achieve the cross-language communication purpose. Nevertheless, the system is only useful for people with different native languages to communicate with each other. The system, in some sense, is being treated as a language translator. Because of its specific design based on speech, it is not suitable for use to translate (non-native) text if the speaker cannot pronounce and speak that language. Furthermore, we

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.

doi:10.4304/jsw.6.9.1664-1671

found that many of the Chinese-English translation systems are designed for Chinese people to look up the meaning of English text. On the other hand, 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 for writing and this is infeasible for non-Chinese users. The types of the input system widely used by Chinese speakers are the Pinyin and the root radical methods. The Pinyin method is based on the Pinyin system that transliterates Chinese ideograms into alphabet. Pinyin system is officially adopted in China. While the root radical method is concerning the smaller component units, referred to as radicals, of a Chinese character. Each radical is associated to a key in the keyboard. In writing, the root radical of the character is identified [11]. This depends on the user's ability to identify the root. It can be a problem for non Chinese-speakers if they do not have any knowledge about the Chinese language.

Based on these considerations, we propose a novel mobile translation framework, *Cyclops – snapshot translation system*. It takes the advantage of using the digital camera of personal hand-held devices and the optical character recognition (OCR) technologies to address the bottleneck of text entry. The interface allows both native and non-native users to access the translation system on mobile devices in the most convenient and natural way. In this paper, we primarily focus on presenting the design model of Cyclops for the translation from Chinese to Portuguese and Chinese to English. That is, in this system, we start with the development of Chinese OCR as our initial stage, since the recognition of non-alphabetic language is more challenge, and this can be further extended to other western languages.

The paper is organized as follows. In section 2, an overview of Cyclops is provided. Section 3 discusses the Chinese recognition module based on peripheral direction contributivity (PDC). The preprocessing of image is described in section 4. Section 5 presents the translation model based on synchronous formalism. The evaluation of the system is conducted and discussed in section 6, followed by a summary of our work to end this paper.

I. 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 display the recognized text together with the translation in target languages. Fig. 1 shows a screenshot of the graphical user interface (GUI) of Cyclops system. 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. The content is dominated by the

Figure 1. User interface of Cyclops – snapshot translation system.

language options on the right side, where the user can ask for the translations 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 Fig. 2. The translation process begins with a preliminary analysis of the given image captured by the digital camera of mobile device. This includes the conversion of image from color into binary, 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 the memory and computation 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. Where the subsequent language analytical modules determine the correct recognition based on different linguistic information and context. Finally the text is translated into the target languages of Portuguese and English.

II. Recognition of Chinese Scripts

As the main input modality, we implemented the character recognition module using the integrated digital camera of a mobile device. In Cyclops, Chinese is chosen

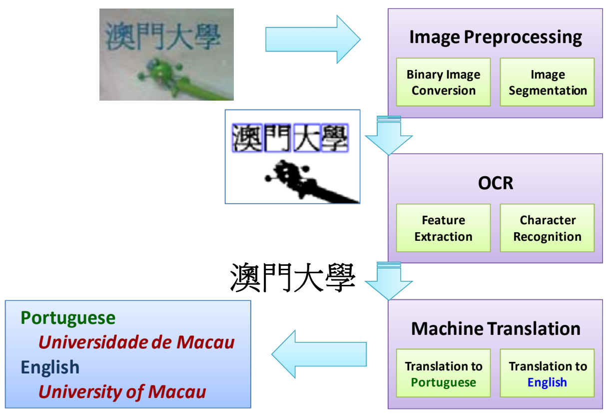

Figure 2. Design model of Cyclops translation system.

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 [12], 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.

A. Feature Representation Using PDC

Chinese characters are made up of strokes in four main directions. 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 [13]. In this work, we adopt the features-based recognition approach by using Peripheral Direction Contributivity (PDC) information to represent the features of character [14][15]. As illustrated in Fig. 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 $< d_0, d_1, ..., d_7 >$, 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 2nd layer (known as 2nd order peripheral feature) along the given scan direction (row or column) and similar for the distance D in 3rd layer, as shown in Fig. 3. 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 [16]. Under the described representation, the total number of feature elements of a character can be formulated as

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

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.

