 

Paper

# Monitoring Emotion by Remote Measurement of Physiological Signals Using an RGB Camera

Genki Okada†, Taku Yonezawa†, Kouki Kurita† and Norimichi Tsumura† (member)

**Abstract** In this paper, we propose a method of emotion monitoring using physiological signals such as RR intervals and blood volumes obtained by analyzing hemoglobin concentrations from facial color images. Emotion monitoring has great potential in areas such as market research, safety, medical and robot systems. The most popular method of emotion monitoring uses physiological signals. However, it is difficult to apply the commonly used methods in practice because special instruments such as electrodes or laser speckle flowgraphy are required to obtain physiological signals. The proposed method uses a simple RGB camera. Using 27 features calculated from the physiological signals obtained from facial RGB images, we classified five emotional states, amusement, anger, disgust, sadness and surprise, with 94% accuracy.

**Keywords**: emotion, physiological signals, RGB camera, remote measurement, heart rate variability.

## 1. Introduction

Better products and services are developed by collecting feedback from consumers and analyzing their trends or preferences. However, is difficult to provide suitable products or services for consumers because the answers in the database may not reflect their real emotions. Some people hide or do not recognize their real emotions in certain situations. Objective detection of the consumer's emotional state could improve the consumer feedback database. Moreover, emotion recognition could be useful for preventing potential accidents or crimes by incorporating it into cars or surveillance cameras. Emotion recognition technique has been the subject of extensive study. Many researchers have attempted to achieve emotion recognition using facial expressions[1, 2, 3], voices[4, 5, 6] and, in particular, physiological signals[7, 8, 9, 10, 11]. Studies on physiological psychology have revealed a strong association between the physiological response of the autonomic nervous system and the human emotional state. Furthermore, physiological signals are less affected by social and cultural differences[12]. It is possible to estimate the emotions that people try to hide or that they cannot even recognize in themselves.

Kashima et al.[7] used laser speckle flowgraphy to measure the blood flow in the facial skin of 16 healthy participants before and after experiencing the five tastes, sweet, sour, salty, umami and bitter. Their results showed unique facial skin blood flow patterns for each taste stimuli. Park et al.[8] used electrodes to measure physiological signals such as skin temperature, electrodermal activity, photoplethysmogram and electrocardiogram in 12 healthy participants before and after they watched movies that elicited seven emotions (happiness, sadness, anger, fear, disgust, surprise and stress). They selected useful features for emotion recognition by means of particle swarm optimization of the features obtained by analysis of the measured physiological signals. The results showed that the seven emotions were classified with around 90% accuracy, thus demonstrating that it is possible to classify emotions using physiological signals. However, these methods are impractical because they use special measuring devices such as laser speckle flowgraphy or contact-type devices. Moreover, the use of contact devices such as electrodes can be uncomfortable and stressful for participants.

Kurita et al.[9] developed a remote heart rate variability (HRV) measurement system using an RGB (Red, Green and Blue) camera to analyze hemoglobin concentrations from facial color images. They identified whether participants were relaxed or stressed by performing a frequency analysis on the HRV. This study demonstrated that it is possible to detect stress without causing unnecessary discomfort to participants. However, this method could not detect the concrete emotions that caused stress.

In this paper, we propose a method of monitoring

Received June 30, 2017; Accepted November 13, 2017
†Graduate School of Advanced Integration Science, Chiba University (Chiba, Japan)

specific emotions using an RGB camera that is practical to use. We used physiological signals such as heart rate variability and blood volume obtained by analyzing facial color images taken before and after the participants watched films that elicited emotions. In Section 2, we describe our no-contact technique for measuring physiological signals using the hemoglobin pigment separation of facial images in the previous study[9]. In Section 3, we explain the features obtained from the analysis of the measured physiological signals. In Section 4, we describe the experiments in which we measured the participants' physiological signals while they were emotionally aroused. In Section 5, we show the emotion classification results using the obtained features. In Section 6, we discuss about our results. Finally, in Section 7, we present our conclusion and future works.

## 2. Method of Remote Measurement of Physiological Signals

Various methods of pulse wave measurement using an RGB camera have been proposed [9, 10, 16]. The pulse wave signal changes with the hemoglobin concentration on the surface of the face. Therefore, in this paper, we treat the change in the average pixel value of the hemoglobin component images obtained using the skin pigment separation on the RGB pixel values of facial images as a pulse wave.

