https://doi.org/10.2352/ISSN.2169-2629.2017.32.147

Emotion Monitoring Using Remote Measurement for Physiological Signals by Camera

Genki Okada[1)]*, Taku Yonezawa*[1)]*, Kouki Kurita*[1)]*, Norimichi Tsumura*[1)]
1) Graduate School of Advanced Integration Science, Chiba University, CHIBA, JAPAN

Abstract

In this paper, we propose the method of emotion monitoring using physiological signals such as RR intervals and blood volumes obtained by analyzing hemoglobin concentration from facial color images. The emotion monitoring has a great potential in areas such as market research, safety measure, medical, and robot systems. Recently, the most popular method of emotion monitoring is by using physiological signals. However, the previous studies are not practical due to using special instruments such as electrodes or laser speckle flowgraphy to obtain physiological signals. Therefore, the proposed method uses simple RGB camera. Using 27 kind of features calculated from the physiological signals from the camera, we classified five different emotional states (amusement, anger, disgust, sadness and surprise). As a result, we could classify emotions with 94% accuracy by selecting features.

1. Introduction

Better products or services are developed by providing feedback from consumer and analyzing trends or preference of consumers. It is difficult to provide suitable products or services for the consumer since the database includes different answers from the consumer's original emotion. Some people cannot tell their real emotion depending on their position and the situation. Objective detection of the consumer's emotional state make it possible to improve the database of feedback from consumer. Moreover, the function of emotion recognition can be the prevention of potential accidents or crime by attaching to cars or surveillance cameras. This technique for emotion recognition has been studied for a long time. Many researchers have attempted to realize emotion recognition using such as facial expressions, voices and physiological signals. Especially, physiological signals attract the most attention in recent years for emotion recognition. In the study of physiological psychology, it is known that there is strong correlation between the physiological response by the action of the autonomic nervous system and human emotional state. Furthermore, physiological signals are less affected by the social and cultural differences [1]. We can estimate the original emotions that the people were trying to hide or that could not be recognized even in themselves.

Hayashi *et al.* [2] measured facial skin blood flow of 16 healthy participants before and after they felt the five taste (sweet, sour, salty, umami and bitter) by the laser speckle flowgraphy. As a result, they showed that facial skin blood flow changed unique to each taste stimuli. Park *et al.* [3] measured physiological signals such as skin temperature, electrodermal activity, photoplethysmography and electrocardiogram of 12 healthy participants before and after they watched the movies that are elicit seven emotions (happiness, sadness, anger, fear, disgust, surprise, and stress) by the electrodes. As a result, they classified seven emotions with around 90% accuracy by selecting useful features for emotion recognition by means of particle swarm optimization to features that were obtained by analyzing the measured physiological signals. In this way, it is possible to classify emotions using the physiological signals. However, these studies are not practical because they used special measuring devices such as laser speckle flowgraphy or contact type devices. Moreover, the contact type devices might be uncomfortable for the participants because external factors such as an electrode make participants feel the stress by giving a burden.

Kurita *et al.* [4] realized remote heart rate variability (HRV) measurement system using RGB camera by analyzing hemoglobin concentration from facial color images. They identified if participants were relaxed or stressed by performing frequency analysis on the heart rate variability. In this study, it become possible to detect stress without causing unnecessary discomfort to the participants. However, this method could not detect the concrete emotion that caused stress.

In this paper, we propose the method of monitoring of the specific emotions using RGB camera which is practical by used. Our method uses physiological signals such as heart rate variability and blood volume obtained by analyzing facial color images before and after the participants watch the films that elicit the emotions. In Section 2, we describes the non-contact physiological signals measurement technique using the hemoglobin pigment separation of the facial image in the prior study [4]. In Section 3, we will explain the features that are obtained by analyzing measured physiological signals. In Section 4, we perform the experiments to measure the physiological signals while we were giving the emotional arousal stimulus to the participants. The results of the classification for emotion using the obtained features are also described in this section. Finally, in Section 5, we describe the conclusion and future tasks.

