Hybrid HMM for Transitional Gesture Classification in Thai Sign Language Translation

Arunee Jaruwanawat, Nopporn Chotikakamthorn, and Worawit Werapan

Faculty of Information Technology &
Research Center for Communications and Information Technology
King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand
(E-mail: s5066025@kmitl.ac.th)

Abstract: A human sign language is generally composed of both static and dynamic gestures. Each gesture is represented by a hand shape, its position, and hand movement (for a dynamic gesture). One of the problems found in automated sign language translation is on segmenting a hand movement that is part of a transitional movement from one hand gesture to another. This transitional gesture conveys no meaning, but serves as a connecting period between two consecutive gestures. Based on the observation that many dynamic gestures as appeared in Thai sign language dictionary are of quasi-periodic nature, a method was developed to differentiate between a (meaningful) dynamic gesture and a transitional movement. However, there are some meaningful dynamic gestures that are of non-periodic nature. Those gestures cannot be distinguished from a transitional movement by using the signal quasi-periodicity. This paper proposes a hybrid method using a combination of the periodicity-based gesture segmentation method with a HMM-based gesture classifier. The HMM classifier is used here to detect dynamic signs of non-periodic nature. Combined with the periodic-based gesture segmentation method, this hybrid scheme can be used to identify segments of a transitional movement. In addition, due to the use of quasi-periodic nature of many dynamic sign gestures, dimensionality of the HMM part of the proposed method is significantly reduced, resulting in computational saving as compared with a standard HMM-based method. Through experiment with real measurement, the proposed method's recognition performance is reported.

Keywords: Sign Language, Gesture Recognition, Hidden Markov Model

1. INTRODUCTION

Sign languages are used in most parts of the world among community of deaf people. A sign language can also be used as a means for communication between a normal and a deaf persons. Research has been carried out as part of an attempt to automatically recognize hand shapes and movements as defined in some sign languages [1]. Although this automated sign language recognition research may find a relatively less commercial and practical value as compared with the speech recognition counterpart, the result should be beneficial to both deaf and normal people in many ways. Not just it can be used to simplify communication between the deaf and other people, result from automated sign recognition research can be applied to improve human-computer interaction efficiency through gesture type of input.

Regardless of the way the hand shape and movement are captured, current sign language translation techniques are differentiated by the recognition approach employed. Examples of feasible approaches include a simple template matching, use of a artificial neural network, and approach based on the Hidden Markov Model (HMM) [4]. Among these approaches, it appears that the HMM principle has received a relatively more attention due to its ability in handling gesture time variation. Another reason for its popularity is its level of success when applied to the speech recognition problem. Example of HMM-based sign language translation research is the work of Vogler and Metaxas [3]. In [3], HMM principle was applied to recognize 53 signs drawn from the American sign language.

Although the HMM principle is an elegant approach for recognition of time-based patterns (such as speech and signs), the method requires high computational complexity when dealing with large patterns. The problem is remedied by the use of sub-word components (phoneme) in a speech recognition system. Similar idea was explored for automatic sign language recognition [2]. However, unlike speech, there is no standard or widely-accepted phoneme-like counterpart in a sign language. In addition, while continuous speech contains no easily-detected boundary points that can be used for word or sub-word segmentation, movement of a signer's hand while performing signs continuously contains pauses and turning points of the movement. It is thus arguable that by exploiting those natural segmental points, use of the HMM method may be reduced. Research works that exploit these features include those of [5]. In [5], such natural segmental points are exploited along with the (roughly) quasi-periodic nature of many dynamic signs as found in Thai sign language. However, certain dynamic signs in Thai sign language are non-periodic [5]. Therefore, those signs are distinguishable from a transitional movement by the method as described in [5].

This paper describes a hybrid Thai sign language translation method. The method is an extension of that as described in [5]. In particular, a HMM module is added to the system in [5] to deal with non-periodic dynamic signs. Use of such a hybrid scheme offers benefits of both the HMM-based and non HMM-based methods. While computational complexity of this hybrid method is slightly higher as compared with that of [5], it can deal with all types of signs: static, periodic, or non-periodic.

