Finger-Spelling Recognition System using Fuzzy Finger Shape and Hand Appearance Features

Kittasil Silanon
Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University
Hatyai, Songkhla, Thailand 90112
Kittasil.silanon@gmail.com

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

***Abstract*— In this paper, we introduce a method for finger-spelling recognition system. The objective is to help the deaf or non-vocal persons to improve their skills on the finger-spelling. Many researches in this field have proposed methods mostly based on hand posture estimation techniques. We propose an alternative flexible method based on fuzzy finger shape and hand appearance analysis. By using depth image, the hand is extracted and tracked using an active contour like method. Its features, such as, finger shape, and hand appearance, have been defined as chain code, which are input to the American finger-spelling recognition system by using a vote method. The performance of the system is tested in real-time environment, which results in around 70% recognition rate.**

Keywords—Finger-Spelling; Hand Posture Estimation; American Finger-Spelling

I. INTRODUCTION

In this paper, we present our research on sign language, especially related to the finger-spelling. The finger-spelling is a basic communication method for deaf and non-vocal persons, in which the hand posture as symbol will represents the alphabets of words of spoken language, such as, names, places, technical words and etc. However, most of these people, especially children, have problems with finger-spelling skills. Usually, the word-level vocabulary signs have been used for communicating with each other, and only 7% to 10% of the finger-spelling is used in the daily life. Evidently, the finger-spelling skills lag far behind the sign language skills. Our research goal to the field is to develop an automatic recognition system for the finger-spelling, in order to help these people to improve their skills. Actually, many systems specific to a language are proposed, for examples, finger-spelling of American (ASL) [11,13,15], British (BSL) [12], Australian (Auslan) [10], Chinese (CSL) [18], Japanese (JSL) [17] etc. Various researches have been proposed, but most of them cannot achieve the critical criteria, such as, accuracy, flexibility, and real time constraint. There are two principle approaches: glove-based and vision-based. The gloves-based methods [14,16] use electronic sensor devices for digitizing hand joint and finger motion, which give the precision of the hand posture that result in high recognition rate in real time, but these methods are very limited by the environment configuration. The vision-based approach consists of two groups of techniques. Firstly, the model-based method [1,5] uses a kinematic hand model to estimate the articulated hand (i.e., joint angle, finger position), leading to a full reconstruction of the articulated hand posture. Secondly, the appearance-based method [4,9,19] uses computer vision techniques to extract important features from images, such as, point, edge, contour or silhouette, for reconstructing the hand posture, and then, for recognizing the finger-spelling.

In this paper, we proposed a vision-based method for hand posture estimation. The method combines both model and appearance-based method using finger shape and hand appearance features, to finally recognize the American finger-spelling. The system consists of four main parts: 1) hand segmentation, to segment the region of interest of the hand form image sequence, 2) key hand posture selection, to determine the key frame representing the hand posture of finger-spelling from image sequences, 3) hand feature definition, to define the finger shapes and hand appearance features as chain code sequence, 4) finger-spelling recognition, to recognize the finger-spelling from hand features by simply using a scored voting method. The paper details the four parts of our method, then the experimentation results and conclusion, respectively.

II. HAND SEGMENTATION

In this step, the method is focused on the segmentation of the region of interest of the hand from the rest of the image. In our experimentation, the hand is simply defined by the closest object to the camera. Since we used depth image, as shown in Fig.1.b, thus the segmentation of the hand from the complex background can be done by using predefined threshold to obtain the hand's region in image, as shown in Fig.1.c.

(a) (b) (c)

Fig. 1. Depth information: (a) image (b) depth image (c) hand region.

ISBN: 978-1-4799-3724-0/14/$31.00 ©2014 IEEE

After hand's region extraction, the hand's features will be estimated and used as hand initialization, for examples, hand's center, fingertips position, palm size, and etc. These initial features of the hand will be used as reference values to compare with the hand gesture changing in finger-spelling recognition system. For the center of the hand, we used the image moment technique to obtain the center point, as shown in Fig.2.a. The palm size is defined by the distance between hand's center and the closest pixel on hand's contour, as shown in Fig.2.b. The fingertip's positions are points located at the curvature of hand's region, as shown in Fig.2.c. We simply label the five fingers by clockwise sorting of points around the palm's center. The detail of our method in this section is published in [7].