2. Method of Remote Measurement of Physiological Signals

Various methods of pulse wave measurement were proposed using a digital camera. The pulse wave signal changes with hemoglobin concentration of the face surface. Therefore, in this paper, we treat the change of the average pixel value of the hemoglobin component images obtained by using the skin pigment separation on RGB pixel values of facial images as pulse wave.

Figure 1 shows the model of human skin. Human skin is multilayer structure that can be roughly divided into the epidermis, dermis and subcutaneous tissue. In practice, the boundary surface of each layer has an irregular shape. However, we treat the boundary surface as plane shape for simplicity. Human skin contains melanin and hemoglobin pigments. Color tone of human skin is greatly affected by these pigments. Melanin pigments exist in epidermis and hemoglobin pigments exist in dermis. Therefore, melanin and hemoglobin pigments can be regarded as being present in the spatially independent by assuming that epidermis is melanin layer and dermis is hemoglobin layer. Light incident on the human skin is divided into surface reflection light and internal reflection light emitted to the outside of the skin after repeated absorption and scattered inside the skin. While surface reflection light represent the color of the light source such as gloss, internal reflection light represent the color of the skin. In this paper, we take the images without surface reflection light by using polarizing plates placed 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 the reflecting light, the observation signal can be represented by the following equation by logarithmic conversion from image space to density space.

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

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

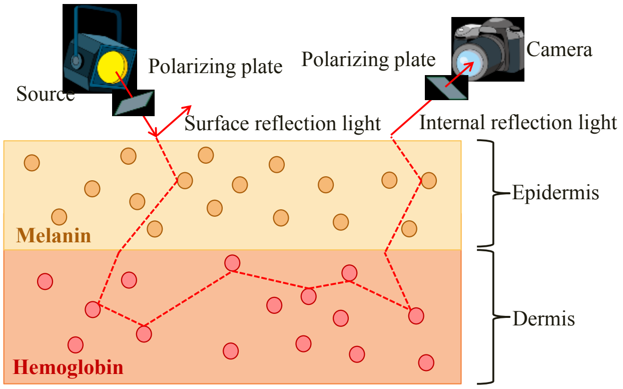


Figure 1 Movement of the light incident on the skin model

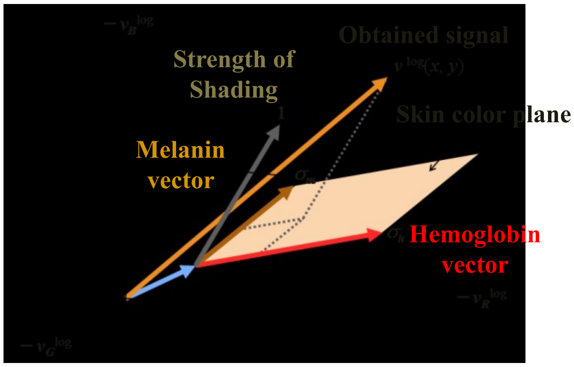
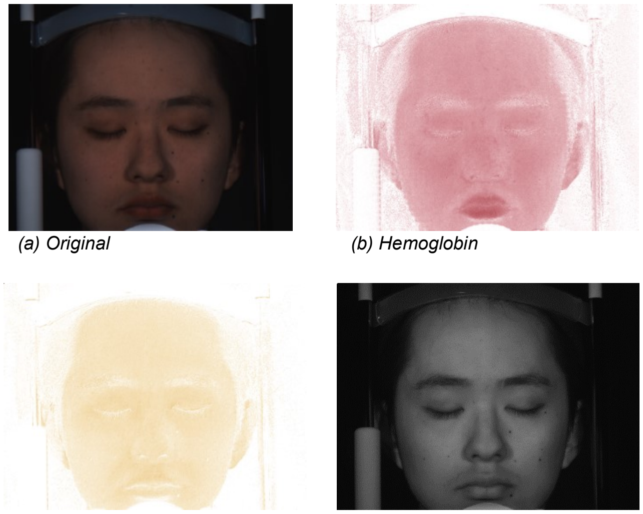


