Multi-classifier Decision: Integration of Multiple Brain Activity-based Classifications

Takahiro Ogawa (member)†, Kento Sugata †, Ren Togo †, Miki Haseyama (member)†

Abstract A novel method that integrates brain activity-based classifications obtained from multiple users is presented in this paper. The proposed method performs decision-level fusion (DLF) of the classifications using a kernelized version of extended supervised learning from multiple experts (KESLME), which is newly derived in this paper. In this approach, feature-level fusion of multiuser electroencephalogram (EEG) features is performed by multiset supervised locality preserving canonical correlation analysis (MSLPCCA). In the proposed method, the multiple classification results are obtained by classifiers separately constructed for the multiuser EEG features. Then DLF of these classification results becomes feasible based on KESLME, which can provide the final decision with consideration of the relationship between the MSLPCCA-based integrated EEG features and each classifier’s performance. In this way, a new multi-classifier decision technique, which depends only on users’ brain activities, is realized, and the performance in an image classification task becomes comparable to that of Inception-v3, one of the state-of-the-art deep convolutional neural networks.

Key words: electroencephalogram, multi-classifier scheme, feature-level fusion, decision-level fusion, image classification.

1. Introduction

Recently, deep neural networks as typified by deep convolutional neural networks (DCNNs) have achieved good results for many computer vision tasks such as image classification. Specifically, DCNN approaches such as AlexNet have significantly improved the accuracy compared to that of previous hand-crafted features such as GIST descriptors, scale-invariant feature transform (SIFT) and pyramid histograms of oriented gradients (PHOG). Further related works on DCNNs continue to sophisticate the network architecture and improve the performance.

Despite explosive growth of deep learning techniques, it is well known that a sufficient number of training samples cannot be prepared in many scenes. For solving this problem, many approaches that enable training from smaller training datasets have been proposed, but they generally assume that other large training datasets different from the target ones can be provided. In this paper, we focus on the most difficult case in which we have only a small number of training samples and try to provide another direction, collaboration with users (also known as workers or labelers) who look at these samples. The simplest solution is the use of the traditional crowd-sourcing technique, but this needs direct decisions from users. On the other hand, brain activity data obtained while watching visual data can be obtained without any decisions. Thus, we aim at realizing a new technique that improves the accuracy by increasing the number of brain activity-based classifiers instead of increasing the number of training samples. Various works related to integration of multiple classifiers have confirmed that a higher performance is achieved by increasing the number of classifiers. Therefore, according to the above reports, it is expected that the integration of multiuser brain activity-based classifications will be another effective cue.

As related works, there have been proposed several methods using multiple features obtained from multimedia contents and those obtained from biological signals of humans such as electroencephalogram (EEG) signals. Therefore, by integrating several classification results obtained from visual features and EEG features extracted from EEG signals recorded while a user stares at the images, the improvement of the classification performance can be expected. However, in most previous works, it becomes difficult to perform successful integration of several classification results with concerning the performance of each classifier. This means that there have been few studies guaranteeing performance improvement based on the integration of the different classifiers.

A novel multi-classifier decision method that integrates multiple brain activity-based classifications is presented in this paper. The proposed method includes two contributions: decision-level fusion (DLF) of the classifications using a
integrates multiuser EEG-based classifications without any
the case of three users or more, the proposed method, which
EEG features are input to KESLME. From experiments in
we introduce MSLPCCA, which integrates multiuser EEG
results obtained from multi-subjects’ EEG data. Furthermore,
are few studies which integrate multiple classification re-
the number of the integrated subjects increases, and there
will not improve the final integration performance even if
MSLPCCA-based feature-level fusion (FLF) + KESLME-based decision-level fusion (DLF)
Final classification result

Fig. 1 Brief overview of the proposed method. The procedures shown in blue boxes correspond to the first stage, and that shown in a green box corresponds to the second stage which is the biggest contribution of this paper.

kernelized version of extended supervised learning from multiple experts (KESLME) and feature-level fusion (FLF) of multiuser EEG features using multiset supervised locality preserving canonical correlation analysis (MSLPCCA)

The proposed method first performs classification via support vector machines (SVMs) by separately using multiple users’ electroencephalogram (EEG) features. Then the proposed method performs DLF of the above multiple classifications, and in this approach, we newly derive KESLME as the biggest contribution of this paper. Since KESLME realizes the final decision from multiple classification results with consideration of both the integrated EEG features and each classifier’s performance, improvement in the classification performance can be guaranteed. Generally, it is expected that the classification performance obtained from each subject’s EEG data is different due to individual differences and some noises and artifacts caused in the EEG observations. Therefore, simple integration without considering the differences of their classification performances will not improve the final integration performance even if the number of the integrated subjects increases, and there are few studies which integrate multiple classification results obtained from multi-subjects’ EEG data. Furthermore, we introduce MSLPCCA, which integrates multiuser EEG features by considering their correlations, and the integrated EEG features are input to KESLME. From experiments in the case of three users or more, the proposed method, which integrates multiuser EEG-based classifications without any visual feature-based classifications, realizes image classification performance comparable to that of Inception-v3, one of the state-of-the-art DCNNs.

