

An EMG Pattern Classification Method Based on a Mixture of Variance Distribution Models

Akira Furui¹, *Student Member, IEEE*, Hideaki Hayashi², *Member, IEEE*, and Toshio Tsuji¹, *Member, IEEE*

Abstract—This paper proposes an electromyogram (EMG) pattern classification method based on a mixture of variance distribution models. A variance distribution model is a stochastic model of raw surface EMG signals in which the EMG variance is taken as a random variable, allowing the representation of uncertainty in the variance. In this paper, we extend the variance distribution model to the multidimensional case and enhance its flexibility for multichannel and processed EMG signals. The enhanced model enables the accurate classification of EMG patterns while considering the uncertainty in the EMG variance. The robustness and applicability of the proposed method are demonstrated through a simulation experiment using artificially generated data and EMG classification experiments using two real datasets.

I. INTRODUCTION

A surface electromyogram (EMG) is a bioelectrical signal generated by muscle contractions that is recorded from the skin surface. Many attempts have been made to use EMG signals for the control of interfaces such as prosthetic hands and information devices.

To enable such applications, it is necessary to extract the user's intention from recorded EMG signals accurately. In general, this is achieved by associating the user's motion with the recorded EMG patterns based on machine learning algorithms. Hudgins *et al.* used a multilayer perceptron (MLP) with back propagation learning to classify multiple limb motions [1]. This neural network-based approach has been widely used because of its high classification ability, but suffers from the problem of overfitting.

The classification ability can be improved by considering the stochastic characteristics of EMG signals. Chan and Englehart focused on this probabilistic approach [2] and demonstrated the potential of Gaussian mixture models (GMMs) in EMG pattern classification by modeling the feature vector of EMG signals based on a GMM. Huang *et al.* also demonstrated that GMMs could classify EMG patterns with equal or better accuracy than an MLP [3].

The stochastic model of EMG signals has been studied in an attempt to extract their features. In one study, raw EMG signals were assumed to follow a Gaussian distribution with a mean of zero [4]. However, more recent studies have indicated that raw EMG signals have non-Gaussian properties [5], [6]. This is believed to reflect the uncertainty involved in the motor unit activity and motor command

(source of muscle activity). Hayashi *et al.* proposed a variance distribution model of raw EMG signals based on the assumption that the EMG signal variance is a random variable, thereby allowing an expression of the uncertainty in the variance [7]. If we could construct a classifier that incorporates a variance distribution model, it would enhance the accuracy of EMG classification.

This paper outlines an EMG pattern classification method based on a mixture of variance distribution models. The variance distribution model included in the proposed method is extended to the multidimensional case and has sufficient flexibility for feature vectors obtained from multichannel EMG signals. Therefore, the proposed method can classify EMG patterns while considering the uncertainty of the EMG variance.

II. PROPOSED CLASSIFICATION METHOD

A. Mixture of Multivariate Variance Distribution Models

The variance distribution model of surface EMG signals proposed by Hayashi *et al.* is a stochastic model for raw EMG signals recorded from a pair of electrodes [7]. EMG pattern classification, however, generally uses EMG signals from multichannel electrodes. In addition, the EMG signals used in such applications are processed by feature extraction techniques such as rectification and smoothing, and their distributions may therefore exhibit skewness and multimodality. Thus, we expand the previous variance distribution model [7] to consider multidimensionality and the flexibility of the processed EMG signals, and propose a mixture of multivariate variance distribution models.

Fig. 1 shows an overview of the mixture of multivariate variance distribution models. The D -dimensional EMG feature vector $\mathbf{x} \in \mathbb{R}^D$ is handled as a random variable that follows a multivariate Gaussian mixture distribution, achieving flexible modeling of the distribution shape. A covariance matrix Σ_k in the Gaussian mixture distribution of component k is also a random variable with a distribution determined by the degree of freedom ν_k and the scale matrix Ψ_k .

