Pattern discrimination method with a boosting approach using hierarchical neural trees

M Okamoto[1*], **K Shima**[2], **Y Matsubara**[1], and **T Tsuji**[2]
[1]Graduate School of Information Sciences, Hiroshima City University, Hiroshima, Japan
[2]Graduate School of Engineering, Hiroshima University, Hiroshima, Japan

The manuscript was received on 31 January 2008 and was accepted after revision for publication on 1 May 2008.

DOI: 10.1243/09596518JSCE563

Abstract: This paper proposes a new pattern classification method using probabilistic neural networks based on a boosting approach. In this method, a log-linearized Gaussian mixture network is used as a weak classifier. The method proposed automatically constructs a suitable classification network from given data. Validity of the proposed method is shown by discrimination results of artificial data and hand shape. The application is confirmed of the proposed method to human interface controlling of home electric appliances using hand shapes.

Keywords: probabilistic neural network, boosting, pattern classification, biological signals

1 INTRODUCTION

Several pattern classification algorithms have been proposed for image discrimination, speech recognition, and data mining [**1**, **2**]. In particular, neural networks (NNs) have been demonstrated as a promising classification tool, since their learning ability allows them to find optimum non-linear relationships between classes and feature patterns from data sets [**3**, **4**]. However, to use NNs effectively as the classifiers for applications, several problems, such as the choice of network structure, learning convergence, and local minima, must be solved.

A probabilistic neural network (PNN), which estimates the probability density function of patterns, has been proved to be an efficient and important method for pattern classification. In particular, Tsuji *et al.* proposed a feedforward PNN, a log-linearized Gaussian mixture network (LLGMN) based on the Gaussian mixture model (GMM) and a log-linear model [**5**]. The LLGMN has been successfully applied to pattern classification of bioelectric signals, e.g. electromyograms [**6**] and electrocardiograms [**5**, **7**], and has been used to develop human interface applications, such as prosthetic devices and electromyography (EMG)-based pointing devices [**8–10**].

However, to estimate the LLGMN parameters, the GMM number of each class must be fixed beforehand. When the GMM number is fixed at an unsuitable value, the LLGMN training cannot avoid convergence to a local minimum for some initial weights and training data. Therefore, better classification performance requires estimation of an optimum LLGMN structure. Several methods, such as information criterion and the variational Bayes approach, have been widely used as the criterion for the structure of a model [**11**, **12**]. In these methods, a suitable learning model can be selected based on discrimination accuracy, likelihood, and model complexity. Although these methods select a suitable model structure, all possible models must be evaluated based on the criterion. Thus, it takes a long time to estimate a suitable model structure.

There has also been growing interest in a boosting approach for the construction of classification systems with simple classifiers [**13–15**]. A general boosting procedure can combine inaccurate and simple classifiers to improve the discrimination accuracy of a classification system. Therefore, this approach eliminates the need for evaluation of unnecessary models.

Corresponding author: Graduate School of Information Sciences, Hiroshima City University, 3-4-1 Ozuka-higashi, Asaminami-ku, Hiroshima City, Hiroshima 731-3794, Japan. email: okamoto@hiroshima-cu.ac.jp

This paper proposes a novel hierarchical classification method that can automatically construct classification models through a learning network. In this method, the LLGMN is utilized in order to create a simple and weak classifier. The proposed method can estimate the number of LLGMNs corresponding to the pattern complexity, according to statistical information obtained from the training data.

The next section provides the details of the LLGMN structure and learning algorithm. Section 3 shows the proposed method for constructing a suitable model using the boosting approach. The results of computer simulation and pattern classification experiments of biological signals are presented in section 4. Finally, the last section concludes this paper.

2 LLGMN [5]

LLGMN is based on a log-linear model and a Gaussian mixture model (GMM). It calculates *a posteriori* probability for the training data. In this method, LLGMN is utilized for partition at the non-terminal node of the hierarchical tree. The structure and learning algorithm of LLGMN are explained below.

2.1 Structure of LLGMN

The structure of LLGMN is shown in Fig. 1. In order to represent a normalized distribution corresponding to each component of GMM as weight coefficients of NN, the input vector $\boldsymbol{x}(\ \Re^D)$ is converted into the modified input vector $\boldsymbol{X}$ as follows:

$$\boldsymbol{X}=\left\{1, x^{\mathrm{T}}, x_1^2, x_1x_2, \ldots, x_2^2, \ldots, x_2x_D, \ldots, x_D^2\right\}^{\mathrm{T}} \quad (1)$$

The first layer of LLGMN consists $H=1+D(D+3)/2$ units, which correspond to the dimension of the input vector $\boldsymbol{X}$, and the identity function is used for the activation function of each unit. The outputs of the first layer multiplied by weight $w_h^{(k,m)}$ are transmitted to the second layer, where $w_h^{(K,M_K)}=0$, and K and M_K denote the number of classes (patterns) and components belonging to class M respectively. In this layer, LLGMN calculates *a posteriori* probability of each Gaussian component $\{k, m\}$. The unit k in the third layer integrates the outputs of M_k units in the second layer

$$^{(2)}I_{k,m}=\sum_{k=1}^{H}{}^{(1)}O_h w_h^{(k,m)} \quad (2)$$

$$^{(2)}O_{k,m}=\frac{\exp\left({}^{(2)}I_{k,m}\right)}{\sum_{k'=1}^{K}\sum_{m'=1}^{M_{k'}}\exp\left({}^{(2)}I_{k',m'}\right)} \quad (3)$$