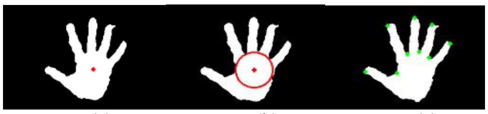

(a) (b) (c)

Fig. 2. Initial features: (a) hand center (b) palm size (c) fingertip position.

III. KEY HAND POSTURE SELECTION

In this section we detail how to select the key frames of the hand posture that represent the finger-spelling alphabets in the image sequences. Note that not every frame in the image sequences is necessary to the recognition system, due to the transition frames (from one alphabet to another) that cannot clearly define the alphabet. The selection condition is that the key frame is selected if the fingertips' position is not changed from the last frame. So, the fingertips must be tracked though the image sequences to determine its changing status. Firstly, the finger area is separated from the palm area. Due to the highest movement of finger during finger-spelling, the finger is then segmented in order to reduce the tracking errors, using its characteristics, stretching and bending fingers, as shown in Fig.3.b.

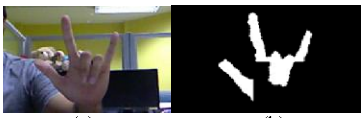

(a) (b)

Fig. 3. Fingers locating: (a) original image (b) finger location.

We assume that the corresponding fingertips in the next frame must be the points located on the contour of the segmented area, which locations do not change much from the current point. The Fig.4.b shows the candidate fingertips, which are calculated with the polygon approximation approach.

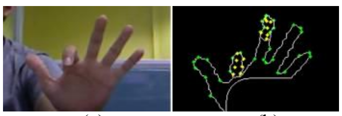

(a) (b)

Fig. 4. Candidate fingertip: (a) original image (b) finger location.

The correspondence of any fingertip between successive frames is then determined by using the concept of active contour [2,8] and the limited searching area. The active contour energy of each candidate fingertip is defined by the energy of continuity, curvature, depth, direction, and distance. The candidate fingertip which has the maximum energy is chosen as its corresponding fingertip in the next frame. The examples of tracking fingertips are shown in Fig.5.

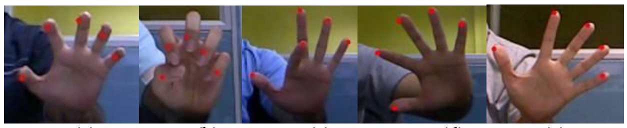

(a) (b) (c) (d) (e)

Fig. 5. Finger tracking: (a) bending (b) closing (c) crossing (d-e) rotation.

Finally, the stable frame will be selected as the key frame of the hand posture. The different position of the fingertip as in (1) between successive frame images is computed in order to detect the temporal difference. The minimum difference is selected as the key hand posture in video. More details of the method are explained in our publication [7].

$$D_k(x,y)=\sum_{i=0}^{n}\left|P_{k-1}(x,y)-P_k(x,y)\right|_i \le \tau_D \qquad (1)$$

IV. HAND FEATURE DEFINITION

The finger shape and hand appearance features have been defined. All of features are converted as feature chain code to discriminate the difference between a set of hand postures.

A. Fuzzy Finger Shape Feature

The finger shape is used to discriminate the difference between a set of hand postures. We define the four basic finger shapes as the following: open, bend, point, and close, as shown in Fig.6.


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

Fig. 6. Finger shape: (a) open (b) bend (c) point (d) close.

