NOVEL IMAGE CLASSIFICATION BASED ON INTEGRATION OF EEG AND VISUAL FEATURES VIA MSLPCCA

*Takuya Kawakami, Takahiro Ogawa and Miki Haseyama**

Graduate School of Information Science and Technology, Hokkaido University
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
E-mail: {kawakami, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

ABSTRACT

This paper presents a novel image classification method based on integration of EEG and visual features. In the proposed method, we obtain classification results by separately using EEG and visual features. Furthermore, we merge the above classification results based on a kernelized version of Supervised learning from multiple experts and obtain the final classification result. In order to generate feature vectors used for the final image classification, we apply Multiset supervised locality preserving canonical correlation analysis (MSLPCCA), which is newly derived in the proposed method, to EEG and visual features. Our method realizes successful multimodal classification of images by the object categories that they contain based on MSLPCCA-based feature integration.

***Index Terms*—** electroencephalogram (EEG), image classification, multimodal scheme, decision-level fusion, canonical correlation analysis.

1. INTRODUCTION

Image classification is an important task for image semantic analysis. Thus, various methods which classify images according to object categories that these images contain have intensively been proposed [1–3]. In order to classify images automatically, these methods utilize visual features extracted from each image. Although the classification accuracy has been improved by using several visual features [4–7], the improvement of the classification performance based on the discovery of the new visual features tends to be saturated. Therefore, it is necessary to introduce a new idea such as solving the problem by using alternative features.

In order to realize image classification based on this approach, we have proposed an image classification method [8] which utilizes both EEG features extracted from EEG signals recorded while a user stares at the images and their visual features. However, this method utilizes feature vectors generated without considering relationships between EEG features and visual features. Therefore, the performance improvement is expected by using vectors based on feature integration considering their relationships.

In this paper, we propose a novel image classification method based on integration of EEG and visual features. The proposed method consists of two stages. In the first stage, our method calculates EEG features and visual features, and classifies images based on Support vector machine (SVM) [9] by using each feature independently. Then, in the second stage, our method performs merging the above classification results, *i.e.,* decision-level fusion. In order to perform the decision-level fusion, we employ a kernelized version of Supervised learning from multiple experts (KSLME) [8]. In this method, we utilize feature vectors generated by integration of EEG and visual features based on a new approach, *i.e.,* the biggest contribution of this paper. Specifically, in order to integrate EEG and visual features with considering relationships between them, we newly derive Multiset supervised locality preserving canonical correlation analysis (MSLPCCA). MSLPCCA enables to apply Supervised locality preserving canonical correlation analysis [10] to more than three variables. Since the conventional method [8] utilizes EEG features and three types of visual features, it is necessary to derive MSLPCCA in order to integrate at least four types of features. Then MSLPCCA enables the feature integration with preserving the locality structure of each variable and using class labels which are generally effective for the classification problem. Consequently, successful image classification based on EEG and visual features becomes feasible by our method.

2. IMAGE CLASSIFICATION VIA MSLPCCA-BASED FEATURE INTEGRATION

The proposed method consists of two stages. In the first stage, we extract the EEG features from EEG signals recorded while a user stares at images, and the visual features are computed from these images. Then we perform image classification based on SVM [9] by inputting EEG and visual features into the classifiers, separately. Thus, multiple classification results are obtained for test data based on each feature. Furthermore, in the second stage, we employ the kernelized decision-level fusion approach, *i.e.,* merging the above classification results, considering their classification accuracy. In this stage, we utilize feature vectors generated by integrating EEG and visual features based on MSLPCCA which is our original method. The details of our method are described below.

2.1. Feature Extraction and Single Feature-based Image Classification

In this subsection, we explain the first stage of our method. Specifically, we explain the EEG features and the visual features used in the proposed method, and the single feature-based image classification method is shown.

EEG Feature Extraction

First, segmentation of each channel's EEG signal is performed at fixed intervals with an overlapped Hamming window. In this paper, f_t ($t = 1, 2, \cdots, F$; F is the total number of EEG segments) denote EEG segments. Next, we compute the EEG features shown in Table 1 from each EEG segment. Note that C and P denote the number of channels of EEG signals and the number of symmetric electrode pairs placed on the scalp, respectively. Thus, the dimension of EEG features becomes $6C + 10P$. In this table, we calculate

*This work was partly supported by Grant-in-Aid for Scientific Research (B) 25280036 from JSPS.

Table 1. Features used for EEG signals in our method. Note that C denotes the number of channels of EEG signals and P shows the number of symmetric electrode pairs placed on the scalp.