The relationship between the input $^{(3)}I_k$ and the output O_k in the third layer is

$$^{(3)}I_k=\sum_{m=1}^{M_k}{}^{(2)}O_{k,m} \quad (4)$$

$$^{(3)}O_k={}^{(3)}I_k \quad (5)$$

The output of the third layer $^{(3)}O_k$ corresponds to *a posteriori* probability $P(k|x)$ of class k given the input vector, and the former can be used to evaluate the ambiguity of a classification result.

This network has the ability of adaptive learning for statistical properties of data. It can discriminate data with complex distributed structure and in comparison to the conventional method [**16**] using normal distribution restricting the parameter.

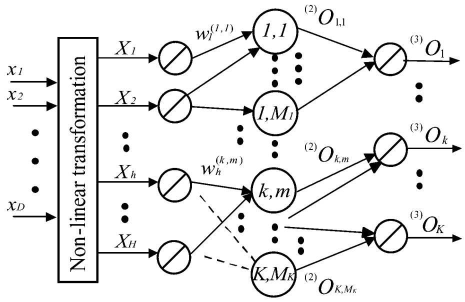


Fig. 1 The structure of LLGMN

3 PROPOSED PATTERN DISCRIMINATION WITH THE BOOSTING APPROACH

In the proposed method, the LLGMNs are used in order to create simple classifiers for the classification of input vectors to produce binary splits. The structure of each classifier is a hierarchical tree using LLGMN as each non-terminal node. By combining classifiers based on a boosting approach, the network can discriminate complex data and calculate *a posteriori* probability for the training data. The structure of the network and the constructing algorithm are explained below.

3.1 Structure of the network

Initially, the network consists of C classifiers, corresponding to the number of classified classes. C is the number of classes of training data. Each classifier achieves a binary classification to calculate *a posteriori* probability of the cth class ($c = 1, 2, \ldots, C$). For binary classification, the parameter of LLGMN K is set as 2. $L_c^{(q)}(x)(c=1, \ldots, C, q=1, \ldots, Q_c)$ is *a posteriori* probability calculated by the classifier, where Q_c is the number of classifiers used for the classification of the cth class added based on the boosting approach. Then, *a posteriori* probability $O_c(x)$ is given as

$$O_c(x) = \max_{q=1, \ldots, Q_c} \left[L_c^{(q)}(x) \right] \tag{6}$$

The structure of the proposed method is shown in Fig. 2. The entropy of outputs is also calculated to present the risk of misclassification. The entropy is defined as

$$H(x) = -\sum_{c=1}^{C} O_c(x) \log O_c(x) \tag{7}$$

If the entropy $H(x)$ is less than the discrimination threshold Te, the class with the largest probability is determined according to Bayes' decision rule

$$Y(x) = \arg \max_{c} \; O_c(x) \tag{8}$$

Otherwise, the determination is suspended.

3.2 Learning of the hierarchical classifier

The structure of the classifier is a hierarchical tree using LLGMN. When the learning of the cth class is performed, the training data are divided into two groups, G_c and $G_{\bar{c}}$, where G_c is a set obtained from the training data belonging to class c and $G_{\bar{c}}$ is the complementary set of G_c. An example of the constructed classifier is shown in Fig. 3.

Consider a training set $\{x^{(n)}, \boldsymbol{T}^{(n)}\}$ ($n = 1, \ldots, N$), where $\boldsymbol{T}^{(n)} = \left(T_1^{(n)}, T_2^{(n)}\right)$. If the input vector $\boldsymbol{x}^{(n)}$ belongs to class c, $T_1^{(n)} = 1$ and $T_2^{(n)} = 0$. An energy function according to the minimum log-likelihood training criterion can be derived as

$$E = \sum_{n=1}^{N} J^{(n)} = -\sum_{n=1}^{N} \sum_{k=1}^{2} T_k^{(n)} \log {}^{(3)}O_k \tag{9}$$

In the training process, modification of the LLGMN weight $\Delta w_h^{(k,m)}$ is defined as

$$\Delta w_h^{(k,m)} = -\eta \sum_{n=1}^{N} \frac{\partial J^{(n)}}{\partial w_h^{(k,m)}} \tag{10}$$

and

$$\begin{aligned} \frac{\partial J^{(n)}}{\partial w_h^{(k,m)}} &= -\frac{\partial}{\partial w_h^{(k,m)}} \left(-\sum_{k=1}^{2} T_k^{(n)} \log {}^{(3)}O_k \right) \\ &= \left({}^{(2)}O_{k,m} - \frac{{}^{(2)}O_{k,m}}{{}^{(3)}O_k} T_k^{(n)} \right) X_h^{(n)} \end{aligned} \tag{11}$$

where $\eta > 0$ is the learning rate.