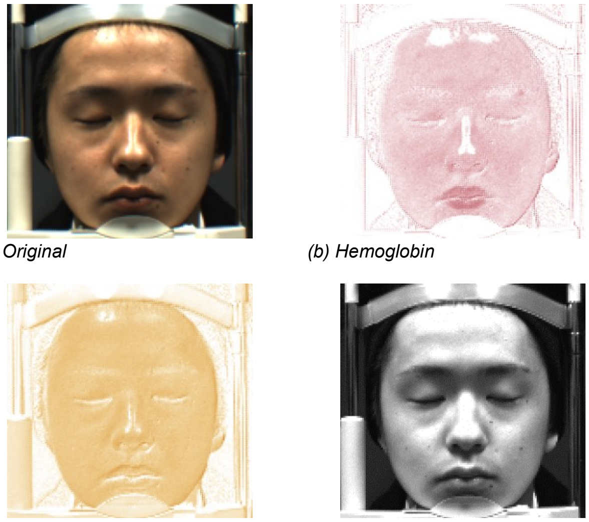
Figure 2 The obtained signals and the three independent signals

Figure 3 (b) and (c) are the extracted melanin and hemoglobin pigments and Figure 3 (d) is shading in whole facial image shown in Figure 3 (a) by independent component analysis. These images are obtained without surface reflection light using polarizing plates. Figure 4 (a) is the facial image taken under fluorescent lights. In the case that 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).



(a) Original
(b) Hemoglobin
(c) Melanin
(d) Shading

Figure 3 The result of skin pigment separation for internal reflection image; (a) Original, (b) Hemoglobin, (c) Melanin, (d) Shading



(a) Original
(b) Hemoglobin
(c) Melanin
(d) Shading

Figure 4 The result of skin pigment separation for fluorescent lamps image; (a) Original, (b) Hemoglobin, (c) Melanin, (d) Shading

The change of the average pixel values in specific region of interest (ROI) in the hemoglobin component images represents the signal of the blood volume change. The peaks of the signal correspond to the peaks of electro cardiogram waveform called R wave. The intervals between R waves are called RR intervals that is important for heart rate analysis. In order to make it easy to the peak detection, the signal was performed detrending [5] and applied a bandpass filter with a Hamming window. The RR intervals were calculated by the peak detection with respect to the filtered signal. Figure 5 shows the change along the time of the average pixel values in the area of forehead and cheek in the hemoglobin component images. Figure 6 shows the detrended and filtered signal.

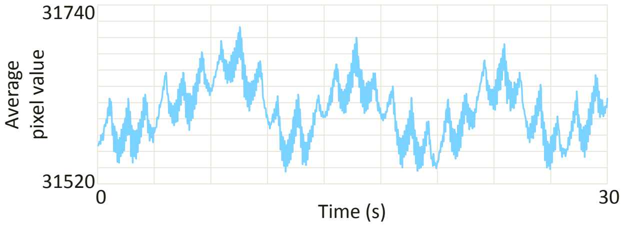


Figure 5 Average pixel values of hemoglobin component images

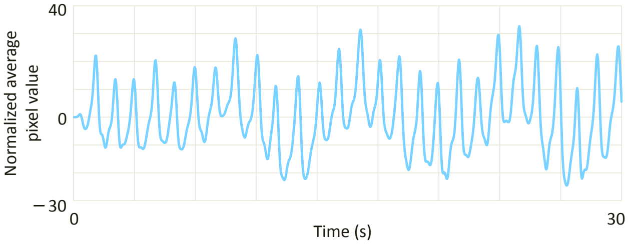


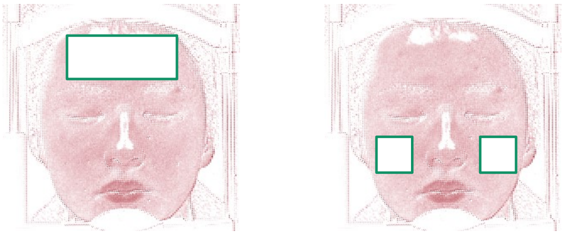
Figure 6 Normalized, detrended and filtered signal

3. Feature Extraction

In this section, we describe the feature values for the emotion classification obtained by analyzing the hemoglobin component images.