For component k , the conditional distribution of the EMG feature vector $\mathbf{x}$ given $\Sigma_k \in \mathbb{R}^{D \times D}$ ($k = 1, \dots, K$; K is the number of components) is expressed via the following multivariate Gaussian distribution:

$$P(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k) = (2\pi)^{-\frac{D}{2}} |\Sigma_k|^{-\frac{1}{2}} \times \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \right], \quad (1)$$

*This work was not supported by any organization

¹A. Furui and T. Tsuji are with the Graduate School of Engineering, Hiroshima University, Higashi-hiroshima, 739-8527 Japan akirafurui@hiroshima-u.ac.jp

²H. Hayashi is with the Department of Advanced Information Technology, Kyushu University, Fukuoka, 819-0395 Japan

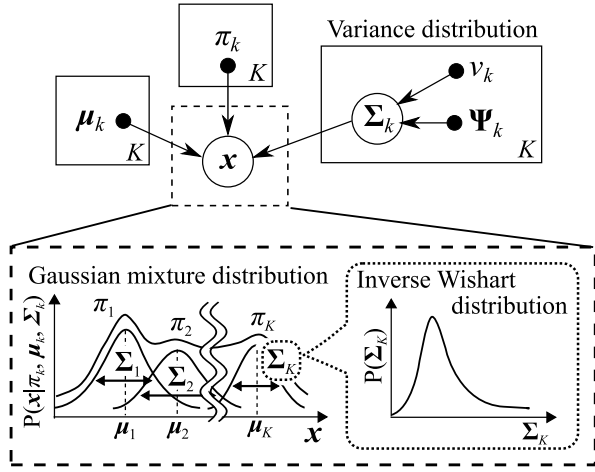


Fig. 1. Overview of the mixture of multivariate variance distribution models. In the model, an EMG feature vector $\mathbf{x}$ is handled as a random variable that follows a Gaussian mixture distribution. A covariance matrix of each component in the Gaussian mixture distribution is also a random variable following the inverse Wishart distribution.

where $\mu_k \in \mathbb{R}^D$ is the mean vector of component k . The distribution of Σ_k is assumed to be the inverse Wishart distribution, which is known as a conjugate prior for the covariance matrix of the multivariate Gaussian distribution:

$$P(\Sigma_k|\nu_k, \Psi_k) = \frac{|\Psi_k|^{\frac{\nu_k}{2}}}{2^{\frac{\nu_k D}{2}} \Gamma_D(\frac{\nu_k}{2})} |\Sigma_k|^{-\frac{\nu_k + D + 1}{2}} \times \exp \left[-\frac{\text{tr}(\Psi_k \Sigma_k^{-1})}{2} \right], \quad (2)$$

where $\nu_k \in \mathbb{R}$ is the degree of freedom and $\Psi_k \in \mathbb{R}^{D \times D}$ is the scale matrix of component k ; these parameters determine the inverse Wishart distribution.

The likelihood of EMG feature vector $\mathbf{x}$ is given by a linear combination of the marginal distributions of $\mathbf{x}$:

$$P(\mathbf{x}) = \sum_{k=1}^K \pi_k P(\mathbf{x}|\nu_k, \mu_k, \Psi_k), \quad (3)$$

where π_k is the mixing coefficient of component k . The marginal distribution of $\mathbf{x}$ is expanded as follows:

$$\begin{aligned} P(\mathbf{x}|\nu_k, \mu_k, \Psi_k) &= \int P(\mathbf{x}|\mu_k, \Sigma_k) P(\Sigma_k|\nu_k, \Psi_k) d\Sigma_k, \\ &= \frac{\Gamma(\frac{\nu_k + 1}{2})}{\Gamma(\frac{\nu_k - D + 1}{2})} \frac{|\frac{1}{\nu_k - D + 1} \Psi_k|^{-\frac{1}{2}}}{[\pi(\nu_k - D + 1)]^{\frac{D}{2}}} (1 + \Delta_k)^{-\frac{\nu_k + 1}{2}}, \end{aligned} \quad (4)$$

where

$$\Delta_k = (\mathbf{x} - \mu_k)^T \Psi_k^{-1} (\mathbf{x} - \mu_k). \quad (6)$$

