SIGN LANGUAGE TRANSLATION USING AN ERROR TOLERANT RETRIEVAL ALGORITHM

Chung-Hsien Wu, Yu-Hsien Chiu, and Kung-Wei Cheng

Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan, R.O.C.

{chwu, chiuyh, kungwei}@csie.ncku.edu.tw

ABSTRACT

This paper addresses an error-tolerant retrieval algorithm for generating sentences from ill-formed sign sequences in Taiwanese Sign Language (TSL). The design methodology is motivated by the kinematics of hand gestures for sign language. In order to increase the input rate and retrieval accuracy, the basic design strategy leads to develop an efficient and effective sign feature retrieval method. In this approach, a Multi-list Code Tree (MCT) data structure for sign feature indexing and an error-tolerant matching algorithm are proposed. For text generation, the optimal path in the word graph, generated from the sign input sequence, is incrementally estimated by using a translation model. Several interface design and evaluation methods are also conducted for empirical study. Evaluation results show that the generation rate and retrieval accuracy for sentence construction are significantly improved.

1. INTRODUCTION

Augmentative and Alternative Communication (AAC) is an area of clinical practice that attempts to compensate for the impairment and disability patterns of individuals with severe expressive disorders. The underlying methodology is to facilitate the disabled with minimum efforts to generate comprehensive message for improving communication disabilities, especially for the profoundly hearing-impaired students. Current research mostly focused on the purpose of developing more efficient user interfaces and of extending the accessibilities [1]. The concept of virtual keyboard is widely used as an alternative interface for generating text. Moreover, the corresponding trade-off design issues between increased motor and decreased cognitive loads were suggested.

Recently, iconic languages have been used successfully in human computer interface and visual programming. However, people with certain degree of disability may lose their ability to sense symbols, voice or the ability to present their ideas. For compensation purpose, alternative methods must be provided to restore their privilege of communication. For examples, Braille is used as a visual aid for the compensation of visual impairment. Icon-based symbols, such as Picture Communication Symbol (PCS), Picosyms and MinSpeak, are alternate methods of communication to help people with difficulty in understanding abstraction of characters [1,4].

In this paper, an error-tolerant sign retrieval algorithm based on kinematic hand configuration information, which consists of handshape movement and location of Taiwanese Sign Language (TSL), is proposed to be an alternative indexing method. More specifically, the study focuses on: 1) investigating the morphology and typology of TSL, 2) organizing the sign features hierarchically as a multilist code tree structure for sign retrieval, 3) developing a sign feature error-tolerant retrieval method for efficient access, and 4) developing a trigger-pair based alignment model for incremental text generation from ill-formed sign feature sequences. The proposed approach aims to improve speech communication ability and activities in daily life for users using TSL. The design concept of sign features takes more considerations to human factors, visual concentration and eye-hand coordination and is easy to be integrated into various interface control units. In addition to speeding up sentence generation rate for individuals, the proposed system has potential to assist people with language disabilities in sentence formation.

2. MULTILIST CODE TREE

The signs in TSL perform functionally similar to the words in natural language. The sub-lexical units in combination with each other make up the morphemes of the language. Stroke first proposed an idea of developing a written sign language code [2]. Parallel to American Sign Language (ASL), the behaviors in TSL are categorized into three types and used as the features for coding a sign:

- **tab**: the location of the sign in relation to the body;
- **dez**: the handshape of the hand;
- **sig**: the movement executed by the hands.

According to Strokeo’s analysis, 21 distinct locations (tab), 50 distinct handshapes (dez), and 12 distinct movements (sig) are used as the basic components of signs [2][3]. Based on this analysis, the lexical entry for each sign can be specified in terms of these types. We define 10 sign features to represent each sign shown in Table 1. This code takes into account the hand orientation, which carries important information in sentence formation, especially for identifying the subject term and word order.