The paper is organized as follows. Section 2 describes major modules in our Thai sign language translation system. The proposed hybrid scheme based on the HMM principle is then described in Section 3. Experimental results are reported in Section 4. Last, discussion and conclusion remarks are also included.

2. THAI SIGN LANGUAGE TRANSLATION SYSTEM

A typical structure of a hand gesture word/sentence in Thai sign language is shown in Fig. 1. From the figure, each sign word/sentence is a sequence of hand gestures. Those gestures are either static or dynamic. A static gesture is a hand of a particular shape without any movement. A dynamic gesture is a hand in motion. A dynamic gesture can be further categorized into either (approximately) quasi-periodic and non-periodic gestures. Among the non-periodic gestures, some of them correspond to transitional movement while some correspond to non-periodic signs. A transitional movement serves as a connection between two consecutive gesture/posture, and conveys no meaning.

Next, consider the system used in this study. Our sign language translation system employs a right-hand instrumented glove, used in combination with a magnetic 6-DOF tracker device. Such devices as described are used here to ensure that data collected and result as obtained are not affected by the accuracy of the acquisition equipment. The architecture of the recognition system is as shown in Fig. 2. The system is composed of 7 modules. The first module performs static gesture detection. If the data obtained from the system's sensors is classified as non-static, the sampled data is passed to the second module to detect hand movement turning points. When this is the case, it is first assumed that the data samples correspond to a moving hand contain segment(s) of transitional movement, or non-periodic gesture, or periodic gesture, or those combination. The outputs of the second module are the sample numbers where change in hand velocity occurs. These turning points are then passed to the 'periodic/non-periodic' gesture classification module. In [5], based on the observation that most meaningful dynamic gestures in Thai sign language can be approximated as quasi-periodic (i.e., those gestures consist of certain hand movement patterns, cyclically repeated), a method was developed to classify samples of hand movement into periodic and non-periodic segments. Most of the data corresponding to non-periodic segments belong to a transitional movement. Some words commonly used in Thai sign language are, however, represented by a non-periodic hand movement [7]. This special case must be handled by the HMM-based transitional/non-periodic gesture classifier. Results from Posture Classifier module, Periodic Gesture classifier module, and the HMM-based classifier module, are then fed to the Word/Sentence Recognition module to arrive at a final recognized word/sentence. In this paper, only the Periodic-gesture and HMM-based modules are discussed in detail.

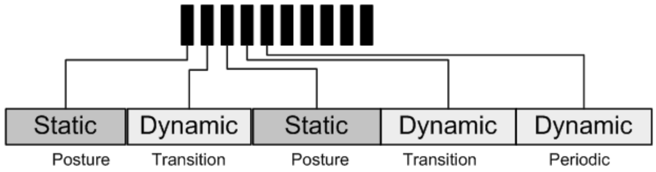

Fig. 1 Typical structure of a hand gesture sequence found in Thai sign language.

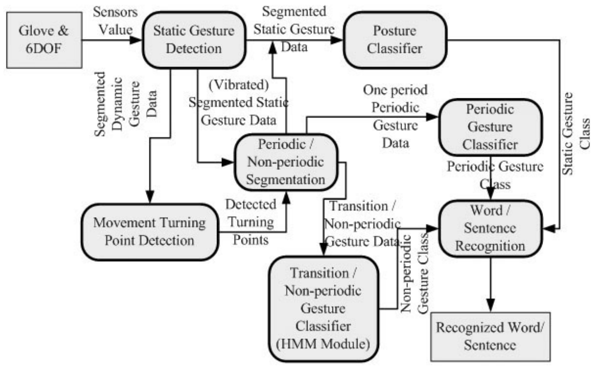

Fig. 2 Block diagram of the Thai sign language translation system

3. HYBRID TRANSITIONAL/NON-PERIODIC GESTURE CLASSIFICATION

3.1 Use of HMM for Non-periodic Gesture Classification

In Fig. 2, HMM is used to classify among non-periodic gesture classes, as well as to identify a transitional gesture. A detailed block diagram of this HMM-based module is shown in Fig. 3. From the figure, each block labeled as 'probability computation' represents a HMM model corresponding to each non-periodic gesture class. In Fig. 3 there are ten HMM models correspond to ten non-periodic gesture classes. The structure of each HMM model is shown in Fig. 4.

