TABLE VII
COMPARISONS OF THE PROPOSED SYSTEM WITH OTHER IDENTIFIERS

Model Name	Final Number of Rules	Final Number of Parameters
Narendra etc.[12]	$\mathcal{N}^3_{2,20,10,1}$	250
L.X.Wang[16]	40	200
proposed FNNS	22	84

can be reduced. A measure of the similarity for fuzzy sets, which indicates the degree to which two fuzzy sets are equal, is also applied to combine similar input linguistic term nodes of a fuzzy neural network. This greatly reduces the number of adjustable parameters. We also derive a new and effective on-line initialization method for choosing the initial parameters of the FNNS. A computer simulation has been presented to illustrate the procedure of the proposed FNNS. The simulation shows that the FNNS indeed yields simpler and more efficient results.

REFERENCES

[1] E. Backer, *Cluster Analysis by Optimal Decomposition of Induced Fuzzy Sets.* Amsterdam: Delft Univ. Press, 1978.

[2] D. Dubois and H. Prade, "A unifying view of comparison indices in a fuzzy set theoretic framework," in *Fuzzy Sets and Possibility Theory: Recent Developments,* R. R. Yager, Ed. New York: Pergamon, 1982.

[3] S. Horikawa, T. Furuhashi, and Y. Uchikawa, "On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm," *IEEE Trans. Neural Networks,* vol. 3, no. 5, pp. 801–806, 1992.

[4] J. S. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.,* vol. 23, no. 3, pp. 665–684, May/June 1993.

[5] J. M. Keller, R. R. Yager, and H. Tahani, "Neural network implementation of fuzzy logic," *Fuzzy Sets Syst.,* vol. 45, pp. 1–12, 1992.

[6] C. C. Ku and K. Y. Lee, "Diagonal recurrent neural networks for dynamic systems control," *IEEE Trans. Neural Networks,* vol. 6, pp. 144–156, Jan. 1995.

[7] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller—Part I & II," *IEEE Trans. Syst., Man, Cybern.,* vol. 20, no. 2, pp. 404–435, 1990.

[8] C. T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Trans. Comput.,* vol. C-40, no. 12, pp. 1320–1336, Dec. 1991.

[9] ——, "A neural fuzzy control system with structure and parameter learning," *Fuzzy Sets Syst.,* vol. 70, pp. 183–212, 1995.

[10] K. Liu and F. L. Lewis, "Some issues about fuzzy logical control," in *Proc. 32nd Conf. Decision and Control,* San Antonio, TX, Dec. 1993, pp. 1743–1748.

[11] A. Nafarieh and J. M. Keller, "A new approach to inference in approximate reasoning," *Fuzzy Sets Syst.,* vol. 41, pp. 17–37, May 1991.

[12] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Networks,* vol. 1, no. 1, pp. 4–27, Mar. 1990.

[13] S. K. Pal and D. K. Dutta Majumder, *Fuzzy Mathematical Approach to Pattern Recognition.* New York: Wiley, 1986.

[14] M. Roubens, "Pattern classification problems and fuzzy sets," *Fuzzy Sets Syst.,* vol. 1, pp. 239–253, 1978.

[15] *Parallel Distributed Processing,* D. E. Rumelhart and J. L. McClelland, Eds. Cambridge, MA: MIT Press, 1986, vol. 1.

[16] L. X. Wang, *Adaptive Fuzzy Systems and Control: Design and Stability Analysis.* Englewood Cliff, NJ: Prentice Hall, 1994.

[17] C. W. Xu and Y. Z. Lu, "Fuzzy modeling identification and self-learning for dynamical systems," *IEEE Trans. Syst., Man, Cybern.,* vol. SMC-17, no. 4, pp. 683–689, July/Aug. 1987.

A Dynamic Gesture Recognition System for the Korean Sign Language (KSL)

Jong-Sung Kim, Won Jang, and Zeungnam Bien

***Abstract*— The sign language is a method of communication for the deaf-mute. Articulated gestures and postures of hands and fingers are commonly used for the sign language. This paper presents a system which recognizes the Korean Sign Language (KSL) and translates into a normal Korean text. A pair of Data-Gloves are used as the sensing device for detecting motions of hands and fingers. For efficient recognition of gestures and postures, a technique of efficient classification of motions is proposed and a fuzzy min–max neural network [4] is adopted for on-line pattern recognition.**

I. INTRODUCTION

Gestures and postures have been used as a means of communication among people for a long time, being interpreted as streams of tokens for a language [1]. They may vary from the stylized lexicon of a traffic cop to the highly developed syntax of a natural language such as the sign language.

The sign language is a method of communication for the deaf-mute. It is understood by means of gestures of both hands and fingers [2].

This paper deals with a system which recognizes the Korean Sign Language (KSL) and translates it into a normal Korean text.