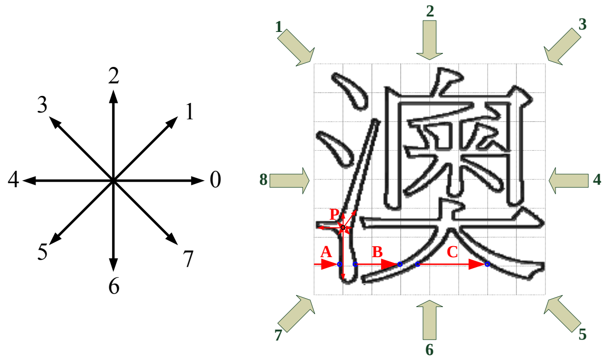

Figure 3. 8-direction DC and PDC scan.

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.

B. Hierarchical Classification

In the recognition module, a 3-level hierarchical matching strategy from coarse to fine is adopted to speed up the classification process. 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. 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 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.

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

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.

III. Character Preprocessing

As the first important step, image and data preprocessing serve the purpose of extracting interested regions, enhancing and cleaning up the images, so that they can be directly and efficiently processed by the feature extraction component. In this section, we review the processing tasks of image binarization, character segmentation and normalization, which are specifically designed for mobile usage and tailored implementation. Different from laptop usage, the OCR systems are intended for the extraction of characters and data from the images of scanned documents or forms which may contain hundreds of characters and different form information. The image captured by the integrated digital camera of mobile device, for our usage, may be narrowed and focused to some specific words or text fragment. And therefore, we may assume that the captured image of text is well positioned with less distortion, e.g. skewed.

A. Image Binarization

The recognition procedure usually requires binarizing the images, which discards most of the noise and replaces the pixels in the characters and the pixels in the background with binary 0s and 1s, respectively. In our algorithm, the extraction of characters is through global thresholding. That uses a single threshold to separate pixels from character or background by comparing the gray level image (that has been converted from original colored image) to the reference value.

$$b(x,y)=\begin{cases}1 & g(x,y)\ge\tau\\ 0 & otherwise\end{cases}\ . \tag{5}$$

where $b(x, y)$ is the pixel of the resulting binary image, $g(x, y)$ is the pixel value of gray level image, and τ is the global threshold. In determination, we use the method of Otsu [17] for threshold selection. It has been recognized as one of the most efficient thresholding techniques.

B. Segmentation of Character

Segmentation takes the role to separate characters from the word or character strings before the character can be recognized by the classifier as described in section 3. Since this process is primarily prepared for the recognition of isolated character. In Cyclops, the segmentation of characters is based on the method of projection analysis [13]. Chinese characters usually have equal size (width), and are easily for us to detect the error of over segmentation caused by the confusion between

(a) (b)

Figure 4. Character segmentation based on projection analysis. (a) Text separated by horizontal projection; (b) characters (and components) separated by vertical projection.

inter-character and within-character space. Fig. 4 (b) illustrates the situation that the characters are formed by multiple components. To tackle this problem, the segmentation procedure consists of three steps: 1) the separation of text from multiline into text segments based on horizontal projection, as shown in Fig. 4 (a); 2) separation of characters (or components) according to the breaks of profile based on vertical projection, illustrated in Fig. 4 (b); and finally 3) recovery of any over-segmented character components by revisiting the segments, comparing the size of each segment against the mean width, and recombining the segments if the size is less than that of the reference size. This simple segmentation algorithm is very effective if the text does not contain any other alphabetic characters.

C. Character Normalization

The task of normalization is to regulate the character size to a 32×32 normalized image, and at the same time, the position and shape of characters are also regulated, so as to reduce the shape variation between the images of same class. Denote the input image and the normalized image as $f(x, y)$ and $u(x', y')$, respectively, the normalized image is generated based on the following linear coordinate mapping and value interpolation

$$\begin{aligned} x' &= \alpha x,\\ y' &= \beta y. \end{aligned} \tag{6}$$

where α and β denote the ratios of scaling, given by

$$\begin{aligned} \alpha &= \frac{w_{nor}}{w_{ori}},\\ \beta &= \frac{h_{nor}}{h_{ori}}. \end{aligned} \tag{7}$$

w_{ori} and h_{ori} are the width and height of the original character, and the width and height of the normalized character are denoted as w_{nor} and h_{nor}. However, this normalization is straightforward and efficient to run on mobile device with satisfactory contribution to fairly recognition accuracy.

IV. Text Post-processing and Translation

This module takes the task to further validate the recognition results and translate the Chinese text into 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 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. 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. At the same while, the syntactic-semantical information is also retrieved from the dictionary for later translation.

In translation module, we use a light version of the transfer-based MT system [18] 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 facilitate the translation in corresponding target language based on user preference. The hand-written (transfer structural) rules are declarative and defined according to the format of CSG formalism as described in [19]. 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.

