Towards Estimating Emotions and Satisfaction Level of Tourist based on Eye Gaze and Head Movement

Dmitrii Fedotov*, Yuki Matsuda†‡, Yuta Takahashi†, Yutaka Arakawa†§, Keiichi Yasumoto†, Wolfgang Minker*
* Institute of Communications Engineering, Ulm University
Albert-Einstein-Allee 43, 89081 Ulm, Germany
Email: {dmitrii.fedotov, wolfgang.minker}@uni-ulm.de
† Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma City, Nara, Japan
Email: yukimat.jp@gmail.com, {takahashi.yuta.to2, ara, yasumoto}@is.naist.jp
‡ Research Fellow of Japan Society for the Promotion of Science
§ JST Presto, Japan

***Abstract*—Following the increase in demand for "smart tourism," various tourist information becomes available. Current tourist guidance systems can provide many possible routes around the city, which are usually based on an optimal distance and time or popularity of the place. To design more enjoyable tourist routes, we should be aware of tourist's perception of the urban environment. In this paper, we propose the methodology of estimating the tourist emotions and satisfaction level by analysing physiological features, such as head movement and eye gaze. Features derived from raw sensor data have the correlation up to 0.58 with emotion and satisfaction labels. This study also shows the differences in feature/label dependencies between different touristic areas and highlights the challenges of tourist satisfaction estimation.**

I. INTRODUCTION

Due to the ubiquity of smart devices, including smartphones and wearable devices, various real-time urban environment information (e.g., congestion degree, roadway traffic volume) becomes available everywhere in the city. This information helps to provide valuable tourist guidance for people, called "Smart Tourism" [1], [2]. Since the emotional status and satisfaction level differ across different users and even for the same user at a different time, the user status (e.g., emotions and satisfaction level) needs to be taken into account to design more context-aware tourist guidance systems.

To collect data on tourist emotions and satisfaction level, online user reviews and questionnaires, such as TripAdvisor[1], are widely used. However, it is difficult to keep users motivated to write the reviews, especially with medium rating values. Thus, to provide reliable information for tourists, it is necessary to collect quantitative data without user reviews. The aim of our research is to be able to estimate context effectively with inter- and intra-user status information.

Many related projects try to estimate emotion/satisfaction level of the people with various methods [3]–[7]. However, many of those have restrictions for applying to the real-world, such as data comprehensiveness and estimation accuracy.

In this paper, we propose the method of estimating tourist emotions and satisfaction level based on observation of the tourist behavior during the sightseeing by using wearable devices. To describe tourist behavior, we focused especially on eye gaze and head movement. The sense of sight is one of the most important sensory systems on sightseeing, and it can be tracked as an eye gaze using cutting-edge technologies. Also, due to the directivity on the sense of sight and hearing, the head movement can be affected by them. Furthermore, intelligent eyewear devices which can measure the eye gaze and head movement will be available as a consumer product in the near future.

Thus, we conducted preliminary experiments in the real-world conditions to confirm the feasibility of estimating emotion/satisfaction level from physiological features. As the experimental field, we have selected two touristic areas: located in Ulm (Germany) and Nara (Japan), which have completely different conditions. As a result, we have found a correlation between physiological data and emotion and satisfaction level and proved the feasibility of such research. The relationship between features and label can differ, depending on the characteristics of sightseeing area.

II. RELATED WORK

In current guidance systems, online user reviews and questionnaires are still widely used to collect tourist emotions and satisfaction level. TripAdvisor collects 5-star rating and comments from tourists as the user reviews about sightseeing spots. In data collecting methods that asks tourists directly, the motivation of posting reviews is essential to guarantee the quantity and quality. However, it also has the risk of skewing the evaluation due to the imbalance of reviewer's ratings. Many research adopt the questionnaire-based survey for measuring the tourist satisfaction [3]. However, methods relied on the questionnaire have problems in sustainability and spatial coverage of the survey.

To know user's emotion status, sensing technologies are often discussed. Resch et al. proposed the emotion collecting system for the urban planning, called "Urban Emotions" [5]. The paper describes that wrist-type wearable devices and social media were used for emotion measurements. However,

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

this approach relied on an assumption that posts on the social media are written in-situ. As the other approach, emotion recognition system based on the acoustic features via dialogue system on the mobile device have been proposed [6]. However, such method has not been able to achieve a realistic accuracy yet. Shapsough et al. describes that emotions can be recognized by using typing behavior on the smartphone [7]. This approach uses machine learning technique and induced high accuracy on emotion recognition. Yet it is not feasible to frequently ask users to type on the smartphone during the sightseeing tour. Ringeval et al. proposed to introduce physiological features in addition to audio-visual and to build a multimodal system that relies on combination of them [8]. As the physiological features, electrocardiogram and electrodermal activity were used. Physiological features have provided lower performance and weaker correlation than audio-visual with continuous emotional labels, but helped to increase the overall performance of multimodal system.