<table>
<thead>
<tr>
<th>Initial Position</th>
<th>Final Position</th>
<th>Movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Hand</td>
<td>Right Hand</td>
<td>dez</td>
</tr>
<tr>
<td>dez</td>
<td>tab</td>
<td>dez</td>
</tr>
<tr>
<td>tab</td>
<td>dez</td>
<td>tab</td>
</tr>
<tr>
<td>dez</td>
<td>tab</td>
<td>sig</td>
</tr>
<tr>
<td>tab</td>
<td>sig</td>
<td>sig</td>
</tr>
</tbody>
</table>

Given the preceding definition, the typology of signs can be further categorized into two main configurations based on distinct types of motor acts as follows: 1) one-handed signs with/without contact to the body; 2) two-handed signs with the
same or different handshapes. These kinds of motor acts are described as follows:

**Type 1**: One-handed signs articulated in succession with a dez, a sig and a tab.

**Type 2**: Two-handed signs articulated in succession with the same dez for two hands, a sig, and a tab.

**Type 3**: Two-handed signs articulated in succession with a primary dez with a sig and a secondary dez with a tab.

In order to retrieve a sign efficiently and accurately, a sign feature database was collected and organized as a multilist code tree (MCT) structure, similar to the trie data structure. Each sign $V^i$ can be represented as an ordered vertex list $V^i = \{f^j\}$ (j=1 to 4), in which $f^j$ represents one of the sign features: tab, dez, and sig. For building the MCT, the sign features, each denoted by a node, are sequentially linked with a downward directed pointer and each terminal node represents a unique sign. In each layer of the MCT, every node has its corresponding sub-tree. The nodes in the same linking layer may correspond to similar signs. This structure is to achieve an efficient and effective access of a sign. An example of the MCT is shown in Fig. 1. In the higher level, the downward directive pointers represent the constraints of morpheme structure. These constraints guide the direction to efficiently retrieve the desired sign. The transversely directive pointers are used to link the possible error patterns between sign feature categories for error-tolerant retrieval. This information is derived by way of empirical interface evaluation stated below. The hierarchical structure is highly correlated to the user’s convention.

![Fig. 1 An example of the multi-list code tree structure](image)

### 3. ERROR-TOLERANT SIGN FEATURE RETRIEVAL

From the analysis of sign typology, three pattern-matching cases are faced in the retrieval process:

- Exactly match a path in MCT
- Partially match some sign features as a result of node deletion, one-handed signing or error inputs
- No match caused by out-of-vocabulary

Formally, while the first feature is identified, the pointer chain provides all secondary features in addition to the primary feature in MCT. The sub-trees can then be searched in parallel to reduce the search time and the time complexity for n input features is equal to $C(n(n+1)/2)$. While this number is much smaller than $n^2$ for exhaustive search. If the successive input features match part of the features in MCT, the related signs will be output to the users. Fig. 2 shows an example for sign retrieval.

![Fig. 2 An example for sign retrieval](image)

Fortunately, the last two states frequently take place in practice. One obvious way to overcome this search issue is to adopt the concept of equivalence class in which an alternate in the same sub-tree may have similar meaning. In this paper, an error tolerant likelihood measure function is defined as follows:

$$LR(V', V^i) = \begin{cases} \sum_{f \in f'} \text{simi}(V', V^i) & \text{if } V'_i = V^i \\ \sum_{f \in f'} \text{simi}(V', V^i) & \text{if only } f^i_s \text{ are different} \\ \end{cases}$$

(1)

where function $\text{simi}(\cdot)$ represents the similarity and takes the path length between two nodes in different sub-trees into account. It is defined as follows:

$$\text{simi}(f^i_s, f^j_s) = \log(N/D)$$

(2)

where $N$ is the number of nodes in the sub-tree; $D$ is the maximum depth. Moreover, the function $RP$ is used to recover the possible feature nodes for the situation of no match. EP represents the error patterns derived from empirical study. Each $RP$ is re-estimated using the mutual information criterion described as follows:

$$RP(V'_i \rightarrow V^i; EP) = \log \frac{P(V'_i \rightarrow V^i; EP)}{P(V'_i \rightarrow EP)P(EP)}$$

(3)

$EP$ is used to retrieve an associated possible dez $V'_i$ from the most frequently wrong-judged candidates dez $V^i$. These error patterns are collected from empirical study of confusion set of sign features and trained using the Apriori algorithm [7] given the support value of 0.01 and the confidence value of 0.1. The $RP$ with high probability are extracted. Finally, The terminal nodes conformed to the criteria stated in equation (1) are...
considered to be the candidates. The successive candidates form a word graph, which are ranked.