For the i^{th} model, let λ^i be the corresponding set of HMM parameters, which are A^i, B^i, π^i. Based on the notation as detailed in [6], A^i is the set of state transition probability, B^i is the set of observation probabilities, and π^i is the set of initial state probabilities. At the training stage, these parameters are estimated. a standard Baum-Welch algorithm [6] is applied here.

During the classification phase, the set of observation data as extracted from a sequence of electronic-glove data samples by the periodic/non-periodic segmentation module, is provided as an input to each of the HMM models. The estimated maximum probability of occurrence using the Viterbi algorithm for each model is compared against others. The one with the highest value, denoted by $p_{\max}$, is chosen and compared against a predefined threshold. If $p_{\max}$ is greater than the threshold, its corresponding gesture class is associated with the input data. Otherwise, the input data is classified as a transitional gesture.

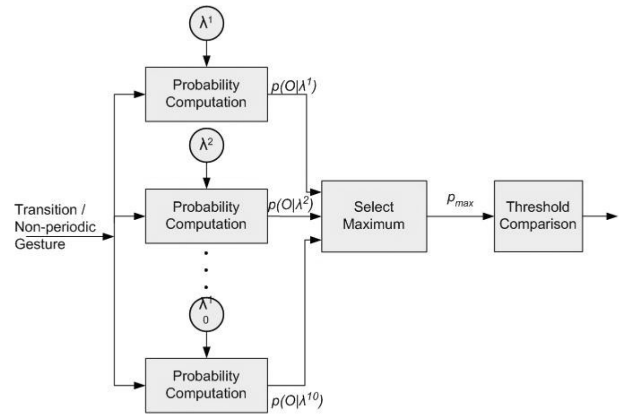

Fig. 3 Detailed block diagram of the non-periodic gesture classifier module.

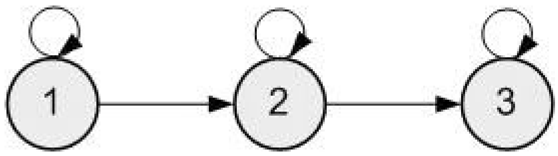

Fig. 4 Structure (topology) of each HMM model representing a non-periodic gesture class

3.2 Algorithm using hybrid non-periodic/periodic and HMM modules.

As described in [5], quasi-periodicity as found in many dynamic gestures is exploited to differentiate them from non-periodic and transitional gestures. A segment of hand gesture data samples is classified as periodic if its spectrum as obtained by Fourier analysis contains a spectral peak that exceeds a pre-defined threshold. This threshold of periodicity is denoted by T_p . By pre-classifying a gesture as periodic or non-periodic using this periodic/non-periodic gesture segmentation module, we reduce the need to perform transitional/non-periodic gesture classification, thus avoiding evaluation of the HMM models. By considering only the gesture classified as dynamic by the previous modules in Fig. 2, the following algorithm describes how the hybrid classifier, which is composed of the periodic/non-periodic segmentation module and the HMM-based non-periodic/transitional gesture classifier module, operates.