III. Estimation Method of Tourist Emotions and Satisfaction Level

In this study, we propose the method to estimate tourist emotions and satisfaction level by observing the tourist behavior with wearable devices during the sightseeing continuously.

A. Workflow

Our goal is to comprehensively estimate tourist emotions and satisfaction level in the cities. We expect that those can be extracted from tourist's various behaviors appeared unconsciously during a tour.

Fig. 1 shows the basic workflow of our proposed approach. The details of each step is explained as follows:

Step 1 - Split the whole tour into sessions: In the tourism domain, it is assumed that a tourist typically requests to provide the guidance information for each sightseeing spot. Thus, as the first step, we split the whole tour into the small periods (sessions) which include at least one sight.

Step 2 - Sensing and ground truthing: The tourists wear the wearable devices during a sightseeing tour. Those collect tourist behavior during the whole sightseeing continuously. At the end of each session, the system asks tourist about current emotions and satisfaction level and records these information as the ground truth. Then, the tourist repeats the same procedure for each session of the tour. Details of features and ground truth are described in the following section.

Step 3 - Building the estimating model: In our future research, we will build on estimation model of emotions and satisfaction level, based on tourist behavior features and ground truth data.

B. Features

In this study, we used physiological features, as they proved their usefulness in emotion recognition systems and can be collected continuously from the natural behavior [8]–[11]. Physiological features can be separated into several groups: heart-related (electrocardiogram, heartbeat), skin- and blood-related (electro dermal activity, blood-pressure), brain-related (electroencephalography), eye-related (eye gaze, pupil size) and movement-related (gestures, gyroscopic data). Some of the features are easily covered in the real-life conditions, e.g., heartbeat and skin response can be collected by smart watches and other wearable devices; other can be measured only in the laboratory environment, e.g., electroencephalography; and some of them can be hard to use in the real-life scenario at the moment, but it may become much easier in the nearest future, e.g., eye-movement with wider usage of smart glasses.

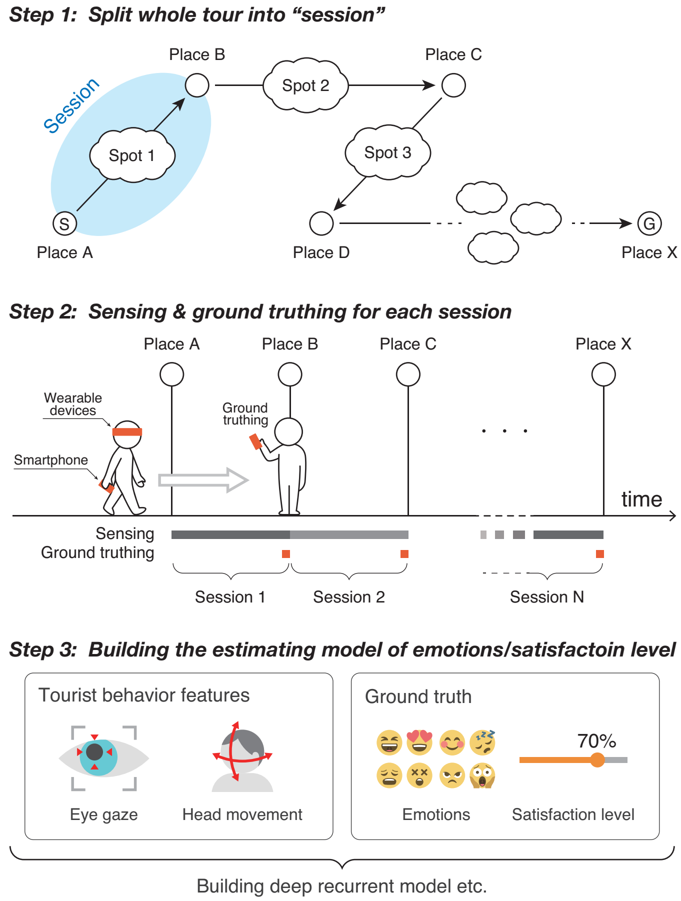


Fig. 1. Workflow to estimate tourist emotions and satisfaction level

In this study, to describe the tourist behavior, we focused especially on eye gaze and head movement. The sense of sight is one of the most important sensory systems in sightseeing, and it can be tracked as an eye gaze using current wearable devices and technologies. Other sensory systems: the sense of hearing, smell, taste and touch are difficult to observe directly from outside of a body. However, due to the directivity on the sense of hearing (and sight), the head movement can be affected by them. Thus, we used the head movement as features in addition to the eye gaze.