A. Constraint Synchronous Grammar

In our translation model, constraint synchronous grammar (CSG) plays an important role, as kernel, to the translation process. The formalism of CSG is defined by means of the syntax of context-free grammar (CFG) to the case of synchronous. The formalism consists of a set of generative productions and each production is constructed by a pair of CFG rules with zero and more syntactic head and link constraints for the non-terminal symbols in patterns. Let L be a context-free language defined over *terminal* symbol V_T and generated by a context-free grammar G using non-terminal symbol V_N disjointed with V_T, starting symbol S, and productions of the form $A \to w$ where A is in V_N and w in $(V_N \cup V_T)^*$. We extend to include the set of *terminal* symbols $V_{T'}$ as the translation in target language, disjoint from V_T, ($V_T \cap V_{T'} = \varnothing$). Let Z as a set of integers, each non-terminal symbol in V_N is assigned with an integer, $\Gamma(V_N) = \{W_\omega \mid W \in V_N, \omega \in Z\}$. The elements of $\Gamma(V_N)$ are indexed non-terminal symbols, and are used for expressing the link relationship of non-terminal symbols between the pair of generative sequences. Let $R = \{r_1, \ldots, r_n \mid r_i \in (\Gamma(V_N) \cup V_{T'}), 1 \le i \le n\}$ be a finite set of rules, and $C = \{c_1, \ldots, c_m\}$ be a finite set of constraints over the associated features of $(\Gamma(V_N) \cup V_T)$, where the features of non-terminal $\Gamma(V_N)$, the syntactic symbols, are inherited from the designated head element during rule reduction. A *target rule* is defined as pair $[r \in R^*, c \in C^*]$ in γ, where $\gamma = R^* \times C^*$ in form of $[r, c]$. $\psi(\gamma_i)$ denotes the number of conjunct features being considered in the associated constraint, and is applied to determine the degree of generalization for a feature constraint. Therefore, the rules, γ_i and γ_j, are orderable, $\gamma_i \prec \gamma_j$, if $\psi(\gamma_i) \ge \psi(\gamma_j)$ (or $\gamma_i \succ \gamma_j$, if $\psi(\gamma_i) < \psi(\gamma_j)$). For $\gamma_i \prec \gamma_j$ ($\psi(\gamma_i) \ge \psi(\gamma_j)$), we say, the constraint of the rule, γ_i, is more specific, while the constraint of γ_j is more general. Then, we consider a set of related target rules working over the symbols, w', on the RHS of production $A \to w'$, the source rule, where $w' \in \Gamma(V_N) \cup V_T$. All of these non-terminals are co-indexed as *link*.

Definition 1. A *target component* is defined as a ordered vector of *target rules* in γ having the form $\sigma = \{\gamma_1, \ldots, \gamma_q\}$, where $1 \le i \le q$ to denote the i-th tuple of σ. The rules are being arranged in the order of $\gamma_1 \prec \gamma_2 \prec \ldots \prec \gamma_q$.

In rule reduction, the association conditions of the target rules are used for investigating the features of corresponding symbols in source rules, similar to that of feature unification, to determine if the active reduction successes or not. At the mean while, this helps in determining the proper structure as the target correspondence.

Definition 2. A *Constraint Synchronous Grammar* (*CSG*) is defined to be 5-tuple $G = (V_N, V_T, P, C_T, S)$ which satisfies the following conditions:

- V_N is a finite set of *non-terminal* symbols;
- V_T is a finite set of *terminal* symbols which is disjoint with V_N;
- C_T is a finite set of *target components*;
- P is a finite set of *productions* of the form $A \to \alpha\ \beta$, where $\alpha \in (\Gamma(V_N) \cup V_T)^*$ and, $\beta \in C_T$, the non-terminal symbols that occur from both the source and target rules are *linked* under the index given by $\Gamma(V_N)$.
- $S \in V_N$ is the initial symbol.