According to a standard KSL dictionary, the 45-year-old Korean Sign Language contains about 6 000 vocabulary words. However, they are formed by combining a relatively small number of basic gestures. Moreover, two types of gestures of hands and fingers are used: one type consists of static postures and the other is dynamic gestures. The former consists of 31 distinct postures expressing the dactylology while the latter is made up with changing patterns, constituting the main body of the KSL and expressing different meanings of vocabulary words.

One may extract features of static postures of 10 fingers by identifying and recognizing the dactylology in the space domain. On the other hand, the recognition of changing patterns of dynamic gestures in the time domain is essential to understand any KSL-based sentences. This means that the recognition of the KSL should be conducted in real-time. For our system, an electronic device, called Data Glove [3], is adopted as an input device in consideration of cost effectiveness of hardware versus real-time processing capability. It is remarked that, in case that an 8-bit gray level vision system is adopted as an input sensing device, the system is required to handle at least 8 Mb/s while, in case of Data Glove, the device needs to handle about 600 b/s. It is also known that the pattern classes of KSL gestures are not linearly separable and that patterns tend to overlap with each other. Therefore, it is desirable to design a pattern classifier in such a way that the amount of mis-classification for those overlapping classes is minimum. Also, the system needs some form of learning capability due to the varying nature of the patterns to handle.

It is remarked that in [6] and [7], neural network based methods were presented for recognition of the American Sign Language (ASL). In the work by Fel [6] were used the back-propagation

Manuscript received August 5, 1994; revised February 26, 1995.

J.-S. Kim and Z. Bien are with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Taejon 305-701, Korea.

W. Jang is with the Agency for Defense Development, Taejon 305-600, Korea.

Publisher Item Identifier S 1083-4419(96)02312-6.

neural networks for recognition of simple gestures (derived from the ASL), but it seems that extensive training is required with explicit specification of the beginning and end of the gesture. For abrogating the need for extensive training in the neural network, a method of relabeling a self organizing map (SOM) is proposed in [7]. This method is proposed to avoid the usual heavy work of retraining when new vocabulary words are added to the system though this method also requires to train the system for recognition of initial classes. Also if the number of new vocabulary words to be added gets increased, the system may fail to show the same success rate.

In this paper, we propose a dynamic gesture recognition method by employing a technique for efficient classification of motions, and applying a fuzzy min–max neural network [4] for on-line pattern recognition.

II. KSL Recognition System

To express a KSL word or sentence, all of the two hands and ten fingers may be used. The right hand performs main motional actions while the left hand is often employed for auxiliary purpose. That is, most of the left hand motions are usually symmetrically the same as those of the right hand.

Since different humans have different hands and fingers in physical dimension, the same form of a gesture made by two different persons may produce nonidentical numerical data when the gestures are measured by ordinary sensing devices. Naturally, some form of transformations may be needed for feature recognition.

The VPL Data-Glove [3], which is an input device for our system, is equipped with two sensor systems for each hand and its fingers. One flex-sensor system consists of fiber optic transducers which measure flexing angles of the finger joints. There is also a "polhemus" sensor system attached to the back of the glove which measures the x, y, z, yaw, pitch, and roll of the hand relative to a fixed source.

Thus, the Data-Glove put on a hand in motion generates 16 types of data. That is, 10 flex angles, 3 position data (x, y, z), and 3 orientation data (roll, pitch, yaw) are obtained from each of the two Data-Gloves. Two sets of the 16 parameters from the right and left hand Data-Gloves, however, are often heavily contaminated by noises generated during electronic sensing and/or caused by signer's unnecessary motions including unconscious hand trembling. It is desired to suppress those irregularly generated noises.

For recognition of the KSL, these 16 kinds of raw data are processed to generate some directional and regional information of the hand motion, which, in turn, are used as features.

In the following, we describe the process of recognizing the KSL in a more detailed manner.

A. Setting of Initial Position

At any instant of time, the hand gesture can be represented by its position and orientation of the hand, and the configuration of the fingers. From the signal point of view, we may say that a sequence of regularly sampled gestures constitute the motion of the hands and fingers. This sequence of gestures will be termed as dynamic gestures. The dynamic gestures as a whole present very complicated forms of data and may not be easily characterized by conventional means, such as image features.

To initiate any dynamic gesture, there should be some starting point in space; that is to say, the hands should start their motions from some initial positions, but the initial positions of the hands for KSL gestures can be different for different people and for different time/places. In fact, for actual KSL speakers, there is no fixed referential initial position of the hands to start dynamic gestures. Therefore, it is needed for the recognition system to be independent of the initial position of a gesture. For this, the initial x, y, and z-axis data are recorded from the Data-Glove and all the subsequent position data of the Data-Glove are calibrated by continuously subtracting the initial position data from the current position data. It is remarked that our coordinate system is determined in reference to the mechanism of the Data-Glove [3] and, thus, the vertically up/down hand motion is called an x-axis motion while the y-axis motion occurs horizontally left/right (refer to Fig. 10).