3.1 Facial Skin Blood Volume

When we feel the taste or the negative emotion, each blood volume in the forehead and cheeks changes differently [2]. Therefore, by setting two ROI on the forehead and cheeks, we can obtain the two average values of hemoglobin component in the area of ROI during detected 10 seconds. Figure 7 shows the region of the set ROI in the hemoglobin component image.



(a) Forehead *(b) Cheeks*
Figure 7 The region of the set ROI; (a) Forehead, (b) Cheeks

3.2 Heart Rate Variability

The heart rate variability that is the variability of successive heartbeat (pulse) intervals is controlled by the sympathetic and parasympathetic of the autonomic nervous system. The features that are used to emotion classification can be obtained by analyzing the RR intervals to estimate the function of the autonomic nervous system.

The time-domain methods can be performed easily because they analyze the RR intervals directly. The average and standard deviation of the RR intervals and heart rate are the features that can be most easily obtained. The standard deviation of the RR intervals reflect the overall change. Meanwhile, the root mean square of successive differences (RMSSD) reflects the short-term fluctuations.

Furthermore, the NN50 that is the number of successive RR intervals having difference more than 50ms and the pNN50 that is the relative amount corresponding to the total number of successive RR intervals are also used as index of parasympathetic.

In addition to these statistical features, there are the geometrical features obtained by analyzing the histogram of RR intervals [6]. Figure 8 shows a histogram of the RR intervals. The RRtri is the integral of the histogram of the RR intervals (the total number of the RR intervals) divided by the maximum value of the density distribution (Y). The TINN (triangular interpolation of NN interval histogram) is the base of the triangle to approximate the histogram of RR intervals (M-N).

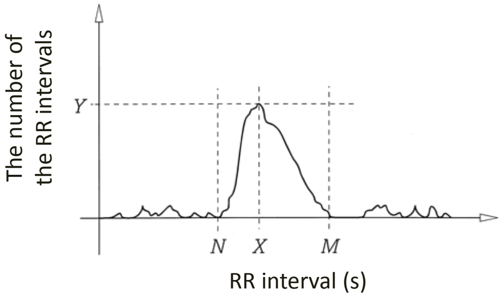


Figure 8 Histogram of RR intervals [6]

The frequency-domain methods analyze power spectral density (PSD) of the RR intervals. The features obtained from the PSD are commonly used as indicator of the action of the autonomic nervous system. In this paper, the PSD is calculated by using fast Fourier transform (FFT) based Welch's periodogram method and autoregressive (AR) model [7].

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

It is reasonable that nonlinear mechanism is assumed to have affected heart rate variability because the control system of the heart is very complex. The nonlinear methods using Poincarè plot are commonly used to analyze heart rate variability. The Poincarè plot is the graph showing the correlation between successive RR intervals. The shape of the plot is used as the features. General method to quantify the shape is to apply the ellipse to the plot as shown in Figure 9. The standard deviation of the minor axis direction represented by the SD1 reflects the short-term fluctuations due to respiratory sinus arrhythmia. Furthermore, the standard deviation of the major axis direction represented by SD2 is the feature due to long-term variations.

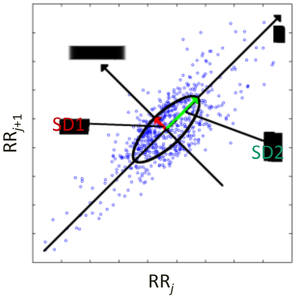


Figure 9 Poincarѐ plot

The influence on the classification would be different for each feature because the unit is different for each feature. In this paper, the obtained 27 features were normalized in the range [0:1].

4. Experiment

In this section, we discuss the experiments for emotion classification using the features obtained by analyzing the physiological signals from the facial images. Seven healthy males in their 20s college students participated in this experiment. Figure 10 shows the experiment environment. The experiments was carried out under fluorescent lights. The RGB camera was placed away 1 meter from the participants and the 27 inch monitor is set away 1.5 meters from the participants. The faces were fixed using the chin rest because it is difficult to obtain the RR intervals accurately if the participants move.

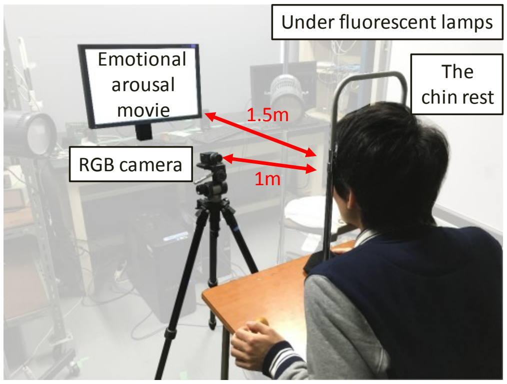


Figure 10 The experiment environment

Prior to the experiment, the participants were introduced the procedure of the experiments and they had an adaptation time to feel comfortable in the experiment system. Their faces were taken for 40 sec prior to the presentation of movies as baseline state and for 33 to 214 sec while the movies were presented, then for 40 sec after presentation of the movies. Participants reported the emotion that they experienced during presentation of the movies and the scene in which the emotion was most strongly expressed. This procedure was conducted on six movies for each emotion.

The obtained RR intervals were analyzed for 30 seconds from the baseline state and the emotional state. The emotional states were determined by subject's report. Skin blood volume was obtained for 10 seconds. The differences between the features from the baseline states and the emotional states was used for the emotion classification.

Various methods was designed to elicit emotions in the laboratory such as music, pictures and movies. Among them, the movies elicit emotions strongly by the dynamic visual and auditory stimulus. In this paper, the movies that have a universal capacity to elicit six emotions (amusement, anger, disgust, fear, sadness and surprise) are used for emotion elicitation [8].

The *k*-nearest neighbor method that is a simple and easy to implement machine learning algorithm. The 70% of the whole features were selected at random for training and the remaining features were used for testing data. The features whose values were out of range of mean ± standard deviation for each feature in the training data were excluded since the accuracy of the *k*-nearest neighbor method is largely reduced by the noise features. In the classification step, *k* nearest training data were selected by calculating the Euclidean distances between the testing data and all the training data in the feature space. The testing data was classified into the majority emotion of *k* training data. In some cases, the number of the most common emotion is more than 1. Therefore, the number of emotions was counted by weighting the inverse of the distance between training and testing data. The classification was repeated 10 times by selecting training data randomly. The classification accuracy was calculated by taking average of success rates of the classifications.

The classification accuracy can be improved by selecting useful features. Individual optimization which is the one of easiest method is used to select features determined by the evaluation of each feature. In this paper, the values of classification accuracy were calculated by excluding one feature at a time from the all features to evaluate each feature. The feature that has low classification accuracy when it is excluded has high influence on the classification accuracy. Therefore, the features having lower nine classification accuracy are used for the emotion classification.

5. Results

Figure 11 shows the classification accuracy computed by using the all features. The highest value of classification accuracy using all features in the six emotions is 52.5% at the time of $k = 4$. The classification accuracy in each emotion excepting fear is more than 50% when $k = 4$. In the case of fear, the accuracy is remarkably low compared to other emotions.