Note that some conventional methods perform the inter-subject analysis of physiological signals including EEG signals. These methods use inter-subject correlation using physiological signals in order to achieve robust analysis for realizing their target tasks on the basis of the fact that the physiological signals of the multiple subjects have common trends. On the other hand, we assume that the classification performance from each subject’s EEG data is different from each other and focus on how such multiple classification results, whose performances are different, should be integrated. Therefore, since we focus on integration of multiple classification results obtained from these multiple subjects’ EEG data, we newly introduce KESLME into our method.

2. Multi-classifier Decision via KESLME-based DLF

The proposed method consists of two stages. In the first stage, the EEG features are extracted from EEG signals recorded while each user is looking at images. SVM-based classifications are performed by separately using each user’s EEG features. Then multiple classification results are obtained for test data based on each user’s EEG features. In the second stage, the multiuser decision, i.e., the final decision, is obtained by KESLME-based integration. In this
approach, we utilize feature vectors generated by integrating the multiuser EEG features based on MSLPCCA. The brief overview of the proposed method is shown in Fig. 1

2.1 EEG Feature Extraction and Classification

The first stage is further divided into the following: feature extraction and single user’s feature-based classification. First, we explain the EEG features used in the proposed method. Segmentation of each channel’s EEG signals is performed at a fixed interval with an overlapped Hamming window. In this paper, \( F_{m,t} (m = 1, 2, \ldots, M; t = 1, 2, \ldots, T; M \text{ and } T \) being the number of users and EEG frames, respectively) denotes the \( m \)th user’s \( t \)th EEG frame. Table 1 shows the EEG features computed from each EEG frame \( F_{m,t} \). Note that \( C \) denotes the number of channels of EEG signals, and \( P \) is the number of symmetric electrode pairs placed on the scalp. Therefore, the dimension of each EEG feature is \( 6C + 10P \). Zero crossing rate is calculated in the time domain, and the other features are computed in the frequency domain by short-time Fourier transform to each channel’s EEG signals. In order to select only features useful for the classification, the dimension of the features is reduced. Although there are many feature or channel selection methods, our previous studies confirmed the effectiveness of dimensionality reduction via the minimum-redundancy and maximum-relevance (mRMR) algorithm. The mRMR algorithm is therefore applied to the EEG features, and we obtain an effective feature set. After this procedure,\( x_{i}^{F_{m,t}} \in \mathbb{R}^{d_{m,t}} (i = 1, 2, \ldots, N; N \) being the number of images in training data) is obtained as an EEG feature vector for each EEG frame \( F_{m,t} \), where \( d_{m,t} \) is the number of the selected features for each EEG frame \( F_{m,t} \).

Next, we explain the classification method based on each user’s EEG features. SVM-based classification is performed for the \( m \)th user’s EEG features at the \( t \)th frame. Since binary classification can be easily expanded into multi-class classification based on one-versus-one or one-versus-all methodologies, we only focus on improvement of the SVM-based binary classification. First, we train classifiers by separately using multiuser EEG feature vectors. Specifically, \( M \times T \) classifiers \( A_{1,1}, A_{1,2}, \ldots, A_{1,T}, A_{2,1}, \ldots, A_{M,T} \) are constructed by using EEG feature vectors \( x_{i}^{F_{1,1}}, x_{i}^{F_{1,2}}, \ldots, x_{i}^{F_{1,T}}, x_{i}^{F_{2,1}}, \ldots, x_{i}^{F_{M,T}} \). Then we can perform the classifications based on EEG by inputting feature vectors extracted from test data into each of the trained classifiers. Finally, \( M \times T \) kinds of classification results are obtained.

2.2 Integration of Multiuser EEG-based Classifications

The second stage includes MSLPCCA-based FLF and KESLME-based DLF for the final decision. Specifically, \( M \times T \) classification results obtained in the first stage are integrated on the basis of KESLME, which is a supervised DLF method newly derived in this paper. This method is an extended version of our previous DLF methods which are also extended versions of those proposed in. Although an unsupervised learning method using the expectation-maximization (EM) algorithm has been introduced in these methods, our proposed DLF framework consists of supervised learning. This means that the complete data are known at the training phase. Thus, KESLME does not need the use of the EM algorithm, and more accurate maximum likelihood estimation becomes feasible. In this stage, we also utilize MSLPCCA to integrate multiuser EEG features for obtaining the input to KESLME.

