Fuzzy Finger Shapes and Hand Appearance Features for Thai letter Finger Spelling

Kittasil Silanon
Department of Computer Engineering,
Prince of Sonkla University,
Phuket, Thailand, 83120
Kittasil.silanon@gmail.com

Nikom Suvonvorn
Department of Computer Engineering,
Faculty of Engineering, Prince of Songkla University,
Hatyai, Thailand, 90112
Nikom.suvonvorn@gmail.com

***Abstract*— Our previous work proposed hand posture estimation technique. The hand region is first extracted using depth image, and then the initial features, such as fingertip, hand center point, and palm size, have been calculated. The concept of active contour using energy function is implemented in order to track fingertip position in the frame image sequence. To discriminate the hand posture sets, a hand feature definition have been established, which is composed of finger shape and hand appearance features. The features are defined as characteristic code to represent hand posture. This technique is applied to the American finger-spelling. Since we focused on Thai letters finger-spelling which are based on American finger-spelling hand postures. Therefore, this paper has introduced a hand posture estimation method for Thai letters finger-spelling. The Hidden Markov Model (HMM) method is used to build a learning model to recognize the sequence of the American finger-spelling hand postures and to provide the Thai letter finger-spelling. The performance of the recognition system can be measured at around 70% recognition rate.**

Keywords*— *Finger-Spelling; Hand Posture Estimation; American finger-spelling; Thai letter finger-spelling.

I. Introduction

The sign language is communicating method for deaf or non-vocal person. For the sign language system, there are two main categories as follow: 1) word-level vocabulary signs, which are the signs of the hand shape, orientation and movement of the hands, arms or body, and facial expressions simultaneously to represent word meanings, 2) finger-spellings, which use only hand shape to spell the letters of the word in a spoken language to represent names, places, technical terms and etc. Usually, the word-level vocabulary signs have been used frequently to communicate each other, but the finger-spelling is used infrequently in daily communications. Therefore, most of deaf and non-vocal persons, especially children, have problems with finger-spelling skills. In order to help these people improve their skills, many systems that specific to finger spelling are proposed, as for example the American (ASL) [8,11,16,18], British (BSL) [12], Australian (Auslan) [9], Chinese (CSL) [23,24], Japanese (JSL) [5,26], and others [2,4,10,13]. Most of researches on finger-spellings will be based on hand posture estimation techniques [1,14,20]. The hand posture estimation technique has two main approaches: 1) using the signal from a glove sensor method, and 2) using a vision-based method to estimate the hand posture. For glove sensor method, all gloves are designed to easily detect hand articulate. However, devices are expensive, troublesome to put on and take off. For vision-based method, 2D image features such as point, contour or silhouette are extracted to estimate hand posture.

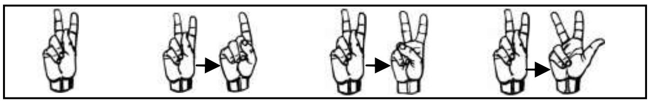
Several methods in finger-spelling recognition have been introduced. However, we have focused on the Thai letters finger-spelling. Saengsri [17] proposed a Thai letters finger-spelling by using the data glove, motion tracker and Neural Network theory to improve the accuracy of the system. Kanjanapatmata [21] presents an image recognition method for the Thai letter using a polar orientation histogram of the hand image and an artificial Neural Network. Veerasakulthong [22] introduced a simple color hand glove and appearance features. Sakulsujirapa [3] presents an appearance features lookup table to analyze the hand posture pattern for identifying Thai letters in finger-spelling. Sirboonruang [25] proposed a method combining the Zernike moment and wavelet moment to capture hand's features and using fuzzy classification algorithm to classify Thai finger-spelling hand postures. Phitakwinai [15] developed the 15 Thai finger-spelling letters and 10 words of the Thai sign language translation system using the scale invariant feature transform (SIFT). Although several approaches are proposed for the Thai letters finger-spelling recognition, however, they cannot achieve the critical criteria, such as, accuracy, flexibility and device constraints. Thus, our research goal is to develop recognition system for Thai letters finger-spelling. We applied a vision-based method for hand posture estimation, which uses both finger shape and hand appearance features, to finally recognize the Thai letter finger-spelling.