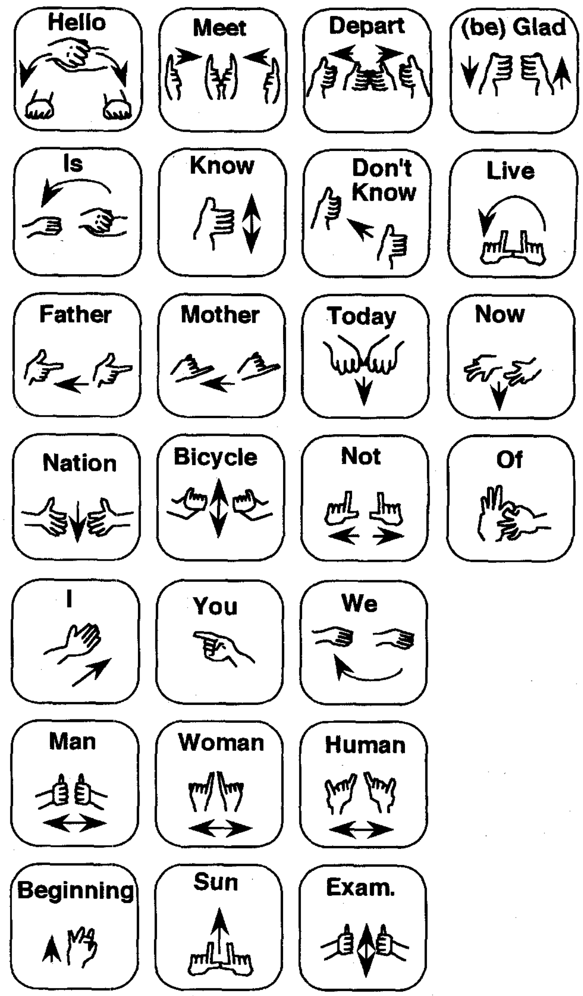


Fig. 1. 25 gestures in the KSL.

B. Partition of Region

The KSL contains about 6 000 gestures but most of the gestures are formed by combining some basic gestures. In this paper, we have selected 25 important basic gestures shown in Fig. 1 for our pilot study.

After analyzing these 25 gestures, one can easily conclude that these gestures contain 10 basic direction types of motion patterns shown in Fig. 2. As mentioned earlier, the input data obtained by the gloves are heavily contaminated by noise. For example, let us consider the trajectories of horizontally left-to-right motion and vertically up-to-down motion shown in Fig. 3. Specifically, a direction type "$\mathbf{D_2}$" is measured by the Data Glove and plotted in Fig. 3(a) while a direction type "$\mathbf{D_5}$" is given in Fig. 3(b).

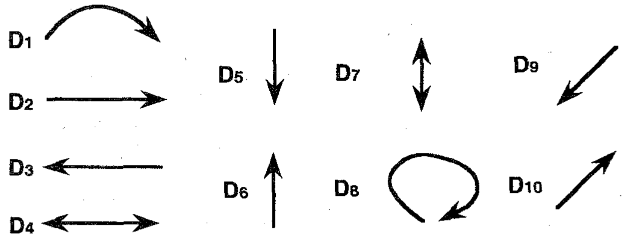
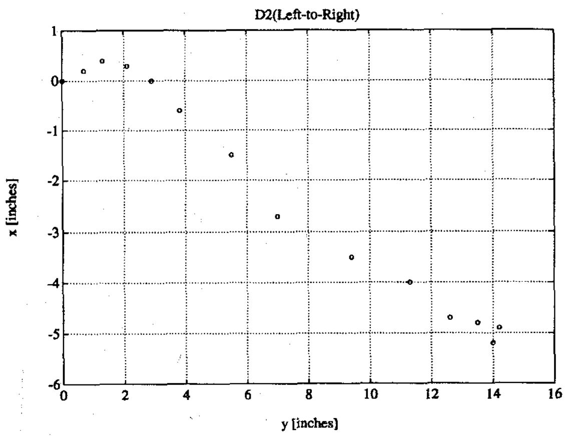
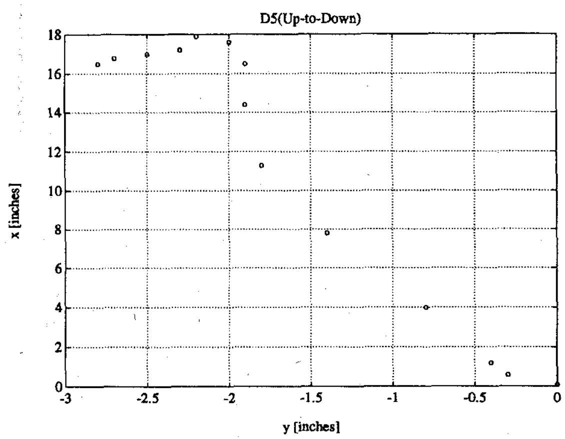


Fig. 2. 10 basic direction types.



