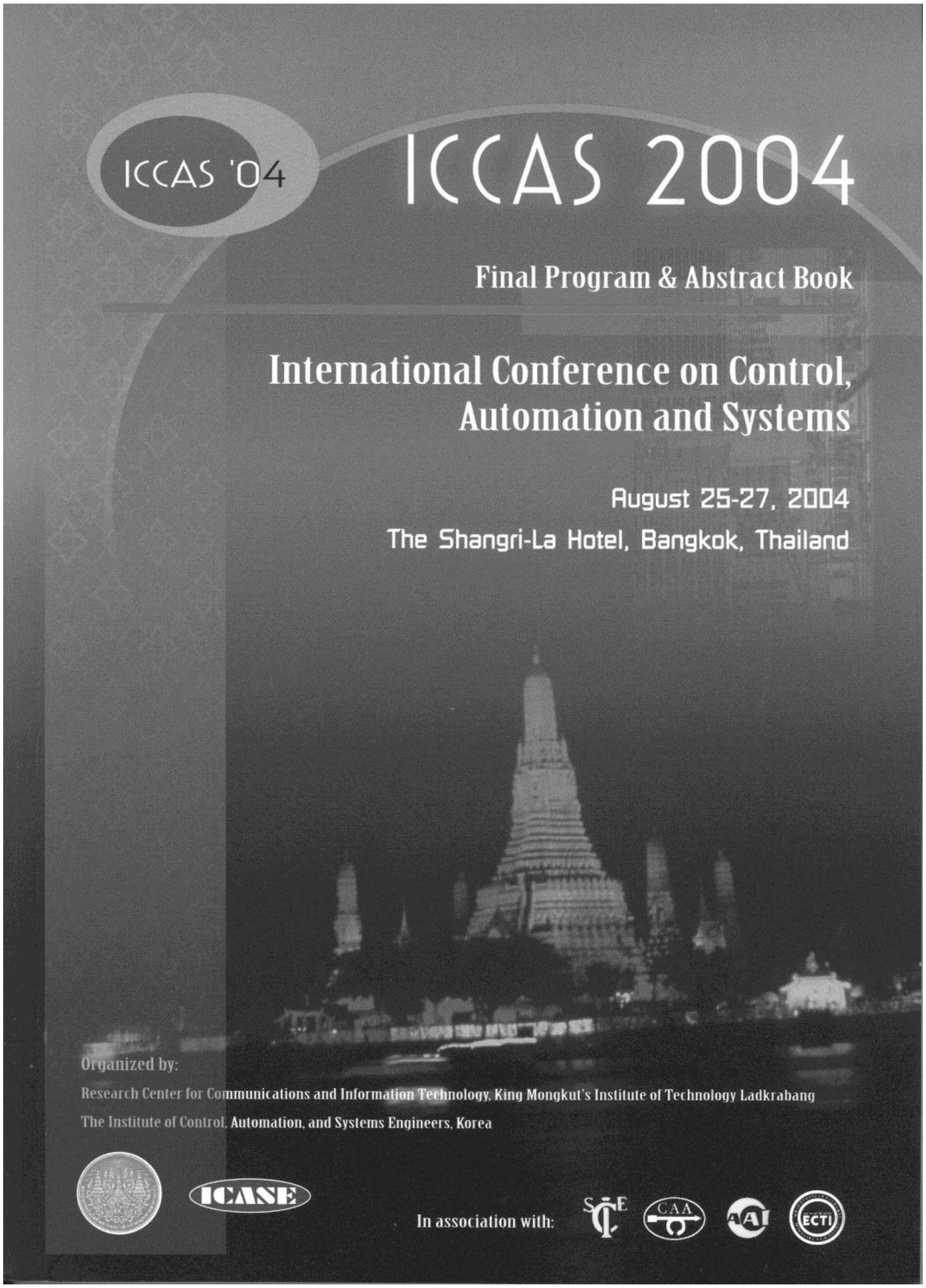
ภาคผนวก ก.