In this step, the fuzzy logic [3,6] is applied in order to classify the finger shapes by using the fingertips characteristics obtained from the last section. The process of fuzzy logic is explained in Fig.7. Firstly, a crisp set of input data are gathered and converted to fuzzy set using linguistic variable, fuzzy linguistic terms and membership functions (μ). This step is called as fuzzification. Subsequently, an inference is made for the evaluation result based on a set of rules. Finally, the resulting fuzzy output is mapped to a crisp output using membership functions, in the defuzzification step. In our system, the linguistic variables are defined as depth (D), distance (Dist) and shape (S). The depths at

ISBN: 978-1-4799-3724-0/14/$31.00 ©2014 IEEE

fingertip positions, and distances between fingertip's positions and center of the hand are the input variables, and the shape is the output variable. For linguistic terms, depth and distance variables are decomposed into a set of "Near", "Middle" and "Far" terms, *D(t) = {Near, Middle, Far}* and *Dist(t) = {Near, Middle, Far}*. The range of each term can be shown in Fig.8. For shape variable, linguistic terms are open, bend and point *S(t) = {Open, Bend, Point}*. We do not include "Close" into shape linguistic terms because "Close" shape will be classified immediately when fingertip position is in the palm region. We use the trapezoidal curve as the type of the membership function of each linguistic term. The fuzzification will be performed to convert all crisp inputs to a fuzzy input set.

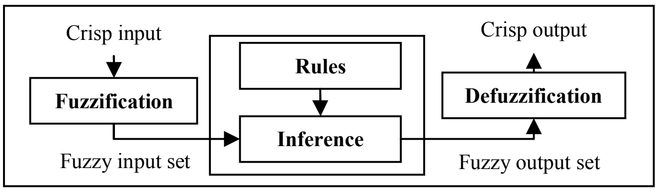
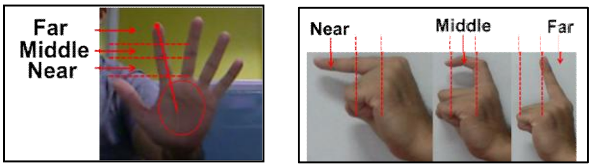


Fig. 7. Fuzzy logic system.



(a) (b)

Fig. 8. Lingusitic terms: (a) distance (b) depth.

The rule is defined for determining the output variables. Our fuzzy rule is defined as a simple IF-THEN rule with a condition and conclusion. Table.1 shows the fuzzy rules for the finger shape classification. Table.2 shows the matrix representation of fuzzy rules. The rows are the depth values, the columns are the distance values, and each cell is the shape output, which corresponds to the specific row and column. For example, the cell (3,2) of the matrix can be described as follows: IF distance is middle AND depth is near THEN shape is point.

TABLE I. FUZZY LOGIC RULE EXAMPLE

Fuzzy Rules
1. IF Distance is Far AND Depth is Far THEN Shape is Open
2. IF Distance is Middel AND Depth is Middle THEN Shape is Bend
3. IF Distance is Near AND Depth is Near THEN Shape is Point

TABLE II. FUZZY MATRIX EXAMPLE

Distance/Depth	**Near**	**Middle**	**Far**
Near	Point	Point	Point
Middle	Point	Bend	Bend
Far	Open	Open	Open

The evaluations of the fuzzy rules and the combination of the results of the individual rules are performed using fuzzy set operations (AND, OR and NOT), this process is called inference. After evaluating the result of each rule and obtaining the final result, the defuzzification will be performed according to the membership function of the fuzzy output variable. For instance, we have the result in Fig.9. at the end of inference.

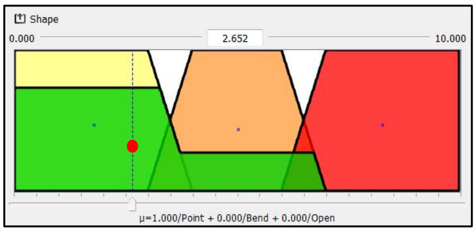


Fig. 9. Fuzzy result.