### Table 1

| Type                        | Wave                  | Dims. |
|-----------------------------|-----------------------|-------|
| Zero crossing rate          | \( 0 \) (4–7Hz)       | \( C \) |
| Content percentage of the power spectrum | \( \text{slow-}\alpha \) (7–9Hz) | \( C \) |
| the mid-\( \alpha \) (9–11Hz) | \( C \) |
| fast-\( \alpha \) (11–13Hz) | \( C \) |
| \( \beta \) (13Hz–)        | \( C \) |
| Power spectrum of the hemispheric asymmetry | \( \text{slow-}\alpha \) (7–9Hz) | \( 2P \) |
| \( \text{mid-}\alpha \) (9–11Hz) | \( 2P \) |
| fast-\( \alpha \) (11–13Hz) | \( 2P \) |
| \( \beta \) (13Hz–)        | \( 2P \) |

(1) MSLPCCA-based FLF

The integrated feature vectors are generated by using MSLPCCA. First, we calculate the average and standard deviation of each EEG feature from the \( m \)th user’s \( T \) EEG frames’ features \( [x_{i}^{F_{m,t}}]_{t=1}^{T} \), and then a \((6C + 10P)\)-dimensional feature vector is obtained. For applying CCA-based methods to the above feature vectors, their dimension should be generally smaller than the number of training samples, i.e., \( N \). Therefore, we apply principal component analysis (PCA) to these vectors to obtain lower-dimensional vectors \( x_{i}^{m} \in \mathbb{R}^{d_{m}}, d_{m} < N \). Finally, we define a matrix \( X^{m} = [x_{1}^{m}, x_{2}^{m}, \ldots, x_{n_{1}}^{m}, x_{m}^{m}, x_{m+1}^{m}, \ldots, x_{n_{1}}^{m}] \) (\( m = 1, 2, \ldots, M \)) for each \( m \)th user, where \( n_{1} \) is the number of positive samples, and the number of negative samples is \( n_{0} = N - n_{1} \).

In MSLPCCA, the \((i, j)\)th element of the \( m \)th similarity matrix \( S_{m} \) is calculated and defined as

\[
S_{ij}^{m} = \begin{cases} 
\exp \left(-\frac{||x_{i}^{m} - x_{j}^{m}||^{2}}{2\sigma_{m}^{2}}\right) & \text{if } x_{i}^{m} \in \mathcal{K}_{x_{j}^{m}}^{m} \text{ or } x_{j}^{m} \in \mathcal{K}_{x_{i}^{m}}^{m} \\
0 & \text{otherwise}
\end{cases}
\]

where \( \mathcal{K}_{x_{i}^{m}}^{m} \) is a set of \( k_{m} \) neighbors of \( x_{i}^{m} \), and these neighbors are defined by the Euclidean distance. The number of
neighboring $k^m$ is determined for each $m$, and its calculation is shown as follows. In the proposed method, we perform the classification using a $k^m$-nearest neighbor method for the training samples based on a leave-one-out scheme and set $k^m$ to the value providing the best classification performance. Furthermore, in Eq. (1), label($x_i^m$) represents the class label of $x_i^m$. These two conditions using $K^m$ and label($\cdot$) introduce the concept of locality preserving and supervised learning, respectively. In addition, $\hat{d}_{\alpha} = \frac{1}{m^2} \sum_{i,j} \sum_{\alpha} \hat{d}_{ij}^m = \frac{1}{m^2} \| x_i^m - x_j^m \|^2$. Next, Laplacian matrices between the $m$th user and $m'$th user are calculated as $L_{mm'} = D_{mm'} - S_{mm'} \circ S_{mm'}$. Note that “$\circ$” represents the Hadamard product, and $D_{mm'} = \text{diag}(D_{1mm'}, D_{2mm'}, \ldots, D_{Nmm'})$ is a diagonal matrix with the $\alpha$th diagonal element $D_{\alpha mm'} = \sum_j S_{ij}^m S_{ij}^{mm'}$. The optimal weight matrices $U^m$ maximizing the correlation between $X^1, X^2, \ldots, X^M$ are calculated by using the Laplacian matrices as

$$
\begin{align*}
(U^1, U^2, \ldots, U^M) &= \arg\max_{U^m} \sum_{m=1}^M \sum_{\alpha=1}^N \sum_{a} U_{am}^m X_{\alpha m}^m L_{\alpha mm'} U_{a m'}, \\
&\quad \text{s.t.} \sum_{m=1}^M U_{am}^m X_{\alpha m}^m X_{\alpha m'}^m U_{a m'} = 1.
\end{align*}
$$

(2)

The weight matrices can be obtained by solving the generalized eigenvalue problem in the same manner as the method$^{12}$.

Then each weight matrix becomes $\hat{U}^m = [u_1^m, u_2^m, \ldots, u_N^m]$, where $u_i^m \in \mathbb{R}^D$ ($i = 1, 2, \ldots, N_r$; $N_r$ being the number of positive eigenvalues) is the $i$th eigenvector. By using the obtained weight matrices $\hat{U}^1, \hat{U}^2, \ldots, \hat{U}^M$, integrated feature vectors $x_i^{FLF}$ are calculated on the basis of the following equation:

$$
x_i^{FLF} = [x_i^1\uparrow^\top, x_i^2\uparrow^\top, x_i^3\uparrow^\top, \ldots, x_i^M\uparrow^\top, \hat{U}^M\uparrow^\top]^\top.
$$

(3)

In this way, MSLPCCA-based FLF considers the class labels and realizes feature integration with the local structure of each set of features being preserved.

