Real-time Recognition System of Korean Sign Language based on Elementary Components

Chan-Su Lee
Dep. of Sensitivity Eng.,
SERI, Korea
chanse@ctrsys.kaist.ac.kr

Gyu-tae Park
Dept. of EE,
KAIST
qpark@ctrsys.kaist.ac.kr

Jong-Sung Kim
Artificial Intelligence Sect.,
ETRI
joskim@etri.re.kr

Zeungnam Bien
Dept. of EE,
KAIST
zbien@ee.kaist.ac.kr

Won Jang
Agency for Defense Development,
Taejon, 305-600, Korea

Sung-Kwon Kim
Samsung Electronics,
Korea

Abstract

Sign language is a method of communication for deaf persons. In communication using hand gesture, sign words and manual alphabets are used together. In this paper is proposed a system which recognizes Korean Sign Language(KSL). KSL is composed of Korean sign words and Korean manual alphabets continuously. To recognize meanings of continuous gestures which have no token of beginning and end, this system segments current motion states using speed and change of speed in motions and state automata. To understand the meaning of a gesture, basic component classifiers using Fuzzy Min-Max Neural Network and fuzzy logic are used. Basic elements of meaning used in this system are 14 hand directions, 23 hand postures, and 14 hand orientations. Meaning of signed gesture is interpreted using basic elements which were recognized by 3 classifiers. This system recognizes 31 Korean manual alphabets and 131 Korean signs in real-time with recognition rate 96.7% for Korean manual alphabets and 94.3% for Korean sign words in case of excluding no recognition case.

Keywords: sign language, gesture segmentation, fuzzy pattern classifier, fuzzy application.

1 Introduction

Human hands are used to act on the world, to grasp and explore objects, to express our ideas. Hand gestures are used for communication among people. In the last several years there has been an increased interest in trying to introduce means of human-to-human interaction to the field of human-computer interaction [1]. The sign language is an important method of communication for deaf-persons. As sign language is well structured code gesture[2], each gesture has assigned meaning and it can be used to express complex meanings by combining basic elements. This paper deals with a system which recognizes the Korean Sign Language(KSL) continuously.

Attempts for recognizing sign language have begun to appear in the literature over the past several years. But those systems have generally limited to manual alphabets[3] or isolated signs[4] and they have small training and test sets. The expandable sign language recognition system without retraining has been attempted using Self Organizing Map[4]. However, this method requires to retrain the system for relabeling signs to recognize. Another approach for expandable sign recognition system is using basic components of signs[5]. The meaning of sign is selected by the result of basic components. In this system, segmentation of continuous gesture into isolated one is critical problem because without proper segmentation to isolated signs proper meaning components cannot be recognized. Recently, Hidden Markov Model(HMM) has been attempted to segment and to recognize continuous gestures[6]. But it has difficulties in expanding vocabulary because it must be retrained for all new sign words and it also has difficulties in recognition of manual alphabet, because manual alphabet has few motions. In this paper, we purpose methods for solving segmentation of continuous gestures and for recognition of manual alphabet and sign language continuously with ex-

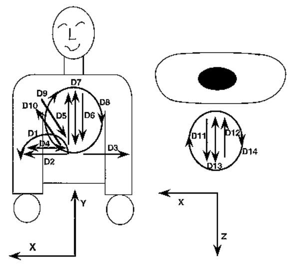


Figure 1. Basic elements of hand direction

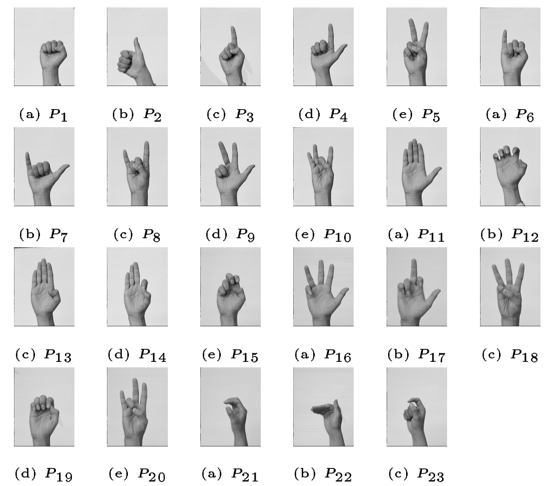


Figure 2. Basic elements of hand posture

pansibility.