บทความและผลงานวิจัยที่ได้รับการตีพิมพ์

1. Arunee Jaruwanawat and Nopporn Chotikakamthorn. “Hybrid HMM for Transitional Gesture Classification in Thai Sign Language” ICCAS 2004, pp. 60



ICCAS2004 August 25-27, The Shangri-La Hotel, Bangkok, THAILAND

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.

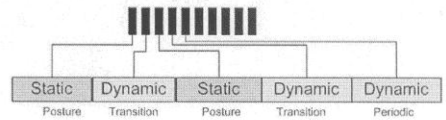


Fig. 1 Typical structure of a hand gesture sequence found in Thai sign language.

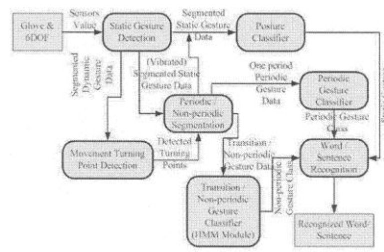


Fig. 2 Block diagram of the Thai sign language translation system

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.

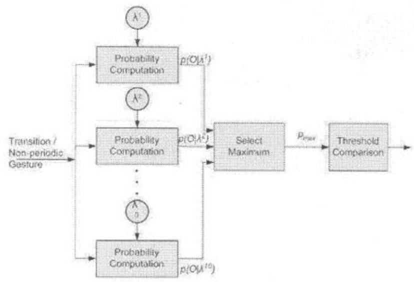


Fig. 3 Detailed block diagram of the non-periodic gesture classifier module.

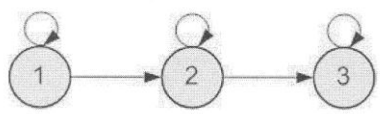


Fig. 4 Structure (topology) of each HMM model representing a non-periodic gesture class

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.

EXP#1

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, $|c^p_{\max}|$, 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 $|c^p_{\max}|$ 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

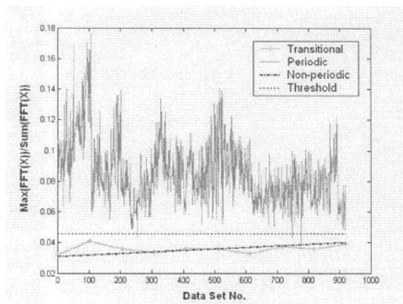


Fig. 5 The normalized peak magnitude for different gesture types, and the so-obtained empirical threshold T_p .

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

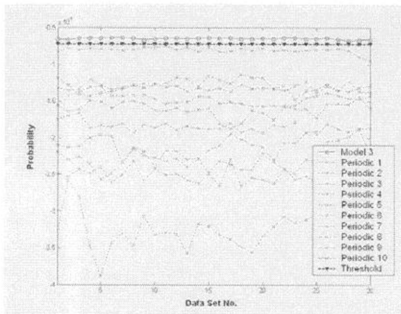


Fig. 6 $p(O \mid \lambda^3)$ corresponding to periodic data sets, compared against those of the non-periodic sign 'beautiful' and its empirical threshold.

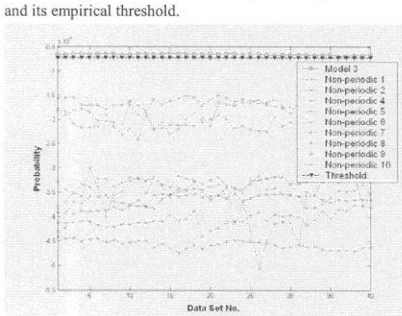


Fig. 7 $p(O \mid \lambda^3)$ corresponding to other non-periodic data sets, compared against those of the non-periodic sign 'beautiful' and its empirical threshold.