(2) KESLME-based DLF

Let $y^m \in \{0, 1\}$ be the label assigned to the feature vector $x^m$ by a classifier $a \in \mathcal{A}$, where $\mathcal{A} = \{A_{m,l=1}^{M,L}, l=1\}$ is a set of $M \times T$ classifiers. Given the actual label $y \in \{0, 1\}$, i.e., ground truth, the classification sensitivity $P_{se}^a$ and specificity $P_{sp}^a$ of classifier $a$ are respectively defined as

$$
P_{se}^a := \text{Pr}[y^a = 1|y = 1],
$$

(4)

$$
P_{sp}^a := \text{Pr}[y^a = 0|y = 0],
$$

(5)

and they can be obtained by applying cross-validation to training samples. In our method, the final decision model, i.e., the discriminating function, is specifically written as follows:

$$
\text{Pr}[y = 1|w, \phi(x^{FLF})] = \sigma \left( w^\top \phi(x^{FLF}) \right),
$$

(6)

where $w$ is a weight, and $\sigma(z) = 1/(1 + e^{-z})$ is a logistic sigmoid function. $\phi(x^{FLF})$ is obtained by mapping the feature vector $x^{FLF}$ into a high-dimensional feature space. Given the $N$ feature vectors $x_i^{FLF}$ within the training data, the weight $w$ is specifically written as $w = \sum_i \alpha_i \phi(x_i^{FLF})$. With the coefficient vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]^\top$, the discriminating function in Eq. (6) is rewritten as follows:

$$
\text{Pr}[y = 1|\alpha, \phi(x^{FLF})] = \sigma \left( \sum_{i=1}^N \alpha_i K(x_i^{FLF}, x^{FLF}) \right),
$$

(7)

where $K(\cdot, \cdot)$ is a kernel function. In order to obtain the final decision model, we estimate the optimal values of the coefficients $\alpha_i$ in the following training phase.

Given the training data $D = \{\phi(x_i^{FLF}), y_i\}_{i=1}^N$ and $L = \{y_i\}_{i=1}^N$, the estimation target is the coefficients $\alpha_i$ in Eq. (7), where $y_i = \{y_i^a\}_{a=1}^A$ is a set of $M \times T$ classification results, and $L$ is a set of actual labels ($y_i$ being the actual label of the $i$th training data). From $D$, the likelihood of the coefficient vector $\alpha$ is defined as

$$
\text{Pr}[D|\alpha] = \prod_{i=1}^N \text{Pr}[y_i|\phi(x_i^{FLF}), \alpha].
$$

(8)

By using classifications $\{y_i\}_{i=1}^N$ and the actual label $L$, each classifier’s sensitivity and specificity are respectively obtained as

$$
P_{se}^a = \frac{1}{n} \sum_{i=1}^n y_i^a,\quad P_{sp}^a = \frac{1}{n} \sum_{i=1}^n (1 - y_i^a)(1 - y_i^0).\quad
$$

(9)

Hence, by using a set of sensitivity $P_{se} = \{P_{se}^a\}_{a=1}^A$ and specificity $P_{sp} = \{P_{sp}^a\}_{a=1}^A$, Eq. (8) is rewritten as follows:

$$
\text{Pr}[D|\alpha] = \prod_{i=1}^N \left[ \text{Pr}[y_i = 1, P_{se} | \phi(x_i^{FLF}), \alpha] \cdot \text{Pr}[y_i = 0, P_{sp} | \phi(x_i^{FLF}), \alpha]\right].
$$

(11)

Assuming that each classifier $a \in \mathcal{A}$ is independent of each other, $\text{Pr}[y_i = 1, P_{se} | \phi(x_i^{FLF}), \alpha]$ can be rewritten as

$$
\text{Pr}[y_i = 1, P_{se}] = \prod_{a \in \mathcal{A}} \left[P_{se}^a \right]^{y_i^a} \left[1 - P_{se}^a \right]^{y_i^0},
$$

(12)

and

$$
\text{Pr}[y_i = 0, P_{sp}] = \prod_{a \in \mathcal{A}} \left[P_{sp}^a \right]^{y_i^a} \left[1 - P_{sp}^a \right]^{y_i^0}.
$$

(13)

