# An Automatic Sign Recognition and Translation System

Jie Yang, Jiang Gao, Ying Zhang, Xilin Chen and Alex Waibel
Interactive Systems Laboratory
Carnegie Mellon University
Pittsburgh, PA 15213
{yang+, jgao, joy, xlchen, ahw}@cs.cmu.edu

## ABSTRACT
A sign is something that suggests the presence of a fact, condition, or quality. Signs are everywhere in our lives. They make our lives easier when we are familiar with them. But sometimes they pose problems. For example, a tourist might not be able to understand signs in a foreign country. This paper discusses problems of automatic sign recognition and translation. We present a system capable of capturing images, detecting and recognizing signs, and translating them into a target language. We describe methods for automatic sign extraction and translation. We use a user-centered approach in system development. The approach takes advantage of human intelligence if needed and leverage human capabilities. We are currently working on Chinese sign translation. We have developed a prototype system that can recognize Chinese sign input from a video camera that is a common gadget for a tourist, and translate the signs into English or voice stream. The sign translation, in conjunction with spoken language translation, can help international tourists to overcome language barriers. The technology can also help a visually handicapped person to increase environmental awareness.

## Keywords
Sign detection, sign translation, vision-based interface, perceptive user interface.

## 1. INTRODUCTION
Individuals communicate with others using a variety of information systems and media in increasingly varied environments. One form of common communication media is a sign. Signs are everywhere in our lives. They suggest the presence of a fact, condition, or quality. They make our lives easier when we are familiar with them, but sometimes they pose problems or even danger. For example, a tourist or soldier might not be able to understand a sign in a foreign country that specifies military warnings or hazards. In this research, we are interested in signs that have direct influence upon a tourist from a different country or culture. These signs include, at least, the following categories:

- Names: street, building, company, etc.
- Information: designation, direction, safety advisory, warning, notice, etc.
- Commercial: announcement, advertisement, etc.
- Traffic: warning, limitation, etc.
- Conventional symbol: especially those are confusable to a foreign tourist, e.g., some symbols are not international.

At the Interactive Systems Laboratory of Carnegie Mellon University, we are developing technologies for automatically detecting, recognizing, and translating signs (Yang, 2001; Gao, 2001). Sign translation, in conjunction with spoken language translation, can help international tourists to overcome language barriers. It is a part of our efforts in developing a tourist assistant system (Yang, 1999). The proposed systems are equipped with a unique combination of sensors and software. The hardware includes computers, GPS receivers, lapel microphones and earphones, video cameras and head-mounted displays. This combination enables a multimodal interface to take advantage of speech and gesture inputs in order to provide assistance for tourists. The software supports natural language processing, speech recognition, machine translation, handwriting recognition and multimodal fusion.

A successful sign translation system relies on three key technologies: sign detection, optical character recognition (OCR), and language translation. At current stage of the research, we focus our efforts on sign detection and translation while taking advantage of state-of-the-art OCR technologies.

Automatic detection of signs from natural scenes is a challenging problem because signs are usually embedded in the environment. The work is related to text detection and recognition from video images, or so called Video OCR. Compared to video OCR tasks, sign detection takes place in a more dynamic environment. The user's movement can cause unstable input images. Non-professional equipment can make the video input poorer than that of other video


*PUI 2001*, November 15-16, 2001, Orlando, FL, USA.

OCR tasks, such as detecting captions in broadcast news programs. In addition, sign detection has to be implemented in real time using limited resources.

Sign translation has some special problems compared to a traditional language translation task. In general, the text used in the sign is short and concise. This characteristic of signs makes the lexical mismatch and structural mismatch become more severe problems. Furthermore, sign translation usually requires context or environment information because sign designers assume a human reader would use such information in understanding signs.