In section 2, After analyzing KSL, basic elements for KSL are defined. In section 3, state automata for segmentation of continuous gesture and pattern classification for recognition of basic elements is presented. In section 4, experimental results of our recognition system for continuous gesture is presented. Lastly, section 5 is remarked as conclusion and future work.

2 Analysis of KSL

KSL are composed of Korean sign words and Korean manual alphabets. According to standard KSL dictionary[7], Korean Sign Language has vocabulary of 6000 sign words and 31 manual alphabets. Most of the signs are formed by combining some basic signs. This paper deals with basic signs, because compound signs which are composed of basic signs could be recognized using basic sign recognizer and proper interpreter. 31 manual alphabets are used to express proper nouns, postpositional words or other words which have no defined gesture. KSL uses Korean sign words and Korean manual alphabets continuously to express a sign sentence. Therefore, sign language recognition system for communication with deaf person is necessary to recognize sign words and manual alphabets continuously. The scope of this paper is to design recognition system which can recognize 131 basic sign languages and 31 manual alphabets continuously in real time.

Basic elements of sign language are composed of cheremes which have similar functions in sign language as phonemes have in the spoken language. After analyzing basic signs, 14 basic direction elements are defined. Figure 1 shows basic elements of hand direction. Considering basic signs and 31 manual alphabets, 23 basic posture elements and 14 basic orientation elements are selected. Figure 2 shows basic elements of hand posture. In this paper, these 14 direction elements, 23 posture elements and 14 orientation elements are used for basic elements for KSL recognition system.

3 Segmentation & Classification for KSL

KSL recognition system has four stage. First stage is data acquisition stage. In this stage this system gets flexure of fingers, position of hand and orientation of hand. 18 flexures of fingers are sensed by CyberGlove^{TM} and x,y,z and roll, pitch, yaw are sensed by Polhemus sensor. In the second stage, state of motion is estimated using motion partition and state automata. For recognition of KSL successively, segmentation of continuous gesture into isolated ones by detecting beginning and end of each gesture is necessary. In the third stage, basic elements of each gesture are recognized by 3 classifiers and a feature extractor. In the fourth stage, meanings of gesture are recognized. Result of direction classifier, post classifier and orientation classifier are used. Overall KSL recognition system are shown in Figure 3.

3.1 Segmentation of motion

As there are no explicit indications for beginning and end of gesture, it is difficult to recognize continuous gesture. But human can catch beginning and end state of motion, that is, there are implicit beginning and end motion state of each sign. We can catch these motion state using motion phase partitioned by speed and change of speed and state automata for KSL. In doing continuous gesture, there are motions with no intention. Movement to change position is one of the

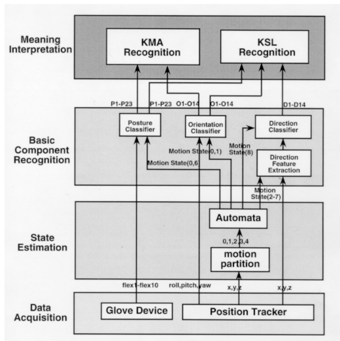


Figure 3. Overall KSL recognition system

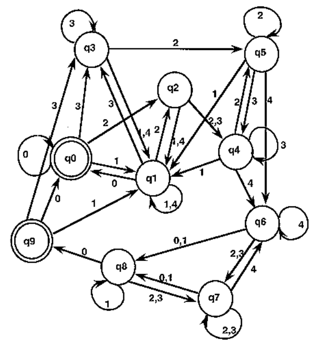


Figure 4. State transition diagram for KSL

Phase (event)	*Condition* speed	Speed Change
Stop(0)	0	+−, 0
Preparation(1)	+	+, -
Stroke(2)	+++	++
Moving(3)	++	+, 0, -
End(4)	+	−

Table 1. Condition for segmentation of motion phase

State	Description	Function
q_0	no motion state	initialize manual alp. rec.
q_1	slow motion in initial	Initialize orientation rec.
q_2	Stroke at starting	feature extraction
q_3	Moving without Stroke	feature extraction
q_4	Moving after Stroke	feature extraction
q_5	Stroke after Moving	feature extraction
q_6	Ending Motion Reduction of speed	feature extraction posture rec.
q_7	Repeatative Moving	feature extraction
q_8	end preparation	direct rec.
q_9	no motion state	KSL rec.