II. Thai Finger-Spelling

The Thai finger-spelling is a usage of hand posture for representing letters, vowels, intonation marks, and numbers to spell the specific names, places, or technical words in Thai language. The Thai finger-spelling was developed by Khun Ying Kamala Kraireuk. It is based on American finger-spelling. The Thai letter finger-spelling is compared to the phonetics of American finger-spelling. Thai finger-spelling will be matched to American finger-spelling whose sound particular character is similar. For example, "ก (Ko kai)" has a similar sound to the "K". Therefore, the hand posture of "K" in American finger-spelling is used to represent "ก (ko kai)" in

978-1-4673-9749-0/16/$31.00

Thai finger-spelling. The combination of American finger-spelling hand posture is extended in order to represent all 42 Thai letters, such as “ข (Kho khai)” = K+1, “ค (Kho khwaiand)” = K+2, “ค (Kho khwai)” = K+3 respectively, as shown in Fig 2.



(a) **(b)** **(c)** **(d)**

Fig 2. Thai finger-spelling example: (a) “ก (ko kai)” (b) “ข (Kho khai)” (c) “ค (Kho khwaiand)” (d) “ค (Kho khwai)”.

III. PROPOSED METHOD

From our previous work [6], we proposed hand posture recognition using finger shape and hand appearance features. The depth image is used to separate hand from other parts of body by predefined depth threshold. The fingertip is tracked by concepts of active contour. Energy of continuity, depth, and direction are calculated to track the position of fingertip in sequence images. The fuzzy logic is used to classify finger shapes as “Open”, “Bend”, “Point”, and “Close” by fuzzy rule definition which based on depth and distance linguistic variables. Moreover, hand appearances features are computed to distinguish difference between hand postures which are similar such as fingers are grouped, separated, or crossed one another, or hand is moved or rotated. The chain code feature is defined to represent hand posture characteristic. To classify hand posture, simple score vote technique estimated similarity between input and predefined hand posture pattern. Our previous work applied to automatic American finger-spelling hand posture recognition system. From the above section, Thai letter finger-spelling system is based on sequence of American finger-spelling hand posture. Since Thai letter finger spelling is the combination of American finger-spelling hand posture, for example “ข (kho khai)” uses “K, 1” or “ค (Kho khwaiand)” uses “K, 2” as shown in Table.I. For real circumstances, “ข (kho khai)” can be “K,K,1,1,1,1,1”, or “K,K,K,K,1,1,1,1,1”, or “K,K,1,1,1,1,1” and so on. Therefore, for this recognition step, a learning-based approach such as the Hidden Markov Model (HMM) [7] is used. The HMM is a stochastic processes which can be used to model any time series data. Therefore, HMM is useful for recognition of Thai finger-spelling, which can be viewed as a series of hand postures in American finger-spelling. The basic elements of HMM can be expressed as follows:

1). A set of learning states (S_i):

$$S = \{S_1, S_2, \ldots, S_N\} \tag{1}$$

The number of states (N) is determined by using the maximum number of hand posture involved in performing a Thai finger-spelling which is three states for hidden states and additional two states for initialization and finalization. Therefore, a five-state model with transitions was chosen for the system.

2). A set of observation symbols (V_i):

$$V = \{V_1, V_2, \ldots V_M\} \tag{2}$$

The observation symbols are represented by the 22 hand postures in American finger-spelling. All observation symbols are listed in Table.II. The hand posture sequence will be converted to an observation symbols sequence for using as input to the learning model. For instance, the “ก (ko kai)” letter uses “K”. This sequence is converted to symbol “8”. The “ข (kho khai)” letter use “K+1”. Thus, the sequence is converted to combination of “8” and “18” and so on.