Therefore, the likelihood in Eq. (11) is rewritten as follows:
The optimal result in the above equation is obtained by ascent-based optimization methodologies. By equating the likelihood, the complete data log-likelihood is newly defined as follows:

$$\ln \Pr[D|\mathcal{L}|\alpha] = \sum_{i=1}^{N} [y_i \ln \rho_i + (1 - y_i) \ln (1 - \rho_i) \delta_i]$$

$$-\frac{\lambda}{2} \alpha^T K \alpha,$$  \hspace{1cm} (19)

where $\lambda \alpha^T K \alpha$ is a regularization term to avoid overfitting, $\lambda$ is a regularization parameter, and $K = [\kappa_1, \kappa_2, \ldots, \kappa_N]$ is a gram matrix. The maximum-likelihood estimator, i.e., the coefficient vector, is found by

$$\hat{\alpha} = \arg \max_{\alpha} [\ln \Pr[D, \mathcal{L}|\alpha]].$$  \hspace{1cm} (20)

The optimal result in the above equation is obtained by the Newton-Raphson method\textsuperscript{15}, one of the famous gradient ascent-based optimization methodologies. By equating the gradient of Eq. (19) to zero,

$$\frac{\partial}{\partial \alpha} [\ln \Pr[D, \mathcal{L}|\alpha]] = 0,$$  \hspace{1cm} (21)

we obtain the optimal result for $\alpha$. Especially, we update the coefficient vector $\alpha$ as follows:

$$\alpha \leftarrow \alpha - \eta H^{-1} g,$$  \hspace{1cm} (22)

where $g$ is the gradient, $H$ is the Hessian and $\eta$ is a step length. The gradient and the Hessian are respectively computed as follows:

$$g = \sum_{i=1}^{N} [y_i - \sigma(\alpha^T \kappa_i)] \kappa_i - \lambda K \alpha,$$  \hspace{1cm} (23)

$$H = -\sum_{i=1}^{N} [\sigma(\alpha^T \kappa_i)][1 - \sigma(\alpha^T \kappa_i)] \kappa_i \kappa_i^T - \lambda K.$$  \hspace{1cm} (24)

Given test data $D_{test} = [\phi(x_{test}^{FLF}), y_{test}]$, the final decision is obtained on the basis of the above trained model, where the set of mutliuser EEG-based classification results is represented as $Y_{test} = \{y_{test}\}_{a \in \mathcal{A}}$. The final decision is obtained by using the probabilistic label calculated from $D_{test}$. First, we respectively calculate

$$\rho_{test} = \sigma(\hat{\alpha}^T \kappa_{test}),$$  \hspace{1cm} (25)

$$\gamma_{test} = \prod_{a \in \mathcal{A}} [P_{test}^{a}|y_{test}|1 - P_{test}^{a}]^{-\gamma_{test}},$$  \hspace{1cm} (26)

and

$$\delta_{test} = \prod_{a \in \mathcal{A}} [P_{test}^{a}]^{-\gamma_{test}}[1 - P_{test}^{a}]^{\delta_{test}},$$  \hspace{1cm} (27)

where

$$\kappa_{test} = [K(x_1^{FLF}, x_{test}^{FLF}), K(x_2^{FLF}, x_{test}^{FLF}), \ldots, K(x_N^{FLF}, x_{test}^{FLF})]^T.$$  \hspace{1cm} (28)

By using $\rho_{test}$, $\gamma_{test}$ and $\delta_{test}$, the probabilistic label $\mu_{test}$ is computed as follows:

$$\mu_{test} = \Pr[y_{test} = 1|\rho_{test}, \phi(x_{test}^{FLF}), \hat{\alpha}]$$

$$\propto \Pr[y_{test}|y_{test} = 1, \hat{\alpha}] \cdot \Pr(\gamma_{test} = 1|\phi(x_{test}^{FLF}), \hat{\alpha})$$

$$= \frac{\gamma_{test} \rho_{test} + \delta_{test} (1 - \rho_{test})}{\gamma_{test} \rho_{test} + \delta_{test} (1 - \rho_{test})}.$$  \hspace{1cm} (29)

Then we obtain the final decision $\hat{y}$ by selecting the label whose probability becomes maximum. This corresponds to the one-versus-all methodology when performing multi-class classification. In this way, the proposed method provides the final decision with consideration of the accuracy of each classifier.

### 3. Experimental Results

We show experimental results to verify the effectiveness of the proposed method. First, we show experimental settings of the performed experiments in 3.1. Next, we show the results of the experiments and their discussions in 3.2. In this subsection, we also show the limitation of the proposed method and its usecase for our future work.

#### 3.1 Experimental Settings

In order to compare the proposed method with one of the state-of-the-art DCNNs, called Inception-v3\textsuperscript{34}, an image classification task was adopted as the experiment. We also adopted conventional hand-crafted feature-based image classification, where GIST\textsuperscript{34}, SIFT\textsuperscript{46}, PHOG\textsuperscript{46} and intensity histogram (IHIST) were used as hand-crafted features.