Table 2. Description of motion state

typical example. Using state automata, meaningless motions can be rejected before recognition.

Comparing each gesture flow with typical one, we can distinguish no intentional gestures from intentional one. Finite state automata(FSA) is composed of finite alphabet, finite state set, state transition function, initial state and set of final states. And it is able to reject or accept some type of string. State automata are modification of FSA from finite alphabets and state sets to countable infinite event sets and state spaces. As continuous gesture can be done infinitely, state automata is suitable. For segmentation of continuous gesture and removal of meaningless gesture from interpretation, state automata[8] is implemented considering speed and change of speed.

For implementation of state automata, event which causing state transition must be given. Three distinct motion of phases typically constitute a gesture[9]:preparation, stroke and retraction. We partitioned motion into 5 phase for KSL. Table 1 show the condition for motion partition. Threshold values for partition of each motion phase are set by experiment. This motion phase is used as event for state automata. Quek[10] developed set of rules for gesture segmentation based on the gesture pattern. In the basis of his rules and investigation of KSL, state for KSL and state transition function for KSL state automata are implemented.

Figure 4 is state transition diagram for KSL state automata. Table 2 describes each state of automata briefly. q_0, q_9 are set of final states. sign words are recognized in state q_9 and manual alphabets are recognized in state q_0. In this state automata, if any sign is not played properly with stroke then automata state cannot reach q_9, and of gesture can be known

3.2 Direction classification using feature extraction and fuzzy logic

Direction classifier recognizes 14 direction elements. Direction classifier is composed of two modules, which are feature extraction and direction classification by fuzzy rule. Feature extraction is executed when state of motion is $q_2 - q_7$ where the hand move exceeding some value of speed. To treat these 3D data effectively, we use feature extraction. Feature is selected to distinguish 14 basic elements of direction effectively Selected features for direction classification are as follow. k is current sampling time step.

(1) *Cumulative length of motion*

This is a cumulative length of motion trajectory over each axes.

$$\begin{aligned} TL_{x,y,z}(k) &= TL_{x,y,z}(k-1) + L_{x,y,z} \\ L_{x,y,z} &= |R_{x,y,z}(k) - R_{x,y,z}(k-1)| \end{aligned} \tag{1}$$

(2) *Cumulative Direction Change in radians*

This is sum of direction change. In circular motion it's value is very large

$$TC_{xy} = TC_{xy}(k-1) \quad +(tan2^{-1}(\frac{y(k)-y(k-1)}{x(k)-x(k-1)}) - tan2^{-1}(\frac{y(k-1)-y(k-2)}{x(k-1)-x(k-2)}) \tag{2}$$

$$TC_{xz} = TC_{xz}(k-1) \quad +tan2^{-1}(\frac{z(k)-z(k-1)}{z(k)-z(k-1)}) - tan2^{-1}(\frac{z(k-1)-z(k-2)}{x(k-1)-x(k-2)}) \tag{3}$$

(3) *Relative end Position*

This is a relative end position of sign gesture from the beginning position of gesture.

$$\begin{aligned} EP_{x,y,z} = \quad & R_{x,y,z}(k) - R_{x,y,z}(0) \\ & R_{x,y,z}(0) : initial\ sampling\ time\ data \end{aligned}$$

(4) *Change of Direction*

This is number of direction changes over each axes

$$CD_x = CD_x + f(x) \tag{4}$$

$$CD_y = CD_y + f(y) \tag{5}$$

$$CD_z = CD_z + f(z) \tag{6}$$

$$f(x) = \begin{cases} 1, & if\ [R_x(k) - R_x(k-1)] \\ & \times[R_x(k-1) - R_x(k-2)] < 0 \\ 0, & otherwise \end{cases}$$

Using those features, we applied fuzzy inference rules for classifying 14 basic directions. Table 3 shows rules for recognizing each hand direction. Membership function is gaussian type and its mean and deviation parameter is determined by fuzzy c-means clustering for each feature and modification by training is used Mamdani's Max-Min inference method is applied to this operation.