We address these challenges at both technology and system levels. We have developed a hierarchical framework for automatic sign detection and applied example-based machine translation (EBMT) (Brown, 1996) technology to sign translation. While developing technologies for automatic sign detection and translation, we would like to utilize human intelligence through nature interaction. A task that is difficult for a computer is not necessarily for a human, or vice versa. For example, if needed, a user can select an area of interest and domain for translation. If multiple signs have been detected within an image, a user can determine which sign is to be translated. The selected part of the image is then processed, recognized, and translated. By focusing only on the information of interest and providing domain knowledge, the approach provides a flexible method for sign translation. It can enhance the robustness of sign recognition and translation, and speed up the recognition and translation process. We have applied the technology to Chinese sign translation. The system can recognize Chinese sign input from a video camera that is a common gadget for a tourist, and translate the signs into English or voice stream.

The organization of this paper is as follows: In section 2, we describe system design. In section 3, we discuss methods for sign detection. In section 4, we induce the application of EBMT technology into sign translation. In section 5 we present an application of the system to Chinese sign translation and discuss evaluation results. In section 6, we conclude the paper.

## 2. SYSTEM DESIGN

An automatic sign recognition and translation system utilizes a video camera to capture the image with signs, detects signs in the image, recognizes signs, and translates results of sign recognition into a target language. Such a system relies on technologies of sign detection, OCR, and machine translation. It is also constrained by available resources such as hardware and software.

### 2.1 System Architecture

We would like to design a system that requires the minimum modification for working on different platforms and environments. In order to achieve such flexibility, we modularize the system into three modules: capture module, interactive module, and recognition and translation module. These modules work in client/server mode and are independent of each other. They can be on the same machine or different machines. Figure 1 shows the system architecture.

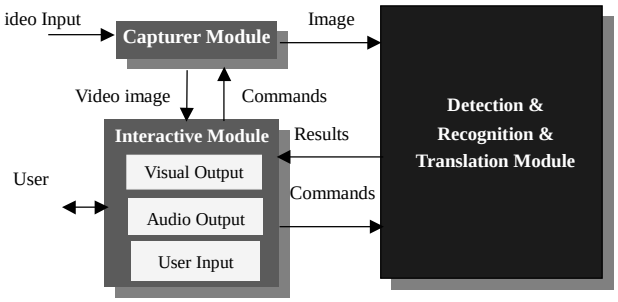


**Figure 1. System architecture.**

The capture module handles video input. It is hardware dependent. We are currently working on windows platform. The module supports both video for window and directX formats. We have tested the module using many different video and digital (DV) cameras through different input channels such as PCI card, PCMCIA card, and USB port.

The video stream or picture is inputted to the recognition and translation module for processing. The recognition and translation module is a key part of the system. The module first performs sign detection and only focuses on the text areas. The sign extraction results are also fed back to a user for potential. These sign regions are further processed and fed into the OCR engine, which recognizes the contents of the sign areas in the original language. Then, the recognition results are sent to the translation module to obtain an interpretation in target language. Under the Interlingual (Hutchins 1986) framework, both the source and target languages can be expanded extensively. Such extension makes a system work for multiple languages.

Interactive module provides an interface between a user and the system. A user-friendly interface is important for a user-centered system. It provides necessary information to a user through an appropriate modality. It also allows a user to interact with the system if needed. In our current system, the interface provides the recognition/translation results and allows a user to select sign regions manually or confirm sign regions extracted automatically. For example, a user can select a sign that he/she is interested to be translated directly if multiple signs have been detected; or in some unfavorable cases the automatic sign extraction algorithms may fail, but a user can still select a sign manually by circling the areas using pointing devices. This function also allows a user to obtain a translation for any part of a sign by selecting the parts where he/she is interested in.

## 2.2 User Interface of the System

A user-friendly interface is important for a user-centered system. It provides necessary information to a user through an appropriate modality. It also allows a user to interact with the system if needed. In our current system, the interface provides the recognition/translation results and allows a user to select sign regions manually or confirm sign regions extracted automatically. For example, a user can select a sign that he/she is interested to be translated directly if multiple signs have been detected; or in some unfavorable cases the automatic sign extraction algorithms may fail, but a user can still select a sign manually by circling the areas using pointing devices. This function also allows a user to obtain a translation for any part of a sign by selecting the parts where he/she is interested. The function of the user interface is summarized in Figure 2.