The Type of EEG Features		Num. of Dimension
Zero Crossing Rate		C
Content percentage of the power spectrum	θ wave (4-7Hz)	C
	slow-α wave (7-9Hz)	C
	mid-α wave (9-11Hz)	C
	slow-α wave (11-13Hz)	C
	β wave (13Hz-)	C
Power spectrum of the hemispheric asymmetry [13]	θ wave (4-7Hz)	$2P$
	slow-α wave (7-9Hz)	$2P$
	mid-α wave (9-11Hz)	$2P$
	slow-α wave (11-13Hz)	$2P$
	β wave (13Hz-)	$2P$

Zero crossing rate [11] in the time domain, and the other features are computed in the frequency domain by applying short-time Fourier transform (STFT) to each channel's EEG signal. The details of EEG features in our method are shown in [12].

Visual Feature Extraction

We utilize four kinds of visual features: Scale invariant feature transform (SIFT) [4], Pyramid histogram of oriented gradients (PHOG) [6], GIST descriptor [7] and the Intensity histogram (IHIST). We calculate feature vectors: $\boldsymbol{x}^{\mathrm{V_S}}$ ($\in \mathbb{R}^{300}$), $\boldsymbol{x}^{\mathrm{V_P}}$ ($\in \mathbb{R}^{3400}$), $\boldsymbol{x}^{\mathrm{V_G}}$ ($\in \mathbb{R}^{512}$) and $\boldsymbol{x}^{\mathrm{V_I}}$ ($\in \mathbb{R}^{256}$) based on each method. Note that we obtain $\boldsymbol{x}^{\mathrm{V_S}}$ by applying BoF approach [14] to extracted 128-dimensional SIFT descriptors. Due to the limitation of pages, we only show the overview of the visual features. The details of each visual feature are shown in [4, 6, 7].

Single Feature-based Image Classification

We explain the method to classify images based on each feature in the first stage. First, since relationships between "stimuli to human beings from the outside" and "which parts of the human brain are affected by these stimuli" are not well-known, we employ the feature selection in order to obtain EEG feature vectors. This means we reduce the dimension of the features shown in Table 1 to select only features useful for the classification. In order to perform the dimensionality reduction, we apply Max-relevance and min-redundancy (mRMR) feature selection algorithm [15] to the EEG features calculated from each segment and obtain an efficient feature set for the image classification. After this procedure, $\boldsymbol{x}_i^{f_t} \in \mathbb{R}^{d^{f_t}}$ ($i = 1, 2, \cdots, N$; N is the number of images included in training data; d^{f_t} is the number of the selected features based on mRMR algorithm for EEG segment f_t) are obtained as EEG feature vectors for each EEG segment f_t ($t = 1, 2, \cdots, F$). As for visual feature vectors, we directly use the vectors $\boldsymbol{x}_i^{\mathrm{V_S}}$, $\boldsymbol{x}_i^{\mathrm{V_P}}$, $\boldsymbol{x}_i^{\mathrm{V_G}}$ and $\boldsymbol{x}_i^{\mathrm{V_I}}$, separately.

In the first stage of our method, we employ SVM as the classifier. Although SVM is a two class classifier, image classification is generally a multi-class problem. Fortunately, since the two class classification can be easily expanded into multi-class classification based on one vs. one approach [16] or one vs. all approach [17], we focus on the improvement of the two class classification performance. We train classifiers by separately using EEG feature vectors calculated from each EEG segment and visual feature vectors. This means multiple classifiers ($F+4$ classifiers) are respectively obtained based on EEG features $\boldsymbol{x}_i^{f_1}, \boldsymbol{x}_i^{f_2}, \cdots, \boldsymbol{x}_i^{f_F}$ and visual features $\boldsymbol{x}_i^{\mathrm{V_S}}, \boldsymbol{x}_i^{\mathrm{V_P}}, \boldsymbol{x}_i^{\mathrm{V_G}}, \boldsymbol{x}_i^{\mathrm{V_I}}$ by using each feature vector for training. Therefore, we can classify images based on EEG and visual features by inputting feature vectors extracted from test data into each trained classifier. Finally, $F + 4$ kinds of classification results are obtained.

2.2. Multiple Feature-based Image Classification

In this subsection, we explain the method to obtain the final classification result in the second stage. In the proposed method, we merge $F + 4$ classification results obtained in the first stage based on KSLME proposed in [8] to determine the final classification result. In the second stage of our method, we newly derive the feature integration method called "MSLPCCA". We utilize this method to generate feature vectors from EEG and visual features and input them into the KLSME. In this subsection, we merge $1, 2, \cdots, F$ EEG segments' classification results and 4 classification results based on visual features. In the proposed method, we regard the $F + 4$ classifiers based on EEG features extracted from each EEG segment and visual features as $F+4$ *annotators*. In order to merge multiple classification results, we focus on the classification accuracy of each annotator and assign higher weights to classification results of annotators which have higher classification accuracy. The details of the second stage are shown below.