Figure 1 illustrates the multilayer structure of human skin, which can be roughly divided into the epidermis and dermis. In practice, the boundary surface of each layer has an irregular shape; however, we treat it as a flat plane for simplicity. Human skin contains melanin and hemoglobin pigments that affect the color tone of the skin. Melanin pigments exist in the epidermis and hemoglobin pigments in the dermis and thus can be regarded as spatially independent. A light incident on the human skin is divided into surface reflection light and internal reflection light that is emitted to the outside of the skin after repeatedly absorbed and scattered inside the skin. Surface reflection light represents the color of the light source, whereas internal reflection light represents the color of the skin. In this study, the images were taken without surface reflection light by placing the polarizing plates in front of the camera and the light source orthogonal to each other. When the modified Lambert-Beer law is assumed to be established with respect to the observation signal that is reflecting the light, the observation signal can be represented by the following equation by logarithmic conversion from the image space to the density space:

$$\boldsymbol{v}^{\log}(x,y) = -\rho_m(x,y)\boldsymbol{\sigma_m} - \rho_h(x,y)\boldsymbol{\sigma_h} + \rho^{\log}(x,y)\boldsymbol{1} + \boldsymbol{e}^{\log} \quad (1)$$

where $\boldsymbol{v}^{\log}$ is the converted observation signal; $(x, y)$ is the pixel location; $\rho_m$ and $\rho_h$ are the concentration of melanin and hemoglobin pigment, respectively; $\boldsymbol{\sigma_m}$ and $\boldsymbol{\sigma_h}$ are the absorption cross section of melanin and hemoglobin pigment, respectively; $p^{\log}$ is a shading parameter for the shape of the skin; $\boldsymbol{1}$ is a vector of the strength of the shading; and elog is the bias vector. Hence, we can regard melanin and hemoglobin pigments as independent signals, as shown in Figure 2. Therefore, it is possible to obtain the distribution of the melanin and hemoglobin pigment concentrations from the RGB values of the facial images.

Figures 3 (b) and (c) show the melanin and hemoglobin pigments and Figure 3 (d) shows the shading extracted by independent component analysis of the whole facial image shown in Figure 3 (a). The images were obtained without surface reflection light using polarizing plates. Figure 4 (a) is the facial image