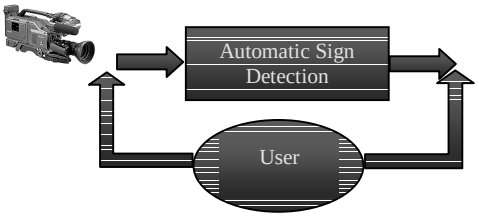


**Figure 2. The function of the user interface.**

User interface design is device dependent. A good interface for one device is not necessarily suitable for other devices. We are currently developing a version of the system for handheld devices and will discuss the problem in our future publications.

## 2.3 A Prototype System

We are currently working on Chinese sign translation. We choose Chinese for several reasons. First, Chinese is major language and very different from European languages. Second, a foreign tourist might have serious language barrier in China because English is not commonly used there. Third, Statistics shows that more people will visit China in the future. Finally, technologies developed for Chinese sign translation can be extended to other languages. The system currently uses commercial OCR software. We will detail the sign detection and translation technologies in section 3 and section 4.

Figure 3 is a screen shot of the current user interface. On the interface, the "unfreeze" button tells the system to input the next scene from a video camera. A user can directly draw on the captured image around the interested areas for recognition/translation, or simply touch on the "Auto Detection" button to let the system do the detection, and then click the detected areas to confirm. The black rectangle indicates the detected sign region. The translation result is overlaid on the image near the location of the sign. Figure 3 shows an example of automatic detection and translation. Figure 4 is an intermediate result of preprocessing/binarization for OCR module, and translation.


**Figure 3. Screen shot of the user interface.**

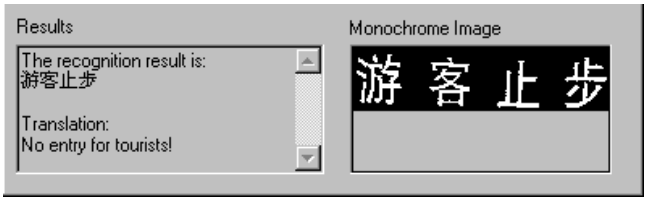


**Figure 4. Preprocessing/binarization for OCR.**

In addition to visual output, the system has audio output. We use Festival (Black 1998) for speech synthesis. Festival is a general multi-lingual speech synthesis system developed at the Center for Speech Technology Research, University of Edinburgh. Festival is a full text-to-speech (TTS) system with various APIs, and an extensive environment for development and research of speech synthesis techniques. For more detailed information, see (Black 1998).

## 3. SIGN DETECTION

Sign detection is related to character or text extraction/ recognition. The previous research falls into three different areas: (1) automatic detection of text areas from general backgrounds (Cui and Huang 1997, Jain and Yu 1998, Li and Doermann 1998, Lienhart 1996, Ohya 1994, Sato 1998, Wong 2000, Wu 1999, Zhong and Jain 1995); (2) document input from a video camera under a controlled environment (Taylor 1999); and (3) enhancement of text areas for character recognition (Lienhart 1996; Ohya 1994, Sato 1998, Watanabe 1998, Wu 1999, Li and Doermann 2000).

The first area is relevant to this research. In this area, most research focus on extracting text from pictures or video, although a few are focused on character extraction from

vehicle license plates (Cui and Huang 1997). Many video images contain text contents. Such texts can be part of the scene, or come from computer-generated text that is overlaid on the imagery (e.g., captions in broadcast news programs). The existing approaches of text detection include: edge filtering (Zhong and Jain 1995); texture segmentation (Wu 1999); color quantization (Jain and Yu 1998); and neural networks and bootstrapping (Lienhart 1996, Li and Doermann 1998). Each of these approaches has its advantage/drawbacks concerning reliability, accuracy, computational complexity, difficulty in improvement, and implementation. Although text with a limited scope can be successfully detected using existing technologies, such as in high quality printed documents, it is difficult to detect signs with varying size, embedded in the real world, and captured using an unconstrained video camera.