2.2.1. MSLPCCA-based Feature Integration

We explain the feature integration method called "MSLPCCA". We generate new integrated feature vectors by applying MSLPCCA to EEG features and visual features. First, we define 5 variables $\boldsymbol{X}^1, \boldsymbol{X}^2, \cdots, \boldsymbol{X}^5$ as $\boldsymbol{X}^r = [\boldsymbol{x}_1^{r(1)}, \boldsymbol{x}_2^{r(1)}, \cdots, \boldsymbol{x}_{n_1}^{r(1)}, \boldsymbol{x}_1^{r(0)}, \boldsymbol{x}_2^{r(0)}, \cdots, \boldsymbol{x}_{n_0}^{r(0)}]$, where $\boldsymbol{x}_i^{r(\cdot)} \in \mathbb{R}^{d^r}$, $n_1 + n_0 = N$, and $\boldsymbol{x}_i^{r(1)}$ and $\boldsymbol{x}_i^{r(0)}$ are respectively feature vectors assigned to positive and negative classes. In the proposed method, 5 variables correspond to $\boldsymbol{X}^{\mathrm{E}}, \boldsymbol{X}^{\mathrm{V_S}}, \boldsymbol{X}^{\mathrm{V_P}}, \boldsymbol{X}^{\mathrm{V_G}}$ and $\boldsymbol{X}^{\mathrm{V_I}}$, respectively. In particular, $\boldsymbol{x}_i^{\mathrm{E}(\cdot)}$ in a variable $\boldsymbol{X}^{\mathrm{E}}$ are generated by $1, 2, \cdots, F$ EEG segments' features. We obtain these feature vectors by calculating the average and standard deviation of each EEG feature from $1, 2, \cdots, F$ EEG segments.

Next, we calculate similarity matrices $\boldsymbol{S}^{X^1}, \boldsymbol{S}^{X^2}, \cdots, \boldsymbol{S}^{X^5}$ in the same way as Locality preserving projection [18]. The (i, j) th component of each matrix is obtained as follows:

$$S_{ij}^{X^r} = \begin{cases} \mathrm{e}^{-\|\boldsymbol{x}_i^{r(\cdot)} - \boldsymbol{x}_j^{r(\cdot)}\|^2 / \sigma_{x^r}^2} & \text{if } \boldsymbol{x}_j^{r(\cdot)} \in \boldsymbol{\Omega}^k_{\boldsymbol{x}_i^{r(\cdot)}} \text{ or } \boldsymbol{x}_i^{r(\cdot)} \in \boldsymbol{\Omega}^k_{\boldsymbol{x}_j^{r(\cdot)}} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In Eq. (1), $\boldsymbol{\Omega}^k_{\boldsymbol{x}_i^{r(\cdot)}}$ is a set of k neighbors of $\boldsymbol{x}_i^{r(\cdot)}$, and these neighbors are defined by the Euclidean distance. In addition, $\sigma_{x^r}^2$ is calculated by $\sigma_{x^r}^2 = \frac{\xi^r}{N(N-1)}$, where ξ^r is the sum of the squared distances between all two vectors in $\boldsymbol{X}^r$. Furthermore, in the proposed method, if two vectors corresponding to (i, j) th component of $\boldsymbol{S}^{X^r}$ have different class labels, we replace the value of $S_{ij}^{X^r}$ as zero. This procedure enables to introduce supervised learning. After this procedure, we define similarity matrices as $\tilde{\boldsymbol{S}}^{X^r}$.

We calculate weight matrices $\boldsymbol{U}^1, \boldsymbol{U}^2, \cdots, \boldsymbol{U}^5$, which maximize the following correlation between variables $\boldsymbol{X}^1, \boldsymbol{X}^2, \cdots, \boldsymbol{X}^5$, by using similarity matrices as follows:

$$\begin{aligned} &\text{Maximize} \sum_{r=1}^{5} \sum_{s=1}^{5} \boldsymbol{U}^{r\top} \boldsymbol{X}^r \boldsymbol{L}^{rs} \boldsymbol{X}^{s\top} \boldsymbol{U}^s, \\ &\text{subject to} \sum_{r=1}^{5} \boldsymbol{U}^{r\top} \boldsymbol{X}^r \boldsymbol{L}^{rr} \boldsymbol{X}^{r\top} \boldsymbol{U}^r = 1. \end{aligned} \tag{2}$$

Weight matrices can be obtained by solving the generalized eigenvalue problem. Thus, each weight matrix is $\hat{\boldsymbol{U}}^r = [\boldsymbol{u}_1^r, \boldsymbol{u}_2^r, \cdots, \boldsymbol{u}_{N_e}^r]$, where $\boldsymbol{u}_l^r \in \mathbb{R}^{d^r}$ is the eigenvector, and N_e is the number of positive eigenvalues. In Eq. (2), $\boldsymbol{L}^{rr} = \boldsymbol{D}^{rr} - \tilde{\boldsymbol{S}}_{X^r} \circ \tilde{\boldsymbol{S}}_{X^r}$ and $\boldsymbol{L}^{rs} = \boldsymbol{D}^{rs} - \tilde{\boldsymbol{S}}_{X^r} \circ \tilde{\boldsymbol{S}}_{X^s}$, respectively. Note that "$\circ$" denotes the Hadamard product, and $\boldsymbol{D}^{rr} = \mathrm{diag}(\sum_i (\tilde{S}_{1i}^{X^r})^2, \sum_i (\tilde{S}_{2i}^{X^r})^2, \cdots, \sum_i (\tilde{S}_{Ni}^{X^r})^2)$ and

$\boldsymbol{D}^{rs} = \mathrm{diag}(\sum_i \tilde{S}_{1i}^{X^r}\tilde{S}_{1i}^{X^s}, \sum_i \tilde{S}_{2i}^{X^r}\tilde{S}_{2i}^{X^s}, \cdots, \sum_i \tilde{S}_{Ni}^{X^r}\tilde{S}_{Ni}^{X^s})$, respectively.

By using generated weight matrices $\hat{\boldsymbol{U}}^1, \hat{\boldsymbol{U}}^2, \cdots, \hat{\boldsymbol{U}}^5$, we obtain feature vectors based on integration of EEG and visual features as follows:

$$\boldsymbol{x}_i^{\mathrm{EV}} = \left[\boldsymbol{x}_i^{1(\cdot)\top}\hat{\boldsymbol{U}}^1, \boldsymbol{x}_i^{2(\cdot)\top}\hat{\boldsymbol{U}}^2, \cdots, \boldsymbol{x}_i^{5(\cdot)\top}\hat{\boldsymbol{U}}^5\right]^\top. \tag{3}$$

MSLPCCA considers the class label of each vector and realizes feature integration with preserving a locality structure of each variable in new feature space. Therefore, we generate feature vectors, which are effective for image classification using both EEG and visual features, in Eq. (3) based on MSLPCCA.

2.2.2. Each annotator's classification accuracy and classification model

We explain the classification accuracy of each annotator and the classification model defined in our method. Let $y^a \in \{0, 1\}$ be the label assigned to the feature vector $\boldsymbol{x}^{\mathrm{EV}}$ by annotator $a \in \mathcal{A}$, where $\mathcal{A} = \{f_1, f_2, \cdots, f_F, \mathrm{V_S}, \mathrm{V_P}, \mathrm{V_G}, \mathrm{V_I}\}$ is a set of annotators, and f and V correspond to Frame (EEG segment) and Visual, respectively. Given the actual label $y \in \{0, 1\}$, *i.e.,* ground truth, the classification accuracy of each annotator, P_{se}^a (sensitivity) and P_{sp}^a (specificity) are respectively defined as follows:

$$P_{se}^a := \Pr[y^a = 1 | y = 1], \tag{4}$$

$$P_{sp}^a := \Pr[y^a = 0 | y = 0]. \tag{5}$$

In our method, a classification model is specifically written as:

$$f_w(\boldsymbol{x}^{\mathrm{EV}}) = \boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}^{\mathrm{EV}}), \tag{6}$$

where $\boldsymbol{w}$ is a weight.

In Eq. (6), $\boldsymbol{\phi}(\boldsymbol{x}^{\mathrm{EV}})$ is obtained by mapping the feature vector $\boldsymbol{x}^{\mathrm{EV}}$ into a high-dimensional feature space. The final classification result $\hat{y}$ is obtained as follows:

$$\hat{y} = \begin{cases} 1 & f_w(\boldsymbol{x}^{\mathrm{EV}}) \geq \tau \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where τ is a predetermined threshold. Given the training dataset $\mathcal{D}$ consisting of N feature vectors $\boldsymbol{x}_i^{\mathrm{EV}}$ $(i = 1, 2, \cdots, N)$, a weight $\boldsymbol{w}$ is specifically written as follows:

$$\boldsymbol{w} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\phi}(\boldsymbol{x}_i^{\mathrm{EV}}), \tag{8}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_N]^\top$. Therefore, by using $\boldsymbol{\alpha}$ in Eq. (8), the discriminating function in Eq. (6) is rewritten as follows:

$$\begin{aligned} f_w(\boldsymbol{x}^{\mathrm{EV}}) &= \boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}^{\mathrm{EV}}) \\ &= \sum_{i=1}^{N} \alpha_i K\left(\boldsymbol{x}_i^{\mathrm{EV}}, \boldsymbol{x}^{\mathrm{EV}}\right), \end{aligned} \tag{9}$$

where $K(\cdot, \cdot)$ is a kernel function of $\boldsymbol{\phi}(\cdot)$, and we specifically employ the Gaussian kernel. In order to determine the discriminating function $f_w(\cdot)$, we have to obtain the coefficients α_i $(i = 1, 2, \cdots, N)$ from training data by using each annotator's classification accuracy defined in Eqs. (4) and (5). The details are shown below.

2.2.3. Training Phase

Given the training data $\mathcal{D}$ consisting of N feature vectors with the classification results by $F + 4$ annotators and their actual labels, $\mathcal{D} = \{y_i, \boldsymbol{\phi}(\boldsymbol{x}_i^{\mathrm{EV}}), \boldsymbol{\mathcal{Y}}_i\}_{i=1}^N$, where y_i is the actual label and $\boldsymbol{\mathcal{Y}}_i = \{y_i^{f_1}, y_i^{f_2}, \cdots, y_i^{f_F}, y_i^{\mathrm{V_S}}, y_i^{\mathrm{V_P}}, y_i^{\mathrm{V_G}}, y_i^{\mathrm{V_I}}\}$ is a set of classification results, the estimation target is the coefficients α_i $(i = 1, 2, \cdots, N)$ in Eq. (9). From the training data $\mathcal{D}$, the likelihood of the coefficient vector $\boldsymbol{\alpha}$ is defined as:

$$\Pr[\mathcal{D}|\boldsymbol{\alpha}] = \prod_{i=1}^{N} \Pr\left[\boldsymbol{\mathcal{Y}}_i | \boldsymbol{\phi}(\boldsymbol{x}_i^{\mathrm{EV}}), \boldsymbol{\alpha}\right]. \tag{10}$$

By using a set of sensitivity $\mathcal{P}_{se} = \{P_{se}^{f_1}, P_{se}^{f_2}, \cdots, P_{se}^{f_F}, P_{se}^{\mathrm{V_S}}, P_{se}^{\mathrm{V_P}}, P_{se}^{\mathrm{V_G}}, P_{se}^{\mathrm{V_I}}\}$ obtained from each annotator and that of specificity $\mathcal{P}_{sp} = \{P_{sp}^{f_1}, P_{sp}^{f_2}, \cdots, P_{sp}^{f_F}, P_{sp}^{\mathrm{V_S}}, P_{sp}^{\mathrm{V_P}}, P_{sp}^{\mathrm{V_G}}, P_{sp}^{\mathrm{V_I}}\}$, the above equation is rewritten as follows:

$$\begin{aligned} \Pr[\mathcal{D}|\boldsymbol{\alpha}] = \prod_{i=1}^{N} \Big\{ &\Pr[\boldsymbol{\mathcal{Y}}_i | y_i = 1, \mathcal{P}_{se}] \times \Pr[y_i = 1 | \boldsymbol{\phi}(\boldsymbol{x}_i^{\mathrm{EV}}), \boldsymbol{\alpha}] \\ &+ \Pr[\boldsymbol{\mathcal{Y}}_i | y_i = 0, \mathcal{P}_{sp}] \times \Pr[y_i = 0 | \boldsymbol{\phi}(\boldsymbol{x}_i^{\mathrm{EV}}), \boldsymbol{\alpha}] \Big\}. \end{aligned} \tag{11}$$

If it is assumed that each annotator $a \in \mathcal{A}$ is independent each other, $\Pr[\boldsymbol{\mathcal{Y}}_i | y_i = 1, \mathcal{P}_{se}]$ can be rewritten as follows:

$$\begin{aligned} \Pr[\boldsymbol{\mathcal{Y}}_i | y_i = 1, \mathcal{P}_{se}] &= \prod_{a \in \mathcal{A}} \Pr[y_i^a | y_i = 1, P_{se}^a] \\ &= \prod_{a \in \mathcal{A}} [P_{se}^a]^{y_i^a} [1 - P_{se}^a]^{1 - y_i^a}. \end{aligned} \tag{12}$$

Similarly, $\Pr[\boldsymbol{\mathcal{Y}}_i | y_i = 0, \mathcal{P}_{sp}]$ can be rewritten as follows:

$$\begin{aligned} \Pr[\boldsymbol{\mathcal{Y}}_i | y_i = 0, \mathcal{P}_{sp}] &= \prod_{a \in \mathcal{A}} \Pr[y_i^a | y_i = 0, P_{sp}^a] \\ &= \prod_{a \in \mathcal{A}} [P_{sp}^a]^{1 - y_i^a} [1 - P_{sp}^a]^{y_i^a}. \end{aligned} \tag{13}$$

Then the likelihood in Eq. (11) is rewritten as

$$\Pr[\mathcal{D}|\boldsymbol{\alpha}] = \prod_{i=1}^{N} [\gamma_i \rho_i + \delta_i (1 - \rho_i)], \tag{14}$$

where

$$\gamma_i = \prod_{a \in \mathcal{A}} [P_{se}^a]^{y_i^a} [1 - P_{se}^a]^{1 - y_i^a}, \tag{15}$$

$$\delta_i = \prod_{a \in \mathcal{A}} [P_{sp}^a]^{1 - y_i^a} [1 - P_{sp}^a]^{y_i^a}, \tag{16}$$

and

$$\begin{aligned} \rho_i &= \Pr[y_i = 1 | \boldsymbol{\phi}(\boldsymbol{x}_i^{\mathrm{EV}}), \boldsymbol{\alpha}] \\ &= \frac{1}{1 + \exp(-\boldsymbol{\alpha}^\top \boldsymbol{\kappa}_i)}, \end{aligned} \tag{17}$$

where $\boldsymbol{\kappa}_i = \left[K(\boldsymbol{x}_1^{\mathrm{EV}}, \boldsymbol{x}_i^{\mathrm{EV}}), K(\boldsymbol{x}_2^{\mathrm{EV}}, \boldsymbol{x}_i^{\mathrm{EV}}), \cdots, K(\boldsymbol{x}_N^{\mathrm{EV}}, \boldsymbol{x}_i^{\mathrm{EV}})\right]^\top$. The maximum-likelihood estimator is found by maximizing the following log-likelihood:

$$\hat{\boldsymbol{\alpha}}_{ML} = \arg\max_{\boldsymbol{\alpha}} \left\{ \ln \Pr[\mathcal{D}|\boldsymbol{\alpha}] \right\}. \tag{18}$$

Let $\mathcal{L} = \{y_1, y_2, \cdots, y_N\}$ be the set of the actual labels, and the complete data log-likelihood can be written as

$$\ln \Pr[\mathcal{D}, \mathcal{L}|\boldsymbol{\alpha}] = \sum_{i=1}^{N} \{ y_i \ln \rho_i \gamma_i + (1 - y_i) \ln (1 - \rho_i) \delta_i \}. \tag{19}$$

In order to maximize this likelihood, the following Expectation-Maximization (EM) algorithm [19] is adopted.

E-step

In the E-step, when the training data $\mathcal{D}$ and the current estimate of the coefficient vector $\boldsymbol{\alpha}$ are given, the conditional expected value of log-likelihood is computed as follows:

$$\mathrm{E}\left\{ \ln \Pr[\mathcal{D}, \mathcal{L}|\boldsymbol{\alpha}] \right\} = \sum_{i=1}^{N} \left\{ \mu_i \ln \rho_i \gamma_i + (1 - \mu_i) \ln (1 - \rho_i) \delta_i \right\}, \tag{20}$$

where μ_i is computed as follows:

$$\begin{aligned} \mu_i &\propto \Pr[\boldsymbol{\mathcal{Y}}_i | y_i = 1, \boldsymbol{\alpha}] \times \Pr[y_i = 1 | \boldsymbol{\phi}(\boldsymbol{x}_i^{\mathrm{EV}}), \boldsymbol{\alpha}] \\ &= \frac{\gamma_i \rho_i}{\gamma_i \rho_i + \delta_i (1 - \rho_i)}. \end{aligned} \tag{21}$$

Table 2. Image Classification Accuracy: These values are the average and standard deviation over all target image categories.

	Conventional Method [8]	Only Visual Features				Only EEG Features	Proposed Method
		SIFT	PHOG	GIST	IHIST		
subject A	79 ± 0.10%	68 ± 0.07%	67 ± 0.07%	67 ± 0.09%	71 ± 0.01%	59 ± 0.30%	**82 ± 0.06%**
subject B	**87 ± 0.05%**					77 ± 0.05%	**87 ± 0.05%**
subject C	**88 ± 0.06%**					72 ± 0.09%	**88 ± 0.02%**
subject D	**94 ± 0.05%**					69 ± 0.02%	93 ± 0.04%
subject E	93 ± 0.04%					71 ± 0.08%	**95 ± 0.02%**

M-step
In the M-step, the coefficient vector $\boldsymbol{\alpha}$ is estimated based on the current estimate μ_i and the training data $\mathcal{D}$ by maximizing the conditional expected value in Eq. (20). Specifically, by solving equation $\frac{\partial}{\partial \boldsymbol{\alpha}}\{\ln \Pr[\mathcal{D}, \mathcal{L}|\boldsymbol{\alpha}]\} = 0$, we obtain the estimated coefficient vector $\boldsymbol{\alpha}$ as follows:

$$\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} - \eta \boldsymbol{H}^{-1}\boldsymbol{g}. \tag{22}$$

In Eq. (22), $\boldsymbol{g}$ is a gradient vector, $\boldsymbol{H}$ is a Hessian matrix and η is a step length. The gradient vector $\boldsymbol{g}$ and the Hessian matrix $\boldsymbol{H}$ are respectively computed as follows: $\boldsymbol{g} = \sum_{i=1}^{N}[\mu_i - \sigma(\boldsymbol{\alpha}^\top \boldsymbol{\kappa}_i)]\boldsymbol{\kappa}_i$ and $\boldsymbol{H} = -\sum_{i=1}^{N}[\sigma(\boldsymbol{\alpha}^\top \boldsymbol{\kappa}_i)][1 - \sigma(\boldsymbol{\alpha}^\top \boldsymbol{\kappa}_i)]\boldsymbol{\kappa}_i \boldsymbol{\kappa}_i^\top$, where $\sigma(\boldsymbol{\alpha}^\top \boldsymbol{\kappa}_i) = \frac{1}{1+\exp(-\boldsymbol{\alpha}^\top \boldsymbol{\kappa}_i)}$.

2.2.4. Testing Phase

Given the test data, the final classification result is obtained as follows. In the previous phase, we essentially solved a regular logistic regression problem with probabilistic labels μ_i. Thus, we obtain the final classification result by applying a threshold to μ calculated from a test data $\{\boldsymbol{\phi}(\boldsymbol{x}^{\mathrm{EV}}), y^{f_1}, y^{f_2}, \cdots, y^{f_F}, y^{\mathrm{V_S}}, y^{\mathrm{V_P}}, y^{\mathrm{V_G}}, y^{\mathrm{V_I}}\}$, where its label is unknown, instead of directly using $\hat{\boldsymbol{\alpha}}_{ML}$. The value of μ is computed by using the estimated coefficient vector $\hat{\boldsymbol{\alpha}}_{ML}$ and γ, δ calculated from the training data. Specifically, $\rho = \frac{1}{1+\exp(-\hat{\boldsymbol{\alpha}}_{ML}^\top \boldsymbol{\kappa})}$ is calculated, where $\boldsymbol{\kappa} = \left[K(\boldsymbol{x}_1^{\mathrm{EV}}, \boldsymbol{x}^{\mathrm{EV}}), K(\boldsymbol{x}_2^{\mathrm{EV}}, \boldsymbol{x}^{\mathrm{EV}}), \cdots, K(\boldsymbol{x}_N^{\mathrm{EV}}, \boldsymbol{x}^{\mathrm{EV}})\right]^\top$. Then, $\gamma = \prod_{a\in\mathcal{A}}[P_{se}^a]^{y^a}[1 - P_{se}^a]^{1-y^a}$ and $\delta = \prod_{a\in\mathcal{A}}[P_{sp}^a]^{1-y^a}[1 - P_{sp}^a]^{y^a}$ are obtained, where P_{se}^a and P_{sp}^a are accuracy of annotator a calculated from training data and y^a is classification result of the test data. Therefore, we obtain the final classification result considering each annotator's accuracy. Then $\mu = \frac{\gamma\rho}{\gamma\rho+\delta(1-\rho)}$ is computed by using ρ, γ and δ. Finally, we obtain the final classification result as follows:

$$\hat{y} = \begin{cases} 1 & \mu \geq \mu' \\ 0 & \text{otherwise,} \end{cases} \tag{23}$$

where μ' is a predetermined threshold. The value of μ is the posterior probability.

3. EXPERIMENTAL RESULTS

In this section, we show experimental results to verify the effectiveness of the proposed method. First, we explain image dataset and EEG signal collection. In this experiment, we utilized Caltech101 dataset [20]. Specifically, we used the images included in "panda", "soccer ball" and "strawberry" in the database for image classification, and the number of images was 35 per category. These 105 images were randomly selected in advance. In this paper, we define images used for image classification as target images. We also used images included in "airplane", "elephant", "joshua tree", "pyramid" and "stapler" in the same database for the non-target images.

Next, we explain how to collect EEG signals in this experiment. In this study, five healthy subjects (subject A, B, C, D and E) participated, and EEG recordings were conducted during staring at images. The age of each subject was 22 or 23 years old. We recorded EEG signals from 12 channels (Fp1, Fp2, F7, F8, T3, T4, C3, C4, P3, P4, O1 and O2) according to the international 10-20 system. Since EEG signals are weak, we amplified these signals by using an amplifier (MEG-6116M, NIHON KOHDEN). We also applied a bandpass filter to recorded EEG signals to avoid artifacts, and set the filter bandwidth to 0.04-30Hz. In addition, the EEG signals were sampled at 2kHz. In this experiment, we collected single-trial EEG signals for each target image by the same experimental procedure as those shown in our previous work [8]. Note that the time length of staring at each image was two seconds (subject A and B) and three seconds (subject C, D and E). In this experiment, we perform image classification by using first one second of EEG signals recorded while a subject stared at target images.

Furthermore, we present the experimental condition. In this experiment, we performed the three class image classification (panda, soccer ball and strawberry) by the object categories that images contain based on one vs. all approach [17]. Therefore, the final classification was determined according to the posterior probability obtained from the testing phase (2.2.4). We followed [2, 3] for our experimental setup. Specifically, we randomly selected 30 training images per class and tested on the remaining images. Then we calculated the classification accuracy which was normalized according to the number of test images per class. We repeated the random selection 10 times and show the average classification accuracy over all classes.

We show the results of image classification in Table 2. As for the proposed method, we vary the number of neighbors k in Eq. (1) and present the best results. In this table, we also show the results of the conventional method [8]. From the obtained results, the proposed method realizes more accurate classification than the conventional method [8]. Therefore, the effectiveness of our method can be verified. Thus, the feature integration based on MSLPCCA is effective for the image classification using both EEG and visual features. In Table 2, we also present the average image classification accuracy by using either EEG or visual features based on SVM. From this table, although accuracy obtained by using each feature separately is not satisfactory, our method realizes successful classification based on collaborative use of all features.

4. CONCLUSION

In this paper, we have proposed a novel image classification method based on MSLPCCA-based feature integration. MSLPCCA enables the feature integration with preserving the locality structure of each variable and using class labels which are generally effective for the classification problem. In our method, this method is applied to EEG and visual features in order to generate feature vectors for the decision-level fusion of the image classification. Experimental results show the effectiveness of the proposed method.

5. REFERENCES

[1] M. Haseyama, T. Ogawa, and N. Yagi, "A review of video retrieval based on image and video semantic understanding," *ITE Transactions on Media Technology and Applications*, vol. 1, no. 1, pp. 2–9, 2013.

[2] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *International journal of computer vision*, vol. 73, no. 2, pp. 213–238, 2007.

[3] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in in*Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2005, vol. 2, pp. 1458–1465.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, vol. 1, pp. 886–893.

[6] Anna Bosch, Andrew Zisserman, and Xavier Munoz, "Representing shape with a spatial pyramid kernel," in *Proceedings of the 6th ACM international conference on Image and video retrieval*, 2007, pp. 401–408.

[7] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.

[8] T. Kawakami, T. Ogawa, and M. Haseyama, "Novel image classification based on decision-level fusion of EEG and visual features," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5915–5919, 2014.

[9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273 –297, 1995.

[10] Jinfeng Yang and Xu Zhang, "Feature-level fusion of fingerprint and finger-vein for personal identification," *Pattern Recognition Letters*, vol. 33, no. 5, pp. 623–628, 2012.

[11] Alexander A. Borbely and HansUlrich Neuhaus, "Sleep-deprivation: Effects on sleep and EEG in the rat," *Journal of comparative physiology*, vol. 133, no. 1, pp. 71–87, 1979.

[12] T. Kawakami, T. Ogawa, and M. Haseyama, "Vocal segment estimation in music pieces based on collaborative use of EEG and audio features," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 1197–1201.

[13] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. H. Chen, "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798 –1806, 2010.

[14] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proceedings of European Conference on Computer Vision*, 2004, vol. 1, pp. 1–22.

[15] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226 –1238, 2005.

[16] T. Hastie and R. Tibshirani, "Classification by pairwise coupling," *The annals of statistics*, vol. 26, no. 2, pp. 451–471, 1998.

[17] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *The Journal of Machine Learning Research*, vol. 5, pp. 101–141, 2004.

[18] X. He and P. Niyogi, "Locality preserving projections," in *NIPS*, 2003, vol. 16, pp. 234–241.

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, pp. 1–38, 1977.

[20] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 2004, pp. 178–178.