(a)

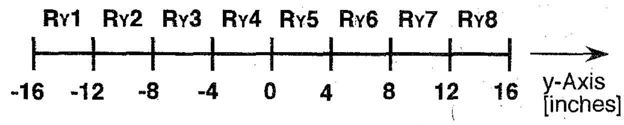


(b)

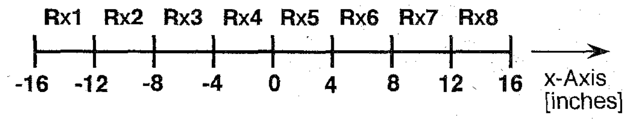
Fig. 3. Trajectories of direction type: $\mathbf{D_2}$ and $\mathbf{D_5}$: (a) a trajectory of left-to-right motion and (b) a trajectory of up-to-down motion.

Analyzing these types of data, we have found that the deviations of measured data around the real (average) value lie within 4 inches. Hence, for efficient filtering and also for reduction of data processing time, we have divided the x-axis and y-axis into 8 regions, respectively, as shown in Fig. 4, taking into account the range of motions. In the paper, motional patterns of hand are represented by these region data.

Fig. 5 shows the profiles of actual y-axis motion data and transformed region data of the gesture "we", which is expressed by a left-to-right circular motion of the hand in the y–z plane. Obviously, such

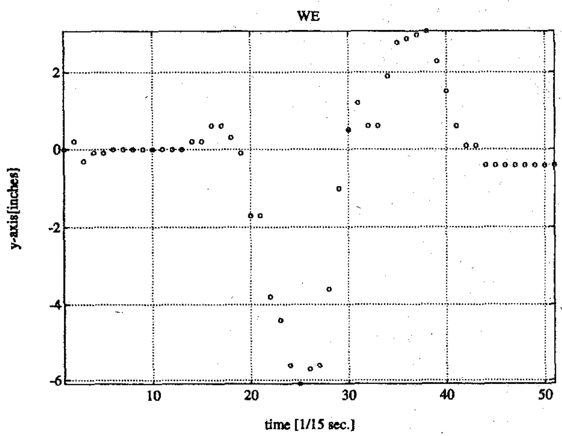


(a)

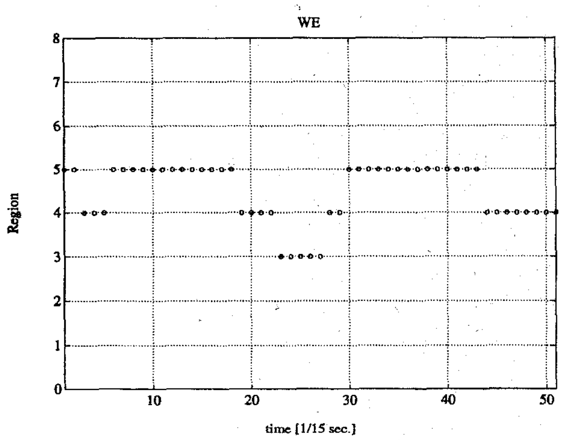


(b)

Fig. 4. Region partition: (a) partition of y-axis region, and (b) partition of x-axis region.



(a)



(b)

Fig. 5. Actual motional data profile: (a) actual motional data profile for gesture "we", and (b) transformed region data profile for gesture "we."

a huge motion data for a gesture as shown in Fig. 5(a) can be reduced to a much smaller set of region data, $R_Y(4,5,4,3,4,5,4)$ and this reduction process enable us to implement a real-time recognition system for dynamic hand motion. Here, $R_Y(4,5,4,3,4,5,4)$ means a sequence of orderly region data, $R_Y4 - R_Y5 - R_Y4 - R_Y3 - R_Y4 - R_Y5 - R_Y4$.