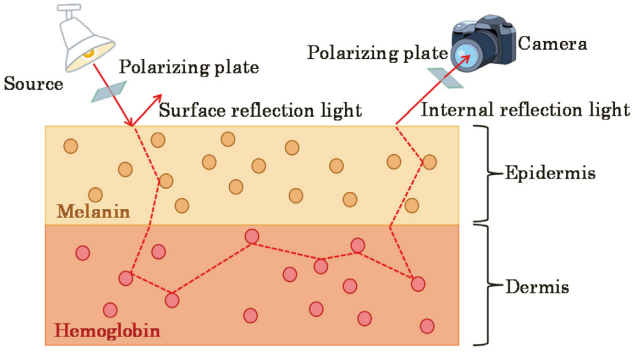


**Fig. 1** Movement of the light incident on skin.

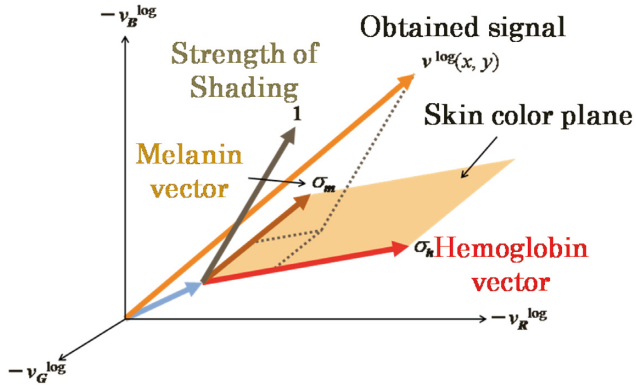


**Fig. 2** Obtained signal and the three independent signals.

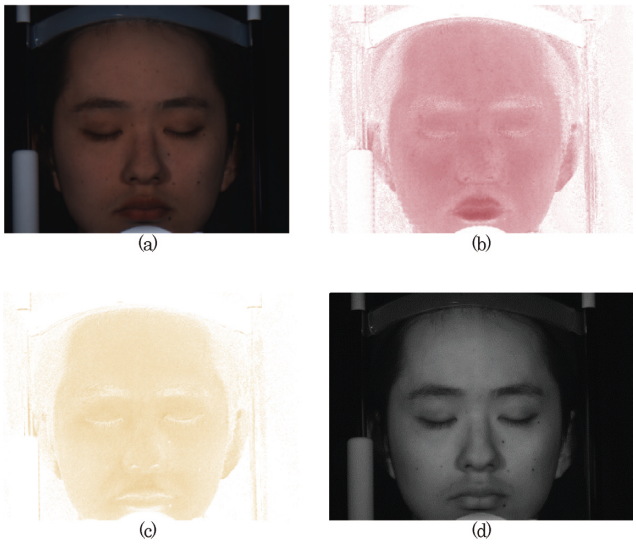


**Fig. 3** Skin pigment separation results for internal reflection image; (a) Original, (b) Hemoglobin, (c) Melanin, (d) Shading.

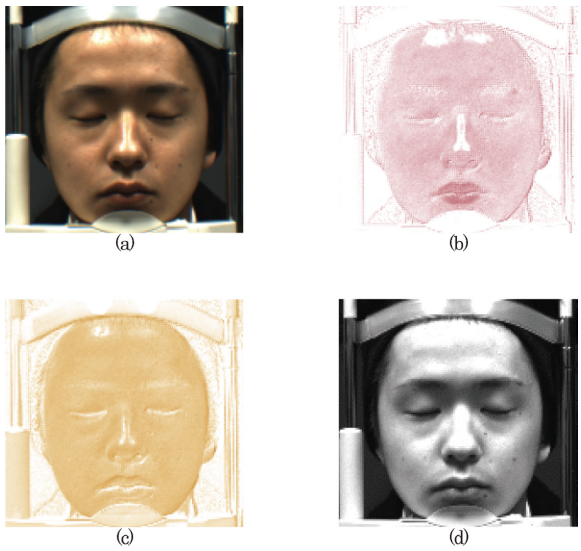


**Fig. 4** Skin pigment separation results for image taken under fluorescent lamps; (a) Original, (b) Hemoglobin, (c) Melanin, (d) Shading.

taken under fluorescent lights. When the facial image contains the surface reflection light, we can also apply skin pigment separation as shown in Figure 4 (b), (c), (d) using each pigment component color vector estimated from the internal reflection image shown in Figure 3 (a).

The change in the average pixel values in the hemoglobin component images in a specific region of interest (ROI) represents the change in the blood volume. Figure 5 shows the selected ROIs for measurement of heart rate variability. The peaks of the signal correspond to the peaks of the electrocardiogram waveform called the R wave. The intervals between R