Direction	Rules
D_1	((T_x=PM or T_y=PL) and (T_x=PS and (T_y=PM or T_Y=PL))) and (R_x=PM or R_x=PL) and R_y=ZO and R_z=ZO and TC_{xy}=PM
D_2	(T_x=PS or T_x=PM) and T_y=ZO and (R_x=PM or R_x=PL) and CD_x=ZO and T_z=ZO
D_3	(T_x=PS or T_x=PM) and T_y=ZO and (R_x=NM or R_x=NL) and CD_x=ZO and T_z = ZO
D_4	(T_x=PM or T_x=PL) and (T_y =ZO or T_y=PS) and T_z=ZO and (R_x=ZO or R_x=PL or R_x= NL) and CD_x = PO and TCP_{xy} = ZO)
D_5	T_x=ZO and (T_y=PS and T_y=PM) and T_z =ZO and (R_y=NM or R_y=NL) and R_z=ZO and CD_x=ZO
D_6	(T_x=ZO and (T_y=PS or T_y=PM) and T_z = ZO and (R_y=PM or R_y=PL) and $R_z = ZO$ and CD_y=ZO
D_7	T_x=ZO and (T_y=PM or T_y=PL) and T_z=ZO and (R_x=ZO or R_y=PL or R_y=NL) and CD_y=PO
D_8	(T_x=PM or T_x=PL) or (T_y=PM or T_y=PL) and (T_z=ZO or T_z= PS) and R_x=ZO and (R_y=ZO or R_y=NM) and TC_{xy}=ZO
D_9	(T_x=PS and T_y=PS) or ((T_x=PS or T_x=PM) and T_y=PM)) or (T_x=PL and T_y=PL) and (R_x=NM or R_x=NL) and (R_y=NM or R_y=NL) and TC_{xy}=ZO
D_{10}	(T_x=PS and T_y=PS) or ((T_x=PS or T_x=PM) and T_y=PM)) or (T_x=PL and T_y=PL) and (R_x=PM or R_x=PL) and (R_y=PM or R_y=PL) and TC_{xy}=ZO
D_{11}	T_x=ZO and (T_y=ZO or T_y=PS) and (T_z=PS or T_z=PM) and (R_z=PM or R_z=PL) and CD_z=ZO
D_{12}	T_x=ZO and (T_y=ZO or T_y=PS) and (T_z=PS or T_z=PM) and (R_z=NM or R_z=NL) and CD_z=ZO
D_{13}	(T_x=ZO or T_x=PS) and ((T_y=ZO or T_y=PS) and (T_z=PM or T_z=PL)) and (T_y=PM and T_z=PL) and (R_z=ZO and R_z=PL or R_z=NL) and CD_z=PO
D_{14}	(T_x=PM or T_x=PL) and (T_y=ZO and T_y=PS and T_y=PM) and (T_z=PM or T_z=PL) and (R_x= NM) and R_z=ZO and TC_{xz}=PL

Table 3. Inference rules for direction classification

3.3 Posture and orientation classification by FMMNN classifier

Posture data are very sensitive to hand size of signer. For same signer, there are shape variations during gesture. To overcome these variations, on line adaptation ability is required for posture classifier. We use Fuzzy Min-Max Neural Network[11] for posture classifier. Fuzzy Min-Max Neural Network can have many hyperboxes for one class, so orientation data which have discontinuities in sensing data can be delt with FMMNN by defining different hyperboxes for each class in discontinuous pointer.

