[Poster Presentation] Measuring the advertisement effect based on emotional responses analysis with crowdsourcing

Kenta MASUI[†]　Genki Okada[‡]　and　Norimichi TSUMURA[†‡]

† Graduate School of Engineering, Chiba University　Yayoi-cho, Inage-ku, Chiba-shi, Chiba, 263-8522 Japan
‡ Graduate School of Advanced Integration Science, Chiba University　Yayoi-cho, Inage-ku, Chiba-shi, Chiba, 263-8522 Japan
E-mail: † k_masui@chiba-u.jp, ‡ tsumura@faculty.chiba-u.jp

Abstract　In this study, we estimate ad liking and purchase intent in the natural environment by using crowdsourcing and the remote measurement of facial expressions and physiological signals. In recent years the market of online video advertisement has expanded, and demand for measuring advertisement effect is increasing. We collected a significant number of videos of Japanese faces watching video advertisements via the Internet. By combining the multiple features of facial expressions and using support vector machine(SVM), we made it possible to estimate the advertisement effect. In addition, it was possible to improve the accuracy more than using only single-mode features.

Keyword　ad liking prediction,　facial expressions,　physiological signals,　RGB camera, remote measurement

1. Introduction

The market size of online video advertisement is rapidly expanding due to the spread of smartphones and social media. While advertisers focus on emotional content, they are struggling to measure the degree to which emotional induction succeeded. Traditionally, consumer tests of video contents, whether self-reports, facial expressions, or physiological signals, have been performed in a laboratory environment, but the subjects might be influenced by irrelevant factors such as the existence of the experimenter and the discomfort of the experimental environment. Since it is difficult to quantify the influence, it is required to measure the degree of emotion induction in a real environment that does not give discomfort to the subjects.

In this paper, we estimate the effectiveness of the effect measurement of the advertisement in subjects in the same situation as usual used by using crowdsourcing and contactless measurement methods of facial expressions and physiological signals. Recently, ad liking and purchase intent can be predicted from viewers' facial expressions and physiological responses [1]. Japanese people have brief facial expressions [2], it is useful to analyze physiological signals such as heart rate, which are effective for ambiguous emotion estimation, and to use multimodal features combined with facial expressions. First, we collected a significant number of videos of Japanese faces watching video advertisements using crowdsourcing. Next, facial expressions and physiological signals such as heart rate and gaze were remotely measured by analyzing the obtained videos. Finally, ad liking and purchase intent were estimated by combining the measured responses as multiple features and using SVM.

2. Data collection via the Internet

The facial video and self-reports were collected by crowdsourcing. We focused on products that are likely to be purchased frequently, not products that require long-term purchase decisions such as automobiles.

In this study, 411 Japanese people participated in a short period. Using crowdsourcing, we were able to gather the data from participants with a wide range of ages, genders, locations, occupations, household compositions, and income, and we were able to reduce the influence of self-selection bias.

3. Acquired features

First, 68 face parts were detected from each frame of the recorded facial video using Kazemi and Sullivan's method [3] and the facial expression feature were calculated. Figure.1 shows the detected face parts.

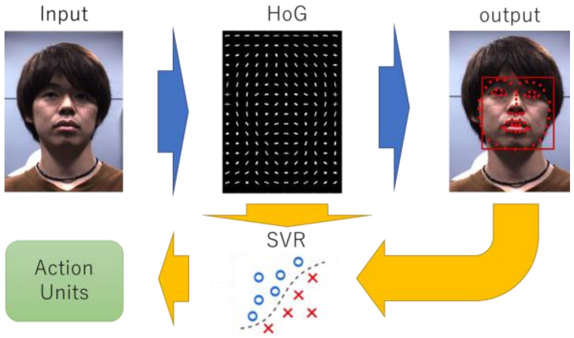


Figure.1 Acquisition of AUs

In addition, we calculated the histogram of oriented gradients (HOG) features of the detected face region. 17 action units (AUs) expressing the movement of the facial muscle are detected by using the coordinates of the facial parts and the HOGs. The time mean value, standard deviation, maximum value, minimum value, median value and entropy of each AUs were calculated.

Translation and rotation of the face were also calculated as features. The time average value, standard deviation, maximum value, minimum value, median value, entropy, skewness, and kurtosis were calculated.

Next, heart rate was measured remotely using the original method based on Poh et al.'s method [4]. Figure.2 shows the flow of the remote heart rate measurement.