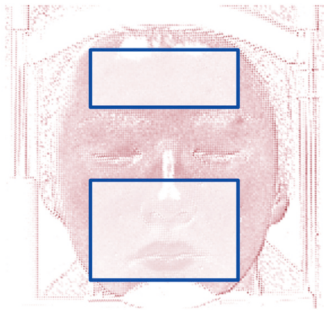

**Fig. 5** The selected ROIs for heart rate variability.

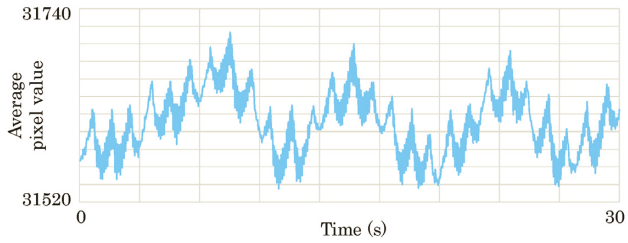


**Fig. 6** Average pixel values of hemoglobin component images.

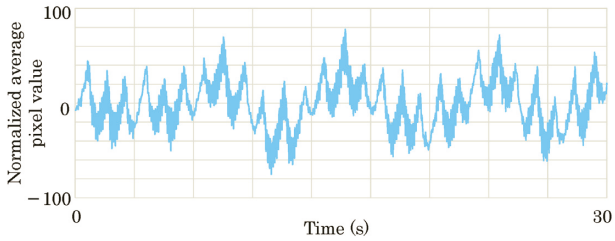


**Fig. 7** Normalized, detrended signal.

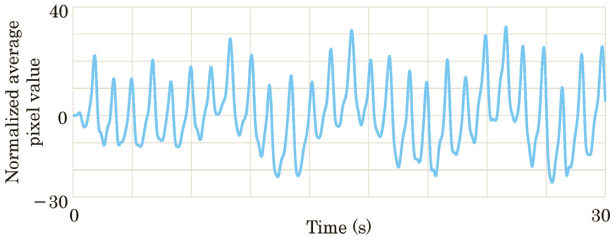


**Fig. 8** Bandpass filtered signal.

waves are called RR intervals and are important for heart rate analysis. To make it easier to detect the peaks, the signal was detrended[13] and a bandpass filter with a Hamming window was applied. The RR intervals were calculated by applying peak detection in the filtered signal. Figure 6 shows the change in the average pixel values over time in the forehead and cheek areas in the hemoglobin component images. Figures 7 and 8 show the detrended and filtered signals.

## 3. Feature Extraction

### 3.1 Heart Rate Variability

HRV is the variability in successive heartbeat (RR) intervals, which is controlled by the sympathetic and parasympathetic parts of the autonomic nervous system. The features used for emotion classification can be obtained by analyzing the RR intervals to estimate the function of the autonomic nervous system. Figure 9 shows the RR intervals obtained by calculating the intervals between the peaks of the filtered signal.

Time-domain methods are easy to perform because they analyze the RR intervals directly. The easiest features to obtain are the average and standard deviation of the RR intervals and the heart rate. The standard deviation of the RR intervals reflects the overall change, while the root mean square of successive differences (RMSSD) reflects the short-term fluctuations.

The NN50, which is the number of successive RR intervals that differ by more than 50 ms, and the pNN50, which is the relative value corresponding to the total number of successive RR intervals, are also used as indications of parasympathetic activity.

In addition to these statistical features, geometrical features are obtained by analyzing the histogram of the RR intervals,[14] shown in Figure 10. The RRtri is the integral of the histogram of the RR intervals (the total number of RR intervals) divided by the maximum value of the density distribution (Y). The triangular interpolation of the NN interval histogram (TINN) is the base of the triangle used to approximate the histogram of RR intervals (M-N).

Frequency-domain methods analyze the power spectral density (PSD) of the RR intervals. The features obtained from the PSD are commonly used as an indicator of autonomic nervous system activity. Here, the PSD is calculated using fast Fourier transform (FFT) based on Welch's periodogram method and autoregressive (AR) model[15].

The high frequency (HF: 0.15-0.4 Hz) component of the HRV reflects the respiratory sinus arrhythmia affected by respiratory and parasympathetic activity. Meanwhile, the low frequency component (LF: 0.04-0.15 Hz) represents the Mayer wave originating from both sympathetic and parasympathetic activity. In this paper, the integral value of HF and LF in the PSD calculated by the FFT and AR method, the percentage of HF and LF in the entire PSD, the normalized values using only LF and HF, and the ratio of LF to HF were used as the features for emotion classification.

It is reasonable to assume that a nonlinear mechanism affects the HRV because the control system of the heart is very complex. Nonlinear methods using Poincarè plots are commonly used to analyze HRV. A Poincarè plot is a graph showing the correlation between successive RR intervals. The some features for emotion classification were obtained by quantifying shape of the plot. A general method of quantifying the shape is to apply an ellipse to the plot, as shown in Figure 11. The standard deviation of the points along the minor axis, represented by SD1, reflects the short-term fluctuations due to respiratory sinus arrhythmia, and the standard deviation of the points along the major axis, represented by SD2, reflects long-term variations.

### 3.2 Facial Skin Blood Volume

The blood volume in the forehead and cheeks shows

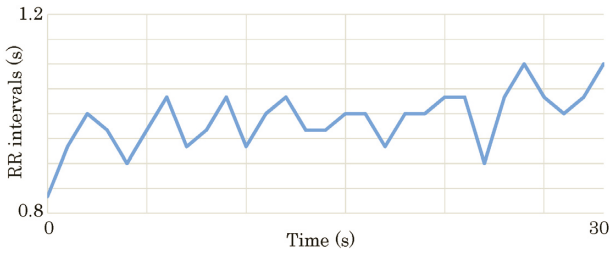


**Fig. 9** RR intervals.

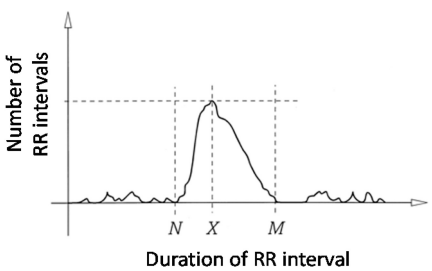


**Fig. 10** Histogram of RR intervals[14].

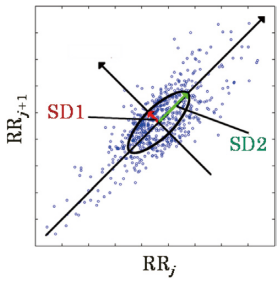


**Fig. 11** Poincarè plot.

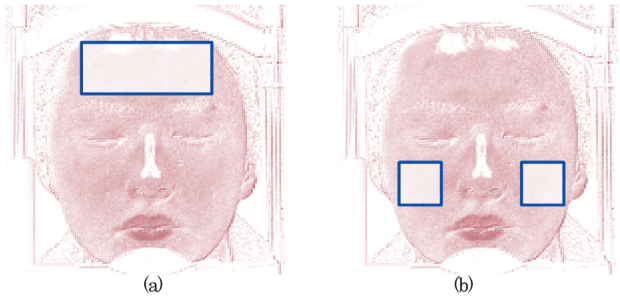


**Fig. 12** The selected ROIs; (a) Forehead, (b) Cheeks.

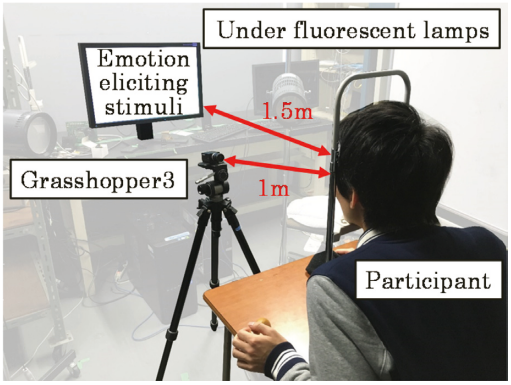


**Fig. 13** The experimental setup.

different changes for each region when we experience a taste or a negative emotion[7]. Therefore, by selecting two ROIs on the forehead and cheeks, we obtained two average values for the hemoglobin component in the ROI over a 10-second period. Figure 12 shows the selected ROIs in the hemoglobin component image.

Each feature has a different influence on the classification because each has a different unit (e.g. ms, beats/ms, %). We obtained 27 features that were normalized in the range [0:1].

## 4. Experiment

In this section, we describe the emotion classification experiment using the features calculated by analyzing the physiological signals obtained from the facial images. Seven healthy male college students in their 20s participated in this experiment. Figure 13 shows the experimental setup. The experiments were carried out under fluorescent lights. The RGB camera [Grasshopper3: Point Grey] which is capable of capturing 1920 x 1200 images at 30 FPS, was placed 1 meter from the participants and the 27 inch monitor was set 1.5 meters from the participants. The participants' faces were fixed using a chin rest because it is difficult to obtain accurate RR intervals if the participant moves.

Before the experiment, the procedure was explained to the participants and they were given time to make themselves comfortable. The images of their faces were taken for 40 seconds prior to the presentation of the movies as the baseline state and for 33 to 214 seconds while the movies were presented, then for 40 seconds after presentation of the movies. Participants reported the emotion that they experienced while watching the movies and the scene in which the emotion was most strongly expressed. This procedure was repeated for each emotion.

The RR intervals were obtained from 30 seconds of data in the baseline state and the emotional state. The emotional states were determined from the participants' reports. Skin blood volume was obtained for 10 seconds from each state. The differences between the features in the baseline states and the emotional states were used for the emotion classification.

Various methods have been designed to elicit emotions in the laboratory, such as music, pictures and movies. Movies elicit strong emotions due to their dynamic visual and auditory stimuli. In this study, we used movies that have a universal capacity to elicit six emotions: amusement, anger, disgust, fear, sadness and surprise[16].

The *k*-nearest neighbor method is a simple and easy to implement machine learning algorithm. Seventy percent of the features were randomly selected for training and the rest were used as testing data. Features with values outside the range of the mean ± standard deviation for each feature in the training data were excluded because the accuracy of the *k*-nearest neighbor method is largely reduced by the noisy features. In the classification step, *k*-nearest training data were selected by calculating the Euclidean distances between the testing data and the training data in the feature space. The testing data were classified into the majority emotion in *k*-nearest training data. In some cases, two or more emotions were equally common. Therefore, the number of emotions was counted by weighting the inverse of the distance between the training and testing data. The classification was repeated 10 times by randomly selecting training data. The classification accuracy was calculated by taking the average classification success rates.