A fuzzy set hyperbox is defined by min pointers($\boldsymbol{V}$) and max pointers($\boldsymbol{W}$) with corresponding membership functions. The initial min-max values($\boldsymbol{V_0}$,$\boldsymbol{W_0}$) of the network are determined by training with sample data. The membership function of the hyperbox is expressed as follows.

$$\begin{aligned} b_j(\ \boldsymbol{A_h}, \boldsymbol{V_j}, \boldsymbol{W_j}\ ) &= \frac{1}{N}\sum_{i=1}^{N}[1 - f(a_{hi} - w_{ji}, \gamma) \\ &- \ f(v_{ij} - a_{hi}, \gamma)] \\ f(x,\gamma) &= \begin{cases} 1 & ,\ x\gamma > 1 \\ x\gamma & ,\ 00 \le x\gamma \le 1 \\ 0 & ,\ x\gamma < 0 \end{cases} \end{aligned}$$

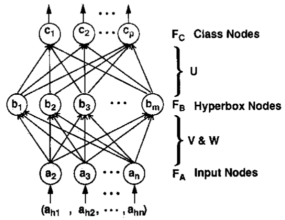


Figure 5. The Structure of Fuzzy Min-Max Neural Network

(a) Hello (b) Meet

Figure 6. Korean sign language for 'Hello' and 'Meet'

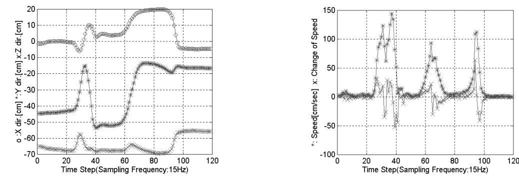
$$\theta = \frac{1}{N}\sum_{i=1}^{N}(max(w_{ij}, x_{hi} - min(v_{ij}, x_{hi}))$$

where, $\boldsymbol{A_j}$ is input for FMMNN classifier. In case of posture classifier, $\boldsymbol{A_j}$ is normalized flexure data from CyberGlove^{TM} and N=10. In case of orientation classifier, is orientation data from Polhemus and N=3. γ is the sensitivity parameter that regulates how fast the membership value decreases as the distance between a_{hi} and b_j increases. And θ defines the size of maximum hyperbox. Hyperbox number is decreased as θ increased Using these initial min-max value and given γ, FMMNN recognize posture and orientation class adapted hyperbox within the limit θ.

4 Experimental Results

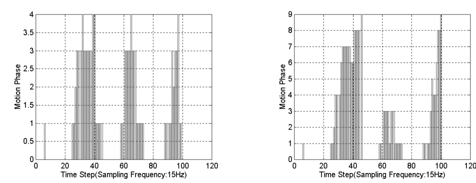
All functions run on a main computer called Indigo2^{TM}. The CyberGlove^{TM} interface unit is connected to a serial port of main computer. In this section one example of two continuous sign will be shown.

Figure 6 shows Korean Sign Language 'hello(an-nung-ha-ship-ni-ga)' and 'meet(man-na-da)'. These two sign words are played continuously. Figure 7 shows acquired data of motion for each axis and speed and change of speed for these 2 signs. Figure 8 shows partitioned motion phase, of which output is event for KSL automata. 'Hello' and 'meet' have different initial position. So in 3 motions for 2 signs, the middle motions is meaningless. But Figure 7 do not give this information. Figure 8 is the output of state automata. According to them end state of motion, second motion is meaningless one. For intended gesture, end of motion state is q_9. This example shows automata for KSL can reject meaningless gesture. Every isolated signs is recognized by basic element classifiers. Basic elements are recognized by direction classifier, posture classifier and orientation classifier. Directions are recognized by using extracted features for 3D movements and fuzzy rule for each direction. Figure 9 shows recognition result of direction for 'meet'. D_3 is maxmum membership value and it's value is more than 0.5. we can know this motion is proper D_3 motion. If max membership value is



(a) 3D movement (b) Speed and change of speed

Figure 7. Sensing data and speed



(a) Motion phase (b) State

Figure 8. Estimated phase and state

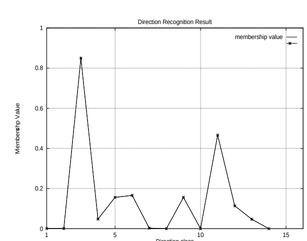
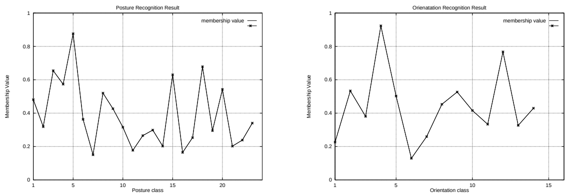


Figure 9. Recognition Result of Direction($D3$)



(a) Postrure(P_5) (b) Orientation(O_4)

Figure 10. Recognition example for manual alphabet(P_5,O_4)

less than 0.5, we reject it's recognition result. Recognition rate for direction classifier is 95.7% in average.