4. INCREMENTAL TRANSLATION

Based on the observation and comparison between written and signed Chinese, several linguistic differences, such as word order, quantifier, function word, conjunction and interrogative, are encountered. These describe our main design issues with respect to structural translation. Motivated by this reason, the trigger pair concept was adopted for extracting highly correlated word pairs. A bilingual training corpus is collected from the textbook in the primary school and the parallel sentences were tagged by the experts. Due to the problem of data sparseness, the part-of-speech class-based concept is adopted. The bilingual translation pattern set TP and grammatical regularities are also trained using the Apriori algorithm.

All the possible fragments in word graph are merged incrementally to form a new keyword or key-phrase sequence. In order to generate well-formed sentences, a statistical translation approach is adopted [5]. Given a signed Chinese sentence $S_i$, it is to be translated into a Chinese sentence $T_j$. Among all possible target strings, the string with the highest probability is estimated using the Bayes’ rule:

$$
\hat{T}_j = \arg \max_{T_j} \{ p(T_j | S_i) \}
$$

$$
= \arg \max_{T_j} \{ p(T_j) \times p(S_i | T_j) \}
$$

(4)

where $p(T_j)$ is the target language model; $p(S_i | T_j)$ is the string translation probability. The translation probability is reformulated as follows:

$$
p(S_i | T_j) = \sum_{K} p(TP_{i}^{K}, WS_{i}^{K}, S_{i}^{K} | T_j^{K})
$$

(5)

where $WS_{i}^{K}$ represents the word sense. In addition to language structure differences, for generating a complete sentence, the function word is needed to be filled into a suitable position. Finally, a target language model is simultaneously used to search the best string with the highest probability. For example, the parsing process accepts the signed sequence (我 (my mother/Na/agent), 媽媽 (mother/Na/agent), 紅 (red/Nd), 蘋果 (apple/Na)), 二 (two/Neu)) to generate a new phrase (我媽媽 (my mother/Na/agent), 兩顆紅色的蘋果 (two red apples/Na)).

When the sign feature sequences are input in succession, the process can incrementally and automatically output the candidate sentences.

5. EXPERIMENTAL RESULTS

5.1. Perplexity Evaluation of Training Corpus

The training corpus is collected from the textbooks from the 1st up to the 5th grade of the primary deaf school in Taiwan. This corpus contains 1531 Chinese sentences with a mean length of 4.9 words. This corpus is further transcribed into the corresponding TSL corpus by TSL experts. 1873 most frequently used TSL words are widely collected from the education and training materials in the deaf schools. These TSL words are also tagged by using our proposed sign feature definition. By adopting the general signing convention in succession with a dez, a dez, a sig, and a tab, an MCT is conducted for the following experiments. In this paper, we adopted the bi-gram language model as the baseline and compared with the model with POS features and the model with translation patterns. Table 2 shows the perplexity comparison of the three models and the improvement using our proposed method.

Table 2. Perplexity evaluation results for the baseline, POS-based, translation pattern (TP) based models.

<table>
<thead>
<tr>
<th></th>
<th>Perplexity</th>
<th>Reduction rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word (Baseline)</td>
<td>5.4</td>
<td>-</td>
</tr>
<tr>
<td>POS Based</td>
<td>4.74</td>
<td>12.22%</td>
</tr>
<tr>
<td>TP based</td>
<td>4.43</td>
<td>17.96%</td>
</tr>
</tbody>
</table>

5.2. Experiments for Handshape Clustering

For developing a sign feature input interface, 85 possible symbols are constructed. The ambiguous information for similar codes in each category is reduced manually. According to the Fitts’ law [6], the movement time $MT$ to select a target of width $W$ that lies at distance or amplitude $A$ is formulated as follows:

$$
MT = a + b \log_{2} \left( A/W + 0.5 \right)
$$

(6)

where parameters $a$ and $b$ are observable constants depending on the user’s response. The design issue shows small $A$ and large $W$ result in small $MT$. In this paper, several experiments were conducted to select suitable classified catalogue for dez and sig. The objective evaluation based on shape similarity and ease of handshape articulation was used. This concept considers the anatomy and physiology of different handshapes. Each finger takes on four different configurations: open, curved, bent and closed. We further applied the formula proposed by [3] to calculate the ease score for clustering. The three criteria used in the ease score are Impendent Extensor Criterion (IEC), Profundus Criterion (PC) and Muscle Opposition in the configuration of selected Fingers Criterion (MOC of SFC). This formula is defined as follows:

$$
Ease\_Score = (IEC + PC) \times (MOC\_of\_SFC)
$$

(7)