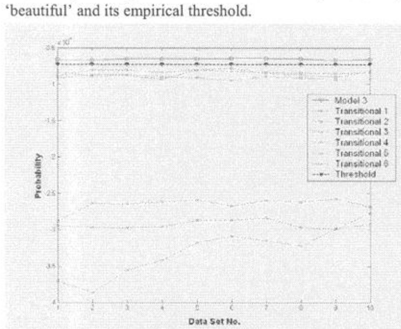


Fig. 8 $p(O \mid \lambda^3)$ corresponding to transitional data sets, compared against those of the non-periodic sign 'beautiful' and its empirical threshold.

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.

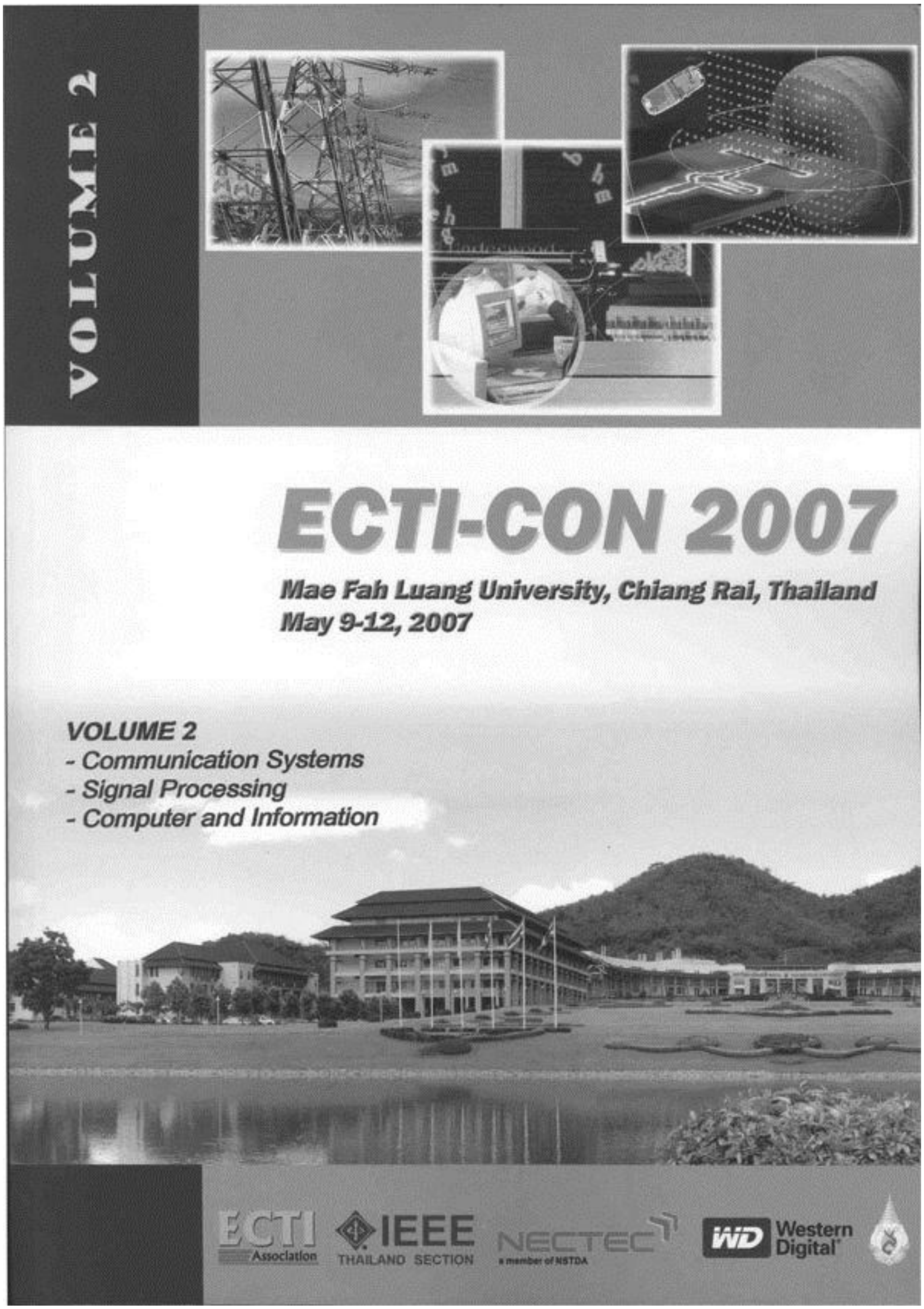
[4] Wu J.,Gao W., Song Y., Liu W. and Pang B.*"A simple sign language recognition system based on data glove"*, Proceedings Fourth International Conference on Signal Processing (ICSP '98), Vol. 2 , 12-16 Oct. 1998, pp. 1257 -1260.