Fully automatic extraction of signs is a challenging problem because signs are usually embedded in the environment. Compared with other object detection tasks, many information sources are unavailable for a sign detection task, such as:

- No motion information: signs move together with background;
- No approximate shape information: text areas assume different shapes;
- No color reflectance information: signs assume different colors.

In order to deal with dynamic environment of sign detection, we propose a new sign algorithm based on an adaptive search strategy. The algorithm embeds the adaptive strategy in a hierarchical structure, with different emphases at each layer as shown in Figure 5.

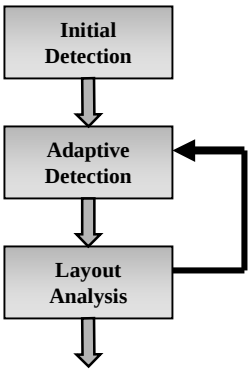


**Figure 5. The block diagram of sign detection algorithm.**

In this three-layer structure, the first layer detects possible sign regions. Since lighting and contrast might vary dramatically in natural scenes, the detection algorithm has to work differently under different circumstances. We utilize a multi-resolution approach to compensate these variations and eliminate noises in the edge detection algorithm; i.e., we apply edge detection algorithm using different scale parameters, and then fuse the results from different resolutions.

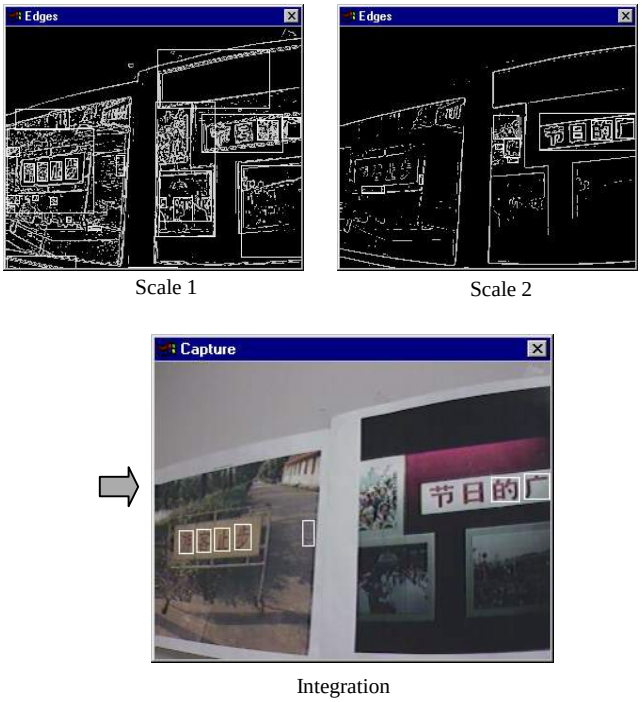


**Figure 6. Initial sign detection using edge detectors.**

Figure 6 illustrates the algorithm. The two signs (in the lower left and upper right position of the image) have different contrasts and lightings, but can be optimally detected using the edge detection algorithm with different scale factors. The sign in the left can be segmented in scale 1, but cannot be detected in scale 2; part of the sign on the right can be detected in scale 2, but is difficult to extract in scale1. The integration result is obtained by combining the detection results from different scales.

The second layer of the framework performs adaptive search. The adaptive search strategy is constrained by two factors: initial candidates detected by the first layer and layout of the signs. More specifically, the search starts from the initial candidate but the search *directions* and *acceptance criterion* are determined by taking the *layout* of signs into account.