Several candidate-clustering configurations were derived. Fig. 3 shows an example of row-column arrangement layout with 2 layers and 5 categories.

![Fig.3. Testing interface for handshape clustering](image-url)
5.3. Experiments for Error Pattern Analysis

The subjective evaluation was conducted to assess the accuracy and error rate of selecting correct handshape associated with a balanced testing vocabulary. Moreover, a formula based on this paradigm was also proposed as follows:

$$T_{av} = \frac{1}{N} \sum_{i=1}^{N} \left( TP_i \left( \mu * P_i \right) + TR_i \left( R_i \right) + TC_i \left( C_i \right) \right)$$

where $T_{av}$ is the average scanning time; $N$ is the total number of selecting steps for each word; $\mu$ is the number of pages; $P_i$, $R_i$ and $C_i$ represents the page, row and column numbers, respectively. In this formula, for each scanning process, the scanning time and location of each step are recorded and aimed to analyze user’s responses. The information of extracting a correct feature associated with the preceding wrong-judged features can provide the development of error-tolerant retrieval mechanism.

According to the subjective and objective evaluation, a PC-based sign feature retrieval interface was developed for function test. Fig. 4 shows the interface for sign feature retrieval. In this paradigm, 52 handshapes are categorized into 5 classes using the Ease Score and are arranged in a two-layer row column layout. 12 movements/orientations are merged into 5 classes without taking the depth information of vision into accounts. In the central region, a virtual robot is proposed with 21 positions. Through the practice evaluation of handshape and movement categorization, 300 balanced testing vocabulary and 15 profoundly deaf students were asked to conduct the evaluation. The average error rate is $1.1 \pm 0.24$ times for correct retrieval of a sign. The statistical testing using one-way ANOVA [8] to analyze the variations between users (inter-effect) and testing patterns (intra-effect) explores the consistency with highly significant.

![Fig. 4. Sign feature retrieval interface](image)

5.4. Experiments for Sentence Construction

For evaluating our proposed retrieval and text generation strategies, word prediction approach is regarded as the baseline system. The scanning time based on equation (8) is estimated. The experimental results show that the scanning time reduction rate for sentence pattern prediction and sign feature retrieval method were 42.86% and 59.74%, respectively. Table 3 shows a comparison of the various prediction strategies in the evaluation of keyword selection. Compared with keystroke saving rate, the previous two methods have better performance than our approach. But, obviously, our approach shows better performance on text generation rate and less cognition load in memorizing symbols.

<table>
<thead>
<tr>
<th>Prediction Strategy</th>
<th>Average Scanning</th>
<th>Improvement Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word prediction</td>
<td>15.4</td>
<td>-</td>
</tr>
<tr>
<td>Sentence pattern</td>
<td>8.8</td>
<td>42.86%</td>
</tr>
<tr>
<td>Sign feature retrieval</td>
<td>6.2</td>
<td>59.74%</td>
</tr>
</tbody>
</table>

6. CONCLUSION

This paper proposed an innovative design methodology for sign feature retrieval and develop a new TSL AAC system for the deaf in need of communication aids. The MCT data structure with error tolerant retrieval algorithm shows the ability of efficiency and robustness. The incremental translation model has the potential to assist people in sentence formation. This prototype design is flexible for the TSL users. Evaluation results show that the generation rate and retrieval accuracy for sentence construction have been significantly improved. In the future, we will make efforts on porting the retrieval interface to a PDA platform and other multi-modal applications.

7. ACKNOWLEDGEMENT

The authors would like to thank the National Science Council, Republic of China, for its financial support of this work, under Contract No. NSC91-2614-H-006-003-F20.

8. REFERENCES