3). The state transition matrix A = $\{a_{ij}\}$, where a_{ij} is the transition probability of taking the transition from state i to state j. Random number will be contained in each element of A. Summation of number in each row has to be 1.

$$a_{ij} = P(q_t = S_j \mid q_{t-1} = S_i), \qquad 1 \le i, j \le N \tag{3}$$

4). The observation symbol probability distribution matrix B = $\{b_j(k)\}$, where $b_j(k)$ give the probability of emitting observation symbol ok from state j. Random number will be contained in each element of B. Summation of number in each row has to be 1.

$$b_j(k) = P(o_t = V_k \mid q_t = S_j), \qquad 1 \le j \le N;\ 1 \le k \le M \tag{4}$$

5). The initial state distribution matrix $\pi = \{\pi_i\}$, where π_i is initial probability of all states in the model. Random number will be contained in each element of π. Summation of number in each row has to be 1.

$$\pi_i = P(q_1 = S_i), \qquad 1 \le i \le N \tag{5}$$

6). The HMM topology, Fully Connected topology, is applied for our system. In this topology, every state can be reached from any state in a finite number of steps.

For training process, the HMM models the 42 Thai letter finger-spellings (λ_k) by the training data input. For recognition, a combination of hand posture in the American finger-spelling system are converted to be observation symbol sequence (O =$V_1, V_2, .., V_T$) and input to each HMM model for calculating the probability $P(O|\lambda_k)$. The models that give the maximum probability will be recognized as result of the input observation symbol sequence

IV. EXPERIMENTS

The 150 sample image sequences of each hand posture are collected. For each hand posture data set, the 50 samples are used to generate each letter model and the 100 samples are used to test the performance of the recognition model. The key hand posture is selected when hand posture is stable in image sequence. The Forward-Backward procedure (HMM) is used to calculate the probability of an input observation sequence. Since HMM does not have a fixed rule for the state number specification, we compare the recognition rate of the HMM process when it use 5 states, 10 states, and the 15 states. Experiments showed the results of the 42 Thai letters recognition model. The aggregate recognition rate of 69.52% for 10 states of HMM is better than the aggregate recognition rate of 68.90% for 5 states of HMM, but the improvement (+0.62%) is marginal, not significant. 15 states of HMM give an even lower rate of 65.88%. We chose to use five states of HMM , give an average recognition rate at 68.90%. As a result, we infer that the alphabet models that use only one hand posture get a quite better result such as "ก" (ko kai), "จ" (cho chan), "ด" (do dek), "บ" (bo baimai), "พ" (po han), "ฟ" (fo fan), "ย" (yo yak), "ร" (ro ruea), "ล" (lo ling), "ว" (wo waen), "ห" (ho hip), "อ" (o ang). There are some exceptions, such as "ต" (to tao) or "T", "ม" (mo ma) or "M", "น" (no nue) or "N" and "ส" (so suea) or "S", which get a poor result (less than 70%) because the fingertip tracking cases for finger overlapping and for adjacency give less precision. Hence, the other letters that are based on these letter groups ("T", "M", "N" and "S") will also give a poor result. These are groups of letters which are based on "N", such as "ง" (ngo ngu) and "ณ" (no nen), the groups of alphabets based on "S" such as "ศ" (so sala) and "ษ" (so rue-si), and groups of letters based on "T" such as "ฏ" (to pa-tak), "ฐ" (tho than), "ฑ" (tho montho), "ฒ" (tho phu-thao), "ถ" (tho thung), "ท" (tho thahan) and "ธ" (tho thong). The same is true of letters which are based on "C", for example "ฉ" (cho ching) or "C+H", "ช" (cho chang) or "C+H+1" and "ฌ" (cho choe) or "C+H+2". These groups of letters have the same problem as mentioned above. HMM with many states can have an over-fitting model problem. An over-fitting model generally occurs when a model is excessively complex, such as having too many training cycles, or parameters relative to the number of observations. The model begins to memorize training data rather than learning to generalize from some trend. So its performance is good on the training examples while the performance on unseen data becomes worse. Therefore, the 15 states of HMM produce quite inferior results compared with HMMs that have fewer training cycles. The application example can be shown in Fig.4 and at [19].

Fig 4. Example of Thai letter finger-spelling recognition system

TABLE I. THAI LETTER FINGER-SPELLING.