LLGMNs are added to avoid the misclassification of training data belonging to $G_{\bar{c}}$. To evaluate the misclassification accuracy of training data belonging to $G_{\bar{c}}$, an evaluation function is defined as

$$F' = \frac{|D(\bar{c}, c)|}{|G_{\bar{c}}|} \tag{12}$$

If F' is greater than the threshold Th', more LLGMNs are added hierarchically, and are trained using a two-class set $D(c, c)$ and $D(\bar{c}, c)$. Then, *a posteriori* probability $L_c^{(q)}(x)$, which is calculated by the qth

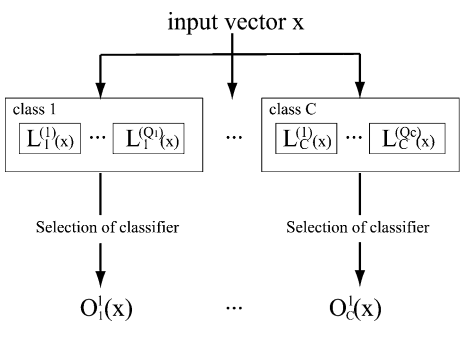


Fig. 2 The structure of the proposed method

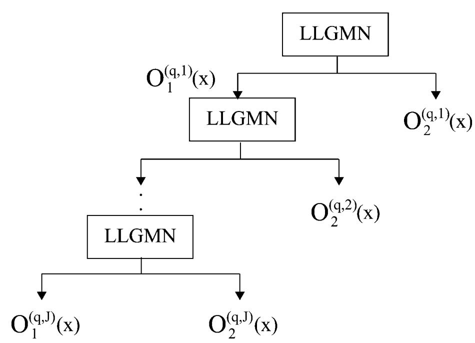


Fig. 3 The structure of classifier

classifier, is defined as

$$L_c^{(q)}(x) = 1 - \sum_{j=1}^{J_q} \left[\left(\prod_{j'=0}^{j-1} {}^{(3)}O_1^{(q,j')}(x) \right)^{(3)} O_2^{(q,j)}(x) \right] \quad (13)$$

where J_q is the number of LLGMNs added to the qth classifier, $O_1^{(q,j)}(x)$ is *a posteriori* probability calculated by the jth LLGMN in the qth classifier, and $O_1^{(q,0)}(x)$ is set to 1. By combining the LLGMN hierarchically to construct a network, the misclassification of data belonging to class c' can be avoided.

3.3 Construction network

In the proposed method, the addition and learning of the classifier is repeated for each class. A classifier is initially trained to classify the training data into G_c and $G_{\bar{c}}$. If $O_1(x) > O_2(x)$, it is considered that $\boldsymbol{x}$ is classified into class c. Then, $D(c, \bar{c})$ is the data set belonging to G_c and is classified into $G_{\bar{c}}$. An evaluation function that considers the training accuracy is defined as follows

$$F = \frac{|G_c| - |D(c, c)|}{|G_c|} \quad (14)$$

If F is greater than the threshold Th, a classifier is added for accurate discrimination. To train a newly added classifier, training data $D(c, \bar{c})$ and $G_{\bar{c}}$ are used. Repeating the addition of classifiers until the evaluation function is less than the threshold Th allows model construction and classifier learning to take place simultaneously.

Through the above training, the model construction and training of the classifier are performed based on a boosting approach.

4 EXPERIMENTS

4.1 Simulation experiments

Firstly, pattern classification experiments on artificial data were conducted for evaluating the performance of the proposed method. A two-dimensional input space consisted of six classes ($C = 6$), each class consisting of five Gaussian sources. Examples of the data are shown in Fig. 4. For each class, 200 samples were generated to train each LLGMN ($M_k = 1$, $K = 2$), and then the trained network was validated using test data (500 samples/class). The values of the parameters Te, Th and Th' were set as 0.8.

For the verification of the classification performance of the proposed method, single LLGMN, support vector machine (SVM) [**2**], and back propagation neural networks (BPNN) [**17**] classifiers were used for the comparison. BPNN had four layers (two hidden layers), the units of which were set as 2, 10, 10, and 4. Also, an SVM having a second-order polynomial kernel was used to perform a two-class classification. By combining two-class classifiers, multiclass classification using SVMs was achieved.

Figure 5 shows the classification results by the proposed method and conventional methods for 10 independent trials (the initial weights and training data were chosen at random). The results clearly indicate that the proposed method achieved the best classification rate among all the four methods. The mean values and standard deviations of the number of added LLGMNs for each class are shown in Fig. 6. For estimating a simple distribution such as a class six, a single LLGMN was used. On the other hand, many LLGMNs were added to the network for the estimation of complex distributions. These results indicate that the proposed method can estimate successfully the suitable class number of each class, and has the advantage that no unnecessary LLGMNs need to be added while evaluating the discrimination accuracy for determining the network structure.

4.2 Pattern classification of finger signal shapes

Next, motion classification experiments using finger-shaped signals were conducted for examining the performance of the proposed method. Three subjects (A, B, and C) participated in the experiments.