In this example, all shaded areas belong to the fuzzy result, which must be defuzzified to obtain a final crisp output. There are different algorithms for defuzzification. In our experimentation, we used the center of gravity (COG) in order to map the fuzzy output to a crisp output, represented with a dot in the Fig.9. The COG algorithm can be written as (2).

$$COG = \frac{\sum_{i=1}^{N} x\mu(x)}{\sum_{i=1}^{N} \mu(x)} \quad (2)$$

B. Hand Appearance Fetaure

Considering only the finger shapes is not sufficient to estimate the correct hand posture used in the finger-spelling. We found that, in general, there are some hand postures having the similar finger shape, but having a different hand appearance. Thus, we define the additional hand appearance features for increasing robustness of the hand features. The hand appearance features are defined by the relative distance and position between fingers in three statuses: "close", "separate" and "cross". Two more features are defined as the hand rotation and hand movement. The hand appearance feature will be combined with finger shape features in order to increase the hand posture discrimination. The set of the hand appearance features in finger-spelling system are shown in Fig.10.

(a) (b) (c) (d) (e)

Fig. 10. Hand Appearance: (a) group (b) separate (c) cross (d) rotation (e) movment.

ISBN: 978-1-4799-3724-0/14/$31.00 ©2014 IEEE

C. *Feature Codes*

To represent the hand posture, the finger shape and hand appearance features are encoded as shown in table.3. In our system, the feature vector is composed of 17 values defined as showed in Fig.11. The first 5 elements (f_1,f_2,f_3,f_4,f_5) indicate the finger shape (thumb, index, middle, ring and pinky respectively). In Fig.11, (f_1,f_2,f_3,f_4,f_5) = (4,1,1,4,4) means that the thumb is "close", index and middle are "open", ring and pinky are "close", respectively. The next 4 elements (f_{12} ,f_{13},f_{14},f_{15}) describe the relation between thumb with index, middle, ring and pinky fingers. In Fig.11, (f_{12},f_{13},f_{14},f_{15}) = (6,6,6,6) indicates that all relations of thumb with other fingers are "separate". The next 3 elements (f_{23},f_{24},f_{25}) describe the relation of index with middle, ring and pinky fingers. The next 2 elements (f_{34},f_{35}) describe the relation of middle with ring and pinky fingers. In Fig.11, (f_{34},f_{35}) = (6,6) indicates that all relations of middle with other fingers are "separate". The element (f_{45}) describes the relation of ring and pinky. In Fig.11, (f_{45}) = (5) indicates that the relation of ring with pinky finger is "close". The next 2 elements (H_r,H_m) represent the rotation and movement of hand. In Fig.11, (H_r,H_m) = (0,0) indicates that there is no rotation or movement for this hand posture. Finally, the vector feature of the hand posture in Fig.11 have the chain codes as P = {4,1,1,4,4,6,6,6,6,6,6,6,6,6,5,0,0}.

TABLE III. FEATURE CODES

Finger shape	Hand appearance
Open = 1	Group = 5
Bend = 2	Separate = 6
Point = 3	Cross = 7
Close = 4	Rotation = 8, No-Rotation = 0
	Movement = 9, No-Movement = 0

Hand posture codes

f_1 = 4	f_2 = 1	f_3 = 1	f_4 = 4	f_5=4
f_{12} = 6	f_{13} = 6	f_{14} = 6	f_{15} = 6	
f_{23} = 6	f_{24} = 6	f_{25} = 6		
f_{34} = 6	f_{35} = 6			
f_{45} = 6				
H_r=0	H_m= 0			

Fig. 11. Hand posture encoding.

V. SCORED VOTING RECOGNITION

The voting and scoring methods are combined and used in our recognition process. This method is simple and effective and can run in real time. Firstly, the *n* patterns of the hand posture chain codes (P_i | i = 1…N) are the pre-defined and the off-line computation. Each pattern represents a finger-spelling that is the chain code of finger shape features and hand appearance features. Secondly, any hand posture, which is captured from the camera, will be detected and encoded into the same type of chain codes. Thirdly, the similarity between the detected pattern and the *n* pre-defined patterns are performed by using a simple scored voting method. This technique will compare each element of the input chain codes with the corresponding element of the pattern chain code. The score will be given if any element has matched. Thus, the hand posture will be recognized to the pattern which gives the maximum score. The method can be used to recognize the different sets of hand posture by using different pre-defined corresponding chain codes for each of the hand posture. The scored voting recognition process can be shown in Fig.12.