As shown in Fig. 6, the region data on x-axis and y-axis motions are sequentially inputted every 1 time unit to 5 cascaded shift

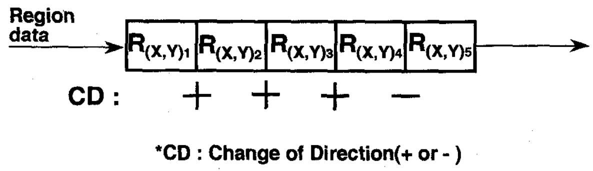


Fig. 6. On-line processing of region data.

registers. It is remarked that, in our system, 1 time unit is equal to 1/15 s. If the current region data are the same as the region data 1 time unit before, the current data are not inputted to the register. To be more specific, let $R_{(X,Y)_i}$, $i = 1, \cdots, 5$, denote the ith register having the region data obtained ith step before. Those 5 cascaded registers contain the most recent information about region data on x-axis and y-axis motions. The current contents of the registers are compared with the contents of 1 step before to determine the direction class of the current gesture shown in Fig. 2. If we denote by $CD_{(X,Y)_k}$, $k = 2, \cdots, 5$ the information on the change of direction obtained at k step before, then we can write

$$CD_{(X,Y)_k} = \begin{cases} +, & \text{if } R_{(X,Y)_k} > R_{(X,Y)_{k-1}} \\ -, & \text{if } R_{(X,Y)_k} < R_{(X,Y)_{k-1}} \end{cases} \qquad k = 2, \cdots, 5. \tag{1}$$

In the above expression, "+" means right/upper motion and "−" means left/down motion. As one can observe from Fig. 2, those 10 essential types of basic motion patterns may be grouped into 4 types. Firstly, the group $\{\mathbf{D_1}, \cdots, \mathbf{D_4}\}$ contains left/right motion and there is no change of direction in x-axis except $\mathbf{D_1}$. Secondly, the group $\{\mathbf{D_5}, \mathbf{D_6}, \mathbf{D_7}\}$ contains mainly up/down motions. Motions in this group may show, however, small change of direction in y-axis due to unnecessary motions such as human's hand trembling. This should be reflected in the expression of CD_Y. Thirdly, $\mathbf{D_8}$ is a circular motion on y–z plane, or y–x plane. This pattern can be recognized by change of direction in y-axis only. Finally, the group $\{\mathbf{D_9}, \mathbf{D_{10}}\}$ is a slant motion and should be described by changes of direction in y-axis and x-axis motions simultaneously.

Specifically, 10 types of direction classes in the x–y plane are described by a sequence of $CD_{(X,Y)}$ data as follows:

$\mathbf{D_1}$: $CD_Y(+, +, +, \times)$ and $CD_X(+, -, \times, \times)$
$\mathbf{D_2}$: $CD_Y(+, +, +, \times)$ and $CD_X(\times, \times, \times, \times)$
$\mathbf{D_3}$: $CD_Y(-, -, -, \times)$ and $CD_X(\times, \times, \times, \times)$
$\mathbf{D_4}$: $CD_Y(+, +, -, -)$ and $CD_X(\times, \times, \times, \times)$
$\mathbf{D_5}$: $CD_Y(+, -, +, \times)$ and $CD_X(-, -, -, \times)$
$\mathbf{D_6}$: $CD_Y(+, -, +, \times)$ and $CD_X(+, +, +, \times)$
$\mathbf{D_7}$: $CD_Y(+, -, +, \times)$ and $CD_X(+, +, -, -)$
$\mathbf{D_8}$: $CD_Y(-, -, +, +)$ and $CD_X(\times, \times, \times, \times)$
$\mathbf{D_9}$: $CD_Y(-, -, \times, \times)$ and $CD_X(-, -, -, \times)$
$\mathbf{D_{10}}$: $CD_Y(+, +, \times, \times)$ and $CD_X(+, +, +, \times)$.

In the above expression, "×" denotes the "don't care" condition. For example, the direction class "$\mathbf{D_1}$" continuously contains region $R_Y(4, 5, 6, 7)$ and $R_X(1, 2, 1)$. Therefore, according to (1), the class $\mathbf{D_1}$ can be represented by $CD_Y(+, +, +, \times)$ and $CD_X(+, -, \times, \times)$.

Because a sign language word or sentence is understood at the end of each motion, the above approach of processing the region data renders no difficulty for real-time processing.

C. Recognition of Postures

In our 25 gestures, 14 types of basic hand postures are included as shown in Fig. 7. It is proposed in this paper that each hand postures be recognized by applying the technique of Fuzzy Min–Max Neural Network [4], which we shall call FMMN network in the following.