B. Posterior Probability Estimation

To classify the vector $\mathbf{x} \in \mathbb{R}^D$ into the given C classes, the posterior probability $P(c|\mathbf{x})$ ($c = 1, \dots, C$) is calculated

as:

$$P(c|\mathbf{x}) = \frac{P(c)P(\mathbf{x}|c)}{\sum_{c=1}^C P(c)P(\mathbf{x}|c)}, \quad (7)$$

$$P(\mathbf{x}|c) = \sum_{k=1}^{K^{(c)}} \pi_k^{(c)} P(\mathbf{x}|\nu_k^{(c)}, \mu_k^{(c)}, \Psi_k^{(c)}), \quad (8)$$

where $P(c)$ is the prior probability of c , and $\pi_k^{(c)}$, $\nu_k^{(c)}$, $\mu_k^{(c)}$, and $\Psi_k^{(c)}$ are the parameters of the mixture of multivariate variance distribution models of class c . For a given $\mathbf{x}$, patterns are classified by selecting the class with the maximum posterior probability.

C. Learning Algorithm

This section describes a learning algorithm for the model parameters. A set of D -dimensional EMG feature vectors $\mathbf{X}^{(c)} = \{\mathbf{x}_n^{(c)} \in \mathbb{R}^D; n = 1, \dots, N\}$ (N is the number of training samples) is used to train class c . The model parameters can be estimated by maximizing the log-likelihood function as follows:

$$\ln L(\mathbf{X}^{(c)}) = \ln \prod_{n=1}^N \sum_{k=1}^{K^{(c)}} \pi_k^{(c)} P(\mathbf{x}_n^{(c)}|\nu_k^{(c)}, \mu_k^{(c)}, \Psi_k^{(c)}). \quad (9)$$

We introduce the expectation-maximization (EM) algorithm to maximize the likelihood function.

To simplify the expression, we define new parameters as:

$$\nu_k'^{(c)} = \nu_k^{(c)} - D + 1, \quad (10)$$

$$\Psi_k'^{(c)} = \frac{1}{\nu_k^{(c)} - D + 1} \Psi_k^{(c)}. \quad (11)$$

The model parameters $[\pi_k^{(c)}, \nu_k'^{(c)}, \mu_k^{(c)}, \Psi_k'^{(c)}]$ are then estimated as outlined below using the EM algorithm.

1) *E-step*: Update the weight parameters as follows:

$$\tau_{k,n}^{(c)} = \frac{\pi_k^{(c)} P(\mathbf{x}_n^{(c)}|\nu_k'^{(c)}, \mu_k^{(c)}, \Psi_k'^{(c)})}{\sum_{k=1}^{K^{(c)}} \pi_k^{(c)} P(\mathbf{x}_n^{(c)}|\nu_k'^{(c)}, \mu_k^{(c)}, \Psi_k'^{(c)})}, \quad (12)$$

$$u_{k,n}^{(c)} = \frac{\nu_k'^{(c)} + D}{\nu_k'^{(c)} + \Delta_k^{(c)}}, \quad (13)$$

where

$$\Delta_k^{(c)} = (\mathbf{x}_n^{(c)} - \mu_k^{(c)})^T (\Psi_k'^{(c)})^{-1} (\mathbf{x}_n^{(c)} - \mu_k^{(c)}). \quad (14)$$

2) *M-step*: Update the mixing coefficient, mean vector, and scale matrix using $\tau_{k,n}^{(c)}$ and $u_{k,n}^{(c)}$.

$$\text{new } \pi_k^{(c)} = \sum_{n=1}^N \frac{\tau_{k,n}^{(c)}}{N}, \quad (15)$$

$$\text{new } \mu_k^{(c)} = \frac{\sum_{n=1}^N \tau_{k,n}^{(c)} u_{k,n}^{(c)} \mathbf{x}_n^{(c)}}{\sum_{n=1}^N \tau_{k,n}^{(c)} u_{k,n}^{(c)}}, \quad (16)$$

$$\text{new } \Psi_k'^{(c)} = \frac{\sum_{n=1}^N \tau_{k,n}^{(c)} u_{k,n}^{(c)} (\mathbf{x}_n^{(c)} - \mu_k^{(c)}) (\mathbf{x}_n^{(c)} - \mu_k^{(c)})^T}{\sum_{n=1}^N \tau_{k,n}^{(c)}}. \quad (17)$$