1. A gesture classified as dynamic by a previous module is first analyzed by the periodic/non-periodic segmentation module. Result of the analysis falls into one of the following cases
   - The gesture is classified as a (vibrated) static gesture. If this is the case, the corresponding captured data is passed to the posture classifier module.
   - The gesture is classified as a periodic gesture. This is the case when the spectral peak of the captured data is greater than T_p . If this is the case, the captured data is passed to the periodic gesture classifier module.
   - The gesture is neither periodic nor static. When this is the case, the captured data is passed to the HMM-based non-periodic/transitional gesture classifier module, and go to the next step.
2. The HMM module is used to classify the captured data. The obtained maximum estimated probability, $p_{\max}$, is compared against a predefined threshold. There are the same numbers of thresholds as the number of HMM models, as denoted by $T_{n,i}$ for the one corresponds to the i^{th} HMM model. The threshold corresponds to the model which achieves maximum estimated probability is used for the comparison. Let's assume that $p_{\max}$ is due to the k^{th} model. Then, if $p_{\max} > T_{n,k}$, the gesture is classified as non-periodic and belongs to the k^{th} non-periodic gesture class. If this is not the case, go to the next step.
3. If $p_{\max} \leq T_{n,k}$, the gesture is classified as ambiguous. This ambiguity means that the gesture is either transitional or periodic. The ambiguity case must be resolved by the subsequent word/sentence recognition module. (Detail of how this can be achieved is, however, outside the scope of this paper.)

4. EXPERIMENTAL RESULT

Three experiments were carried out in this study. Details of the experiments are reported below.

<u>EXP#1</u>

First we experimented with periodic/non-periodic segmentation by using the corresponding module as shown in Fig. 2. The experiment was carried out to find appropriate threshold T_p , as used in the algorithm of Section 3. A total of 880 gesture data sets were recorded by asking a signer to perform 10 different isolated signs drawn from the Thai sign language dictionary. Among 34 signs, they consist of 14 periodic signs, 10 non-periodic signs, and 10 static signs. The data sets corresponding to static signs also contain the same amount of transitional gesture segments. Each data set was put into the system in Fig. 2. The data sets classified as non-static were further classified by the periodic/non-periodic segmentation module. In this module, for each data set, the normalized peak magnitude, $\left|c^{p}_{\max}\right|$, of its Discrete Fourier Transform (DFT) coefficients was computed (The normalization was performed by dividing the DFT spectral peak by the total sum of all DFT coefficient magnitudes). The result, categorized by the gesture types, is shown in Fig. 5.

From Fig. 5, T_p = 0.046 was chosen empirically. By using this threshold value, it was found that 0.33% of the periodic gesture data sets have their $\left|c^{p}_{\max}\right|$ below the threshold, while all non-periodic and transitional gesture data sets were correctly classified as not periodic by using the threshold. This largely eliminates the need to perform HMM module evaluation for almost all periodic gestures. And because

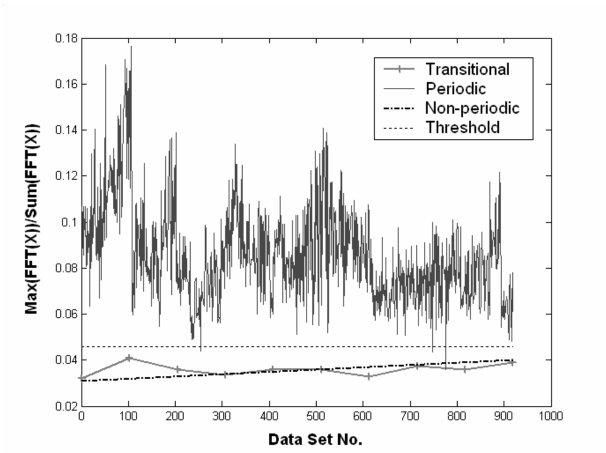

Fig. 5 The normalized peak magnitude for different gesture types, and the so-obtained empirical threshold T_p .

periodic gestures account for about 20% of all gestures found in the Thai sign language, by avoiding the need to perform more computing-intensive HMM module evaluation, this results in reduced system computational complexity.