B. Bilingual Syntax Modelling

In the application of language translation, the syntax relationship between source and target languages are described by CSG productions. The first component (in right hand side of productions) represents the sentential patterns of source language, while the second component represents the translation patterns in target language. Unlike other synchronous formalisms, the target component of production consists of one or more generative rules associated with zero or more controlled conditions. It uses the linguistic features of non-terminal symbols of source language to distinguish the possible generation correspondences in target translation. In such a way, the source components in CSG are generalized by leaving the task of handling the features constraints in target component. This helps to compact and reduce the grammar size in mobile application. In production (8), it has two generative rules associated with the sentential pattern of the source $NP_1\ VP\ NP_2\ PP\ NP_3$.

$$\begin{aligned} S \rightarrow\ & NP_1\ VP^*\ NP_2\ PP\ NP_3 \\ & \{[NP_1\ VP^1\ NP_3\ VP^2\ NP_2;\ VP_{cate} = vb1, \\ & \quad VP_{s:sem} = NP_{1sem},\ P_{io:sem} = NP_{2sem}, \\ & \quad VP_{o:sem} = NP_{3sem}], \\ & [NP_1\ VP\ NP_3\ NP_2\ ;\ VP = vb0, VP_{s:sem} = NP_{1sem}, \\ & \quad VP_{io:sem} = NP_{2sem}]\} \end{aligned} \tag{8}$$

While the generative rules, $NP_1\ VP^1\ NP_3\ VP^2\ NP_2$ and $NP_1\ VP\ NP_3\ NP_2$, are maintained in vector. Each of which is associated with set of control conditions. The determination of the suitable generative rule is based on these associated constraints. The one satisfying all the conditions determines the relationship between the source and target sentential pattern. Take the first constraint as example, the condition expression: $VP_{cate} = vb1$, $VP_{s:sem} = NP_{1sem}$, $VP_{io:sem} = NP_{2sem}$, $VP_{o:sem} = NP_{3sem}$ specifies if the senses of the first, second and the third nouns (NPs) in the input strings matched to that of the subject, direct and indirect objects governed by the verb, VP, with the category type of $vb1$. Once the condition gets satisfied, the source structure can be successfully recognized and reduced. Meanwhile, the corresponding structure of target pattern, $NP_1\ VP^1\ NP_3\ VP^2\ NP_2$, is determined also.

The syntactic relationship between the source and target patterns is established by the given "*subscripts*". For clarity, the index subscripts are explicitly stated to the non-terminal symbols for case of multiple occurrences, e.g. NPs in the production: $S \rightarrow NP_1\ VP\ NP_2\ PP\ NP_3$ $[NP_1\ VP^*\ NP_3\ NP_2]$. Otherwise symbols those appear only once in both the source and target rules, such as VPs, are implicitly linked to give the synchronous rewriting. Linked non-terminal must be derived from a sequence of synchronized pairs. Consider the production: $S \rightarrow NP_1\ VP\ NP_2\ PP\ NP_3$ $[NP_1\ VP^*\ NP_3\ NP_2]$, the second NP (NP_2) in the source rule corresponds to the third NP (NP_2) in the target rule, the third NP (NP_3) in source rule corresponds to the second NP (NP_3) in target pattern, while the first NP (NP_1) and VP correspond to each other in both source and target rules. The asterisk "*" indicates the head element, and its usage is to propagate all the related features/linguistic information of the head symbol to the reduced non-terminal symbol in the left hand side. This achieves the property of features inheritance in CSG formalism.

C. Expressiveness of CSG

Different languages often have structural differences between each other, including the syntactic order between the languages, discontinued constituents, and constituents that may vanish or appear in the target language translation. All these issues can be handled by CSG. The ordering of the constituents is modeled easily by using the subscripts and the sequence defined in CSG production rule. The discontinuity between words in different languages is solved by defining non-terminal symbols that appear in the source but not the target pattern or vice-versa. As an example, the preposition (PP) in the source rule does not show up in any of the target rules in (8). On the other hand, consider the following bilingual sentence: "她把兩支鋼筆借給了佩德羅" / "Ela emprestou ao Pedro duas canetas" (She lent two pens to Peter). Suppose that this sentence is going to reduce to the symbol S in (9).

$$\begin{aligned} S \rightarrow\ & NP_1\ PP\ NP_2\ VP^*\ NP_3\ \{ \\ & [NP^1\ VP\ a\ NP^3\ NP^2]\ ;\ VP_{cat} = vb1,\ PP = \text{把}", \\ & \quad VP_{s:sem} = NP_{1sem},\ VP_{o:sem} = NP_{2sem}, \\ & \quad VP_{io:sem} = NP_{3sem} \\ & [NP^1\ VP\ NP^2\ em\ NP^3]\ ;\ \ldots\} \end{aligned} \tag{9}$$

In this case, there is a discontinuity of the words in Chinese, where "把" and "借給了" should be associated with the verb "emprestou" (to lend). Since the sentence matches the source pattern of (9), it will be associated with the target pattern $NP^1\ VP\ a\ NP^3\ NP^2$. Similarly, the consideration of constituents that are disappeared or shown in the target syntactical pattern is handled in a similar way.