In this experiment, we utilized the Caltech-101 dataset\textsuperscript{46} and used images included in the categories “panda”, “soccer ball” and “strawberry” in the database for image classification. We randomly selected 35 images per category in advance. Thus, the number of images used for image classification was 105. These images were defined as target images. Images included in “airplane”, “elephant”, “joshua tree”, “pyramid” and “stapler” in the same dataset were used as non-target images.
In this experiment, eight subjects participated, i.e., $M = 8$, and EEG recordings were conducted while a subject was looking at images. The biological signals were recorded from 12 channels ($Fp1$, $Fp2$, $F7$, $F8$, $T3$, $T4$, $C3$, $C4$, $P3$, $P4$, $O1$ and $O2$; $C = 12$ and $P = \frac{12}{2}$) according to the international 10–20 electrode system shown in Fig. 2. Since EEG signals were weak, the signals were amplified by using an amplifier (MEG-6116M, NIHON KOHDEN). A band-pass filter was also applied to recorded EEG signals to avoid artifacts, and the bandwidth was set to 0.04–30Hz. The sampling rate of the EEG signals was 2000Hz. The segmentation interval and size of the overlapped Hamming window were 0.1 sec and 0.05 sec, respectively, and then $T$ became 19. Single-trial EEG signals were collected for each target image through the same experimental procedure as that shown in our previous works\(^{(35)}\)\(^{(36)}\).

During the experiment, each subject was instructed to sit comfortably on a chair and keep relaxing. Furthermore, each subject stared at images by a display with resolution of $1920 \times 1080$ without blinking, where we also let them reduce their eye movements as far as possible for avoiding artifacts. Each subject watched the images while sitting on a chair about one meter away from the display. For easily recognizing the objects included in viewed images, we previously let them know what kinds of categories were used for the experiments.

As explained above, we sequentially showed each subject two kinds of images, target images used for the image classification and non-target images not used for the image classification, where the target images correspond to “panda”, “soccer ball” and “strawberry”, and the non-target images correspond to “airplane”, “elephant”, “joshua tree”, “pyramid” and “stapler”. Furthermore, EEG signals were collected while subjects stared at the above images in a scheme similar to an oddball paradigm, and each subject was instructed to count the number of images shown as the target images. The number of times each subject performed the task was the same as the number of the target image categories, i.e., three. Each subject stared at each image in three seconds, and a black scene in two seconds was inserted as silence between two images for avoiding the effect of the previous images.

Since the target images each have one of the three category labels, the one-versus-all methodology\(^{(28)}\) was adopted. Our experimental settings followed\(^{(35)}\)\(^{(36)}\), i.e., 30 training images were randomly selected per category, i.e., $N = 90$ and $n_t = 30$, and test images were the remaining 15 ($= 5 \times 3$) images. In addition, we repeated the random selection ten times. Table 2 shows the average image classification accuracy of the above ten trials. In this table, the kernel functions of SVM and KESLME were Gaussian kernels. Moreover, the above kernel functions’ parameters, the regularization parameters in SVM and $\lambda$ in KESLME were determined to the values, which provided the best performance based on the grid search\(^{(37)}\) using the five-fold cross-validation. Specifically, the kernel function parameters were determined from $2^{-15}$, $2^{-13}$, $\cdot$, $2^1$, $2^3$, and the regularization parameters were determined from $2^{-5}$, $2^{-3}$, $\cdot$, $2^{13}$, $2^{15}$. In addition, $\lambda$ was determined from the range of $-21$, $-19$, $-17$, $\cdot$, $3$, $5$, $7$. The dimension $d_{\text{Fm}}^m$ after the mRMR algorithm, the dimension $d_{\text{smp}}^m$ after PCA, $P_{\text{sp}}^m$ and $P_{\text{sp}}^t$ were determined by applying two-fold cross-validation to the training dataset. The searching range of $d_{\text{Fm}}^m$ was 10, 20, $\cdot$, 90, 100, and that of $d_{\text{smp}}$ was 5, 10, 15, 20, 30, 35, 40, 70, 80, 89.

As described above, the determination of the parameters was performed in different manners. As shown in the previous section, we determined $K_0$ based on the leave-one-out cross-validation manner. On the other hand, the other parameters were determined based on the two-fold or five-fold cross-validation manners. Generally, it is suitable to determine all of the parameters based on the leave-one-out cross-validation manner, i.e., the same manner, for obtaining the best performance, but it was quite time-consuming. Therefore, in this paper, we determined several parameters based on the two-fold or five-fold cross-validation manners for reducing the computation complexity. Therefore, the performance of the proposed method can be improved by using the leave-one-out cross-validation manner, and the performance verification of our method was not unfair since we did not use advantageous conditions for our method. Note that all of the parameters were determined from only the training data.