[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.

2. Arunee Jaruwanawat and Nopporn Chotikakamthorn. "Continuous Thai Sign Language Recognition using HMM with Transition Models" ECTI Conference 2007 Volume 2, pp. 991-994



Continuous Thai Sign Language Recognition using HMM with Transition Models

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

***Abstract*-This paper addresses the problem of continuous Thai sign language recognition. Sign language is typically composed of hand shapes and gestures formed by the movement of hands, arms and perhaps body also. One of the problems in performing continuous sign recognition is the existence of hand and arm movement in transition between two signs. This spontaneous transition movement conveys no meaning. However, not taking it into account results in poor recognition performance. This papers proposes the inclusion of algorithmic-based HMM transition movement models within a HMM-based sign language recognition system. A transition movement is represented by a single-state HMM model, of which its model parameters are obtained algorithmically based on the HMM parameters of the two signs involved in the transition. This avoids the need to collect extensive training data for modeling all possible transition movements in a conventional manner. In addition, to realize continuous sign language recognition, in this paper a pruning strategy for reducing the complexity of a Viterbi search is also described. Based on the experiment result, inclusion of such model within a HMM-based recognizer was found to offer improved recognition performance.**

***Keywords*- Sign Language, Hidden Markov Model, Transition Movement**

I. Introduction

People with hearing and speech impairments communicate within their communities through the use of a sign language. Use of this specialized language, however, is limited to those who have acquired knowledge of the language. This largely prevents deaf people from communicating with those outside their communities. Also, while currently there exists a speech recognition system, allowing a computer user to input textual data through voice, there is not yet an equivalent product for deaf people. A sign language recognition system has been seen as an enabling technology to allow for natural human-computer interaction, as well as to assist in communications between those who understand the sign language as those who do not.

Previous works in continuous sign language recognition, with the emphasis on problems related to transition movement, include those of [1-6]. In [1], equipped with an electronic glove device and a 6-DOF tracking sensor, a method was proposed to segment continuous sentences such that transition movement can be removed before the segmented data being processed by a HMM-based recognizer. This 'hard' pre-segmentation strategy is, however, suboptimal as compared to a 'soft' segmentation inherent in a general HMM-based pattern recognition method, where segmentation and recognition are jointly performed to achieve optimal result. In [2, 4], transition movements are first extracted. This extracted data is then used to perform data clustering using a Dynamic Time Warping (DTW) method. The so-obtained clusters are used to form HMM models for transition movements. The problem with this method based on conventional HMM model construction is its need for extensive training data to sufficiently cover possible transition movement patterns.

In [3], a large vocabulary sign language recognition is considered. Due to a large sign database, extensive Viterbi search is required. The solution suggested in [3] is based on a two-stage classification strategy. A sign gesture input is first classified into one of the clusters of similar signs. Then, a Viterbi search is performed on a small subset of HMM models corresponding to signs belonging to that cluster. Although efficient computation is obtained, this two-stage classification method is suboptimal for the search is performed over a subset of a set of all feasible solutions. It also results in poorer recognition rate as compared with a direct search over all feasible HMM models [1].