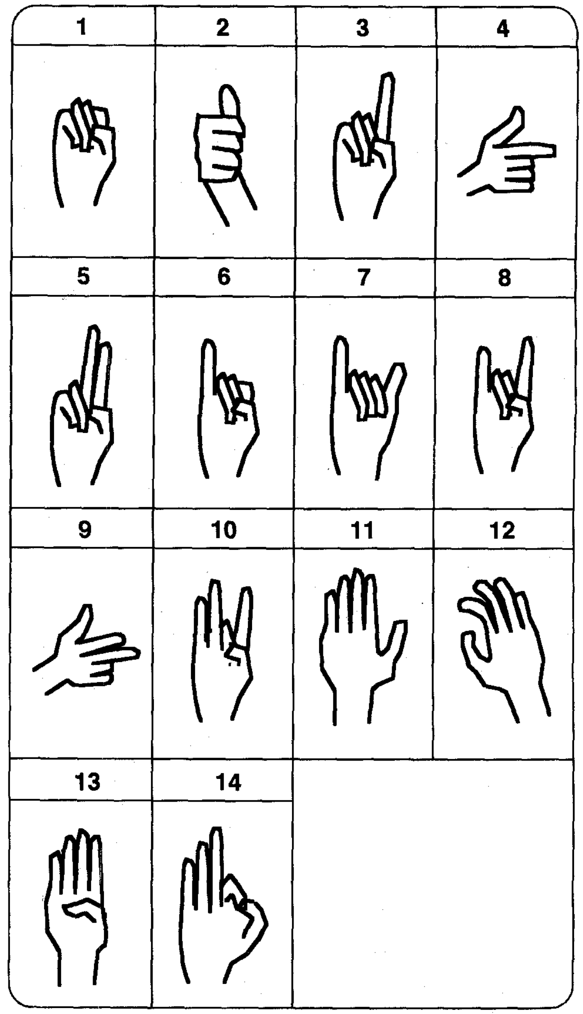


Fig. 7. 14 basic postures in the KSL.

The FMMN network requires no pre-learning about posture class and has on-line adaptability.

A fuzzy set hyperbox [4] is a 10-dimensional box in our study defined by a min point (V) and a max point (W) with a corresponding membership function. The initial min–max values (V, W) of the network are determined based on experimental data of many individuals whose flex angles may show wide range of varying values.

The two flex angles of each finger (linearly scaled to lie between zero and one) are inputted to FMMN network shown in Fig. 8. We use the same function to describe a hyperbox as suggested in [4]. The membership function of the hyperbox is expressed as follows:

$$B_j(fa_h, V_j, W_j) = \frac{1}{10} \sum_{i=1}^{10} [1 - f(fa_{hi} - w_{ji}, \gamma) - f(v_{ji} - fa_{hj}, \gamma)],$$

where

$$f(x, \gamma) = \begin{cases} 1, & x\gamma > 1 \\ x\gamma, & 0 \le x\gamma \le 1 \\ 0, & x\gamma < 0 \end{cases}$$

$$fa_{f1}: fa_{h1,3,\cdots 9}$$
$$fa_{f2}: fa_{h2,4,\cdots,10}.$$

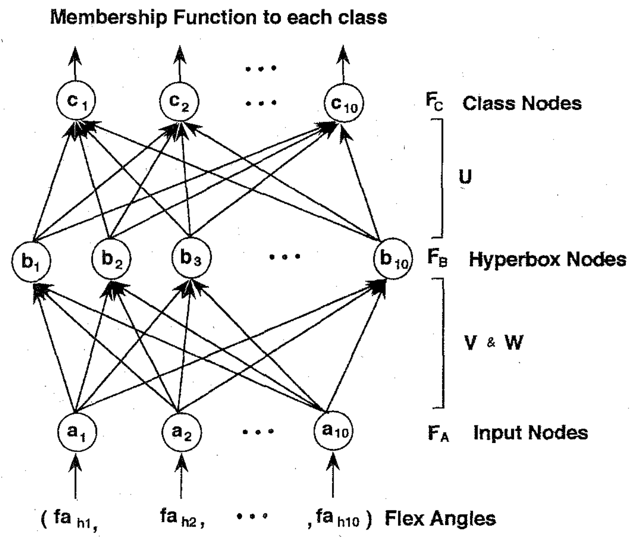


Fig. 8. The structure of fuzzy min–max neural network.

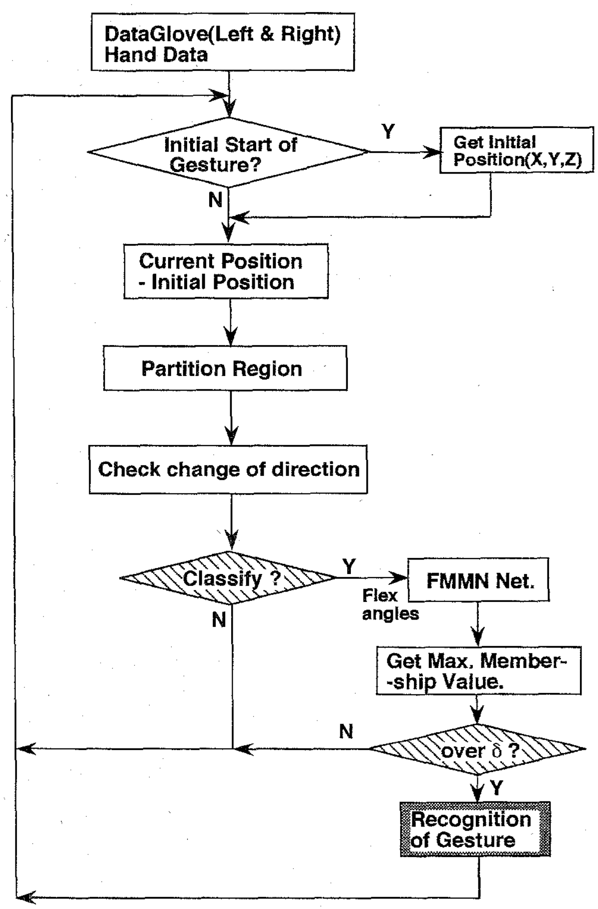


Fig. 9. The flow chart of the KSL recognition.