4.2.1 Experimental conditions

The subjects were asked to perform 31 types of motions ($C = 31$). The motions are shown in Fig. 7. Five shape signal channels ($D = 5$) were rectified and digitized using an analogue-to-digital (A/D) converter (sampling frequency of 167 Hz). Five shape sensors (Measurand Corp.) were attached to each finger of the right hand. These sensors are one-degree-of-freedom (DOF) measuring devices. The attached sensors are shown in Fig. 8. One end of each sensor was fixed to the wrist of the subject and the other ends were fixed to the corresponding tips of fingers. Also, for measuring the angle of the finger, the sensors were passed through the tubes that were fitted to the fingers (see Fig. 8). In order to fix the sensors to easy-to-use positions for each subject the exact positions of sensors were not specified,

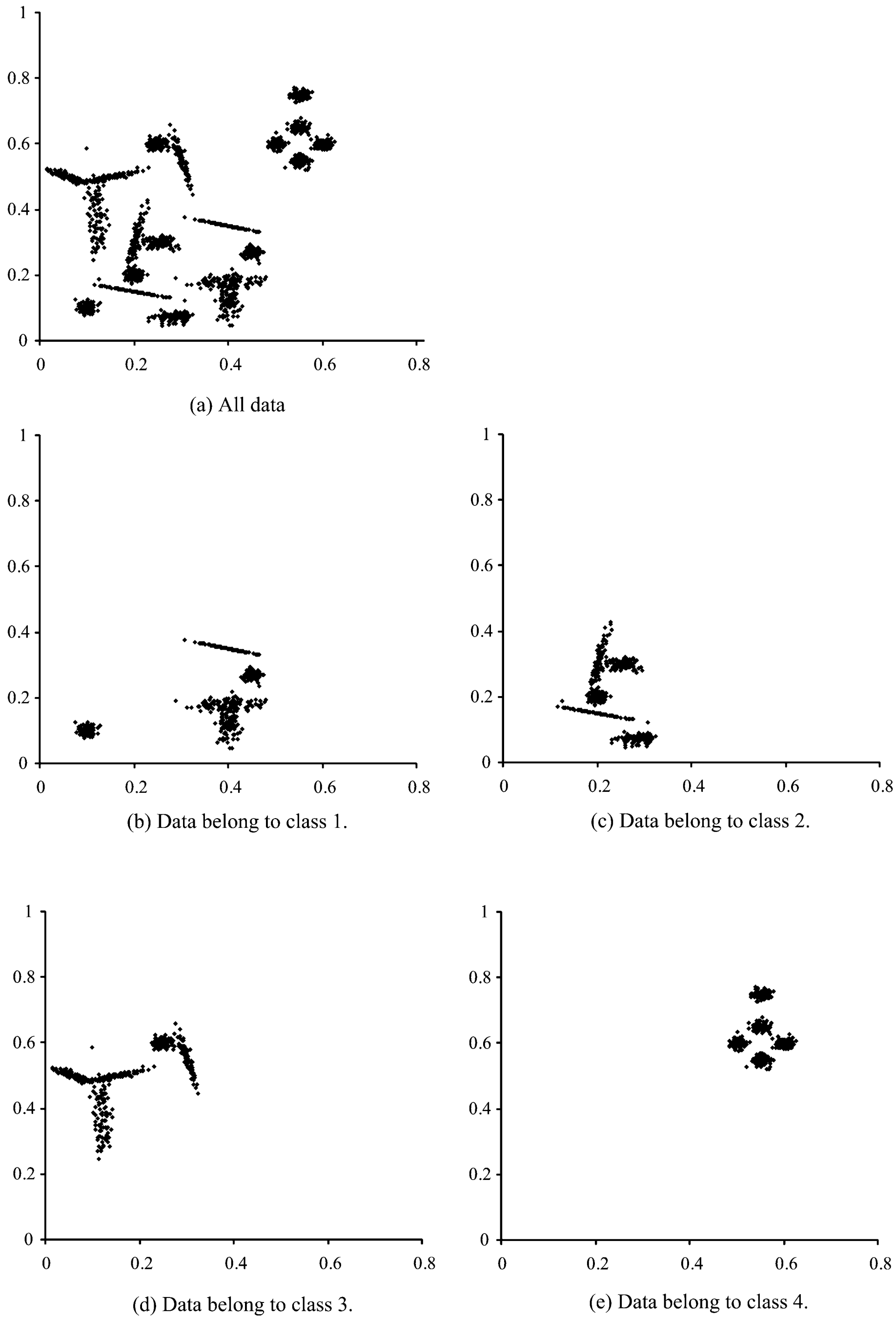


Fig. 4 Artificial data used for discrimination experiments

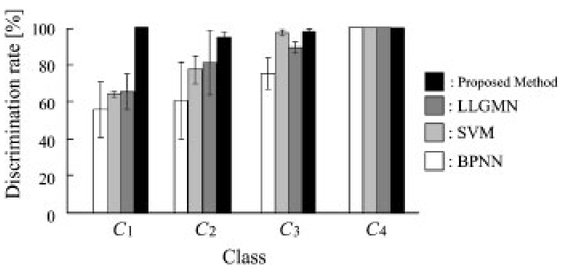


Fig. 5 Classification results

because these sensors are utilized as user-friendly interface devices in the next experiment. The measured signals $S_d(n)$ were normalized as follows for obtaining a maximum value of 1

$$N_d(n) = \frac{S_d(n) - S_d^{\mathrm{st}}}{S_d^{\max} - S_d^{\mathrm{st}}} \tag{15}$$

where S_d^{st} is the mean value of $S_d(n)$ measured when the hand is relaxed and $S_d^{\max}$ is the mean of the maximum value of each channel. The normalized