Most research works in sign language recognition deal with languages used by people of large population such as American and Chinese sign languages. However, it is known that a sign language, like its spoken language counterpart, is different from one region to another. For Thai sign language, early works include those of [5-6]. In [5, 6], a Thai sign language recognition system based on a statistical-based estimator was described. The method performs a pre-segmentation of input data from an electronic glove into static and dynamic gestures. Recognition is performed from the classified data using a Bayesian estimator. The method can perform recognition of signs with certain regularity in the movement patterns. However, it cannot deal with signs of irregular patterns, as well as transition movements. A solution to automated Thai sign language recognition based on HMM models was suggested in [5]. Use of HMM models allows signs of arbitrary movement patterns to be handled in a unified manner. However, the method in [5] was applied on a single-sentence basis. In [6] a method was developed to classify periodic gestures as found in Thai sign language and those with non-periodic nature, of which transitional gestures form a large part of them. However, the method in [6] can not differentiate between some non-periodic and yet meaningful gestures from those due to transition movements. No known work has been done on recognition of continuous Thai sign language containing both periodic and non-periodic gestures.

In this paper, a problem of continuous Thai sign language recognition is considered. In particular, the problem due to the existence of a transition movement is addressed. It is known that, by neglecting such spontaneous movement in modeling the recognition system, recognition performance is degraded [3]. The method described in this paper is based on a HMM model. Unlike those in [2, 4], this paper proposes an algorithmic-based method for construction of a transition model. This avoids the need for collection and processing of extensive training data for transition movements. Also, it can be used to handle transition movement between any two arbitrary signs. To address complexity in performing Viterbi search on continuous signs, a pruning strategy based on the detection of static signs is also described.

This paper first introduces a general framework of a Thai sign language recognition system in Section 2. Then, a detailed discussion on modeling a transition movement is described in Section 3. A Viterbi search strategy is also given. Experimental result is provided in Section 4, with concluding remarks given in Section 5.

II. THAI SIGN LANGUAGE RECOGNITION SYSTEM

Like other sign languages, Thai sign language is composed of hand shapes and gestures formed by the movement of hands, arms and body. Sometimes, facial expression also plays a crucial role in sign interpretation. Here, only part of the gestures formed by hand shapes and movement of hands and arms are considered. In this study (see Figure 1), a sign gesture is roughly classified into static (posture) and dynamic gestures. Dynamic gestures are further classified into sign movements and transition movements.

- Static gesture (posture) is a sign of which there is no hand or arm movement. Only the hand posture that defines the meaning of the sign.
- Dynamic gesture contains a hand/arm movement. It is further divided into
  - dynamic signs, which are meaningful hand movement, and
  - transition movements, which occur during the transition from one sign to another and convey not meaning.

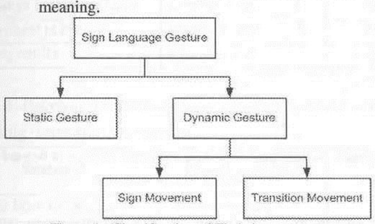


Figure 1. Classification of Thai sign gestures

An overall view of the Thai sign language recognition system, which contains the work described here as its part, is illustrated in Figure 2.

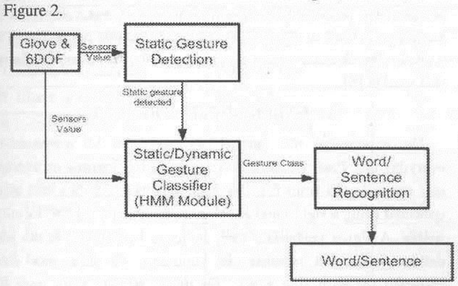


Figure 2. Block diagram of the continuous Thai sign language recognition system.