While most signs in western languages are in the horizontal direction, Chinese signs in both horizontal and vertical directions are commonly used. One reason might be that Chinese language is rather character based than word based. Some special signs are designed in specific shapes for aesthetics reasons. We will ignore these layouts at this stage. In addition to directions, we can use shape and color criteria to discriminate different signs. Using these heuristics, we designed searching strategy and criterion

under the constraints above, which we called the *syntax* of sign layout. In fact, it plays the similar role as syntax in language understanding when parsing sentences, except that it is used to discriminate different layouts to assist the adaptive searching of sign regions.

In a simple case, the color or intensity of a sign does not change significantly. In such a case, we can use color information of the initially detected text to extract characters with similar attributes from the searched areas, based on the layout syntax. In some situations, however, colors within a sign region change dramatically due to lighting sources or the quality of the sign. In those situations, we can still use the approximate location, direction, and size of the sign characters to extract the characters from the background, but we cannot use the color information of the initially detected text directly.

Most OCR algorithms separate text from background using a "binarization" process. The hidden assumption of such approach is that both color of background and text are in uniform distributions. This assumption is often invalid in sign extraction. As shown in Figure 7, due to the lighting and noise conditions in natural scenes, the color compositions of signs differ in different locations.

We use a Gaussian mixtures model to segment signs. The number of Gaussian mixtures is crucial in appropriately modeling the background. Inappropriate selection of Gaussian mixtures will result in errors in text detection. We determine the number of Gaussian mixtures by considering if the characters can be extracted from the background, and the characters have to satisfy the layout syntax.


(a)


(b)


(c)

**Figure 7. Adaptive character extraction using color modeling: (a) Original sign, (b) Color space modeled by two Gaussian mixtures, (c) Color space modeled by three Gaussian mixtures.**

The number of Gaussian mixtures can be changed in each adaptation area. Figure 7 shows an example of sign extraction using the Gaussian mixture models. We are interested in extracting the sign from the area within the white rectangle. Figure 7 (a) is the original sign, and Figure 7 (b) and Figure 7 (c) are segmentation results using two and three Gaussian mixtures, respectively. The rightmost character is confused with the background if using only two Gaussin mixtures, but can be extracted using three Gaussian mixtures.

The objective of layout analysis is to align characters in an optimal way, so characters belong to the same sign will be aligned together. Chinese text layout has some unique features. Figure 8 is an example of a Chinese sign. Each character in the sign is composed of several connected sub-components, sometimes the sub-components align in the same way as characters in a sign. This case is very common in Chinese signs. However, such a sign poses a big problem to automatic sign layout analysis: how can a system know if a text region is a character or only segment of a character, without recognition of the whole text area? We use layout analysis techniques, which utilize various heuristics, to deal with this problem.

A major contribution of the new framework lies in its ability to refine the detection results using local color and layout information. We are considering incorporating more information to this framework to further enhance the detection rate. The detailed algorithm can be found in the Technical Report (Gao, 2001).


**Figure 8. Chinese characters are multi-segment.**

Figure 9 shows an example of a detection result. White rectangles indicate detected sign regions. In our Chinese sign database, signs were taken by a high resolution digital camera and printed out on papers. During the test, the signs are caught by a video camera and detected in real time. In Figure 9, the system captures two signs with different backgrounds from a non-favorite angle. The system can still successfully detect all the signs. This demonstrates that the detection framework provides considerable flexibility to allow the detection of slanted signs and signs with non-uniform character sizes.


**Figure 9. An example of sign extraction result.**

## 4. SIGN TRANSLATION

Sign translation is different from a traditional language translation task in some aspects. The lexical requirement of a sign translation system is different from an ordinary machine translation (MT) system, since signs are often filled with abbreviations, idioms, and names, which do not usually appear in formal languages. The function of a sign requires it be short and concise. The lexical mismatch and structural mismatch problems become more severe in sign translation because shorter words/phrases are more likely to be ambiguous due to insufficient information from the text to resolve the ambiguities. Furthermore, sign translation is sensitive to the domain of the sign: lexical in different domains has different meaning. However, domain identification is difficult because the signs are concise and provide few contexts. For structural matching, the system needs to handle ungrammatical language usage, which is common in signs.