Because there is no closed form expression for the degree of freedom $\nu_k'^{(c)}$, we introduce two determination approaches

TABLE I
PARAMETERS FOR THE EXPERIMENTAL DATA GENERATION

	ν_k	μ_k	Ψ_k	
Class 1	10.0	0.5 -1.0	2.5 0.0	0.0 2.5
Class 2	3.0	2.5 2.5	1.0 0.6	0.6 1.0
Class 3	3.0	0.5 3.0	2.0 0.0	0.0 0.05

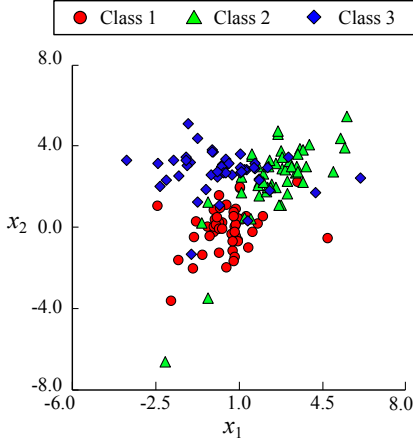


Fig. 2. Scattergram of a dataset used in the simulation experiment. The influence of dispersion in the variance results in each class having outliers.

for this parameter. One fixes the degree of freedom for each component in advance, and the other estimates the degree of freedom from the learning data by maximizing the log-likelihood $\ln L(\mathbf{X}^{(c)})$ based on a line search during the *M-step*. Finally, the log-likelihood $\ln L(\mathbf{X}^{(c)})$ is evaluated and steps 1)–2) are repeated until the calculation converges.

Using this procedure, the parameters of the proposed model corresponding to class c can be trained.

III. SIMULATION EXPERIMENT

A. Method

To verify that the proposed method can handle uncertainty in the variance of the input vectors, a simulation experiment was performed using two-dimensional ($D = 2$) data with three classes ($C = 3$). The data were artificially generated using the mixture of multivariate variance distributions with a single component ($K^{(c)} = 1$). Table I indicates the parameters for each class used in the generation process. An example of a dataset used in the experiments is shown in Fig. 2.

In the experiment, the training and test sets for each class consisted of 50 and 5000 samples, respectively. The average classification rate was calculated by changing the combination of training/test samples in each set 20 times. The average classification rate was compared with that of a GMM. Both the proposed method and the GMM had 1 component. The degree of freedom $\nu_k^{(c)}$ in the proposed method was estimated from the training data.

B. Results and Discussion

Fig. 3 shows color-coded posterior probabilities for each class given by the proposed method and the GMM using

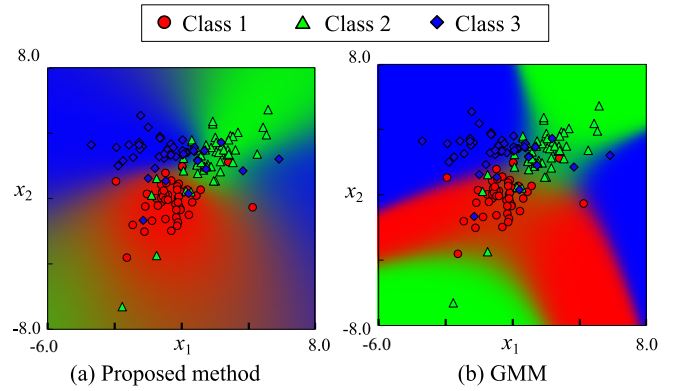


Fig. 3. Color-coded posterior probabilities for each class given by (a) the proposed method and (b) the GMM. Red, green, and blue areas indicate posterior probabilities close to 1.0 for classes 1, 2, and 3, respectively.

inputs in the range $-6.0 \leq x_1 \leq 8.0$ and $-8.0 \leq x_2 \leq 8.0$. The red, green, and blue areas show that the posterior probabilities are close to 1.0 for classes 1, 2, and 3, respectively. The classification rates using the proposed method and the GMM are $87.60 \pm 0.32\%$ and $84.18 \pm 5.38\%$, respectively, and a paired *t*-test indicates this is a significant difference ($p < 0.01$).