From the figure, the system is composed of four main parts. The first one is responsible for acquisition of hand shapes and movements. Here, an electronic glove with a 6-DOF electro-magnetic sensor is used as an acquisition device. The second part is the HMM-based gesture recognizer. The third part is the static gesture detector. The result of the detection is used to assist the HMM recognizer in selecting feasible Viterbi paths (pruning). And the last part is the word/sentence recognizer. This part uses the result from the gesture recognizer, along with a sign language model, to produce the final recognized word/sentence. In this paper, the focus is mainly on the second and, to a lesser extent, the third parts of the system.

III. CONTINUOUS THAI SIGN LANGUAGE RECOGNITION WITH ALGORITHMIC-BASED TRANSITION MODELS

A. Modeling the transition movement

Here, an HMM model is adopted for the recognition of both static and dynamic signs. As pointed out by previous work [2, 4], transition movement should be incorporated in the model to improve recognition rate. Instead of directly training the model using data of transition movements, in this paper we propose that a transition movement is modeled using a single state HMM of which its model parameters are obtained algorithmically at run-time. Thus, given two signs (the last states of the word models A and B respectively) connected by a particular transition movement, the series of interconnected HMM models is described by Figure 3.

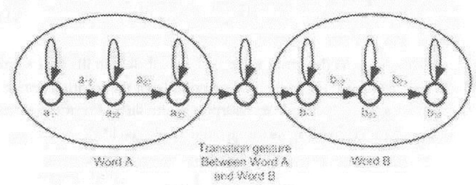


Figure 3. Series of HMM states, connected by a transition state, to model continuous signing.

In our study, a single-state HMM transition model is constructed at run-time based on HMM parameters corresponding to the last state of Word A and the first state of Word B. To explain the algorithm, some notations for HMM models are first introduced. Let (π, A, μ_i, U_i) be an HMM model parameters corresponding to the first word (Word A in Figure 3). Here, π is the initial probability of being in the first state. **A** is a matrix of state transition probabilities. In addition, μ_i is the means vector of the i^{th} state outcome and U_i is the corresponding covariance matrix. Similarly, (π', A', μ_i', U_i') are HMM parameters corresponding to the second word (Word B in Figure 3). In addition, we define $(\bar{\pi}, \bar{A}, \bar{\mu}, \bar{U})$ as model parameters corresponding to the single-state transition model. These parameters are obtained below.

First, because each transition model is designed for a particular pair of sign words, its initial probability can be set to $1 - a_{33}$, where a_{33} is the probability that the system, once reaching the last state in

Word A, remains on the same state. Next, because the model contains a single state, $\overline{A}$ is scalar. It is set to $1 - \pi'$. The model parameters related to the outcome probability is obtained according to the following equations.

$$\overline{\mu} = \frac{\mu_3 + \mu'_1}{2} \tag{1}$$

$$\overline{U}_{11} = \frac{U_{3,11} + U'_{1,11}}{2} \tag{2}$$

$$\overline{U}_{12} = \overline{U}_{21}^{T} = 0 \tag{3}$$

$$\overline{U}_{22} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_N^2 \end{bmatrix} \tag{4}$$

From Eq. (2), $\overline{U}_{11}$ is the $M \times M$ top-left submatrix of $\overline{U}$, where M is the number of data points obtained from an electronic glove. Similarly, $U_{3,11}$ and $U'_{1,11}$ are the $M \times M$ top-left submatrice of U_3 and U'_1 respectively. In Eqs (3, 4), $\overline{U}_{12}$, $\overline{U}_{21}$, and $\overline{U}_{22}$ are the remaining submatrice of $\overline{U}$ (see Figure 4, for an illustration). The size N of $\overline{U}_{22}$ is equal to the number of data points obtained from a 6-DOF electromagnetic location sensor (N = 6). The variance parameters in Eq. (4) are obtained from the following equation.