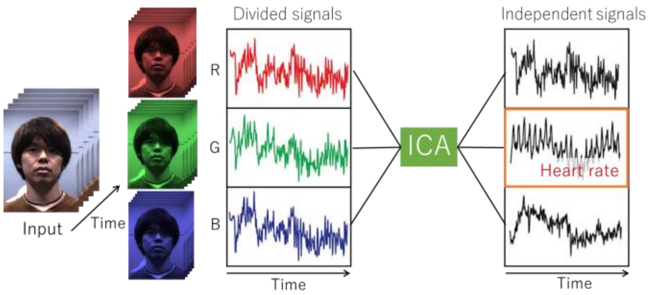


Figure.2 Extraction of heart rate signals

HRV, which is the variability of the continuous the interbeat intervals (IBI), is controlled by the sympathetic and parasympathetic nerves of the autonomic nervous system. The function of the autonomic nervous system can be evaluated by analyzing IBI in the time domain analysis and nonlinearly analysis. Time domain analysis and nonlinear analysis were executed on the estimated heart beat signal. After that, ten kinds of feature values were calculated.

Finally, we used the coordinates of the detected face parts to detect eye areas including the eyelids, iris, and pupil. When the positions of the eyes and the pupil are detected, the gaze vector of the eye is individually calculated for each eye. Time mean value, standard deviation, maximum value, minimum value, median value, entropy, kurtosis, skewness of each of the left and right gaze were calculated.

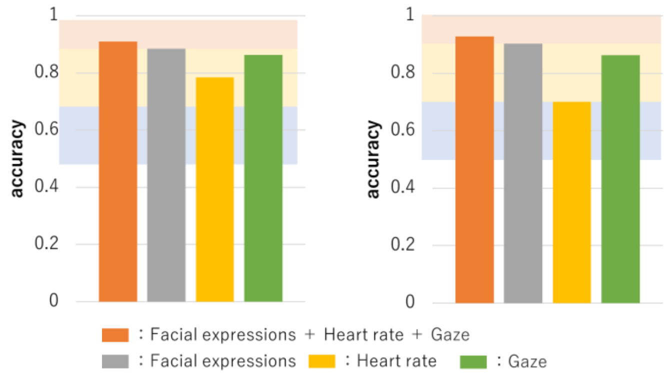
4. Classification

As the label for estimating the effectiveness of the advertisement, the purchase intent and the ad liking answered by the subject are used. The difference of purchase intent between preliminary survey and post-survey were calculated, and the amount of change in purchase intent due to advertisement viewing was acquired. For both purchase intent and ad liking, the data are divided into two groups to create a two-class classification problem . Classification is performed by an SVM using the calculated features and labels. In classification, the number of samples in each class is not uniform. Therefore, we solved this imbalance problem by oversampling using SMOTE[5]. In addition, ReliefF [6] was used for feature selection in order to reduce of calculation.

5. Results

Figure.3(a) shows the result of purchase intent estimation obtained by changing the threshold of SVM. When all three features were used, the estimation accuracy was the highest. Next, Figure.3(b) shows the result of ad liking estimation obtained by changing the threshold of SVM. As with purchase intent estimation, the estimation accuracy was the highest when all three feature values were used.

In purchase intent, the estimation accuracy of only the heart rate is higher than the ad liking. The difference between these results seems to support the finding of a previous study [1], which is that facial expression analysis is effective for stronger emotions such as ad liking while heart rate is more beneficial for subtle reactions and emotions such as purchase intent.



(a)Purchase intent (b)Ad liking

Figure.3 Estimation accuracy

6. Conclusion and future work

We estimated ad liking and purchase intent by acquiring and analyzing multiple features from face expressions when subjects watch video advertisements. We collected a significant number of data that is the facial responses when viewing video advertisements using crowdsourcing. Along with the face video, affective responses to video advertisements and products were collected by self-reports as labeling in SVM. Next, facial expression, heart rate, and gaze were measured from the collected facial video. Next, we estimated the ad liking and the purchase intent by combining each features. As a result, combining features indicates that the accuracy of both advertising effects will be higher than using only a single features.

Future work is to clarify the relationship between the acquired features and the contents of the video advertisements.

References

[1] P. Pham and J. Wang. "Understanding emotional responses to mobile video advertisements via physiological signal sensing and facial expression analysis." *IUI '17 Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pp. 67–78, (2017)

[2] J. M. Girard and D. McDuff. "Historical heterogeneity predicts smiling: evidence from large-scale observational analyses." *12th IEEE International Conference on Automatic Face and Gesture Recognition*, pp.719–726, (2017)

[3] V. Kazemi and J. Sullivan. "One millisecond face alignment with an ensemble of regression trees." *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874 (2014).

[4] M. Z. Poh, D. J. McDuff, and R. W. Picard. "Advancements in noncontact, multiparameter physiological measurements using a webcam." *IEEE Transactions on Biomedical Engineering*, 58(1):pp.7–11, (2011)

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. "SMOTE: Synthetic minority over-sampling technique." *Journal of Artificial Intelligence Research*, 16:pp.321–357, (2002)

[6] M. Robnik-Sikonja and I. Kononenko. "Theoretical and empirical analysis of ReliefF and RReliefF." *Machine Learning*, 53:pp.23–69, (2003)