Moreover, imperfect sign recognition makes sign translation more difficult. Though in many cases human being can correctly "guess" the correct meaning using context knowledge even with erroneous input, for MT systems, it is still a difficult problem. In summary, with the challenges mentioned above, sign translation is not a trivial problem that can be readily solved using the existing MT technology.

In the existing MT techniques, the knowledge based MT system works well with grammatical sentences, but it requires a great amount of human effort to construct its knowledge base, and it is difficult for such a system to handle ungrammatical text which appears frequently in signs.

On the other hand, Statistical MT and Example Based Machine Translation (EBMT) (Brown 1996) enhanced with domain detection is more appropriate to a sign translation task. This is a data-driven approach. What EBMT needs are a bilingual corpus and a bilingual dictionary where the latter can be constructed statistically from the corpus. Matched from the corpus, EBMT can give the same style of translations as the corpus. Our translation system is based on this approach. In addition, we can use a database search method to deal with names, phrases, and symbols related to tourists.

We start with the EBMT software (Brown 1996, Brown 1999). The system is used as a shallow system that can function using nothing more than sentence-aligned plain text and a bilingual dictionary. Given sufficient parallel texts, the dictionary can be extracted statistically from the corpus (Brown 1997). In the translation process, the system looks up all matching phrases in the source-language and performs a word-level alignment on the entries containing matches to determine a (usually partial) translation. Portions of the input for which there are no matches in the corpus do not generate a translation.

Because the EBMT system does not generate translations for 100% of its input text, a bilingual dictionary and phrasal glossary are used to fill any gaps. Selection of a "best"' translation is guided by a trigram model of the target language and a chart table (Hogan and Frederking 1998). In this work, we extended the EBMT system to handle Chinese input by adding support for the two-byte encoding used for the Chinese text (GB-2312).

The segmentation of Chinese words from character sequences is important for translation of Chinese signs. This is because the meaning of a Chinese sentence is based on words, but there is no explicit tags around Chinese words. A module for Chinese word segmentation is included in the system. This segmentor uses a word-frequency list to make segmentation decisions.

We tested the EBMT based method using 50 randomly selected signs from our database, assuming perfect sign recognition in the test. We first tested the system using a Chinese-English dictionary from the Linguistic Data Consortium (LDC), and a statistical dictionary built from the HKLC (Hong Kong Legal Code) corpus. As a result, we only obtained about 30% reasonable translations. We then trained the system with a small corpus of 670 pairs of bilingual sentences (Kubler 1993), the accuracy is improved from 30% to 52% on 50 test signs. It is encouraging that the improvement in EBMT translations is obtained without requiring any additional knowledge resources. We will further evaluate the improvement of the translation quality, when we combine words into larger chunks on both sides of the corpus.

We have not yet taken full advantage of the features of the EBMT software. In particular, it supports equivalence classes that permit generalization of the training text into templates for improved coverage. We intend to test automatic creation of equivalence classes from the training corpus (Brown 2000) in conjunction with other improvements reported herein. In addition, we will take advantage of domain information to further improve the translation quality.

## 5. SYSTEM EVALUATION

We have evaluated the prototype system for sign detection and translation. We have built up a database containing about 800 Chinese signs taken from China and Singapore. We tested our sign detection algorithm using 50 randomly selected signs from our sign database. Table 1 gives the automatic sign detection results. By properly selecting parameters, we can control the ratio of miss detection and false alarms. Presently, such parameters are selected according to user's preferences, i.e. acceptability of different types of errors from users' point of view.