In Fig. 3(a), the posterior probability given by the proposed method resembles the distribution shape of the experimental data, indicating that the proposed method can learn the appropriate decision boundary separating each class. Although the GMM can also adequately learn the experimental data, the spread of its decision regions is different from that of the proposed method (Fig. 3(b)). This is because the outliers in the experimental data cause an increase in variance for each class in the GMM. This tendency is notable in class 2, and the decision region of class 2 in the GMM is heavily affected by outliers and becomes wider than that of the proposed method. In contrast, the proposed method considers the dispersion in variance by introducing the variance distribution, which seems to restrain the expansion of the decision regions. This is confirmed by the fact that the proposed method achieves better classification performance than the GMM. These results demonstrate the robustness of the proposed method against outliers caused by the dispersion in variance.

IV. EMG CLASSIFICATION EXPERIMENT

A. Method

To evaluate the suitability of the proposed classification method for real EMG data, an experiment was conducted using two EMG datasets. In this experiment, we evaluated two forms of the proposed classification method. The first one, referred to as type 1, uses the same degree of freedom $\nu_k^{(c)}$ for each component and each class. The second one, referred to as type 2, uses a trained degree of freedom $\nu_k^{(c)}$ estimated from the training samples. The number of components $K^{(c)}$ and the degree of freedom $\nu_k^{(c)}$ for type 1 were determined using 10-fold cross-validation ($K^{(c)}$ ranging from 1 to 5 and $\nu_k^{(c)}$ ranging from 2 to 10). The number of components

$K^{(c)}$ for type 2 was also determined using 10-fold cross-validation. The results given by the proposed method (types 1 and 2) were compared with those of a GMM and an MLP. The number of components (from 1 to 5) for the GMM and the number of nodes (from D to $D + 10$) in the hidden layer of the MLP were determined based on 10-fold cross-validation.

Datasets I and II were taken from [8] and [9], and can be downloaded from Dr. Khushaba's website [10]. For dataset I, eight subjects (aged 20–35 years old) were recruited and seated on an armchair, facing a personal computer with a steering wheel attached to the desk. They then performed fourteen classes of finger movements ($C = 14$) over four trials ($T = 4$). Dataset II contains measurements from eight subjects (aged 20–35 years old) seated on an armchair with their arm supported and fixed in one position. The subjects performed fifteen classes of finger and hand movements ($C = 15$) over three trials ($T = 3$). In both datasets, EMG signals were recorded using eight-channel electrodes ($D = 8$) at 4 kHz and digitized using a 12-bit A/D converter. To extract feature vectors, the EMG signals were then rectified and smoothed using a second-order Butterworth low-pass filter with a cut-off frequency of 1 Hz.

The average classification accuracy was calculated for each subject based on leave-one-trial-out cross-validation. In this validation process, $T - 1$ trials were used as the training set and the remaining trial was used as the test set. Because it is difficult to procure many training samples in real-world applications, 1% of the available samples were randomly selected from the training set and used for the training samples.

B. Results and Discussion

Fig. 4 shows the classification accuracy for each method. Statistical test results based on a paired t -test with Bonferroni-Holm adjustment are also given.

From Fig. 4, the proposed method (type 1) shows significantly higher classification accuracy than the other methods for both datasets. In contrast, the accuracy of the proposed method (type 2) using the trained degree of freedom is significantly lower than that of type 1 and the GMM. Note that the degree of freedom in the proposed method controls the tolerance of the dispersion in EMG variance. Hence, one possible explanation is that the proposed method (type 2) overfitted the dispersion in the variance of the training data, and lost its generalization ability for the test data.