$$\sigma_i^2 = \{|\mu_{3,M+i} - \mu'_{1,M+i}| / \alpha\}^2 \tag{5}$$

Here, $\mu_{3,M+i}$ is the mean value at the third state in the first word model, corresponding to the ith data point from the location sensor. In addition, $\mu'_{1,M+i}$ should be interpreted similarly. From the same equation, α is a scaling factor, arbitrarily chosen as 6.

$$\left[\begin{array}{c|c} \overline{U}_{11} & \overline{U}_{12} \\ \hline \overline{U}_{21} & \overline{U}_{22} \end{array}\right]$$

Figure 4. Decomposition of $\overline{U}$ into four submatrice

B. Viterbi search strategy

Generally, pattern recognition by HMM model involves a search for a best path (i.e., a sequence of HMM states) using a Viterbi algorithm. However, for continuous sign language recognition, with medium to large vocabulary support, the search complexity and space is increased significantly. To maintain the search complexity and memory space requirement to a practically realizable level, certain search strategy must be applied to a standard Viterbi algorithm. Here, by observing that many words in Thai sign language are composed at least partly of static signs, those static moments are detected and utilized (along with some heuristic rules) as part of the pruning strategy in an attempt to keep the number of active paths to an acceptable figure. The search strategy used here is summarized in Figure 5.

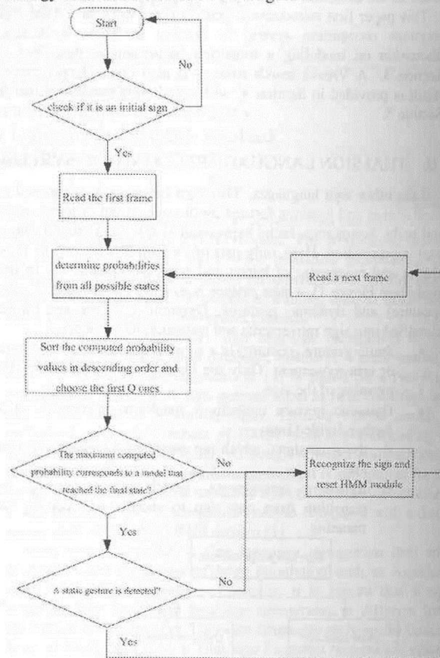


Figure 5. A flow chart describing a pruning strategy, to be used with a standard Viterbi algorithm

From Figure 5, the parameter Q is the maximum number of feasible Viterbi paths that can be kept at any time instance. Therefore, from time to time, rules must be enforced to select Q best paths from all candidates. Here, in addition to a simple accumulated probability sorting, it is proposed that a detection of static sign is used to significantly reduce infeasible paths. Detail of a static sign detector is not given here. It is, however, similar to that used in [6].

IV. EXPERIMENTAL RESULT

The experiment was carried out based on 13 sentences of everyday use. Each sentence is composed of 2-3 words on average and was selected from [7]. For each sentence, 20 data sets were collected using a right-hand electronic glove and a 6-DOF location sensor. A signer performed each sentence by starting in the pre-defined initialized position. In summary, 40 data sets were collected for each sign word. Of these, 30 sets were used for training and the remaining data sets were used for testing.

Two HMM-based sign language recognition techniques were applied to the same data sets. The first one employed the proposed transition models as part of the HMM models. The second one did not model the transition movement. For the latter model, each HMM sign model was simply connected to the next one. The results are shown in Tables 1 and 2. From the tables, it is seen that the proposed method offers noticeable performance improvement. It is noted, however, that this is achieved without additional work on training the transition movement.

