HUMAN EMOTION RECOGNITION USING MULTI-MODAL BIOLOGICAL SIGNALS BASED ON TIME LAG-CONSIDERED CORRELATION MAXIMIZATION

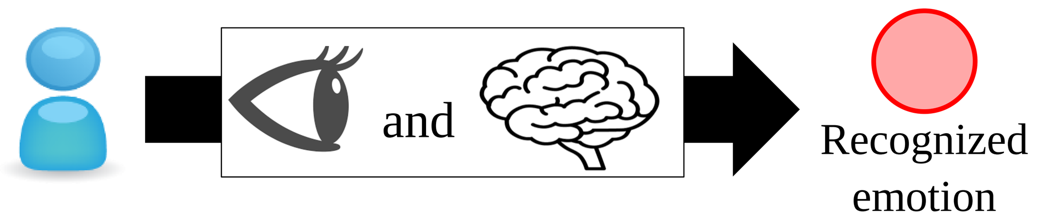
22-27 MAY 2022
ICASSP 2022
SINGAPORE
【PAPER ID: 8946】

Yuya Moroto, Keisuke Maeda, Takahiro Ogawa and Miki Haseyama   Hokkaido University, Japan
E-mail: moroto@lmd.ist.hokudai.ac.jp

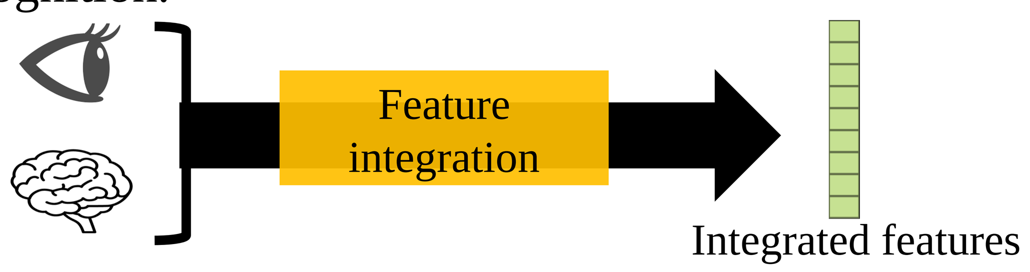
BACKGROUND

Human emotions are well-known to play an important role not only for human communications, but also for human-computer communications.

To make computers recognize human emotions, multi-modal biological signals, which are eye gaze and brain activity, have been focused [9-12].

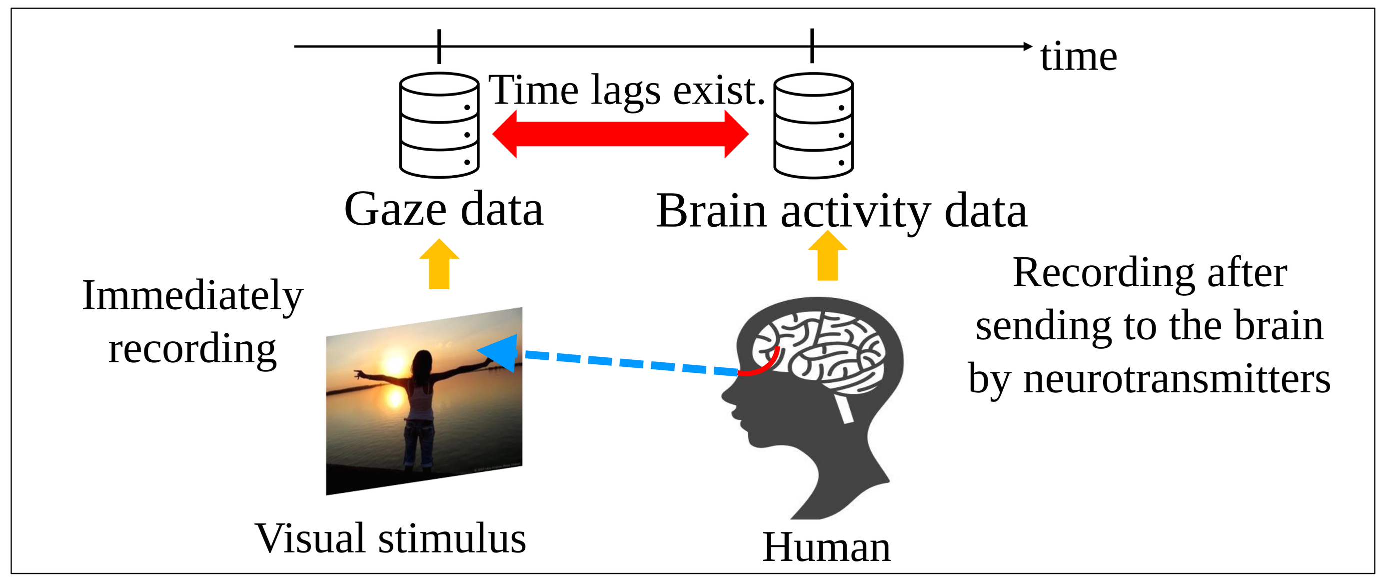


Previous works have integrated these signals for more accurate human emotion recognition.