Letter	Posture	Letter	Posture
ก (ko kai)	K	ธ (tho thong)	T, H, 1
ข (kho khai)	K,1	น (no nu)	N
ค (kho khwai)	K, 2	บ (bo baimai)	B
ฆ (kho ra-khang)	K, 3	ป (po pla)	P, 1
ง (ngo ngu)	N, G	ผ (pho phueng)	P, 2
จ (cho chan)	J	ฝ (fo fa)	F, 1
ฉ (cho ching)	C, H	พ (pho phan)	P
ช (cho chang)	C, H, 1	ฟ (fo fan)	F
ซ (so so)	S, P	ภ (pho sam-phao)	P, 3
ฌ (cho choe)	C, H, 2	ม (mo ma)	M
ญ (yo ying)	Y, 1	ย (yo yak)	Y
ฎ (do cha-da)	D, 1	ร (ro ruea)	R
ฏ (to pa-tak)	T, 5	ล (lo ling)	L
ฐ (tho than)	T, 2	ว (wo waen)	W
ฑ (tho montho)	T, 4	ศ (so sala)	S, 1
ฒ (tho phu-thao)	T, 3	ษ (so rue-si)	S, 2
ณ (no nen)	N, 1	ส (so suea)	S
ด (do dek)	D	ห (ho hip)	H
ต (to tao)	T	ฬ (lo chu-la)	L, 1
ถ (tho thung)	T, 1	อ (o ang)	A
ท (tho thahan)	T, H	ฮ (ho nok-huk)	H, 1

TABLE II. OBSERVATION SYMBOLS.

Posture	Symbol	posture	Symbol	Posture	Symbol
A	1	L	9	Y	17
B	2	M	10	1	18
C	3	N	11	2	19
D	4	P	12	3	20
F	5	R	13	4	21
H	6	S	14	5	22
J	7	T	15		
K	8	W	16		

V. CONCLUSION

We presented a method that enables the estimation of the hand posture for a Thai alphabet finger-spelling recognition system. From previous work, a depth image was used for robust hand region segmentation, and for removing the complex background. The active contour concept calculates the energy function to track the fingertip's position in the frame sequence. The finger shape and hand appearance features were proposed to represent different hand posture sets. The finger shape features are based on fuzzy logic. The hand appearance features consisted of finger relation, hand rotation and hand movement. The simple score vote is computed to compare the similarities between input hand posture and pre-defined hand posture. Since Thai alphabet finger-spelling is based on combination of hand posture sequence in American finger-spelling, therefore, the learning-based method, Hidden Markov Model (HMM), is use to build 42 Thai letter models that recognize the sequence of the American finger-spelling hand postures and provide the Thai alphabet finger-spelling. The letter hand posture recognition result is performed with 5, 10, and 15 states of HMM, and provides the best average recognition rate of 70% for 10 stages. The method does not only apply to American finger-spelling or Thai alphabet finger-spelling. We expect that our method can be applied to other applications, as for example games, robot controlling or visual input devices. However, our main further work is to increase the speed and tracking accuracy of the fingertips. This technique could be better improved, if the post-tracking process such as Kalman filters or particle filtering has been applied. Moreover, there is other topologies of HMM that could be investigated, for example left-to-right, linear or Bakis topologies. If these techniques are improved, then the error should be decreased and greatly impact on the final performance of the system.

ACKNOWLEDGMENT

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, office of the Higher Education Commission.

REFERENCES

[1] A. Erol, et al., Vision-based hand pose estimation: A review , Computer Vision and Image Understanding 108 (2007) 52-73.

[2] A. Park, S. Yun, J. Kim, S. Min and K. Jung, Real-Time Vision-based Korean Finger Spelling Recognition System, Engineering and Technology 34, 293-298, (2008).

[3] B. Sakulsujirapa, M. Karnjanadecha, A. Choksuriwong, Thai Sign Language Recognition Based on Finger Pattern Analysis, In International Conference on Embedded Systems and Intelligent Technology (ICESIT2011), Phuket, Thailand, 9th - 11th February 2011, pages 16.

[4] F. Isikdogan, S. Albayrak, Automatic recognition of Turkish fingerspelling, Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on , vol., no., pp.264-267, 15-18 June 2011.

[5] K. Fujimura, L. Xia, Sign recognition using depth image streams, Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on , vol., no., pp.381-386, 2-6 April 2006.