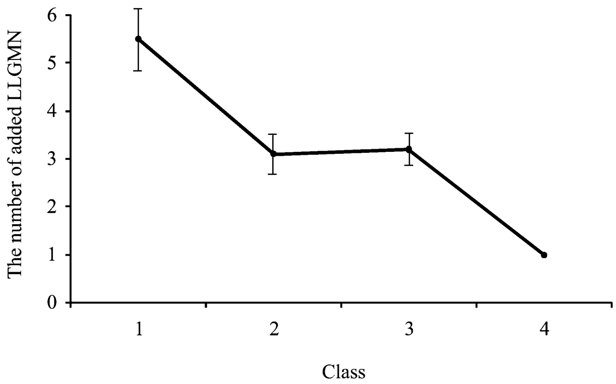


Fig. 6 The number of added LLGMN

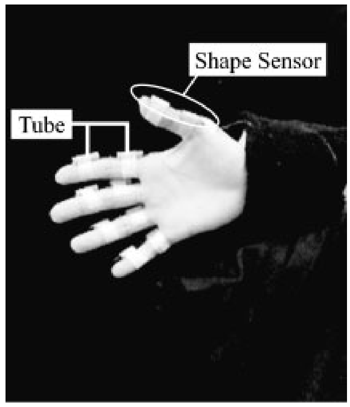


Fig. 8 Shape sensors attached to fingers

signals were compared with a prefixed threshold M_d to determine whether the subject changed the motion of the hand. In addition, signals $N_d(n)(d = 1, \ldots, 5)$ are normalized to make the sum of all D channels equal to 1 as follows

$$s_d(n) = \frac{N_d(n)}{\sum_{d=1}^{D} N_d(n)} \tag{16}$$

The values of the parameters Th and Th' were set as 0.8 and M_d was set as 0.5.

In this experiment, the shape signals measured beforehand were selected using the present proposed method.

4.2.2 *Pattern classification results*

The mean values and standard deviations of the classification rates are shown in Fig. 9. BPNN had four layers (two hidden layers), the units of which were set as 5, 10, 10, and 32. Moreover, 32 SVMs were used for the classification. As shown in the figure, the classification results of the proposed method are similar to those of SVM and single LLGMN for the case of subjects A and B. In the case of subject C, however, the classification results of other methods degrade more than that of the proposed method.

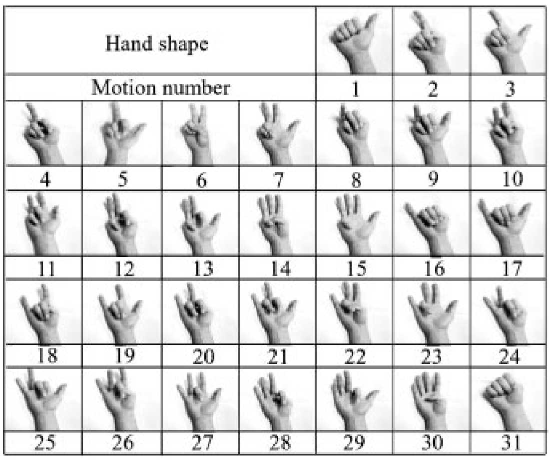


Fig. 7 The 31 pattern of hand shape

Table 1 shows an example of the number of added classifiers and LLGMNs in the network of subject C. Here, it is inferred that a better classification is achieved by adding the classifiers and LLGMNs.

Figure 10 shows the mean values and standard deviations of the classification rates of each type of motion for subject C. From this figure, it can be clarified that the classification rate has improved overall by the proposed method. For example, Fig. 11 shows the mean values and standard deviations of the signal patterns of motions 1 and 31. This figure shows that the patterns of motion 1 are similar to those of motion 31. As a result, a single LLGMN cannot accurately identify the difference between these patterns. However, the proposed method can estimate the distribution of each type of motion accurately using more than one LLGMN. For example, the patterns of motions 8 and 9 overlap (see Fig. 12); a more suitable structure can be constructed using the proposed method by combining the LLGMNs. It is clear that by adding LLGMNs to a network for the estimation of the distribution, the proposed method can achieve a more accurate classification than a single LLGMN.

4.3 Application on the human interface

An experiment was conducted to test the operation of the proposed method. In this experiment, a Bio-Remote system was used for controlling the electric appliances. This system was manipulated according to the user's intention determined from the biological signals.