$$\begin{aligned} NP \rightarrow\ & NP_1\ NP_2\ \{\ [NP^2\ de\ NP^1]\ ; \\ & NP_{1sem} = place,\ NP_{2sem} = institution\} \end{aligned} \tag{10}$$

For instance, in production (10), it often happens that the preposition "的" (of) in Chinese is vanished. However, it is necessary to add the word "de" (of) in the target pattern to have a correct translation. For example, in the bilingual sentence: "澳門大學" / "Universidade de Macau" (University of Macau), although "的" (of) in Chinese is vanished, it still requires the preposition "de" (of) in the target sentence.

D. Translation as Parsing CSG

CSG formalism is parsed by a modified version of generalized LR algorithm [21] that takes the features constraints and the inference of the target structure into consideration. The main reason for choosing this algorithm is due to the considerable efficiency over the Earley's parsing algorithm [22] which requires a set of computations of LR items at each stage of parsing [21]. Furthermore, the parsing table used is extended by adding features constraints and the target rules into the actions table [18]. In application to the translation of language, the syntactic constituents and corresponding lexical items together with the necessary linguistic information between the source and target sentences are modeled and described by the set of CSG productions. The recognition of input sentence and the transformation into target language based on the formalism can be easily realized by the parser component, and is known as synchronous parsing. The process can be formally described as:

Definition 3. A set P of productions is said to *accept* an input string s iff there is a derivation sequence Q for s using source rules of P, and any of the constraint associated with every *target component* in Q is satisfied. Similarly, P is said to *translate* s iff there is a synchronized derivation sequence Q for s such that P accepts s, and the link constraints of associated *target rules* in Q are satisfied. The derivation Q then produces a translation t as the resulting sequence of terminal symbols included in the determined target rules in Q.

Hence, to the translation of an input text, it essentially consists of three steps. For an input sentence s, the structure of sentence is analyzed by using the rules of source components from the synchronous productions; by using the augmented generalized LR parsing algorithm as described. Secondly, the link constraints that are determined during the rule reduction process are propagated to the corresponding target rules R (as selection of target rules) to construct a target derivation sequence Q. And finally, based on the derivation sequence Q, translation of the input sentence s is generated by referencing the set of generative rules R that attached to the corresponding constituent nodes in the parsed tree, to realize the translation in target language.

In Cyclops system, CSG is used to implement the translation module, as a light version of translation engine tailored for the hand-held device. The translation module uses a CSG parser and a bilingual dictionary to achieve the translation process. No other linguistic analytical components, such as morphological analyzer and word class tagger, are used.

V. 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

TABLE I. Distribution of testing data with different fonts and styles

	Ming (normal)	Kai (normal)	Song (normal)	Ming (italic)	Total number
Phrases	25	25	25	25	100

TABLE II. 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 III. 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.

us to analyze and evaluate the overall performance of Cyclops under different operational environment and conditions. 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 Fig. 7 (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. Fig. 7 (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.

VI. 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. Among them, the kernel theory and the design techniques of Chinese OCR and machine translation based on the CSG formalism are described. The preliminary empirical results show that the proposed system is feasible and achieves an average translation result of 88% in selected domain.