Table 1. Recognition performance using the method with algorithmic-based transition models

Sentence	Word sequence	Accuracy (%)
1. I was born in Denmark.	Born-Denmark	100
2. I met him in Denmark.	Meet-Him-Denmark	80
3. Please, sit down.	Please-Sit	100
4. Your mom is fine.	Mother-Fine	95
5. Today is Monday.	Today-Monday	95
6. I take a rest on Friday.	Friday-Rest	80
7. I have ten relatives.	Relative-Ten	85
8. I have ten goats.	Goat-Ten	85
9. I study English on Monday.	Monday-Study-English	90
10. I'll go swimming in the day after tomorrow.	Swim-The day after tomorrow	100
11. Orange has a sour taste.	Orange-Sour	90
12. Elephant is big.	Elephant-Big	80
13. Frog eat fly.	Frog-Eat-Fly	90

Table 2. Recognition performance using the method without modeling transition movement.

Sentence	Word sequence	Accuracy (%)
1. I was born in Denmark.	Born-Denmark	75
2. I met him in Denmark.	Meet-Him-Denmark	70
3. Please, sit down.	Please-Sit	85
4. Your mom is fine.	Mother-Fine	80
5. Today is Monday.	Today-Monday	85
6. I take a rest on Friday.	Friday-Rest	70
7. I have ten relatives.	Relative-Ten	70
8. I have ten goats.	Goat-Ten	75
9. I study English on Monday.	Monday-Study-English	75
10. I'll go swimming in the day after tomorrow.	Swim-The day after tomorrow	80
11. Orange has a sour taste.	Orange-Sour	75
12. Elephant is big.	Elephant-Big	70
13. Frog eat fly.	Frog-Eat-Fly	75

V. CONCLUSION

In this paper, the problem of modeling transition movements for Thai sign language recognition has been addressed. The transition models, based on algorithmic construction at run-time, have been described. It has been verified through the experiment that the use of the proposed transition model offers improved performance over the standard HMM-based method without transition movement modeling. The improvement is attained without the need for additional training work for modeling these spontaneous movements.

REFERENCES

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

[2] G. Wen, F. Gaolin, Z. Debin and C. Yiqiang, *"Transition Movement Models for Large Vocabulary Continuous Sign Language Recognition"*, Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, 17-19 May 2004 Page(s):553 - 558

[3] Y. Guilin, Y. Hongxun, L. Xin and J. Feng, *"Real Time Large Vocabulary Continuous Sign Language Recognition Based on OP/Viterbi Algorithm"*, Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, Volume 3, 20-24 Aug. 2006 Page(s):312 - 315

[4] G. Fang, W. Gao and D. Zhao, *"Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models"*, Systems, Man and Cybernetics, Part A, IEEE Transactions on Volume 37, Issue 1, Jan. 2007 Page(s):1 – 9

[5] C. Siam and V. Wuttichai, *"Isolated Thai Sign Language Recognition System using Hidden Markov Models"*, Proceedings of the 25th National Electrical Engineering Conference (EECON-25), Hadyai, Thailand, November 21-22, 2002.

[6] Worawit Werapan, *"Gesture Segmentation and Recognition Techniques for Automated Thai Sign Language Translation"*. Master of Science in Information Technology School of Graduate Studies, King Mongkut's Institute of Technology Ladkrabang, 2005.

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

ประวัติผู้เขียน

ชื่อผู้เขียน	นางสาวอรุณี จารุวนาวัฒน์
วัน เดือน ปีเกิด	วันที่ 29 กรกฎาคม พ.ศ. 2523
วุฒิการศึกษาระดับปริญญาตรี	วิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์
สถานที่สำเร็จการศึกษา	มหาวิทยาลัยบูรพา
ปีที่สำเร็จการศึกษา	2545
ประสบการณ์การทำงาน	พฤศจิกายน 2544 – เมษายน 2545 ตำแหน่งผู้ช่วยนักวิจัย ประจำห้อง Multimedia and Virtual Research Lab. สำนักวิจัยการสื่อสารและเทคโนโลยีสารสนเทศ เมษายน – พฤษภาคม 2550 ตำแหน่งผู้ช่วยนักวิจัย คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง