MVLFDA-BASED VIDEO PREFERENCE ESTIMATION USING COMPLEMENTARY PROPERTIES OF FEATURES

Akira Toyoda[†]*, Takahiro Ogawa*[‡] *and Miki Haseyama*[‡]

[†]School of Engineering, Hokkaido University
N-13, W-8, Kita-ku, Sapporo, Hokkaido, 060-8628, Japan
[‡]Graduate School of Information Science and Technology, Hokkaido University
N-14, W-9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan
E-mail: {toyoda, ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

ABSTRACT

This paper presents a new method to estimate users' video preferences using complementary properties of features via Multiview Local Fisher Discriminant Analysis (MvLFDA). The proposed method first extracts multiple visual features from video frames and electroencephalogram (EEG) features from users' EEG signals recorded during watching video. Then we calculate EEG-based visual features by applying Locality Preserving Canonical Correlation Analysis (LPCCA) to each visual feature and EEG features. The EEG-based visual features reflect users' preferences since the correlation between visual features and EEG features which reflect users' preferences is maximized. Next, MvLFDA, which is newly derived in this paper, integrates multiple EEG-based visual features. Since MvLFDA explores complementary properties of different features, it can be expected that the features obtained by integrating multiple EEG-based visual features are more effective for users' preference estimation than each EEG-based visual feature. The biggest contribution of this paper is the new derivation of MvLFDA. Then successful estimation of users' video preferences becomes feasible using features obtained by MvLFDA.

***Index Terms*—** individual preference, electroencephalogram (EEG), canonical correlation analysis, fisher discriminant analysis.

1. INTRODUCTION

A huge amount of video exists on the Internet through the development of various video sharing services. Then appropriate recommendation services enable users to easily find their favorite video. From this background, various recommendation methods have been proposed [1–5], and many methods exploit low-level visual features which are directly extracted from video in order to predict users' preferences [4, 5]. However, since low-level visual features are not generally related to individual preferences, it is considered that the direct use of visual features has the limitation for estimating users' preferences. Therefore, we need alternative approaches for individual preference estimation in order to overcome this limitation.

Recently, some studies have attempted to estimate human emotions and preferences by analyzing electroencephalogram (EEG) signal changes in response to visual and auditory stimuli [6–14]. Particularly, Frantzidis *et al.* proposed a methodology for classifying EEG signals recorded during passive viewing emotional pictures [7], and Hadjidimitriou *et al.* reported discrimination between users' EEG signals during listening to liked or disliked musical pieces [13, 14]. According to these studies, it is confirmed that EEG signals reflect users' emotions and preferences for visual and auditory stimuli. In addition, there have been proposed approaches which use the features extracted from users' EEG signals during listening to musical pieces in order to transform their audio features to new audio features for accurate estimation of favorite musical pieces [15, 16]. The new features were obtained by applying Canonical Correlation Analysis (CCA) [17] and its variants to EEG and audio features. Since EEG responses are also evoked by visual stimuli, it becomes feasible to transform original visual features to new features, *i.e.*, EEG-based visual features by applying the above computational methods to visual features. It can be expected that the obtained new features also reflect users' preferences since the correlation between EEG features reflecting users' preferences and original visual features is maximized by CCA.

In general, video frames are represented by multiple visual features. Each visual feature summarizes visual characteristics of video frames and has complementary properties to the other visual features. As the simplest method for dealing with multiple visual features, concatenating their feature vectors into a long vector is commonly used. However, it is possible that each EEG-based visual feature which is obtained by applying the aforementioned computation to each visual feature also has complementary properties to the other EEG-based visual features. Then it is difficult for this simplest scheme to make use of the complementary properties adequately. Therefore, it is necessary to introduce a method which can integrate multiple EEG-based visual features, exploring the complementary properties like [18–22], which learn combination coefficients for each feature in order to find complementary properties of different features.

This paper presents a new method to estimate users' video preferences using complementary properties of EEG-based visual features. First, we obtain EEG-based visual features by applying Locality Preserving Canonical Correlation Analysis (LPCCA) [23] to the two modal features which are extracted from EEG signals and video frames. Furthermore, for more accurate preference estimation, we newly derive a novel method; Multiview Local Fisher Discriminant Analysis (MvLFDA), which learns a common space over

This work was partly supported by JSPS KAKENHI Grant Numbers JP17H01744, JP15K12023.

multiple features and explores complementary properties of different features. MvLFDA is an extended version of Local Fisher Discriminant Analysis (LFDA) [24] to treat multiple features based on Multiview Spectral Embedding (MSE) [18]. The derivation of MvLFDA is the biggest contribution of this paper. Consequently, our method realizes successful preference estimation based on the above non-conventional approaches.

2. ESTIMATION OF USERS' VIDEO PREFERENCES

Our method consists of three stages. In the first stage, we extract EEG features from EEG signals recorded while users are watching video and their visual features. In the second stage, a projection matrix which transforms the visual features into the features reflecting users' preferences is calculated by applying LPCCA to the EEG and visual features. In the third stage, we transform multiple EEG-based visual features into a common space by applying MvLFDA. The estimation of video preferences then becomes feasible in this common space. The details of each stage are shown below.

2.1. Feature Extraction

This subsection shows the EEG features and the visual features used in our method.

2.1.1. EEG Feature Extraction

We compute EEG features which are shown in **Table 1**. First, each channel's EEG signal is divided into segments (hereafter EEG segments) with an overlapped Hamming window. The EEG features are computed for each EEG segment based on our previously reported method [15]. Then the means and standard deviations of the features over all EEG segments are computed, and these statistics are redefined as EEG features. In addition, we apply min-Redundancy and Max-Relevance (mRMR) algorithm [25] to EEG features and labels (Like (L) or Dislike (D)), which are given by users to each sample and represent users' preferences for them, in order to select only EEG features related to users' preferences. Finally, we obtain EEG feature vectors $\boldsymbol{x}$ ($\in \mathbb{R}^{d^x}$), where d^x denotes the dimension of EEG feature vectors after applying the mRMR algorithm.

2.1.2. Visual Feature Extraction

We use GIST descriptor (GIST; 960 dimension) [26], Histogram of Oriented Gradients (HOG; 1296 dimension) [27], HSV histogram (HSV; 64 dimension), Gabor texture (Gabor; 30 dimension), and Scale Invariant Feature Transform (SIFT; 100 dimension) [28] as visual features. In addition, we adopt Convolutional Neural Network (CNN) whose architecture is proposed by [29] and weights are predefined by using ImageNet [30], and then use the outputs of hidden layer's neurons (CNN features; 4096 dimension) based on [31]. These features are extracted from each frame of the target video. Then, for each element of these features, the mean over all frames of the target video is computed and redefined as the visual feature. Finally, we define vth visual feature vector $\boldsymbol{y}^{(v)} \in \mathbb{R}^{d^{y^{(v)}}}$ ($v = 1, 2, \cdots, V$), where $d^{y^{(v)}}$ is the dimension of vth feature vector, and V is the kinds of visual features, *i.e.*, $V = 6$.

2.2. Multimodal Feature Fusion via LPCCA

In this subsection, we explain the calculation method of the EEG-based visual features via LPCCA. LPCCA can project visual features into the new feature space where the correlation between EEG and visual features is maximized. First, from the EEG and visual feature vectors obtained in the previous subsection, we define $\boldsymbol{X} = [\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_n] \in \mathbb{R}^{d^x \times n}$ and $\boldsymbol{Y}^{(v)} = [\boldsymbol{y}_1^{(v)}\ \boldsymbol{y}_2^{(v)}\ \cdots\ \boldsymbol{y}_n^{(v)}] \in \mathbb{R}^{d^{y^{(v)}} \times n}$, where n is the number of samples. Note that it is assumed that these matrices are centered. Furthermore, the column vectors of $\boldsymbol{X}$ and $\boldsymbol{Y}^{(v)}$ respectively correspond to $\boldsymbol{x}$ and $\boldsymbol{y}^{(v)}$ obtained in the previous subsection. In LPCCA, we calculate similarity matrices $\boldsymbol{A}_X \in \mathbb{R}^{n \times n}$ and $\boldsymbol{A}_Y^{(v)} \in \mathbb{R}^{n \times n}$, where (i, j) th elements of $\boldsymbol{A}_X$ and $\boldsymbol{A}_Y^{(v)}$ are defined as follows:

$$A_X^{i,j} = \begin{cases} \mathrm{e}^{-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 / t_x} & \boldsymbol{x}_i \in \mathrm{LN}(\boldsymbol{x}_j) \text{ or } \boldsymbol{x}_j \in \mathrm{LN}(\boldsymbol{x}_i) \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

$$A_Y^{(v)\,i,j} = \begin{cases} \mathrm{e}^{-\|\boldsymbol{y}_i^{(v)} - \boldsymbol{y}_j^{(v)}\|^2 / t_y^{(v)}} & \boldsymbol{y}_i^{(v)} \in \mathrm{LN}(\boldsymbol{y}_j^{(v)}) \text{ or } \boldsymbol{y}_j^{(v)} \in \mathrm{LN}(\boldsymbol{y}_i^{(v)}) \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where $t_x = \frac{1}{n(n-1)} \sum_{i,j=1}^n \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2$ and $t_y^{(v)} = \frac{1}{n(n-1)} \sum_{i,j=1}^n \|\boldsymbol{y}_i^{(v)} - \boldsymbol{y}_j^{(v)}\|^2$. Then $\mathrm{LN}(\boldsymbol{x}_i)$ and $\mathrm{LN}(\boldsymbol{y}_i^{(v)})$ denote k-nearest neighbors of $\boldsymbol{x}_i$ and $\boldsymbol{y}_i^{(v)}$, respectively. By using the matrices $\boldsymbol{A}_X$ and $\boldsymbol{A}_Y^{(v)}$, we define the following matrices: $\boldsymbol{A}_{XY}^{(v)} = \boldsymbol{A}_X \circ \boldsymbol{A}_Y^{(v)}$, $\boldsymbol{A}_{XX} = \boldsymbol{A}_X \circ \boldsymbol{A}_X$ and $\boldsymbol{A}_{YY}^{(v)} = \boldsymbol{A}_Y^{(v)} \circ \boldsymbol{A}_Y^{(v)}$, where "$\circ$" is an operator calculating the Hadamard product. Furthermore, we compute the following matricies: $\boldsymbol{L}_{XY}^{(v)} = \boldsymbol{D}_{XY}^{(v)} - \boldsymbol{A}_{XY}^{(v)}$, $\boldsymbol{L}_{XX} = \boldsymbol{D}_{XX} - \boldsymbol{A}_{XX}$, and $\boldsymbol{L}_{YY}^{(v)} = \boldsymbol{D}_{YY}^{(v)} - \boldsymbol{A}_{YY}^{(v)}$, where $\boldsymbol{D}_{XY}^{(v)} = \mathrm{diag}[\sum_{j=1}^n (A_{XY}^{(v)\,1,j}), \sum_{j=1}^n (A_{XY}^{(v)\,2,j}), \cdots, \sum_{j=1}^n (A_{XY}^{(v)\,n,j})]$ and $A_{XY}^{(v)\,i,j}$ is (i, j) th element of $\boldsymbol{A}_{XY}^{(v)}$. The matrices $\boldsymbol{D}_{XX}$ and $\boldsymbol{D}_{YY}^{(v)}$ are then obtained in the same manner. In addition, we compute vectors $\hat{\boldsymbol{w}}_x$ and $\hat{\boldsymbol{w}}_y^{(v)}$ which project matrices $\boldsymbol{X}$ and $\boldsymbol{Y}^{(v)}$ into the common latent space by solving the following optimization problem:

$$(\hat{\boldsymbol{w}}_x, \hat{\boldsymbol{w}}_y^{(v)}) = \arg \max_{\boldsymbol{w}_x, \boldsymbol{w}_y^{(v)}} \frac{\boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{L}_{XY}^{(v)} \boldsymbol{w}_y^{(v)}}{\sqrt{\boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{L}_{XX} \boldsymbol{w}_x} \sqrt{\boldsymbol{w}_y^{(v)\mathrm{T}} \boldsymbol{L}_{YY}^{(v)} \boldsymbol{w}_y^{(v)}}}. \tag{3}$$

By solving Eq. (3) via the eigenvalue problem, several results of $\hat{\boldsymbol{w}}_x$ and $\hat{\boldsymbol{w}}_y^{(v)}$ can be obtained. In our method, we align projection vectors $\hat{\boldsymbol{w}}_y^{(v)}$ in columns to obtain the projection matrix $\hat{\boldsymbol{W}}_{CCA}^{(v)} = [\hat{\boldsymbol{w}}_{y,1}^{(v)}\ \hat{\boldsymbol{w}}_{y,2}^{(v)}\ \cdots\ \hat{\boldsymbol{w}}_{y,d^{\hat{y}^{(v)}}}^{(v)}]$, where $d^{\hat{y}^{(v)}}$ denotes the dimension of the projected visual feature vectors, *i.e.*, the EEG-based visual feature vectors, and $d^{\hat{y}^{(v)}} < d^{y^{(v)}}$. Then the calculation of the EEG-based

Table 1. EEG features. C : number of channels of EEG signals and C_P : number of symmetric electrode pairs placed on the scalp.

DESCRIPTION		DIMENSION
Zero Crossing Rate		C
Content Percentages of Each Band	θ wave (4-8Hz)	C
	slow α wave (8-10Hz)	C
	fast α wave (11-13Hz)	C
	α wave (8-13Hz)	C
	slow β wave (13-19Hz)	C
	fast β wave (20-30Hz)	C
	β (13-30Hz)	C
	γ wave (30-49Hz)	C
Power Spectrum of The Hemispheric Asymmetry	θ wave (4-8Hz)	$2C^p$
	slow α wave (8-10Hz)	$2C^p$
	fast α wave (11-13Hz)	$2C^p$
	α wave (8-13Hz)	$2C^p$
	slow β wave (13-19Hz)	$2C^p$
	fast β wave (20-30Hz)	$2C^p$
	β (13-30Hz)	$2C^p$
	γ wave (30-49Hz)	$2C^p$
TOTAL		$9C + 16C^P$

visual feature vectors becomes feasible by multiplying the projection matrix $\hat{\boldsymbol{W}}_{CCA}^{(v)}$ into visual feature vectors as follows:

$$\hat{\boldsymbol{y}}^{(v)} = \hat{\boldsymbol{W}}_{CCA}^{(v)\mathrm{T}} \boldsymbol{y}^{(v)}. \tag{4}$$

2.3. Multiview Feature Fusion via MvLFDA

In this subsection, we explain the calculation method of the common space over all features via MvLFDA. MvLFDA can fuse multiple EEG-based visual features while exploring the complementary properties in these features. First, from the EEG-based visual features obtained in the previous subsection, we define $\hat{\boldsymbol{Y}}^{(v)} = [\hat{\boldsymbol{y}}_1^{(v)}\ \hat{\boldsymbol{y}}_2^{(v)}\ \cdots\ \hat{\boldsymbol{y}}_n^{(v)}]$. MvLFDA finds a low-dimensional embedding of $\hat{\boldsymbol{Y}}^{(v)}$, *i.e.*, $\boldsymbol{Z} = [\boldsymbol{z}_1\ \boldsymbol{z}_2\ \cdots\ \boldsymbol{z}_n]\ (\in \mathbb{R}^{d^z \times n})$ over all features, where d^z denotes the dimension of the common space. Next, given the means of all samples $\boldsymbol{\mu}$ and samples in class c ($c \in \{L, D\}$) $\boldsymbol{\mu}_c$ in the common space $\boldsymbol{Z}$, within-class scatter matrix $\boldsymbol{S}_w^{(v)}$ and between-class scatter matrix $\boldsymbol{S}_b^{(v)}$ are defined as follows:

$$\begin{aligned}\boldsymbol{S}_w^{(v)} &= \sum_{c\in\{L,D\}} \sum_{j:label(\boldsymbol{z}_j)=c} (\boldsymbol{z}_j - \boldsymbol{\mu}_c)(\boldsymbol{z}_j - \boldsymbol{\mu}_c)^{\mathrm{T}} \\ &= \sum_{i,j=1}^{n} A_w^{(v)\ i,j} (\boldsymbol{z}_i - \boldsymbol{z}_j)(\boldsymbol{z}_i - \boldsymbol{z}_j)^{\mathrm{T}} \\ &= \boldsymbol{Z}\boldsymbol{L}_w^{(v)}\boldsymbol{Z}^{\mathrm{T}}, \end{aligned} \tag{5}$$

$$\begin{aligned}\boldsymbol{S}_b^{(v)} &= \sum_{c\in\{L,D\}} (\boldsymbol{\mu}_c - \boldsymbol{\mu})(\boldsymbol{\mu}_c - \boldsymbol{\mu})^{\mathrm{T}} \\ &= \sum_{i,j=1}^{n} A_b^{(v)\ i,j} (\boldsymbol{z}_i - \boldsymbol{z}_j)(\boldsymbol{z}_i - \boldsymbol{z}_j)^{\mathrm{T}} \\ &= \boldsymbol{Z}\boldsymbol{L}_b^{(v)}\boldsymbol{Z}^{\mathrm{T}}, \end{aligned} \tag{6}$$

where label$(\cdot) \in \{L, D\}$. In addition, $A_w^{(v)\ i,j}$ and $A_b^{(v)\ i,j}$ are defined as follows:

$$A_w^{(v)\ i,j} = \begin{cases} \frac{A^{(v)\ i,j}}{n_c} & \mathrm{label}(\hat{\boldsymbol{y}}_i^{(v)}) = \mathrm{label}(\hat{\boldsymbol{y}}_j^{(v)}) = c \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

$$A_b^{(v)\ i,j} = \begin{cases} A^{(v)\ i,j}(\frac{1}{n} - \frac{1}{n_c}) & \mathrm{label}(\hat{\boldsymbol{y}}_i^{(v)}) = \mathrm{label}(\hat{\boldsymbol{y}}_j^{(v)}) = c \\ \frac{1}{n} & \text{otherwise} \end{cases}, \tag{8}$$

where n_c is the number of samples in class c, and $A^{(v)\ i,j}$ is defined as follows:

$$A^{(v)\ i,j} = \begin{cases} \mathrm{e}^{-\|\hat{\boldsymbol{y}}_i^{(v)} - \hat{\boldsymbol{y}}_j^{(v)}\|^2 / \hat{t}_y^{(v)}} & \hat{\boldsymbol{y}}_i^{(v)} \in \mathrm{LN}(\hat{\boldsymbol{y}}_j^{(v)}) \text{ or } \hat{\boldsymbol{y}}_j^{(v)} \in \mathrm{LN}(\hat{\boldsymbol{y}}_i^{(v)}) \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

where $\hat{t}_y^{(v)} = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} \|\hat{\boldsymbol{y}}_i^{(v)} - \hat{\boldsymbol{y}}_j^{(v)}\|^2$. Furthermore, we define $\boldsymbol{D}_w^{(v)}$ and $\boldsymbol{D}_b^{(v)}$ as follows:

$$\boldsymbol{D}_w^{(v)} = \mathrm{diag}[\sum_{j=1}^{n} (A_w^{(v)\ 1,j}),\ \sum_{j=1}^{n} (A_w^{(v)\ 2,j}),\ \cdots,\ \sum_{j=1}^{n} (A_w^{(v)\ n,j})], \tag{10}$$

$$\boldsymbol{D}_b^{(v)} = \mathrm{diag}[\sum_{j=1}^{n} (A_b^{(v)\ 1,j}),\ \sum_{j=1}^{n} (A_b^{(v)\ 2,j}),\ \cdots,\ \sum_{j=1}^{n} (A_b^{(v)\ n,j})]. \tag{11}$$

In Eqs. (5) and (6), the matrices $\boldsymbol{L}_w^{(v)}$ and $\boldsymbol{L}_b^{(v)}$ are computed as follows:

$$\boldsymbol{L}_w^{(v)} = \boldsymbol{D}_w^{(v)} - \boldsymbol{A}_w^{(v)}, \quad \boldsymbol{L}_b^{(v)} = \boldsymbol{D}_b^{(v)} - \boldsymbol{A}_b^{(v)}, \tag{12}$$

where (i, j) th elements of the similarity matrices $\boldsymbol{A}_w^{(v)}$ and $\boldsymbol{A}_b^{(v)}$ are $A_w^{(v)\ i,j}$ and $A_b^{(v)\ i,j}$, respectively. By using $\boldsymbol{L}_w^{(v)}$ and $\boldsymbol{L}_b^{(v)}$, the optimization problems for each feature are defined as follows:

$$\arg\max_{\boldsymbol{Z}} \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}_b^{(v)}\boldsymbol{Z}^{\mathrm{T}}) - \gamma \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}_w^{(v)}\boldsymbol{Z}^{\mathrm{T}}), \quad \text{s.t. } \boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} = \boldsymbol{I}, \tag{13}$$

where Eq. (13) is rewritten as follows:

$$\begin{aligned}\arg\max_{\boldsymbol{Z}} \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}_b^{(v)}\boldsymbol{Z}^{\mathrm{T}}) - \gamma \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}_w^{(v)}\boldsymbol{Z}^{\mathrm{T}}) &= \arg\max_{\boldsymbol{Z}} \mathrm{Tr}(\boldsymbol{Z}(\boldsymbol{L}_b^{(v)} - \gamma\boldsymbol{L}_w^{(v)})\boldsymbol{Z}^{\mathrm{T}}) \\ &= \arg\max_{\boldsymbol{Z}} \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}^{(v)}\boldsymbol{Z}^{\mathrm{T}}). \end{aligned} \tag{14}$$

In Eqs. (13) and (14), $\boldsymbol{L}^{(v)} = \boldsymbol{L}_b^{(v)} - \gamma\boldsymbol{L}_w^{(v)}$, and γ is a trade-off parameter. Then we impose weighted factors to the obtained objective functions for each feature in order to well explore complementary properties of different features. By summing over all features, the new optimization problem is defined as follows:

$$\arg\max_{\boldsymbol{Z},\boldsymbol{\alpha}} \sum_{v=1}^{V} \alpha_v \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}^{(v)}\boldsymbol{Z}^{\mathrm{T}}) - \eta\|\boldsymbol{\alpha}\|^2, \ \text{ s.t. } \boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} = \boldsymbol{I}, \ \sum_{v=1}^{V} \alpha_v = 1, \ \alpha_v \geq 0, \tag{15}$$

where the elements of $\boldsymbol{\alpha}$ are α_v, and the second term is needed for regularization. In Eq. (15), we iteratively update $\boldsymbol{Z}$ and $\boldsymbol{\alpha}$ to obtain the optimal result of $\boldsymbol{Z}$ as follows.

[Update of $\boldsymbol{Z}$]

We fix $\boldsymbol{\alpha}$ and obtain the optimal solution of $\boldsymbol{Z}$ by solving the following optimization problem:

$$\arg\max_{\boldsymbol{Z}} \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}^{\mathrm{T}}), \ \text{ s.t. } \boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}} = \boldsymbol{I}, \tag{16}$$

where $\boldsymbol{L} = \sum_{v=1}^{V} \alpha_v \boldsymbol{L}^{(v)}$. Then the solution of $\boldsymbol{Z}$ in (16) is obtained as follows:

$$\boldsymbol{Z} = [\boldsymbol{\varphi}_1\ \boldsymbol{\varphi}_2\ \cdots\ \boldsymbol{\varphi}_{d^z}], \tag{17}$$

where $\boldsymbol{\varphi}_k$ ($k = 1, 2, \cdots, d^z$) are the generalized eigenvectors of $\boldsymbol{L}$, which are associated to the generalized eigenvalues $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_{d^z}$.

[Update of $\boldsymbol{\alpha}$]

We fix $\boldsymbol{Z}$ and compute $\boldsymbol{\alpha}$ by solving the following Lagrangian function:

$$F(\boldsymbol{\alpha}, \lambda) = \sum_{v=1}^{V} \alpha_v \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}^{(v)}\boldsymbol{Z}^{\mathrm{T}}) - \eta\|\boldsymbol{\alpha}\|^2 - \lambda(\sum_{v=1}^{V} \alpha_v - 1), \tag{18}$$

where we set the derivative of $L(\boldsymbol{\alpha}, \lambda)$ with respect to α_v and λ to zero as follows:

$$\begin{aligned}\frac{\partial F(\alpha,\lambda)}{\partial \alpha_v} &= \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}^{(v)}\boldsymbol{Z}^{\mathrm{T}}) - 2\eta\alpha_v - \lambda = 0, \qquad v = 1,\ 2,\ \cdots, V, \\ \frac{\partial F(\alpha,\lambda)}{\partial \lambda} &= \sum_{v=1}^{V} \alpha_v - 1 = 0. \end{aligned} \tag{19}$$

From Eq. (19), α_v can be obtained as follows:

$$\alpha_v = \frac{V\mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}^{(v)}\boldsymbol{Z}^{\mathrm{T}}) - \sum_{v=1}^{V} \mathrm{Tr}(\boldsymbol{Z}\boldsymbol{L}^{(v)}\boldsymbol{Z}^{\mathrm{T}}) - 2\eta}{2V\eta}. \tag{20}$$

Consequently, we obtain the optimal common space $\boldsymbol{Z}$. Since feature vectors in the common space learn complementary properties in the multiple EEG-based visual features, it can be expected that these vectors reflect users' preferences more accurately. Furthermore, we

compute the projection matrix $\boldsymbol{W}_{FDA}^{(v)}$ ($\in \mathbb{R}^{d^{\hat{y}^{(v)}} \times d^{z}}$) which projects the EEG-based visual features into the common space as follows:

$$\boldsymbol{Z} = \boldsymbol{W}_{FDA}^{(v)\mathrm{T}} \hat{\boldsymbol{Y}}^{(v)}. \tag{21}$$

In order to obtain optimal $\boldsymbol{W}_{FDA}^{(v)}$, *i.e.*, $\hat{\boldsymbol{W}}_{FDA}^{(v)}$, we substitute $\boldsymbol{W}_{FDA}^{(v)\mathrm{T}} \boldsymbol{Y}^{(v)}$ for $\boldsymbol{Z}$ in Eq. (16) as follows:

$$\arg\max_{\boldsymbol{W}_{FDA}^{(v)}} \mathrm{Tr}(\boldsymbol{W}_{FDA}^{(v)\mathrm{T}} \hat{\boldsymbol{Y}}^{(v)} \boldsymbol{L} \hat{\boldsymbol{Y}}^{(v)\mathrm{T}} \boldsymbol{W}_{FDA}^{(v)}). \tag{22}$$

From Eq. (22), $\hat{\boldsymbol{W}}_{FDA}^{(v)} = [\boldsymbol{\psi}_1^{(v)}\ \boldsymbol{\psi}_2^{(v)}\ \cdots\ \boldsymbol{\psi}_{d^z}^{(v)}]$ can be obtained, where $\boldsymbol{\psi}_k^{(v)} (k = 1, 2, \cdots, d^z)$ are the generalized eigenvectors of $\hat{\boldsymbol{Y}}^{(v)} \boldsymbol{L} \hat{\boldsymbol{Y}}^{(v)\mathrm{T}}$, which are associated to the generalized eigenvalues $\zeta_1^{(v)} \geq \zeta_2^{(v)} \geq \cdots \geq \zeta_{d^z}^{(v)}$. By multiplying the projection matrix obtained in this subsection and the previous subsection, *i.e.*, $\hat{\boldsymbol{W}}_{FDA}^{(v)}$ and $\hat{\boldsymbol{W}}_{CCA}^{(v)}$, into visual feature vectors extracted from unknown data, we can transform the feature vectors to the EEG-based feature vectors and project EEG-based feature vectors into the common space which captures users' preferences more accurately. This means that if these two projections are calculated once, our method can perform this transformation for new samples without newly recording users EEG signals. Finally, estimation of video preferences can be realized based on a trained estimator in the common space.

3. EXPERIMENTAL RESULTS

In this section, we show experimental results to verify the effectiveness of the proposed method. First of all, we collected movie trailers of four genres from YouTube[1]. The genres are "action", "comedy", "drama", and "horror". In our experiment, we used eight movie trailers for each genre, *i.e.*, the total number of movie trailers is 32. The experimental task then consisted of (1) relaxing video clip period (15s); (2) a fixation white cross on a black background (15s); (3) the movie trailer period; (4) the subjective rating, which was assessed by a value of four levels, *i.e.*, 4 (like very much), 3 (like), 2 (do not like), 1 (do not like at all). Therefore, visual feature vectors could be grouped with respect to two classes, *i.e.*, "*L*" and "*D*". Class "*L*" consisted of the visual feature vectors corresponding to movie trailers rated 4 or 3. On the other hand, class "*D*" consisted of the visual feature vectors corresponding to movie trailers rated 2 or 1.

In our experiment, EEG signals were recorded from four healthy subjects. We recorded EEG signals from 12 electrodes according to the international 10-20 system. We also applied a band-pass filter to recorded EEG signals to avoid artifacts, where the filter bandwidth was set to 0.04-100Hz.

For evaluating the performance of our method, we used six comparative methods: the method using only visual features and applying MvLFDA to these visual features (C1), the method concatenating multiple visual features to one long vector and applying LPCCA to EEG features and the long vector (C2), and the methods applying LPCCA to EEG features and each visual feature and integrating multiple EEG-based visual features via simply concatenating them to one long vector (C3), MSE [18] (C4), Supervised MSE (SMSE) [21] (C5), and Multi-view Discriminant Analysis (MvDA) [32] (C6), where C2 and C3 applied LFDA to the long vector after concatenating multiple features. We use Support Vector Machine (SVM) [33] as the estimator for all methods. Then we employed the leave-one-out method and used F-measure as the evaluation measure. In addition, we performed one sided paired t-tests in order to evaluate the statistical significance of the experimental results. Results of our experiment and t-tests are shown in **Fig. 1** and **Table. 2**, respectively, where "PM" represents our proposed method. Comparing with the results of our proposed method and C1, we can confirm EEG-based visual features obtained via LPCCA are effective for preference estimation. In addition, comparing with the results of our method and all comparative methods, we can verify the effectiveness of our method using MvLFDA. Therefore, our method realizes successful estimation.

[1]https://www.youtube.com/

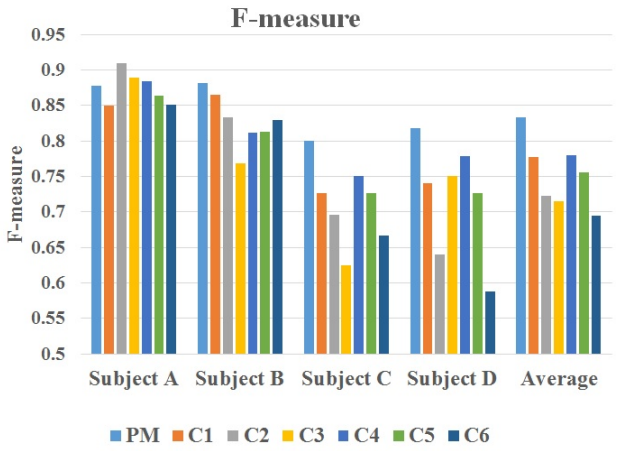

Fig. 1. The result of experiment.

Table 2. The p-value of one sided paired t-tests performed between the experimental result of our proposed method and each comparative method.

C1	C2	C3	C4	C5	C6
0.0254	0.0939	0.0574	0.0486	0.0170	0.0469

4. CONCLUSION

In this paper, we have presented a new method to estimate users' video preferences using complementary properties of EEG-based visual features. The proposed method first extracts multiple visual features and computes their EEG-based visual features suitable for representing users' preferences by LPCCA. Next, MvLFDA projects multiple EEG-based visual features into a common space which is effective for preference estimation. The experimental results have shown the effectiveness of the proposed method.

5. REFERENCES

[1] J. Zhang and P. Pu, "A recursive prediction algorithm for collaborative filtering recommender systems," in *Proceedings of the ACM International Conference on Recommender Systems*, pp. 57-64, 2007.

[2] G. Lekakos and P. Caravelas, "A hybrid approach for movie recommendation," *Multimedia Tools and Applications*, vol. 36, no.1-2, pp. 55-70, 2008.

[3] P. Cui, Z. Wang, and Z. Su, "What videos are similar with you?: Learning a common attributed representation for video recommendation," in *Proceedings of the ACM International Conference on Multimedia*, pp. 597-606, 2014.

[4] Y. Deldjoo, M. Elahi, M. Quadrana, and P. Cremonesi, "Toward building a content-based video recommendation system based on low-level features," in *Proceedings of International Conference on Electronic Commerce and Web Technologies*, pp. 45-56, 2015.

[5] Y. Deldjoo, M. Elahi, P. Cremonesi, F. Garzotto, P. Piazzolla, and M. Quadrana, "Content-based video recommendation system based on stylistic visual features," *Journal on Data Semantics*, pp. 1-15, 2016.

[6] H. Alicia and F. Claude, "Predicting the three major dimensions of the learner's emotions from brainwaves," *International Journal of Computer Science*, vol. 2, no. 3, pp. 187-193, 2007.

[7] C. Frantzidis, C. Bratsas, M. Klados, E. Konstantindis, C. Lithari, A. Vivas, C. Papadelis, E. Kaldoudi, C. Pappas, and P. Bamidis, "On the classification of emotional biosignals evoked while viewing affective pictures: an integrated data-mining-based approach for healthcare applications," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 309-318, 2010.

[8] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211-223, 2012.

[9] S. Wang, Y. Zhu, G. Wu, and Q. Ji, "Hybrid video emotional tagging using users' EEG and video content," *Multimedia Tools and Applications*, vol. 72, no. 2, pp. 1257-1283, 2014.

[10] S. Koelstra, A. Yazdani, M. Soleymani, C. Mühl, J. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras, "Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos," in *Proceedings of International Conference on Brain Informatics*, pp. 89-100, 2010.

[11] A. Yazdani, J. Lee, J. Vesin, and T. Ebrahimi, "Affect recognition based on physiological changes during the watching of music videos," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 1, pp. 7:1-7:26, 2012.

[12] J. Moon, Y. Kim, H. Lee, C. Bae, and W. Yoon, "Extraction of user preference for video stimuli using EEG-based user responses," *Journal of ETRI*, vol. 35, no. 6, pp. 1105-1114, 2013.

[13] S. Hadjidimitriou and L. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 12, pp. 3498-3510, 2012.

[14] S. Hadjidimitriou and L. Hadjileontiadis, "EEG-based classification of music appraisal responses using time-frequency analysis and familiarity ratings," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 161-172, 2013.

[15] R. Sawata, T. Ogawa, and M. Haseyama, "Human-centered favorite music estimation: EEG-based extraction of audio features reflecting individual preference" in *Proceedings of the IEEE International Conference on Digital Signal Processing (DSP)*, pp. 818-822, 2015.

[16] R. Sawata, T. Ogawa, and M. Haseyama, "Novel favorite music classification using EEG-based optimal audio features selected via KDLPCCA," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.759-763, 2016.

[17] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321-377, 1936.

[18] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 6, pp. 1438-1446, 2010.

[19] B. Xie, Y. Mu, D. Tao, and K. Huang, "m-SNE: Multiview stochastic neighbor embedding," *IEEE Transactions on Image Processing*, vol. 41, no. 4, pp. 1088-1096, 2011.

[20] W. Liu and D. Tao, "Multiview Hessian Regularization for Image Annotation," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2676-2687, 2013.

[21] S. Liu, L. Zhang, W. Cai, Y. Song, Z. Wang, L. Wen, and D.D. Feng, "A supervised multiview spectral embedding method for neuroimaging classification," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 601-605, 2013.

[22] L. Liu, M. Yu, and L. Shao, "Multiview alignment hashing for efficient image search," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 956-966, 2015.

[23] T. Sun and S. Chen, "Locality preserving CCA with applications to data visualization and pose estimation," *Image and Vision Computing*, vol. 25, no. 5, pp. 531-543, 2007.

[24] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," in *Proceedings of International Conference on Machine learning*, vol. 23, pp. 905-912, 2006.

[25] 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.

[26] 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.

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

[28] D.G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, pp. 1150-1157, 1999.

[29] H. Peng, F. Long, and C. Ding, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.

[30] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 248-255, 2009.

[31] J, Donahue, Y, Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proceedings of International Conference on Machine Learning*, pp. 647-655, 2014.

[32] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 188-194, 2016.

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