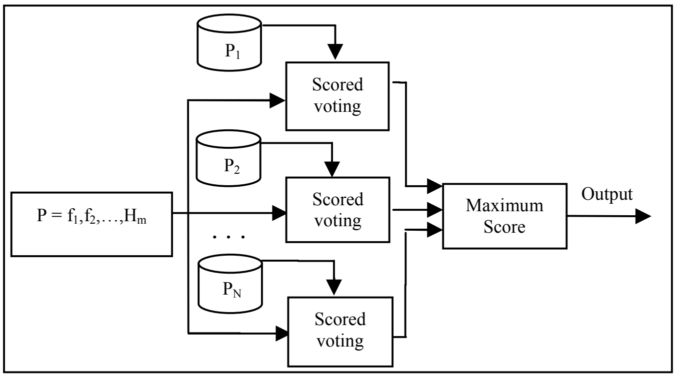


Fig. 12. Scored voting recognition process.

VI. EXPERIMENT RESULT

We applied our method to the American finger-spelling. There are 31 patterns of the pre-defined chain code of hand postures, consisting of 26 American-English alphabets (A-Z) and 5 hand postures for 1-5 numbers. Each hand posture is tested at 100 rounds. All hand postures and the feature chain code are shown in Fig.15 and table 4. The result of the recognition is shown in Fig.14, and the real-time example of our system in shown in Fig.13, available at: www.youtube.com/watch?v=qW7i6TyVBaM&feature=youtu.be

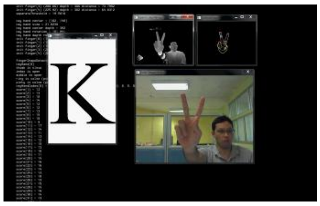


Fig. 13. Real-time American finger-spelling recognition system.

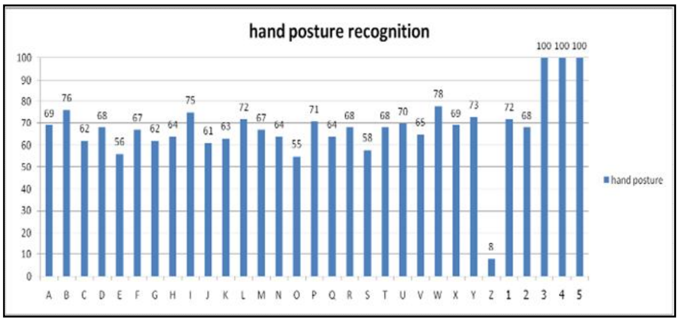


Fig. 14. Real-time American finger-spelling recognition system.

As the result shown in Fig.15, the best cases (100% recognition rate) of the system are the hand postures of the numbers “3”, “4” and “5”. The hand postures of these 3 cases are not much different from initial hand posture (Open hand). The movement of finger does not much change. Thus, finger with immediate changing direction that increases the tracking error. For the general cases (65% recognition rate), “A-Y” and “1-2” hand postures, most of these hand postures have only some finger movements continuously, and does not change much from the previous frame. Hence, our finger tracking using active contour concept can perform more accurately tracking process gives less error. The worst case (less than 10% recognition rate) is the hand posture of “Z” alphabet. We found that this hand posture has the rapid hand movement

VII. CONCLUSION

We proposed an automatic recognition system for finger-spelling using hand posture estimation. Depth image is used for robust hand region segmentation and removing the complex background. The active contour concept is applied for fingertip tracking. The finger-shape and hand appearance features are proposed. The scored voting method is introduced for the n-pattern hand posture chain codes recognition. Our system is tested with the American finger-spelling including 26 hand postures for the A-Z alphabets and 5 hand postures for the 1-5 numbers. The recognition rate (not including “Z” alphabet hand posture) is performed in average at around 70%. The system can be performed in real time. We expect that our method can be applied to the other hand posture applications, such as, game, robotic controlling, visual input device, another language finger-spelling, and etc. However, our main future work is to increase the tracking accuracy of the fingertips that greatly effect to the final performance of the system.