EXP#2

In the second experiment, the classification performance of the HMM module was evaluated. Ten commonly-used non-periodic gesture signs were taken from the Thai sign language dictionary. Each sign was performed by the same signer for forty times. The resulting 40 data sets for each sign were divided into 30 data sets used for HMM training, and the remaining 10 data sets for testing. All ten HMM models are based on a 3-state topology as shown in Fig. 4. The recognition rate for all ten gesture classes is shown in Table 1. From the table, it was found that all data sets were correctly classified. This perfect result is perhaps due to the limited number of gestures used in the experiment, however. Thus, more extensive experiment may be needed to confirm the result.

Table 1 Recognition rate for 10 non-periodic signs

Sign	Recognition Rate (%)
Turn left	100
Turn right	100
Beautiful	100
Italy	100
Nepal	100
Myanmar	100
Weather	100
Write	100
Fly	100
Walk on a rough surface	100

Because non-periodic signs account for a small percentage of all signs in the Thai sign language, evaluation of the HMM module to classify those non-periodic signs will not significantly increase the system's overall computational complexity.

EXP#3

In this experiment, a total of 500 data sets drawn from all dynamic gesture types (periodic, non-periodic, and transitional) were used to obtain the empirical value of $T_{n,k}$. The 500 data sets consist of

- 420 data sets from 14 periodic signs
- 60 data sets from 6 transitional signs
- 400 data sets from 10 non-periodic signs.

Classification error was categorized into

- Missed classification: the rate at which the non-periodic gestures were incorrectly classified (as periodic or transitional gesture), by using the empirical thresholds as shown in Table 2.
- False classification: the rate at which either periodic or transitional gesture was classified, based on the empirical thresholds as shown in Table 2, as one of the non-periodic gestures.

Figs. 6-8 show the HMM probability values $p(O \mid \lambda^3)$ for the third non-periodic sign ('beautiful') corresponding to periodic, non-periodic (excluding data sets belonging to the sign 'beautiful'), and transitional gesture data sets. The empirical value of $T_{n,3}$ is also shown in all figures.

Table 2 Classification error rates, and empirical values of $T_{n,k}$

Sign	Threshold	Missed Classification (%)	False Classification (%)
Turn left	-7.13×10^3	0	0
Turn right	-1.16×10^4	0	0
Beautiful	-7.26×10^3	0	0
Italy	-1.42×10^4	0	1.67%
Nepal	-9.37×10^3	0	0
Myanmar	-8.22×10^3	0	0
Weather	-1.36×10^4	0	0
Write	-9.34×10^3	0	0
Fly	-1.36×10^3	0	0
Walk on a rough surface	-1.98×10^4	0	0

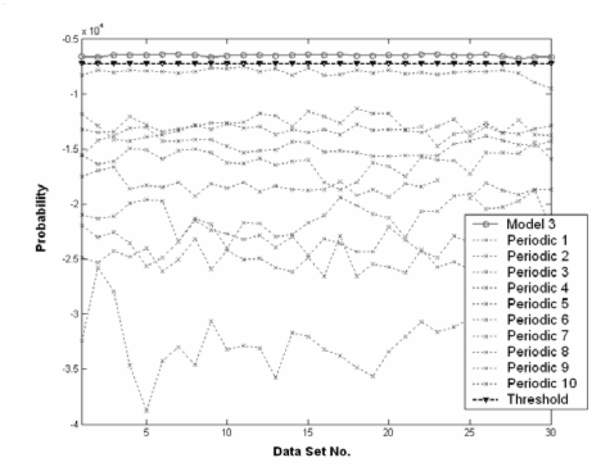

Fig. 6 $p(O \mid \lambda^3)$ corresponding to periodic data sets, compared against those of the non-periodic sign 'beautiful' and its empirical threshold.