The classification accuracy can be improved by selecting useful features. Individual optimization is one of the easiest methods for evaluating and selecting features. We calculated the classification accuracy by excluding one feature at a time from the rest to evaluate each one individually. Lower classification accuracy

when a certain feature is excluded means that the feature has a strong influence on the classification accuracy. Therefore, the nine features with lower classification accuracy used for the emotion classification.

## 5. Results

Figure 14 shows the classification accuracy computed using all of the features. The highest accuracy using all features for the six emotions was 52.5% when $k$ = 4. The accuracy for every emotion except fear was more than 50% when $k$ = 4. The accuracy for fear was remarkably lower than for the other emotions.

The nine features with lower accuracy determined by the individual optimization results is the average of the RR intervals, the absolute power of HF in the FFT method, the skin blood volume of the cheeks, the absolute power of LF in the AR method, the standard deviation of heart rate, pNN50, the ratio between HF and LF in the AR method, TINN and SD2.

As shown in Figure 15, the highest classification accuracy obtained using the nine features selected by individual optimization and excluding fear was 94% when $k$ = 4. Each emotion was classified with around 90% accuracy when $k$ = 4 or $k$ = 5.

## 6. Discussion

The classification accuracy using all of the features for the six emotions was similar to that in the previous research using contact-type measurement equipment. The low accuracy for fear seems to be due to the movie used to elicit fear. The movie was scenes from the end of the suspenseful psychological thriller *Silence of the Lambs*, chosen as the fear arousal movie in 1995. Some participants could not understand the story line. Furthermore, the movie includes two scenes in which the tense female police officer finds it difficult to open a door and she progresses slowly toward the dark basement. The former might elicit amusement about somebody making a mistake and the latter might elicit tension. Consequently, the movie might not have elicited fear in the participants.

Therefore, we classified the five emotions (excluding fear) using the nine selected features. The standard deviations of most of the selected features were low. The features were different from those selected in the previous study, probably because of the difference in the methods used, such as the number of participants and

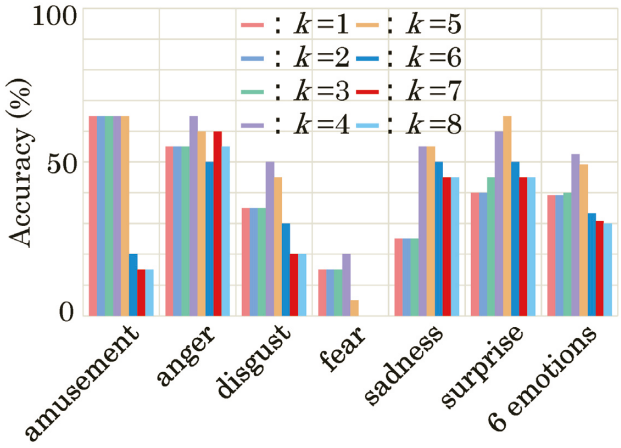


**Fig. 14** Accuracy when using all features.

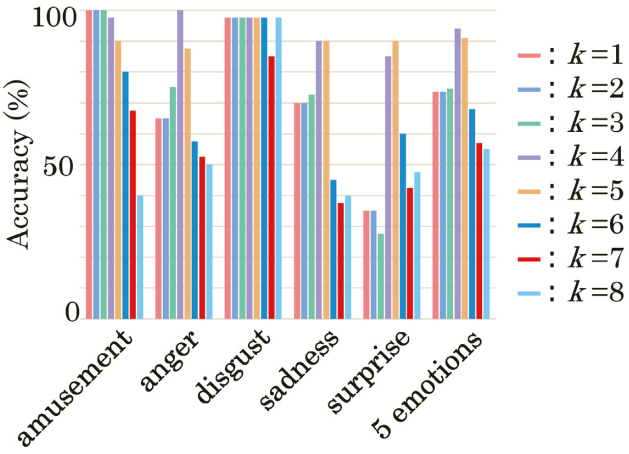


**Fig. 15** Accuracy when using selected features, excluding fear.

features, the movies used for emotional arousal, the measurement equipment and feature selection. The accuracy improved considerably to around 90% when $k$ = 4 or 5. Too small or too large a value of $k$ reduces the classification accuracy due to the strong effects of noise features.