ACKNOWLEDGMENT

We would like to thank the National Research University Project of Thailand’s Office of the Higher Education Commission for financial support.

REFERENCES

[1] C. C. Lien, “A scalable model-based hand posture analysis system,” Machine Vision and Applications, 2005, Volume 16, Number 3, Page 157.

[2] D.J. Williams and M. Shah, “A Fast algorithm for active contours and curvature estimation,” In: CVGIP:Image Understanding, vol.55, no.1, pp.14-26(January 1992).

[3] E. Cox, “Fuzzy fundamentals,” Spectrum, IEEE, vol.29, no.10, pp.58-61, Oct. 1992

[4] G. Panin, S. Klose, A. Knoll, “Real-Time Articulated Hand Detection and Pose Estimation,” Advances in Visual Computing, Lecture Note in Computer Science, vol 5876, pp.1131-1140,2009.

[5] J. Lee and T.L Kunii, "Model-based analysis of hand posture," Computer Graphics and Applications, IEEE , vol.15, no.5, pp.77,86, Sep 1995.

[6] J.M. Mendel, “Fuzzy logic systems for engineering: a tutorial,” Proceeding of the IEEE, vol.83, no.3, pp.345-377, Mar 1995.

[7] K. Silanon and N. Suvonvorn, “Fingertips Tracking Based Active Contour for General HCI Application,” (DaEng-2013) Lecture Note in Electrical Engineering, vol.285, 2014, pp 309-316.

[8] M. kass, A. Witkin, D. Terzopoulos, “Snake: Active contour models,” In: International journal of computer vision, vol.1, no.4, pp.321-331 (1988).

[9] N. Shimada, K.. Kimura, Y. Shirai, “Real-time 3D hand posture estimation based on 2D appearance retrieval using monocular camera,” Proc. ICCV workshop on Recognition, Analysis, and Tracking of Face and Gesture in Real-Time System, Vancouver, Canada, pp.23-30.

[10] P. Goh and 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.

[11] R. Feris, M. Turk, R. Raskar, K. Tan, 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 and 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. Ricco and C. Tomasi, “Fingerspelling Recognition through Classification of Letter-to-Letter Transitions,” 9th Asian Conference on Computer Vision, pages 214--225, September 2009.

[14] S.S. Fels and G.E. Hinton, “Glove-talk: A neural network interface between a data-glove and a speech synthesizer,” IEEE Transactions on Neural Networks, vol.4, pp.2-8, 1993.

[15] T.B. Dinh, V.B. Dang, D.A. Duong, T.T. Nguyen, D.D. Le, “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.

[16] T. Baudel and M.B. Lafon, “CHARADE: Remote Control of Object using FreeHand Gesture,” Communications of the ACM, vol 36, pp.29-35, 1993.

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

[18] W. Jiangqin and G. Wen, “The recognition of finger-spelling alphabet for chinese sign language,” the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction.

[19] Z. Pan, Y. Li, M. Zhang, C. Sun, K. Guo, X. Tang, S.Z. Zhou, “A real-time multi-cue hand tracking algorithm based on computer vision,” Virtual Reality Conference (VR), 2010 IEEE, vol.,no.,pp.219-222, 20-24 March 2010.