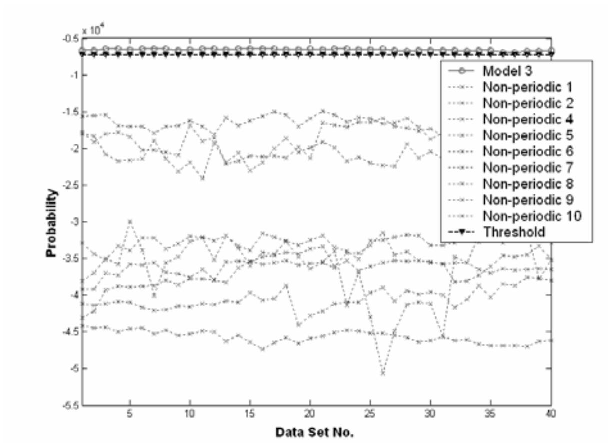

Fig. 7 $p(O \mid \lambda^3)$ corresponding to other non-periodic data sets, compared against those of the non-periodic sign 'beautiful' and its empirical threshold.

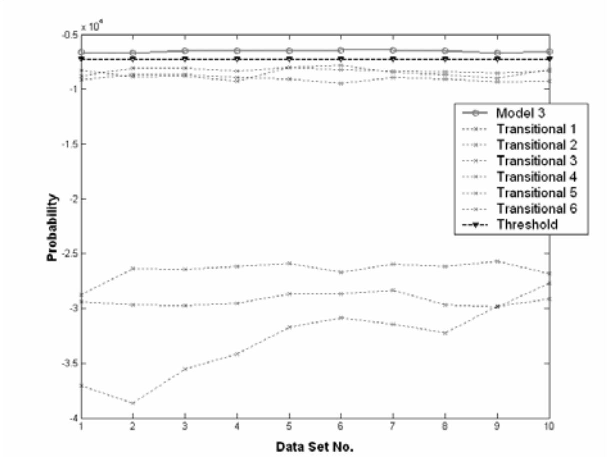

Fig. 8 $p(O \mid \lambda^3)$ corresponding to transitional data sets, compared against those of the non-periodic sign 'beautiful' and its empirical threshold.

5. CONCLUSION

In this paper, a hybrid approach to automated Thai sign language translation has been proposed. The proposed method combines the HMM-based classifier with another two non-HMM-based classifiers developed in previous work. Because the HMM module is used for classifying a relatively small number of signs, the hybrid method retains the advantage of the non-HMM-based method in terms of computational saving. In addition, use of the HMM module here makes possible the transitional/non-periodic gesture classification. This classification of non-periodic gesture has been a major weakness in our previous Thai sign language translation system. As a result of this hybrid scheme, the system can now distinguish both periodic and non-periodic gestures from transitional movement. Future work includes more experiments based on a larger number of data sets.

REFERENCES

[1] Gaolin F. and Wen G., *"A SRN/HMM system for signer-independent continuous sign language recognition"*, Proceedings. Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 20-21 May 2002, pp. 297 -302.

[2] Hermann Hienz, Britta Bauer, and Karl-Friedrich Kraiss, "*HMM-Based Continuous Sign Language Recognition Using Stochastic Grammars"*, In: Brattfort A. et al (Eds.): Gesture-Based Communication in Human Computer Interaction GW 99, 1999, pp. 185-196.

[3] Vogler, C. and Metaxas, D., "*ASL recognition based on a coupling between HMMs and 3D motion analysis* ", Sixth International Conference on Computer Vision, 4-7 Jan. 1998, pp. 363 -369.

[4] Wu J.,Gao W., Song Y., Liu W. and Pang B.*"A simple sign language recognition system based on data glove"*, Proceedings Fourth International Conference on Signal Processing (ICSP '98), Vol. 2 , 12-16 Oct. 1998, pp. 1257 -1260.

[5] W. Werapan, and N. Chotikakamthorn "*Gesture Segmentation Technique for Thai Sign Language Translation*", The Second International Symposium on Communications and Information Technology, Chonburi, Thailand, 2002, pp. 483-486.

[6] Rabiner, L.R., *"A tutorial on hidden Markov models and selected applications in speech recognition"*, Proceedings of the IEEE, Vol. 77 , Issue: 2 , Feb. 1989, pp. 257 - 286*"*

[7] The National Association of the Deaf in Thailand, *The Thai Sign Language Dictionary: Revised and Expanded Edition*, Thai Watana Panich, 1990.