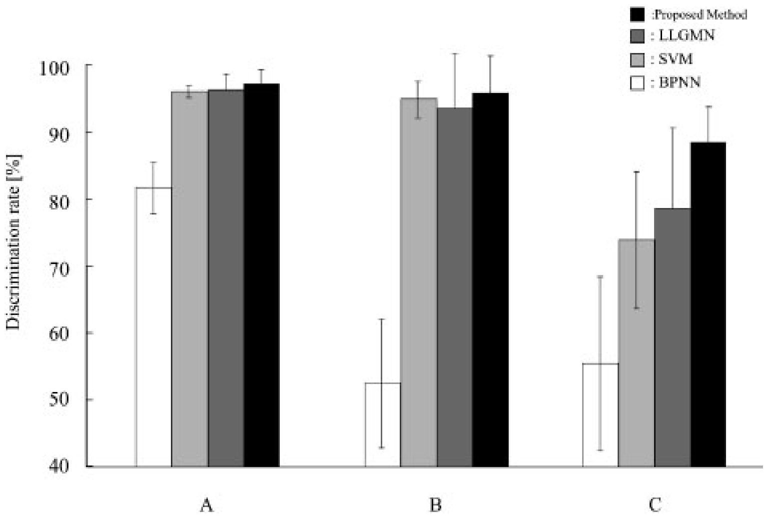


Fig. 9 Discrimination results for three subjects

In general, it is difficult to discriminate the user's intentions from the biological signals. Therefore, if necessary, the user can manipulate various applications with residual functions that combine input channels using this system. The proposed method can discriminate various hand shapes from the biological signals. An experiment was conducted, using a healthy person as the subject, for verifying the validity of the proposed method.

The function of the control system for electric home appliances using hand shapes is shown in Fig. 13 [**19**]. In this system, the discrimination results are sent to the main unit, and the infrared signals corresponding to the electric home appliances are transmitted directly via the infrared light-emitting diode (LED) of the main unit to the appliances. In this experiment, operations corresponding to the

Table 1 The number of added classifiers and LLGMNs

Class	Classifiers	LLGMNs
1	2	5
2	1	1
3	1	1
4	3	3
5	2	2
6	1	1
7	1	1
8	4	4
9	2	6
10	3	3
11	2	4
12	2	2
13	2	3
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	2	6
20	1	1
21	1	1
22	1	1
23	1	1
24	1	1
25	1	1
26	2	4
27	3	4
28	1	1
29	1	1
30	1	1
31	2	7

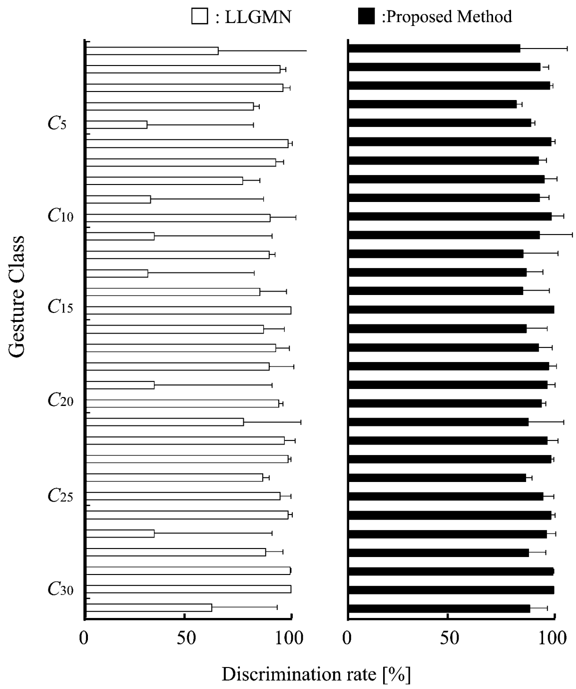


Fig. 10 Discrimination results of finger motion for subject C

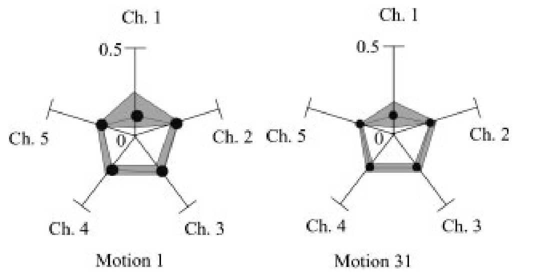


Fig. 11 Rader charts of the hand gesture pattern of gesture 1 and gesture 31 for subject C. The line indicates the mean value of each channel

user's motions are executed until the same discrimination occurs 150 times.

Examples of the operations corresponding to discriminated motion are shown in Table 2 [**19**]. In

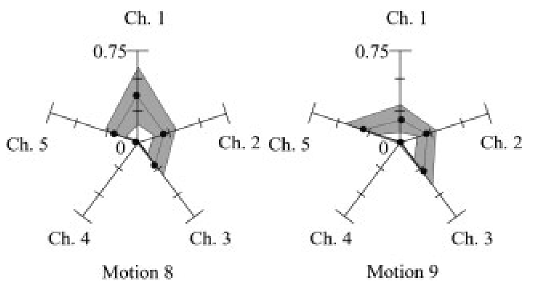
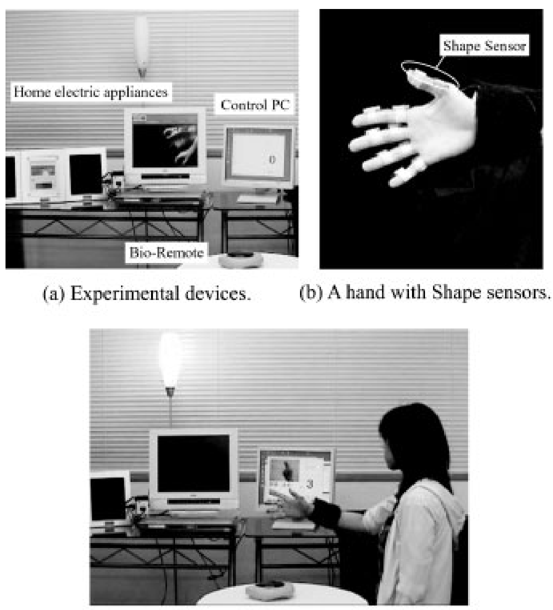


Fig. 12 Rader charts of the hand gesture pattern of gesture 8 and gesture 9 for subject C. The line indicates the mean value of each channel



(a) Experimental devices. (b) A hand with Shape sensors.