ACKNOWLEDGMENT

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] W. J. Hutchins, "Compendium of Translation Software - Directory of Commercial Machine Translation Systems and Computer-Aided Translation Support Tools," (5th ed.): EAMT (2009)
[2] N. Ward, "Machine Translation in the Mobile and Wearable Age," In: Proceedings of MT Roadmap Workshop at TMI-2002. Keihanna, Japan (2002)
[3] Hand Held Electronic Dictionaries, http://en.wikipedia.org/wiki/Electronic_dictionary
[4] Mobile Translation, http://en.wikipedia.org/wiki/Mobile_translation
[5] MobileTran, http://www.tranexp.com/win/MobileTran.htm
[6] IdiomaX, http://www.idiomax.com/mobile-translator.aspx
[7] A. Waibel, A. Badran, A. W. Black, R. Frederking, D. Gates, A. Lavie, et al., "Speechalator: Two-Way Speech-to-Speech Translation on a Consumer PDA," In: Proceedings of the European Conference on Speech Communication and Technology, pp. 369-372. GENEVA (2003)
[8] R. Hsiao, A. Venugopal, T. Kohler, Y. Zhang, P. Charoenpornsawat, A. Zollmann, et al., "Optimizing Components for Handheld Two-way Speech Translation for an English-Iraqi Arabic System," In: Proceedings of InterSpeech, pp. 765-768. Pittsburgh, USA (2006)
[9] B. O'Riordan, K. Curran, D. Woods, "Investigating Text Input Methods for Mobile Phones," Journal of Computer Science, vol. 1, no.2, pp. 189-199 (2005)
[10] F. Wong, Y. H. Mao, "Framework of Electronic Dictionary System for Chinese and Romance Languages," Automatique des Langues (TAL), vol. 44, no.2, pp. 225-245 (2003)
[11] K. Svoboda, R. West, "Using Affordances in Electronic Chinese/English Dictionaries for Non Chinese-Speackers," Carleton University Cognitive Science (2002)
[12] C. Y. Suen, S. Mori, H. C. Rim, P. S. P. Wang, "Intriguing aspects of oriental languages," Int. J. of Pattern Recognition & Artificial Intelligence, vol. 12, no.1, pp. 5-29 (1998)
[13] M. Cheriet, N. Kharma, C.-L. Liu, C. Y. Suen, "Character Recognition Systems - A Guide for Students and Practioners," John Wiley & Sons, Inc., Hoboken, New Jersey (2007)
[14] T. Wakahara, Y. Kimura, M. Sano, "Handwritten Japanese Character Recognition Using Adaptive Normalization by Global Affine Transformation," In: Proceedings of 2001 Sixth International Conference on Document Analysis and Recognition, pp. 424-428 (2001)
[15] C. Y. Suen, S. Mori, S. H. Kim, C. H. Leung, "Analysis and Recognition of Asian Scripts - the State of the Art," In: Proceedings of Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), pp. 866-878. Edinburgh, Scotland (2003)
[16] T. Akiyama, N. Hagita, "Automated entry system for printed documents," Pattern Recognition, vol. 23, no.11, pp. 1141-1154 (1990)
[17] N. Otsu, "A Threshold Selection Method from Gray-Scale Histogram," IEEE Trans. Systems, Man, and Cybernetics, vol. 8, pp. 62-66 (1978)
[18] F. Wong, D. C. Hu, Y. H. Mao, M. C. Dong, Y. P. Li, "Machine Translation Based on Constraint-Based Synchronous Grammar," Natural Language Processing – IJCNLP 2005, Vol. 3651, pp. 612-623. Jeju Island, Republic of Korea: Springer Berlin / Heidelberg (2005)
[19] F. Wong, "The Research and Implementation of Portuguese to Chinese Machine Translation System Based on Constraint Synchronous Grammar," PhD Thesis, Tsinghua University, Beijing, China (2005).
[20] K. Ronald, "The Formal Architecture of Lexical-Functional Grammar," Journal of Information Science and Engineering, Vol. 5, pp. 305-322, 1989.
[21] M. Tomita, "An efficient augmented-context-free parsing algorithm," Computational Linguistics, Vol. 13, No. 1-2, pp. 31-46, 1987.
[22] J. Early, "An Efficient Context-Free Parsing Algorithm," Communications of the Association for Computing Machinery, Vol. 13, No. 2, pp. 94-102, 1970.

Fai Wong received the B.Sc. and M.Sc. degrees in Software Engineering from the University of Macau in 1995 and 1999, and Ph.D. degree in Automation from Tsinghua University in 2005. He is currently an Assistant Professor in the Department of Computer and Information Science at University of Macau, with a secondary appointment in the Instituto de Engenharia de Sistemas e Computadores de Macau. His active and diverse research interests span areas of natural language processing and computational linguistic. His current research interests include machine translation, snapshot translation and machine learning to natural language processing. He is a member of IEEE and ACM.

Sam Chao, received the B.Sc., M.Sc. and PhD degrees in Software Engineering from the University of Macau in 1994, 1999 and 2008 respectively. Since 1996, she has been with the Department of Computer and Information Science at the University of Macau, currently is an Assistant Professor. She has participated in the management and development of more than twelve local and international commercial and research projects. Her current research focuses are data mining and machine learning technology, hybrid intelligent systems and knowledge acquisition in language and bioinformatics. She is a member of IEEE.

Wai Kit Chan received his B.Sc. degree in Software Engineering, 2009, from the Department of Computer and Information Science of University of Macau. His current research interest includes Chinese character recognition in large scale. He is currently a M.Sc. student in Computer and Information Science at University of Macau.