**Table 1. Test results of automatic detection on 50 Chinese signs**

| Detection without missing characters | False alarms | Detection with missing characters |
|---|---|---|
| 45 | 8 | 5 |

We have tested the robustness of the detection algorithms by changing conditions, such as image resolution, camera view angle, and lighting conditions. The algorithms worked fairly well for low resolution images (e.g., from 320 x 240 to 80 x 60). The algorithms can also handle signs with significant slant, various character size and lighting conditions. More detailed experimental results can be found in the Technical Report (Gao, 2001).

We also tested the EBMT based sign translation method. We assumed perfect sign recognition in our tests. For 50 randomly selected signs, some examples of sign translation are given in Table 2. These examples are selected from reasonable translation results.

**Table 2. Examples of translation from Chinese to English**

| *Original Chinese* | *English Translation* |
|---|---|
| 行人出口 | Pedestrian exit |
| 儿童医院 | Child hospital |
| 查询电话 | Inquiry telephony |
| 各种车辆 | Various kinds vehicle |
| 古生物化石陈列室 | Old living creature fossil exhibit room |

Figure 10 illustrates the error analysis of the translation result. It is interesting to note that 40% of errors come from mis-segmentation of Chinese words. Obviously, there is a significant room for improvement in word segmentation. The improvements in proper name and domain detection can also enhance the accuracy of the system significantly.

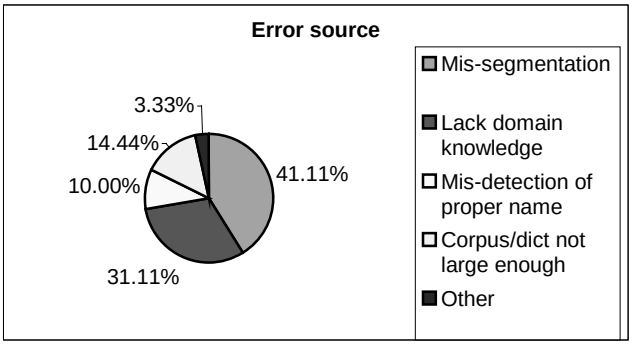


**Figure 10. Error analysis of the translation experiment.**

## 6. CONCLUSIONS

We have presented an automatic sign recognition and interpretation system in this paper. Sign recognition and translation can assist international tourists to overcome language barriers and help a visually handicapped person to increase environmental awareness. While developing technologies for automatic sign detection, recognition, and translation, we have attempted to take advantage of human intelligence in system development. We have proposed a framework for automatic detection of signs from natural scenes. The framework considers critical challenges in sign extraction and can extract signs robustly under different conditions (image resolution, camera view angle, and lighting). We have extended EBMT method to Chinese sign translation and demonstrated its effectiveness and efficiency. We are further improving the robustness and accuracy of our system.

There is a big room to further improve the sign detection and translation methods. For example, it is possible to eliminate false detection by combining sign detection with OCR. The confidence of the sign extraction system can be improved by incorporating the OCR engine in an early stage. We are particularly interested in enhancing translation quality, including for an imperfect OCR system.

## ACKNOWLEDGEMENTS

We would like to thank Dr. Ralf Brown and Dr. Robert Frederking for providing initial EBMT software. We would like to thank William Kunz and Jing Zhang for their help on developing the interface for the prototype system. We would also like to thank other members in the Interactive Systems Labs for their inspiring discussions and support. This research is partially supported by DARPA under TIDES project.

## References

[1] Black, A.W., Taylor, P., and Caley, R., Festival, *www.cstr.ed.ac.uk/projects/festival.html*, The Centre for Speech Technology Research (CSTR) at the University of Edinburgh, 1998.

[2] Brown, R.D., Adding Linguistic Knowledge to a Lexical Example-Based Translation System. *Proceedings of the Eighth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-99)*, pp. 22-32, Chester, England, August, 1999.

[3] Brown, R.D., Automated Dictionary Extraction for "Knowledge-Free" Example-Based Translation. *Proceedings of the Seventh International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-97)*, pp. 111-118, Santa Fe, New Mexico, July, 1997.