### 3.2 Results and Discussions

From Table 2, although accuracy obtained by using each hand-crafted feature is not satisfactory, Inception-v3 pool3 feature-based SVM and our method realize successful clas-
sification based on the ImageNet-based pretrained effective network and the collaborative use of multiuser EEG features, respectively. In this table, we also show the results of the multimodal scheme-based method\textsuperscript{17} with KESLME, i.e., “our previous framework”\textsuperscript{17}+KESLME”. Our previous framework which uses both of the visual and EEG data obtained from each subject, i.e., multimodal data, also does not focus on the same task. Nevertheless, we have performed the comparison with our previous framework in such a way that the condition between the previous approach and our method becomes similar. The results of “single subject (M = 1)+KESLME”, which is an integration method of single subject EEG-based classifications without MSLPCCA is also shown in the table. The obtained results show that the proposed method realizes more accurate classification than does the single user-based method and our previous framework\textsuperscript{17}. Thus, the proposed method based on the multi-classifier scheme is one of the effective cues to performance improvement.

Table 2 Image classification accuracy: the second column shows the average accuracy and the third column shows the standard deviation between combinations with respect to subjects. The methods in the first four rows used hand-crafted features, and the method in the fifth row used DCNN-based features. In the sixth row, the above four hand-crafted features and single subject EEG features were used. In the seventh row, KESLME-based DLF was applied to a subject’s T classifications and the feature obtained without using MSLPCCA. The results of our method are the others (rows 9-15). Our method used only EEG features of multiple subjects.

| Method | Acc. | Std. |
|--------|------|------|
| GIST\textsuperscript{3}+SVM\textsuperscript{18} | 0.6733 | — |
| SIFT\textsuperscript{4}+SVM\textsuperscript{18} | 0.6800 | — |
| PHOG\textsuperscript{5}+SVM\textsuperscript{18} | 0.6733 | — |
| LHST\textsuperscript{6}+SVM\textsuperscript{18} | 0.7067 | — |
| Inception-v3 (pool3)\textsuperscript{8} | 0.9867 | — |
| Our previous framework\textsuperscript{17}+KESLME | 0.9117 | 0.0345 |
| Single subject (M = 1)+KESLME | 0.7517 | 0.1023 |
| Multi subjects+MSLPCCA\textsuperscript{17}+KESLME | — | — |

| M = 2 (sC2 = 28) | 0.8690 | 0.0704 |
| M = 3 (sC3 = 56) | 0.9352 | 0.0399 |
| M = 4 (sC4 = 70) | 0.9730 | 0.0177 |
| M = 5 (sC5 = 94) | 0.9871 | 0.0095 |
| M = 6 (sC6 = 28) | 0.9957 | 0.0052 |
| M = 7 (sC7 = 8 ) | 0.9992 | 0.0024 |
| M = 8 (sC8 = 1 ) | 1.0000 | — |

Next, we discuss the limitations and future work of our study. Although the proposed method enables successful integration of the classification results obtained from multiple EEG-based classifiers, it does not use any visual information. Visual features obtained through middle layers in recent deep neural network frameworks such as AlexNet\textsuperscript{38}, VGG16-19\textsuperscript{40} and Inception-v4\textsuperscript{41} are useful for the classification even if target categories are not the same as those of generic objects. This means that the integration of not only EEG-based classification results but also those obtained from the visual features of the pre-trained deep neural network framework may improve the performance of the proposed method.

Finally, we show the usecase of the proposed method. In data collecting stages of new target classification problems, there are quite few training samples. This becomes significant when performing the classification for newly defined labels especially in expert fields such as the fields of medical and social infrastructures. For example, in the medical field, although there exist many attempts for preparing image databases of medical images, it is difficult to collect sufficient number of training images for rare cases. Furthermore, in the field of social infrastructures, classification of distress images according to their kinds and severities is desired. However, since their patterns are various, the number of training data available for each class becomes quite small. Then even though there have been proposed many useful techniques such as transfer learning\textsuperscript{42,43}, the number of the training samples with high data diversity is still too small to achieve their successful training.

In the above expert fields, multiple experts tend to view the same images for checking them, where their discriminant abilities are different from each other. Therefore, we collect their EEG data obtained when viewing such images and aim at performing unconscious crowd sourcing without obtaining their direct decisions. If we collect decisions from each expert with their EEG data, it forces large efforts and burdens for them. Thus, we aim at performing the classifi-
cation from only their EEG data in both of the training and test phases. It should be noted that since the ability of each expert is different as described above, the classification performances from their EEG data also become different. In our method, we newly derive the new DLF for integrating the classification results with considering the difference of their classification performances. This is the main contribution of our paper.

The usefulness of the proposed method becomes significant when it is applied to expert fields. Therefore, its verification using image data obtained from such fields is necessary. This will be also addressed in our future work.

4. Conclusion

In this paper, we have proposed a novel integration method based on KESLME of multiple EEG features to obtain a multi-classifier decision. KESLME is newly derived and provides the final decision with consideration of the relationship between the MSLPCCA-based integrated EEG features and the performance of each classifier. In an image classification task, we confirmed that the performance of the proposed method is comparable to that of one of the state-of-the-art DCNNs.

5. ACKNOWLEDGEMENT

This work was partly supported by the MIC/SCOPE #181601001 and JSPS KAKENHI Grant Number JP17H01744.