Here, γ is a sensitivity parameter that regulates the speed at which the membership values decrease when an input pattern is separated from the hyperbox core. And fa_{f1} is the flex angle of the inner joints of fingers and fa_{f2} is the flex angle of the outer joints of fingers. If γ increases, the membership function becomes more crisp. In this system, γ of fa_{f1} is smaller than γ of fa_{f2} because fa_{f2} is more sensitive to the change of flex angles. Given an input posture, the output of this network is the membership function values for 14 posture classes and the class with the maximum value of membership is classified as the designated posture if the value is above a given threshold (δ). If the maximum value is below the threshold (δ), no decision is made.

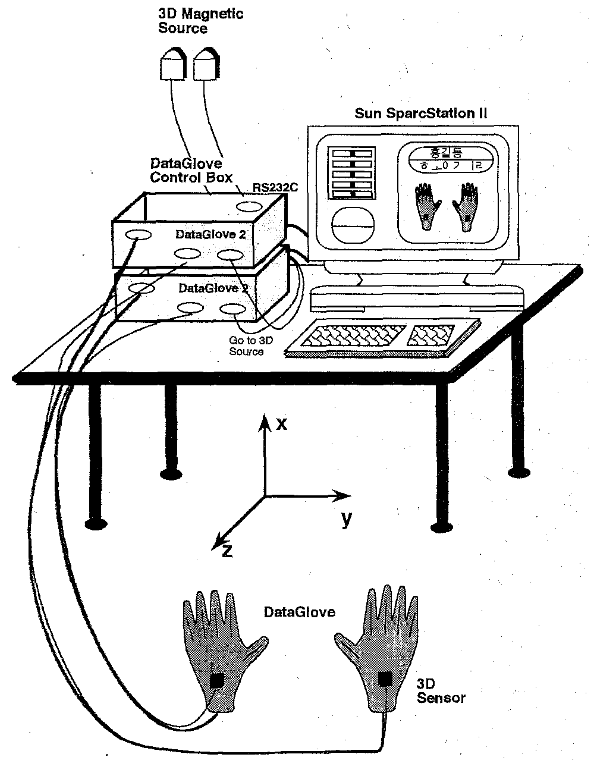


Fig. 10. The Configuration of the KSL Recognition System

D. Recognition of the KSL

When a signer performs a motion for a gesture, many data are generated from the Data-Glove and inputted to the system. Region partition transforms these raw data set into a set containing small number of data and this small number of region data are used to recognize the direction class. An input gesture is identified as one of the given direction classes, and then, the hand shape of that motion is recognized by the posture recognition method. With these two stages, the input pattern is recognized as one of the gesture classes on-line.

The flow chart of the recognition algorithm is given in Fig. 9.

III. Experimental Results

Fig. 10 shows the configuration of the KSL recognition system organized for our experiment. All the functions are run on a Sun Sparc-Station II (Sun 4/75) as main computer and the Data-Glove sends 32 data sets to the main computer via RS232C with 9,600 Baud rates.

The system function can be divided into 4 levels: in the first level, a signer who wears the Data-Glove sets the starting position. Then, the data from Data-Glove are transformed as region data. In the second level, one of 10 direction classes is determined by examining the change of direction $[CD_{(X,Y)}]$. In the third level, its posture class