By using several biological signals, human emotion recognition has been improved.

In the case of the human emotion recognition for visual stimuli, humans acquire information through the eyes and process it in the brain. Thus, there is a time lag between gaze data and brain activity data [19].

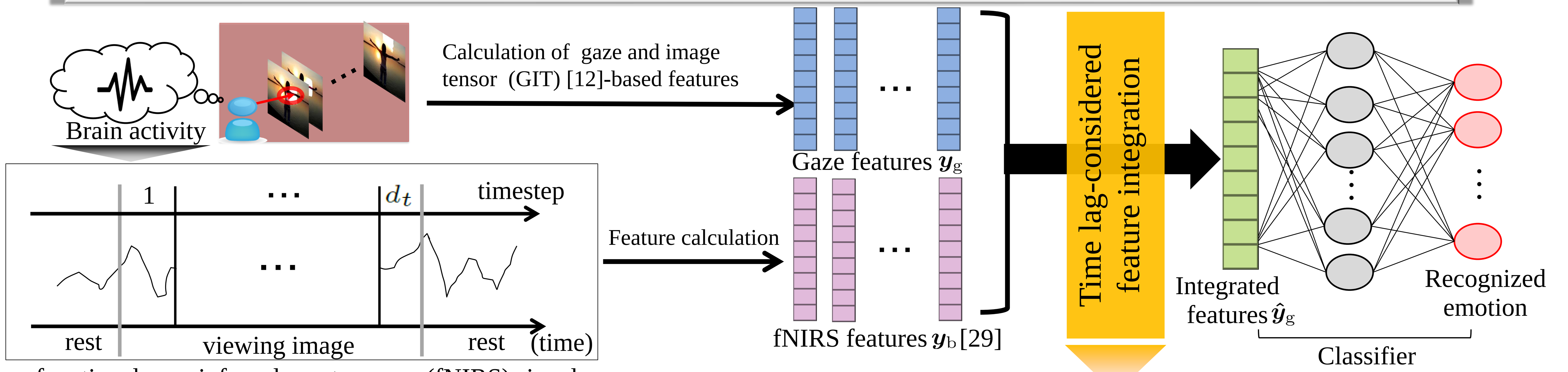


Previous studies just integrate multi-modal information without considering time lags, which may lead to the decrease in accuracy..

By considering such time lags, the integration with the correct correspondence is expected to be realized.

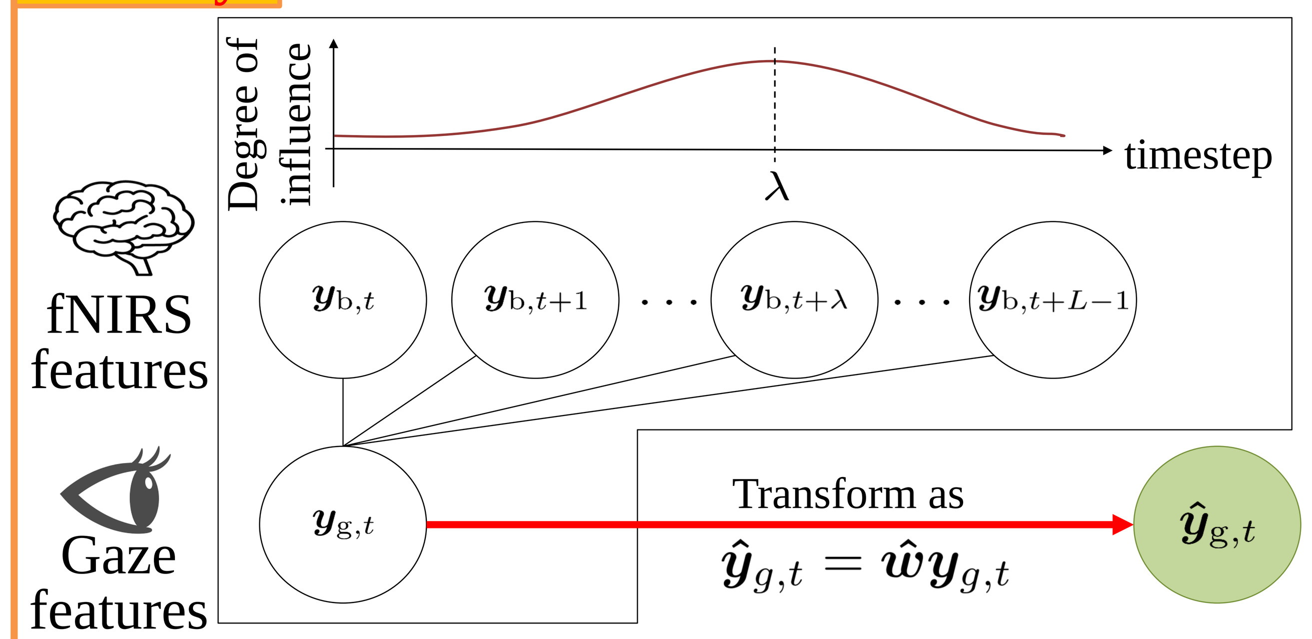
Proposed Method

Multi-modal human emotion recognition based on time lag-considered canonical correlation analysis



functional near-infrared spectroscopy (fNIRS) signals

Novelty



For realizing the time lag-considered correlation maximization, we assume that the time lag follows a Poisson distribution.

⇒We calculate the transform vector $\hat{w}$ considering the time lags.

$$\hat{\boldsymbol{w}} = \arg\max_{\boldsymbol{w}} \boldsymbol{w}_{\mathrm{gaze}}^{\top} \sum_{n=1}^{N} \boldsymbol{C}_n^{\mathrm{b}} \boldsymbol{w}_{\mathrm{brain}} \quad \text{s.t.} \ \boldsymbol{w}_{\mathrm{gaze}}^{\top} \boldsymbol{C}_n^{\mathrm{gaze}} \boldsymbol{w}_{\mathrm{gaze}} = \boldsymbol{w}_{\mathrm{brain}}^{\top} \boldsymbol{C}_n^{\mathrm{brain}} \boldsymbol{w}_{\mathrm{brain}} = 1$$

$$\boldsymbol{C}_n^{\mathrm{b}} = \boxed{\frac{1}{\sum_{l=0}^{L} e^{-\lambda}\lambda^l/l!} \sum_{l=0}^{L} \frac{e^{-\lambda}\lambda^l}{l!}} \boldsymbol{Y}_{\mathrm{gaze},n,l} \boldsymbol{Y}_{\mathrm{brain},n,0}^{\top} \qquad \boldsymbol{Y}_{p,n,l} = [\boldsymbol{y}_{p,n,L-l}, \boldsymbol{y}_{p,n,L+1-l}, \ldots, \boldsymbol{y}_{p,n,d_t-l}] \quad (l = 0, 1, \ldots, L-1)$$

Time-lags are represented by introducing weights following the Poiasson distribution into CCA

N : number of samples
C_n^p: variance matrix of modality p
$\boldsymbol{w}$: transform vector
d_t : number of timesteps
$p = \{\mathrm{gaze}, \mathrm{brain}\}$

$y_{g,t}$: gaze features at timestep t
$y_{b,t}$: fNIRS features at timestep t
$\hat{y}_{g,t}$: transformed gaze features at timestep t
L : hyperparameter deciding the number of timesteps that influence visual stimuli on fNIRS features
λ : shape parameter of the Poisson distribution

We realize the feature integration considering the time lags for human emotion recognition.

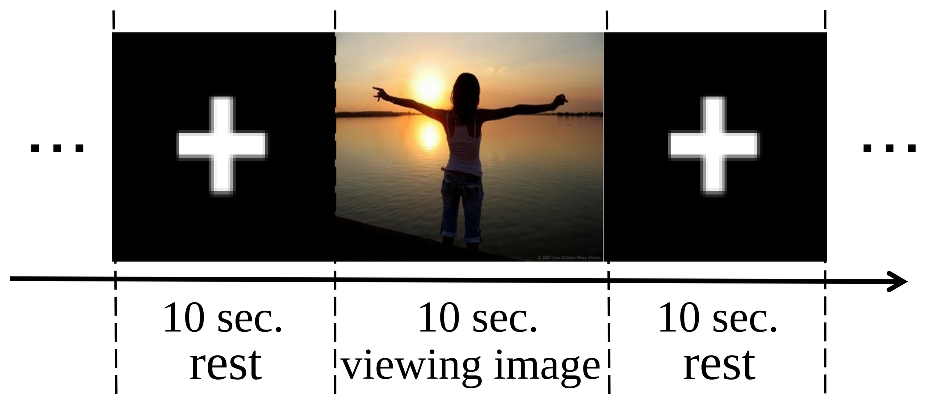
EXPERIMENTAL RESULTS

Dataset

80 images included in an art photo dataset [30] (Training : 64, Test : 16)

Settings

Number of subjects : 10
Instrument : Tobii Eye Tracker 4c for gaze data
LIGHTNIRS for brain activity data
Ground Truth : Subject feedbacks (positive/negative)
Evaluation Metrics : F1-score, Accuracy
fNIRS Features: statistical and wavelet transform-based features [29]
Hyperparameters: L, λ were set to 5, 1
Dimensions of Features: 440 (fNIRS), 1440 (Gaze), 50 (Integrated)



Experimental design

Comparative Methods

- Two methods that used gaze or fNRIS features as the abbreviation studies (Abbreviations 1 and 2).
- Five methods proposed in [9-12, 32]. These methods adopted different feature integration methods as the right table.