Basic elements for posture and orientation classifier are recognized by FMMNN. FMMNN for posture classifier is set to θ=0.085, γ=6, then 79 hyperboxes are generated in this case. For single signer, the recognition rate is 96.5%. Orientation classifier is also trained with 10 sample data for each basic orientation class. FMMNN orientation classifier is set to θ=0.03,γ=4. The recognition rate is 97.8% for careful gesture. Korean manual alphabets are recognized by posture and orientation classifier. Figure 10 shows recognition result of posture classifier and orientation classifier for Korean manual alphabet. Knowing that element posture type is 'P_5' and element orientation type is 'O_6', interpreter for manual alphabet can recognize it's meaning. The meaning of this manual alphabet is ''(U). KSL is recognized by posture classifier, orientation classifier and direction classifier. For sign gesture of 'meet', direction is 'D_3', posture class is 'P_3' and orientation class is 'O_6'. Comparing recognized results with these basic elements, this system can easily interpret the meaning of motion. Recognition rate of 131 KSL is 80.1% in average in considering rejection as error. But if excluding no recognition case, the recognition rate is up to 95% because many of ambiguous gestures are rejected recognition by automata and by elementary classifier related to fuzzy membership value. Only by adding code of basic elements for new gesture in interpreter, this system can recognize another gesture without any retraining.

5 Conclusion and Future work

In this paper we presented the real-time recognition system for Korean sign words and Korean manual alphabets. State automata is suggested for segmentation and feature extraction. Fuzzy rules are used for direction classification. For posture and orientation recognition, FMMNN is used. Gesture meaning is recognized by interpreter which using results of basic element classifiers.

Automatic fuzzy rule generation for direction pattern classifier is for future study.

References

[1] T. S. H. at el., "Hand gesture modeling, analysis, and synthesis," in *Proc. of Int. Workshop on Automatic Face - and Gesture - Recognition*, vol. 2, pp. 73 – 79, 1995.

[2] D. Morris, *Man Watching: A field Guide to Human Behaviour*. Grafton, 1968.

[3] T. Takahashi and F. Kinsino, "Hand gesture coding based on experiments and using a hand gesture interface devices," *SIGCHI*, no. 2, pp. 67 – 73, 1991.

[4] M. B. Waldron and S. Kim, "Isolated asl sign recognition system for deaf persons," *IEEE Trans. Rehabilitation Engineering*, vol. 3, pp. 261 – 271, Sept. 1995.

[5] J. S. Kim, W. Jang, and Z. Bien, "A dynamic gesture recognition system for the korean sign language(ksl)," *IEEE Trans. on Systems, Man, and Cybernatics*, pp. 354 – 359, April 1996.

[6] T. Starner and A. Pentland, "Visual recognition of american sign language using hidden markov models," in *Proc. of Int. Workshop on Automatic Face - and Gesture - Recognition*, pp. 189 – 194, 1992.

[7] S.G.Kim, *The Korean Sign Language Dictionary(written in korean)*. Oh-Sung Press, 1993.

[8] C. G. Cassandras, *Discrete Event Systems*. Boston: IRWIN, 1993.

[9] A. Kendon, *Relationship between Verbal and Nonverbal Communication*, ch. Gesticulation and speech: Two aspects of the process if utterance. Mouton, 1980.

[10] K. H. Quek, "Toward a vision-based hand gesture interface," in *Int. Proc. of the VRST'94*, pp. 17 – 31, 1994.

[11] P. Simpson, "Fuzzy min-max neural networks-part 1 : Classification," *IEEE Trans. on Neural Networks*, vol. 3, pp. 776 – 786, Sep. 1992.