## 7. Conclusion and future works

We obtained the physiological signals from facial RBG images by extracting hemoglobin concentrations while the participants watched movies selected to elicit emotion. The physiological signals were used as the features for emotion classification. Moreover, we accurately classified five emotions using the features selected by the individual optimization method.

In our future works we aim to improve the accuracy of fear classification using different experimental stimuli or by building the correspondence to the participant's movement.

## References


1) M. Yeasin, B. Bullot and R. Sharma, "Recognition of facial expressions and measurement of levels of interest from video," in IEEE Transactions on Multimedia, 8, 3, pp.500-508, June 2006
2) P. Lucey et al., "Automatically Detecting Pain in Video Through Facial Action Units," in IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), 41, 3, pp.664-674, June 2011
3) A. Chakraborty, A. Konar, U.K. Chakraborty and A. Chatterjee, "Emotion Recognition From Facial Expressions and Its Control Using Fuzzy Logic," in IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans, 39, 4, pp.726-743, July 2009
4) R. Cowie et al., "Emotion recognition in human-computer interaction," in IEEE Signal Processing Magazine, 18, 1, pp.32-80, Jan 2001
5) Chul Min Lee and S.S. Narayanan, "Toward detecting emotions in spoken dialogs," in IEEE Transactions on Speech and Audio Processing, 13, 2, pp.293-303, Mar. 2005
6) G. Zhou, J.H. L. Hansen and J.F. Kaiser, "Nonlinear feature based classification of speech under stress," in IEEE Transactions on Speech and Audio Processing, 9, 3, pp.201-216, Mar 2001
7) Kashima H and Hayashi N. "Basic taste stimuli elicit unique responses in facial skin blood flow." PLoS ONE 6: e28236(2011)
8) B.-J. Park, E.-H. Jang, S.-H. Kim, C. Huh and J.-H. Sohn, "Seven emotion recognition by means of particle swarm optimization on physiological signals: Seven emotion recognition," in Proc. 9th IEEE ICNSC, Apr. 2012, pp.277-282
9) Kurita K, Yonezawa T, Kuroshima M and Tsumura N, "Non-Contact Video Based Estimation for Heart Rate Variability Spectrogram using Ambient Light by Extracting Hemoglobin Information," Color and Imaging Conference, Volume 2015, Number 1, Oct. 2015, pp.207- 211
10) M.Z. Poh, D.J. McDuff and R.W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," in IEEE Transactions on Biomedical Engineering, 58, 1, pp.7-11, Jan. 2011
11) G. de Haan and V. Jeanne, "Robust Pulse Rate From Chrominance-Based rPPG," in IEEE Transactions on Biomedical Engineering, 60, 10, pp.2878-2886, Oct. 2013
12) O. Alaoui-Ismaili, O. Robin, H. Rada, A. Dittmar and E. Vernet-Maury, "Basic emotions evoked by odorants: comparison between autonomic responses and self-evaluation," Physiology and Behavior, Vol.62, pp.713-720(1997)
13) M.P. Tarvainen, P.O. Ranta-aho and P.A. Karjalainen, "An advanced detrending method with application to hrv analysis," Biomedical Engineering, IEEE Transactions on, 49, 2, pp.172-175(2002)
14) Task force of the European society of cardiology and the North American society of pacing and electrophysiology. Heart rate variability - standards of measurement, physiological interpretation and clinical use. Circulation, 93(5):1043{1065, Mar. 1996
15) S.L. Marple. Digital Spectral Analysis. Prentice-Hall International(1987)
16) Sato, W., Noguchi, M. & Yoshikawa, S.: Emotion elicitation effect of films in a Japanese sample. Soc. Behav. Personal., 35: 863-874(2007)


**Genki Okada** received his BE degree from Chiba University. He is currently a graduate student at Chiba University.

**Taku Yonezawa** received his BC and ME degree from Chiba University.

**Kouki Kurita** received his BC and ME degree from Chiba University.

**Norimichi Tsumura** received his B.E., M.E., and Dr. Eng. degrees in Applied Physics from Osaka University in 1990, 1992, and 1995, respectively. He is currently an associate professor in the Department of Information and Image Sciences, Chiba University (since February 2002).