References

1) A. S. Razavi et al., “CNN features off-the-shelf: An astounding baseline for recognition,” in Proc. the IEEE Conf. Computer Vision and Pattern Recognition Workshops, 2014, pp. 806–813.
2) A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
3) A. Oliva and A. Torralba, “Modeling the shape of the scene: A holistic representation of the spatial envelope,” Int. Journal of Computer Vision, vol. 42, no. 3, pp. 145–175, 2001.
4) D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
5) A. Bosch, A. Zisserman, and X. Munoz, “Representing shape with a spatial pyramid kernel,” in Proc. ACM Int. Conf. Image and Video Retrieval, 2007, pp. 401–408.
6) K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
7) C. Szegedy et al., “Going deeper with convolutions,” in Proc. the IEEE Conf. Computer Vision and Pattern Recognition, 2015, pp. 1–9.
8) C. Szegedy et al., “Rethinking the inception architecture for computer vision,” in Proc. the IEEE Conf. Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
9) L. Shao, F. Zhu, and X. Li, “Transfer learning for visual categorization: A survey,” IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 5, pp. 1019–1034, May 2015.
10) V. C. Raykar et al., “Learning from crowds,” Journal of Machine Learning Research, vol. 11, no. 7, pp. 1297–1322, 2010.
11) P. G. Ipeirotis, F. Provost, and J. Wang, “Quality management on Amazon Mechanical Turk,” in Proc. ACM SIGKDD Workshop on Human Computation, 2010, pp. 64–67.
12) D. R. Karger, S. Oh, and D. Shah, “Iterative learning for reliable crowdsourcing systems,” in Advances in Neural Information Processing Systems, 2011, pp. 1953–1961.
13) G. Yang, Y. Lin, and P. Bhattacharya, “A driver fatigue recognition model using fusion of multiple features,” in Proc. IEEE Int. Conf. Systems, Man and Cybernetics, vol. 2, 2005, pp. 1777–1784.
14) Q. Zhang and M. Lee, “A hierarchical positive and negative emotion understanding system based on integrated analysis of visual and brain signals,” Neurocomputing, vol. 73, no. 16, pp. 3264–3272, 2010.
15) P. Smyth et al., “Inferring ground truth from subjective labelling of videos images,” in Proc. Advances in Neural Information Processing Systems, 1995, pp. 1085–1092.
16) M. Soleymani, M. Pantic, and T. Pun, “Multimodal emotion recognition in response to videos,” IEEE Trans. on Affective Computing, vol. 3, no. 2, pp. 211–223, 2012.
17) T. Kawakami, T. Ogawa, and M. Haseyama, “Novel image classification based on integration of EEG and visual features via MSLPCCA,” in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, 2015, pp. 957–961.
18) C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, pp. 273–297, 1995.
19) J. P. Demschowski et al., “Audience preferences are predicted by temporal relia-

Paper » Multi-classifier Decision: Integration of Multiple Brain Activity-based Classifications
42) V. Behbood et al., “Transfer learning using computational intelligence: A survey,” Knowledge-Based Systems, vol. 80, pp. 14-23, 2015.
43) K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” Journal of Big Data, vol. 3, no. 1, pp. 1-40, 2016.

Takahiro Ogawa received his B.S., M.S. and Ph.D. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2003, 2005 and 2007, respectively. He is currently an assistant professor in the Graduate School of Information Science and Technology, Hokkaido University. His research interests are multimedia signal processing and its applications. He has been an Associate Editor of ITE Transactions on Media Technology and Applications. He is a member of the IEEE, EURASIP, IEICE, and Institute of Image Information and Television Engineers (ITE).

Kento Sugata received his B.S. and M.S. degrees in Electronics and Information Engineering from Hokkaido University, Sapporo, Japan in 2016 and 2018, respectively. He is a member of the IEEE.

Ren Togo received his B.S. in Health Sciences from Hokkaido University, Japan in 2015 and M.S. degrees in the Graduate School of Information Science and Technology, Hokkaido University, Japan in 2017. He is currently pursuing a Ph.D. degree at the Graduate School of Information Science and Technology, Hokkaido University. His research interests are medical image processing and its applications. He is a member of the IEEE.

Miki Haseyama received her B.S., M.S. and Ph.D. degrees in Electronics from Hokkaido University, Japan in 1986, 1988 and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University as an associate professor in 1994. She was a visiting associate professor of Washington University, USA from 1995 to 1996. She is currently a professor in the Graduate School of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development into semantic analysis. She has been a Vice-President of the Institute of Image Information and Television Engineers, Japan (ITE), an Editor-in-Chief of ITE Transactions on Media Technology and Applications, a Director, International Coordination and Publicity of The Institute of Electronics, Information and Communication Engineers (IEICE). She is a member of the IEEE, IEICE, Institute of Image Information and Television Engineers (ITE) and Acoustical Society of Japan (ASJ).