(c) An operation scene for home electric appliances.

Fig. 13 Operation of home electric appliances using discrimination results of hand shape

Table 2 Example of command allocation for home electric appliances

Motion number	Object	Command
1	Light	On
2		Off
3	TV	Switch
4		CH up
5		CH down
6	CD player	Power on/off
7		Play
8		Stop
9		Volume up
10		Volume down
⋮	⋮	⋮

a usual Bio-Remote, various operations can be performed by repeating a command selection. However, from this figure, it can be inferred that each operation corresponds to a single hand shape motion. In this experiment, the shape signals measured beforehand were identified by the proposed method.

An example of the subject's operation is shown in Fig. 14. In this figure, five channels of the normalized signals, discrimination results, and control commands are plotted. The grey areas indicate that the Bio-Remote is not operated.

From these experimental results, it can be inferred that the subject could operate the home electric appliances by changing the shape of his or her fingers. It should be noted that there was no malfunction and that the home electric appliances could be operated according to the subject's intent, confirming that using the proposed system, the subject can control various electric appliances simply by moving his or her fingers.

5 CONCLUSION

In this paper, a novel hierarchical probabilistic neural network based on a boosting approach is proposed. In the proposed method, the structure of the classification network is constructed by adding LLGMNs as classifiers to estimate the distribution of training data. By evaluating the structure based on classification accuracy, the addition of unnecessary LLGMNs can be avoided.

Experimental results on the artificial data set and hand shape signals prove the feasibility of the proposed method. Comparison experiments of the proposed method and single LLGMN were conducted, and the high classification performance of the proposed method was confirmed. It has been shown that the proposed method is suitable for classification of complex data, since the required

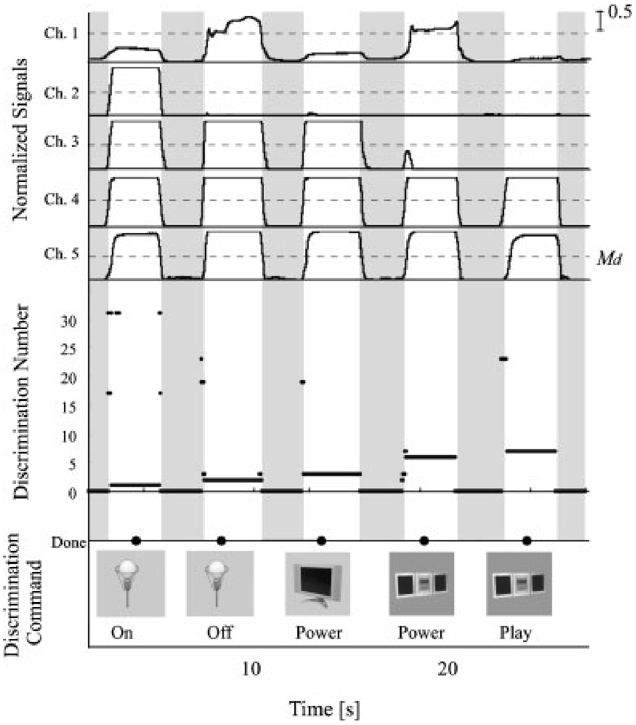


Fig. 14 An example of the experimental results during the hand shapes of the electric appliances by the subject

classifiers will automatically be added in the network in order to perform an accurate classification. Furthermore, assignment of the hand shapes directly to the operation command of a home electric appliance confirmed the feasibility of direct operation of home electric appliances using the Bio-Remote.

In addition to establishing the classification ability of the proposed PNNs, future study must focus on various theoretical aspects. The proposed method should also be applied to other data.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to Ms Juri Murakami for collecting the experimental data.