The above results reveal that the proposed method can classify EMG patterns with higher accuracy than GMM and MLP when the degree of freedom is fixed. This is because the proposed classification method is naturally derived from the stochastic model of EMG, thereby allowing the handling of uncertainty in the EMG signals. In this experiment, the degree of freedom in the proposed method (type 1) was determined by cross-validation. However, if we could theoretically determine the optimum value of the degree of freedom, the accuracy of the proposed method may be further improved.

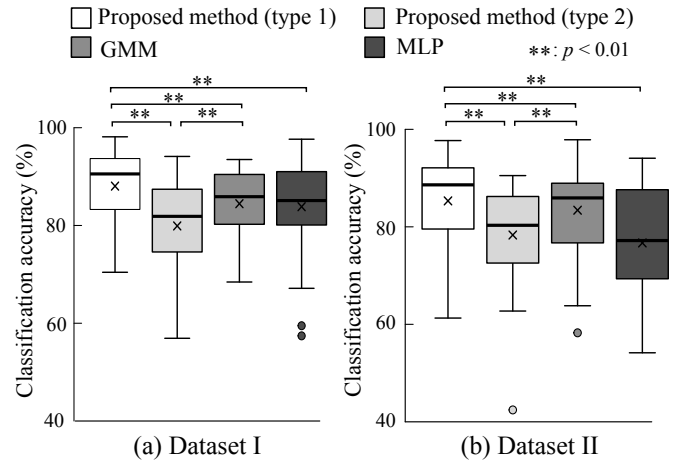


Fig. 4. Box plots of EMG classification results. “x” in the box plot denotes the average accuracy of each method.

V. CONCLUSION

This paper has proposed an EMG pattern classification method based on a mixture of variance distribution models. The results of simulation experiments demonstrate that the proposed method is robust against outliers generated from the dispersion in variance. The applicability of the proposed method to real EMG signal classification was demonstrated by the results of an EMG classification experiment.

In future research, we will examine the optimal settings of the hyperparameters in the proposed method, and also conduct further evaluation experiments using different types of data.

REFERENCES

- [1] B. Hudgins, P. Parker, and R. Scott, “A new strategy for multifunction myoelectric control,” *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, 1993.
- [2] A. D. C. Chan and K. B. Englehart, “Continuous classification of myoelectric signals for powered prostheses using Gaussian mixture models,” *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (IEEE Cat. No.03CH37439)*, vol. 3, pp. 2841–2844, 2003.
- [3] Y. Huang, K. Englehart, B. Hudgins, and A. Chan, “A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses,” *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1801–1811, 2005.
- [4] N. Hogan and R. W. Mann, “Myoelectric signal processing: optimal estimation applied to electromyography - part i: derivation of the optimal myoprocessor,” *IEEE Trans. Biomed. Eng.*, vol. BME-27, no. 7, pp. 382–395, 1980.
- [5] E. A. Clancy and N. Hogan, “Probability density of the surface electromyogram and its relation to amplitude detectors,” *IEEE Trans. Biomed. Eng.*, vol. 46, no. 6, pp. 730–739, 1999.
- [6] K. Nazarpour, A. H. Al-Timemy, G. Bugmann, and A. Jackson, “A note on the probability distribution function of the surface electromyogram signal,” *Brain Res. Bull.*, vol. 90, pp. 88–91, 2013.
- [7] H. Hayashi, A. Furui, Y. Kurita, and T. Tsuji, “A variance distribution model of surface EMG signals based on inverse gamma distribution,” *IEEE Trans. Biomed. Eng.*, vol. 64, no. 11, pp. 2672–2681, 2017.
- [8] R. N. Khushaba, S. Kodagoda, D. Liu, and G. Dissanayake, “Muscle computer interfaces for driver distraction reduction,” *Comput. Methods Programs Biomed.*, vol. 110, no. 2, pp. 137–149, 2013.
- [9] R. N. Khushaba and S. Kodagoda, “Electromyogram (EMG) feature reduction using mutual components analysis for multifunction prosthetic fingers control,” *2012 12th Int. Conf. Control. Autom. Robot. Vision, ICARCV 2012*, pp. 1534–1539, 2012.
- [10] R. N. Khushaba, “Dr. rami khushaba - home.” [Online]. Available: <http://www.rami-khushaba.com/>