[4] Brown, R.D., Automated Generalization of Translation Examples. *Proceedings of the Eighteenth International Conference on Computational Linguistics (COLING-2000)*, pp. 125-131, 2000.

[5] Brown, R.D., Example-based machine translation in the pangloss system. *Proceedings of the 16th International Conference on Computational Linguistics*, pp. 169-174, 1996.

[6] Cui, Y. and Huang, Q., Character Extraction of License Plates from Video. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 502-507, 1997.

[7] Gao, J., Yang, J., Zhang, Y., Waibel, A., Text Detection and Translation from Natural Scenes, Technical Report CMU-CS-01-139, Computer Science Department, Carnegie Mellon University, June, 2001.

[8] Gao, J. and Yang, J., "An Adaptive Algorithm for Text Detection from Natural Scenes," Proceedings of Computer Vision and Pattern Recognition (CVPR 2001).

[9] Hogan, C. and Frederking, R.E., An Evaluation of the Multi-engine MT Architecture. Machine Translation and the Information Soup: *Proceedings of the Third Conference of the Association for Machine Translation in the Americas (AMTA '98)*, vol. 1529 of Lecture Notes in Artificial Intelligence, pp. 113-123. Springer-Verlag, Berlin, October.

[10] Hutchins, John W., Machine Translation: Past, Present, Future, Ellis Horwood Limited, England, 1986.

[11] Jain, A.K. and Yu, B., Automatic text location in images and video frames. *Pattern Recognition*, vol. 31, no. 12, pp. 2055-2076, 1998.

[12] Kubler, Cornelius C., "Read Chinese Signs". Published by Chheng & Tsui Company, 1993.

[13] Li, H. and Doermann, D., Automatic Identification of Text in Digital Video Key Frames, *Proceedings of IEEE International Conference of Pattern Recognition*, pp. 129-132, 1998.

[14] Li, H. and Doermann, D., Superresolution-Based Enhancement of Text in Digital Video. *ICPR*, pp. 847-850, 2000.

[15] Lienhart, R., Automatic Text Recognition for Video Indexing, *Proceedings of ACM Multimedia 96*, pp. 11-20, 1996.

[16] Ohya, J., Shio, A., and Akamatsu, A., Recognition of characters in scene images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 214-220, 1994.

[17] Sato, T., Kanade, T., Hughes, E.K., and Smith, M.A., Video OCR for digital news archives. *IEEE Int. Workshop on Content-Based Access of Image and Video Database*, 1998.

[18] Taylor, M.J., Zappala, A., Newman, W.M., and Dance, C.R., Documents through cameras, *Image and Vision Computing*, vol. 17, no. 11, pp. 831-844, 1999.

[19] Waibel, A., Interactive Translation of Conversational Speech, *Computer*, vol. 29, no. 7, 1996.

[20] Watanabe, Y., Okada, Y., Kim, Y.B., and Takeda, T., Translation camera, *Proceedings Fourteenth International Conference on Pattern Recognition*, pp. 613-617, 1998.

[21] Wong, E. K. and Chen, M., A Robust Algorithm for Text Extraction in Color Video, *Proceedings of IEEE Int. Conference on Multimedia and Expo (ICME2000)*, 2000.

[22] Wu, V., Manmatha, R., and Riseman, E.M., Textfinder: an automatic system to detect and recognize text in images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pp. 1224-1229, 1999.

[23] Yang, J., Yang, W., Denecke, M., and Waibel, A., Smart sight: a tourist assistant system. *Proceedings of Third International Symposium on Wearable Computers*, pp. 73-78. 1999.

[24] Yang, J., Gao, J., Yang, J., Zhang, Y., Waibel, A., Towards Automatic Sign Translation, Proceedings of Human Language Technology 2001.

[25] Zhong Y., Karu, K., and Jain, A.K., Locating Text in Complex Color Images, Pattern Recognition, vol. 28, no. 10, pp. 1523-1536, 1995.