	Features		Time	
	Gaze	fNIRS	Change	Lag
Abbreviation 1		✓	✓	
Abbreviation 2	✓		✓	
Deep CCA [10]	✓	✓		
BDAE [9]	✓	✓		
BLSTM [11]	✓	✓	✓	
MVAE [32]	✓	✓	✓	
CCA with GIT [12]	✓	✓	✓	
Our method	✓	✓	✓	✓

Quantitative Evaluation

We confirmed the following effectiveness.

- **Abbreviation 1 and 2 vs Ours**
  ⇒ Effectiveness of using the multi-modal signals
- **Deep CCA and BDAE vs Ours**
  ⇒ Effectiveness of considering time changes
- **BLSTM, MVAE and CCA with GIT vs Ours**
  ⇒Effectiveness of considering time lags

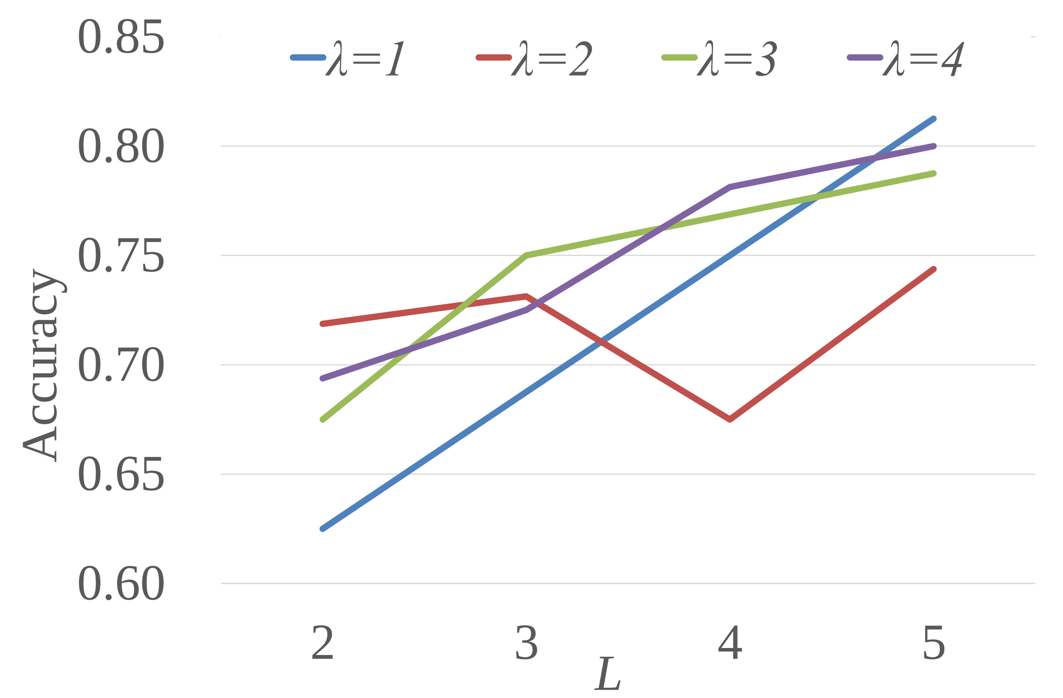
	F1-score	Accuracy
Abbreviation 1	0.65	0.52
Abbreviation 2	0.76	0.77
Deep CCA [10]	0.53	0.58
BDAE [9]	0.55	0.57
BLSTM [11]	0.44	0.44
MVAE [32]	0.52	0.57
CCA with GIT [12]	0.67	0.74
Our method	**0.78**	**0.81**

We verified that our method was effective for the human emotion recognition.

Hyperparameter Confirmation

We confirmed the accuracy changes with hyperparameters in our method.

- For any λ, the accuracy is best at L = 5.
- It is the highest value when λ = 1 that means that the peak of the time lag is one second.
- Our results can be close to the results reported in previous studies [19, 34] in the field of brain computing.



The human cognition process can be well represented using the time lag in the proposed method.

Acknowledgement : This work was partly supported by the KAKENHI Grant Number JP17H01744 and JP21H03456