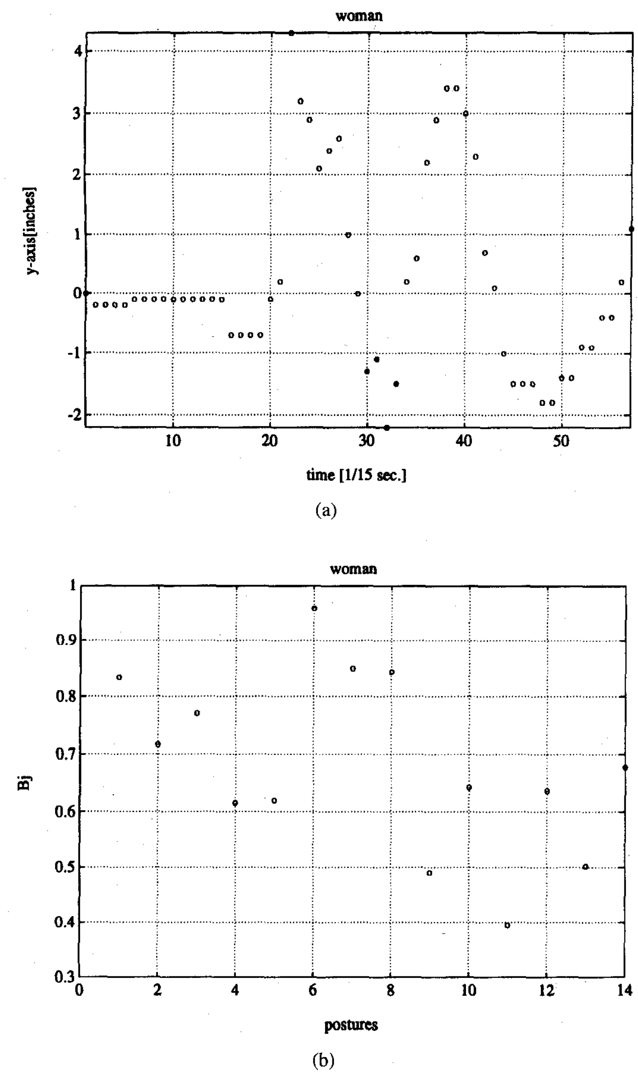


Fig. 11. Recognition result of gesture "woman": (a) actual motion data profile for the gesture "woman" and (b) membership values for the each posture.

is identified with one of the 14 postures. Then in the final level, the result of a dynamic gesture is classified by using the above two results and is displayed as texts via on-line graphics.

Fig. 11 shows a recognition result of gesture "woman", for which the two hands are moving to make a right/left motion of hands with unfolded little finger. In Fig. 11(a) is shown the motion profile along y-axis sampled at every 1/15 s. In this case, the direction class is determined to $\mathbf{D_2}$ at the 30th time unit by examining the change of direction as described above. After the direction class is decided, the membership values about 14 postures are obtained as shown in Fig. 11(b) by inputting flex angles to the FMMN network. In this Fig. 11(b), the number of class (horizontal axis) is assigned to the corresponding posture numbers in Fig. 7. In the network, $\theta = 0.2$, $\gamma_1 = 4$, $\gamma_2 = 10$, and $\delta = 0.8$ are used. As shown in Fig. 11(b), the 6th posture has the maximum value of B which is greater than δ and therefore, we conclude that the result of recognition is the gesture "woman".

Many experiments have been conducted with 25 different sign languages and we have found that the success rate of our method in classification reaches up to almost 85% of the given words. Taking into account the fact that the deaf-mute who use gestures often misunderstand each other [2], we may say this success rate seems acceptable as a pilot study. We find that abnormal motions in the gestures and postures, and errors of sensors are partly responsible for the observed mis-classification.

IV. Concluding Remarks

In this paper, a dynamic gesture recognition method is proposed, for which a new technique for efficient classification of motions is employed, and a fuzzy min–max neural network is applied for on-line pattern recognition. Currently, expandability of the classifier is under investigation when new patterns (words) are added along with further refinement study to achieve higher success rate.

References

[1] D. J. Sturman, "Whole-hand input," Doctoral dissertation of MIT, Feb. 1992.
[2] W.-J. Son, *Guide to the Korean Sign Language.* Seoul: Chang-Rok Press, 1988 (written in Korean).
[3] *Data-Glove Model 2+ Operation Manual,* VPL Research Inc., CA, 1992.
[4] P. Simpson, "Fuzzy min–max neural networks—Part 1: Classification," *IEEE Trans. Neural Network,* vol. 3, pp. 776–786, Sept. 1992.
[5] ——, "Fuzzy min–max neural networks—Part 2: Clustering," *IEEE Trans. Fuzzy Systems,* vol. 1, no. 1, pp. 32–45, Feb. 1993.
[6] S. S. Fels and G. E. Hinton, "Glove-Talk: A neural network interface between a data-glove and a speech synthesizer," *IEEE Trans. Neural Networks,* vol. 4, no. 1, pp. 2–8, Jan. 1993.
[7] M. B. Waldron and S. Kim, "Increasing manual sign recognition vocabulary through relabeling," in *Int. Joint Conf. on Neural Network 1994,* Orland, June 1994, pp. 2885–2889.
[8] Abe, "A classifier using fuzzy rules extracted directly from numerical data," in *FUZZ-'93 International Conf. on Fuzzy System,*, Mar. 1993, pp. 1191–1198.
[9] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms.* New York: Plenum Press, 1981.
[10] J. Bezdek and S. K. Pal, *Fuzzy Models for Pattern Recognition.* IEEE Press, 1992.