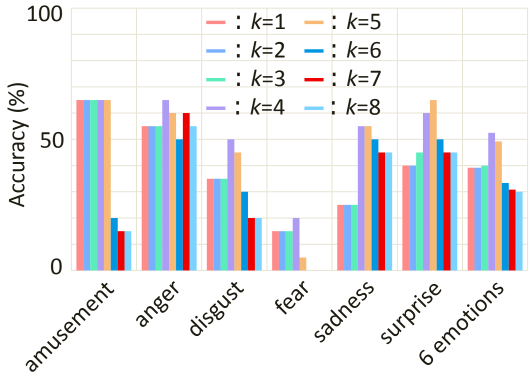


Figure 11 The accuracy when using all features

The nine features having lower accuracy determined by the result of individual optimization is the average of the RR intervals, the absolute powers of HF in FFT method, the skin blood volume of the cheeks, the absolute powers of LF in AR method, the standard deviation of heart rate, the corresponding relative amount of the number of successive RR intervals differing more than 50 ms, the ratio between HF and LF in AR method, the base of the triangle to approximate the histogram of RR intervals and the standard deviation of the major axis direction of the ellipse applied to the Poincarѐ plot.

As shown in figure 12, the classification accuracy obtained by using nine features selected by individual optimization and excepting fear is 94% where $k = 4$ is used since the highest accuracy is achieved. Each emotion is classified with around 90% accuracy when $k = 4$ or $k = 5$.

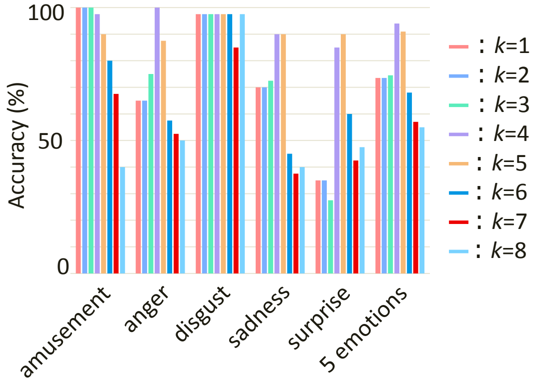


Figure 12 The accuracy when using selected features and excepting fear

7. Discussion

The classification accuracy using all features in the 6 emotions is similar to that in the previous research using contact type measurement equipment. The low accuracy of fear seems to be observed due to the movie that elicits fear. The movie that is a scene of Silence of the Lambs was chosen as fear arousal movie in 1995. The scene is the end of the suspenseful psychological thriller. Therefore, some participants could not understand the story line while watching the movie. Furthermore, the movie includes the scene that tense female police officer cannot open the door well and that she progresses slowly dark basement. The former might elicit amusement by somebody else's mistake and the latter might elicit tension. Consequently, the movie might not have elicited fear with precision.

Therefore, we classified the 5 emotions with the exception of fear using selected nine features. The standard deviations of the most of the selected features have low value. These features are different from the selected features in the previous research. This dissimilarity seems to be caused by the difference of the method between the present and the previous research such as the number of participants and features, emotional arousal movies, measurement equipment and feature selection. The accuracy has widely been improved to around 90% when $k = 4$ or 5. Too small or too large number of k lower the classification accuracy because classification accuracy strongly affected by the noise features.

8. Conclusion and Future works

We obtained the physiological signals from the facial RBG images by extracting hemoglobin concentration while the participants watched the emotion elicitation movies. The physiological signals were used for emotion classification as the features. Moreover, we classified 5 emotions accurately using the features selected by result of individual optimization method.

Our future works are the improvement of the accuracy of fear by using different stimulus from this experiment or building a more robust database and the correspondence to participant’s movement.

References

[1] 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.

[2] Kashima H and Hayashi N. "Basic taste stimuli elicit unique responses in facial skin blood flow." PLoS ONE 6: e28236, 2011.

[3] 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.

[4] 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, October 2015, pp. 207-211

[5] 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, vol. 49, no. 2, pp. 172–175, 2002.

[6] 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, March 1996.

[7] S.L. Marple. Digital Spectral Analysis. Prentice-Hall International, 1987.

[8] Sato, W., Noguchi, M. & Yoshikawa, S. : Emotion elicitation effect of films in a Japanese sample. Soc. Behav. Personal., 35 : 863-874, 2007.

Author Biography

Genki Okada was born in Tokyo, Japan, on 24 April 1993. He received the B.E degrees in Chiba University in 2016. He is interested in the color image processing, computer vision, computer graphics and biomedical optics.