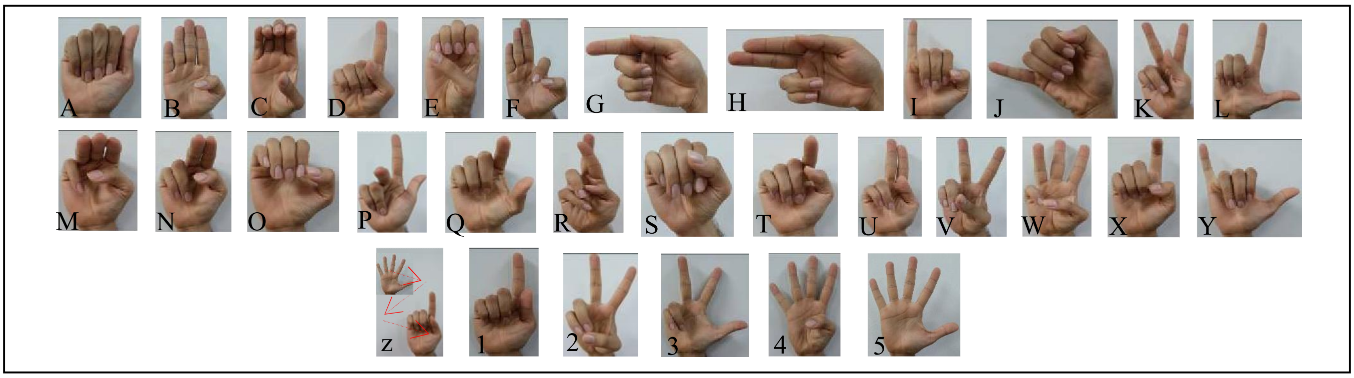


Fig. 15. Hand posture sets.

TABLE IV. FEATURE CODES

Hand posture	F_1	F_2	F_3	F_4	F_5	F_{12}	F_{13}	F_{14}	F_{15}	F_{23}	F_{24}	F_{25}	F_{34}	F_{35}	F_{45}	H_r	H_m
A	4	4	4	4	4	6	6	6	6	5	6	6	5	6	5	0	0
B	4	1	1	1	1	6	6	6	6	5	6	6	5	6	5	0	0
C	4	2	2	2	2	6	6	6	6	5	6	6	5	6	5	0	0
D	4	1	4	4	4	6	5	6	6	6	6	6	5	6	5	0	0
E	4	4	4	4	4	6	6	6	5	5	6	6	5	6	5	0	0
F	4	4	1	1	1	5	6	6	6	6	6	6	5	6	5	0	0
G	4	1	4	4	4	6	6	6	6	6	6	6	5	6	6	8	0
H	4	1	1	4	4	6	6	6	6	5	6	6	6	6	5	8	0
I	4	4	4	4	1	5	6	6	6	5	6	6	5	6	6	0	0
J	4	4	4	4	1	5	6	6	6	5	6	6	5	6	6	8	0
K	4	1	1	4	4	6	6	6	6	6	6	6	6	6	5	0	0
L	1	1	4	4	4	6	6	6	6	6	6	6	5	6	5	0	0
M	4	3	3	3	4	6	6	6	5	5	6	6	5	6	6	0	0
N	4	3	3	4	4	6	6	5	6	5	6	6	6	6	5	0	0
O	4	4	4	4	4	5	6	6	6	5	6	6	5	6	5	0	0
P	1	1	3	4	4	6	6	6	6	6	6	6	6	6	5	0	0
Q	3	3	4	4	4	6	6	6	6	6	6	6	5	6	5	0	0
R	4	1	1	4	4	6	6	6	6	7	6	6	6	6	5	0	0
S	4	4	4	4	4	7	5	6	6	5	6	6	5	6	5	0	0
T	4	3	4	4	4	6	5	6	6	6	6	6	5	6	5	0	0
U	4	1	1	4	4	6	6	6	6	5	6	6	6	6	5	0	0
V	4	1	1	4	4	6	6	5	6	6	6	6	6	6	5	0	0
W	4	1	1	1	4	6	6	6	5	6	6	6	6	6	6	0	0
X	4	2	4	4	4	6	5	6	6	6	6	6	5	6	5	0	0
Y	1	4	4	4	1	6	6	6	6	5	6	6	5	6	6	0	0
Z	4	1	4	4	4	6	5	6	6	6	6	6	5	6	5	0	9
1	4	1	4	4	4	6	6	6	6	6	6	6	5	6	5	0	0
2	4	1	1	4	4	6	6	5	5	6	6	6	5	6	5	0	0
3	1	1	1	4	4	6	6	6	6	6	6	6	6	6	5	0	0
4	4	1	1	1	1	6	6	6	6	6	6	6	6	6	6	0	0
5	1	1	1	1	1	6	6	6	6	6	6	6	6	6	6	0	0

ISBN: 978-1-4799-3724-0/14/$31.00 ©2014 IEEE