REFERENCES

1 **Gardner, E.** The space of interactions in neural network model. *J. Physics A: Mathematical and General.*, 1988, **21**, 257.
2 **Cristianini, N.** and **Shawe-Taylor, J.** *An Introduction to support vector machines*, 2000 (Cambridge University Press, Cambridge).
3 **Bishop, C. M.** *Neural networks for pattern recognition*, 1995 (Oxford University Press, Oxford).
4 **Hiraiwa, A., Shimohara, K.,** and **Tokunaga, Y.** EMG pattern analysis and classification by neural network. In IEEE Informational Conference on System, *Man and Cybernetics*, 1989, pp. 1113–1115.
5 **Tsuji, T., Fukuda, O., Ichinobe, H.,** and **Kaneko, M.** A log-linearized Gaussian mixture network and its application to EEG pattern classification. *IEEE Trans. on Systems, Man and Cybernetics – Part C: Applications and Reviews*, 1999, **29**(1), 60–72.
6 **Tsuji, T., Shigeyoshi, H., Fukuda, O., Ichinobe, H.,** and **Kaneko, M.** Bio-mimetic control of an extrenally powered prothetic forearm based on EMG signals (in Japanese). *Trans. Jap. Soc. Mech. Engrs*, 2000, **66**, 2764–2771.
7 **Fukuda, O., Tsuji, T.,** and **Kaneko, M.** Pattern classification of EEG signals using a log-linearlized Gaussian mixture neural network. In Proceedings of the IEEE International Conference on *Neural Network*, 1995, pp. 2479–2484.
8 **Fukuda, O., Tsuji, T.,** and **Kaneko, M.** An EMG controlled robotic manipulator using neural networks. In Proceedings of the IEEE International Workshop on *Robot and human communication*, 1997, pp. 442–447.
9 **Fukuda, O., Tsuji, T.,** and **Kaneko, M.** An EMG-controlled pointing device using a neural netowork. In Proceedings of the 1999 IEEE International Conference on *Systems, Man, and Cybernetics*, Tokyo, 1999, **4**, pp. 63–68.
10 **Fukuda, O., Tsuji, T., Kaneko, M.,** and **Ohtsuka, A.** A human-assisting manipulator teleoperated by EMG signals and arm motions. *IEEE Trans. on Robotics and Automn.*, 2003, **19**, 210–222.
11 **Akaike, H.** A new look at the statistical model identification. *IEEE Trans. Autom. Control*, 1974, **19**, 716–723.
12 **Watanabe, K.** and **Watanabe, S.** Lower bounds of stochastic complexities in variational Bayes learning of Gaussian mixture models. In Proceedings of the IEEE Conference on *Cybernetics and intelligent systems (CIS04)*, 2004, pp. 99–104.
13 **Schapire, R. E.** The strength of weak learnability. *Mach. Learning*, 1990, **5**(2), 197–227.
14 **Freund, Y.** Boosting a weak learning algorithm by majority. *Information and Computation*, 1995, **121**(2), 256–285.
15 **Domingo, C., Tsukiji, T.,** and **Watanabe, O.** Partial Occam's razor its applications. *Information Processing Lett.*, 1997, **64**, 179–185.
16 **Perlovsky, L. I.** and **McManus, M. M.** Maximum likelihood neural networks for sensor fusion and adaptive classification. *Neural Networks*, 1991, **4**, 89–102.
17 **Rumelhart, D. E., McClelland, J. L.,** and **Williams, R. J.** Learning internal representations by error propagation. *Parallel Distributed Processing*, 1986, **1**, 318–362.
18 **Tsuji, T., Shima, K., Funabiki, A., Shitamori, S., Shiba, K., Fukuda, O.,** and **Otsuka, A.** A new

manipulation method for environment control systems using shape sensors (in Japanese). *Life Support*, 2006, **18**(4), 5–12.

19 **Okamoto, M., Murakami, J., Shiba, K.,** and **Tsuji, T.** A tree-based hierarchical neural network and its application to hand shape discrimination. *Trans. Soc. Instrument and Control Engrs* (in Japanese), 2007, **43**(2), 128–134.

APPENDIX

Notation

C	number of classes
D	length of input vector
$D(c, c')$	data set belonging to class c and classified into class c'
E	energy function for training process
F	evaluation function for boosting
F'	evaluation function for classifier
G_c	set of data belonging to class c
$G_{\bar{c}}$	complementary set of G_c
H	number of units of the first layer
$H(x)$	entropy
${}^{(2)}I_{k,m}$	input of the second layer
${}^{(3)}I_k$	input of the third layer
J_q	number of LLGMN in qth classifier
$J^{(n)}$	log-likelihood
K	number of units of the third layer
$L_c^{(q)}(x)$	output of classifier
M_d	threshold
M_k	number of components
$N_d(n)$	normalized signal
$O_c(x)$	output of proposed network
${}^{(1)}O_{k,m}$	output of the first layer
${}^{(2)}O_{k,m}$	output of the second layer
${}^{(3)}O_k$	output of the third layer
${}^{(3)}O_k^{(q,j)}$	output of jth LLGMN in qth classifier
$P(k\|x)$	*a posteriori* probability of k
Q_c	number of classifiers
$s_d(n)$	processed signal to make the sum of all channels equal to 1
$S_d(n)$	measured signals
$S_d^{\max}$	mean of the maximum value of each channel
S_d^{st}	mean value of Sd(n)
T_e	threshold for discrimination
$T^{(n)}$	outputs of training data
Th	threshold for F
Th'	threshold for F'
$w_h^{(k,m)}$	weight of LLGMN
x_i	element of input vector x
x	input vector of LLGMN
$x^{(n)}$	inputs of training data
$\boldsymbol{X}$	converted input vector of LLGMN
$\boldsymbol{X}^{(n)}$	input vector converted from $x^{(n)}$
$Y(x)$	determined class
$\Delta w_h^{(k,m)}$	modification of the weight
η	learning rate