[6] K. Silanon, N. Suvornvon, Finger-spelling recognition system using fuzzy finger shape and hand appearance features. Fourth International Conference on Digital Information and Communication Technology and its Applicationsm DICTAP 2014, Bangkok, Thailand, May 6-8, 2014. pp. 419-424

[7] L.R. Rabiner, “A tutorial on hidden markov models and selected application in speech recognition,” Proc. IEEE 77, 1989, pp 267-293.

[8] M. Zhenyao, N. Ulrich, Real-time Hand Pose Recognition Using Low-Resolution Depth Images, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006 , vol.2, pp. 1499- 1505, 2006.

[9] P. Goh, E.J. Holden, Dynamic Fingerspelling Recognition using Geometric and Motion Features, Image Processing, 2006 IEEE International Conference on , vol., no., pp.2741-2744, 8-11 Oct. 2006.

[10] R. Akmeliawati, et al., Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network, Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE , vol., no., pp.1-6, 1-3 May 2007.

[11] R. Feris, M. Turk, R. Raskar, T. Karhan; G. Ohashi, Exploiting Depth Discontinuities for Vision-Based Fingerspelling Recognition, Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on , vol., no., pp. 155, 27-02 June 2004.

[12] S. Liwicki, M. Everingham, , “Automatic recognition of fingerspelled words in British Sign Language,” Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on , vol., no., pp.50-57, 20-25 June 2009.

[13] S. Min, et al., Simple Glove-Based Korean Finger Spelling Recognition System, Lecture Notes in Computer Science 4705, 1063. (2007).

[14] S.C.W. Ong, S. Ranganath, Automatic sign language analysis: a survey and the future beyond lexical meaning, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.27, no.6, pp. 873- 891, Jun 2005.

[15] S. Phitakwinai, S. Auephanwiriyakul, N. Theera-Umpon, Thai Sign Language Translation Using Fuzzy C-Means and Scale Invariant Feature Transform, Lecture Notes in Computer Science, Vol 5073, pp. 1107 – 1119, 2008.

[16] S. Ricco, C. Tomasi. Fingerspelling, Recognition through Classification of Letter-to-Letter Transitions, 9th Asian Conference on Computer Vision, pages 214--225, September 2009.

[17] S. Saengsri, V. Niennattrakul, C.A. Ratanamahatana, TFRS: Thai finger-spelling sign language recognition system, Digital Information and Communication Technology and it's Applications (DICTAP), 2012 Second International Conference on , vol., no., pp.457,462, 16-18 May 2012.

[18] T. B. Dinh, et al., Hand gesture classification using boosted cascade of classifiers, Proceedings of the 4th IEEE International Conference on Research, Innovation and Vision for the Future, RIVF'06, p. 139-144, 2006

[19] Thai Alphabets finger spelling, July 15, 2014, Available at (online): http://www.youtube.com/watch?v=tn7J0k1KZis&feature=youtu.

[20] V.I. Pavlovic, R. Sharma, T.S. Huang, Visual interpretation of hand gestures for human-computer interaction: a review, Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.19, no.7, pp.677,695, Jul 1997

[21] W. Kanjanapatmata, Sign language gesture recognition for Thai consonant using artificial neural Networks, King Mongkut's University of Technology North Bangkok, Bangkok, 2005.

[22] W. Veerasakulthong, Thai Hand Sign Recognition, M. Science. Thesis, National Institute of Development Administration, Bangkok, Thailand, 2007.

[23] W. Jiangqin, G. Wen, The recognition of finger-spelling letters for Chinese sign language, the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction, 2001.

[24] X. Teng, et al., A hand gesture recognition system based on local linear embedding, Journal of Visual Languages & Computing Volume 16, Issue 5, October 2005, Pages 442–454.

[25] Y. Sribooruang, P. Kumhom, K. Chamnongthai, Hand posture classification using wavelet moment invariant, Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2004. (VECIMS). 2004 IEEE Symposium on , vol., no., pp.78,82, 12-14 July 2004.

[26] Y. Tabata, T. Kuroda, Finger spelling recognition using distinctive features of hand shape, In: 7th ICDVRAT with Art Abilitation. Maia, Portugal: 2008:287-292.