In the context of our study, we used three devices to record the features in real-time: Android smartphone (GPS-data, accelerometer data, gyroscope, magnetic field), mobile eye tracking headset Pupil with two 120 Hz eye cameras

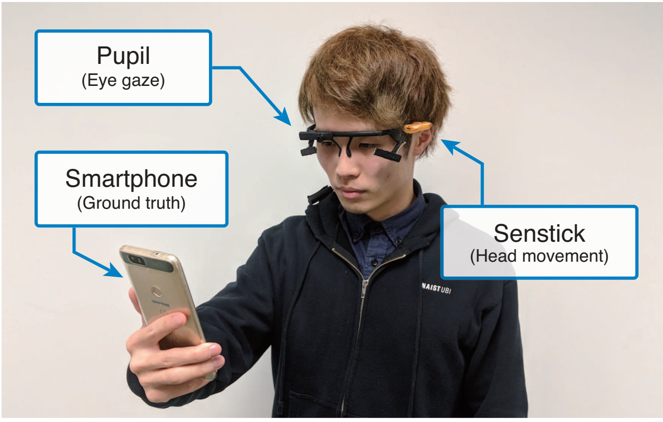


Fig. 2. Devices used on preliminary experiments: Pupil [12], SenStick [13] and smartphone

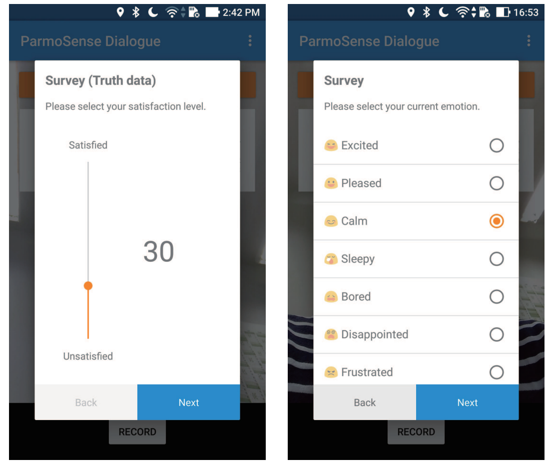


Fig. 3. Android application for collecting ground truth data of Satisfaction Level (left) and Emotion Status (right)

[12] (eye gaze, pupil features) and sensor board SenStick [13] mounted on an ear of eye tracking device (accelerometer, gyroscope, magnetic field, brightness, UV level, air humidity, temperature, air pressure) shown in Fig. 2.

In near future, intelligent eyewear devices which can measure the eye gaze and head movement will be available as a consumer product.

C. Ground truth

We collect ground truth data by using the Android application shown in Fig. 3. Tourist can manually enter the ratings of the session (labels) using two scales: the emotional status and satisfaction level.

Satisfaction level: The tourists can choose the current satisfaction level between 0 (fully unsatisfied) and 100 (fully satisfied). A neutral satisfaction level is 50 and it should approximately represent the state of the participant at the beginning of the experiment.

Emotion status: To represent emotion status of tourists, we adopt two-dimensional map defined on Russell's circumplex space model [14]. Fig. 4 shows the representation of the emotion status. We divided this map into nine categories as follows: excited (0), happy/pleased (1), calm/relaxed (2), sleepy/tired (3), bored/depressed (4), disappointed (5), distressed/frustrated (6), afraid/alarmed (7) and neutral (8).

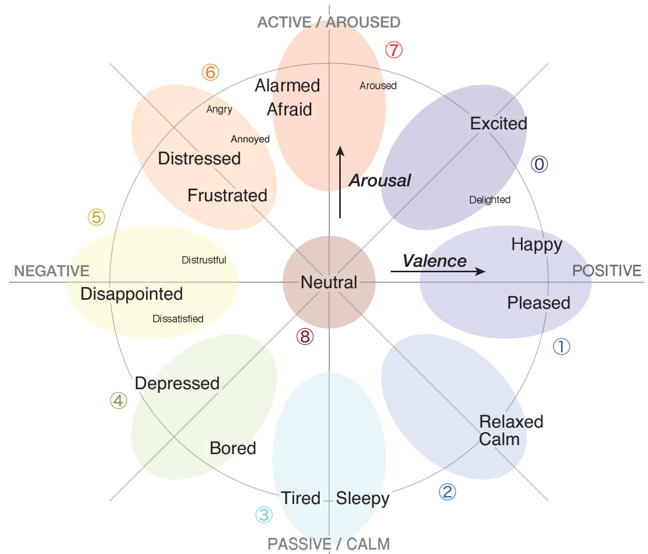


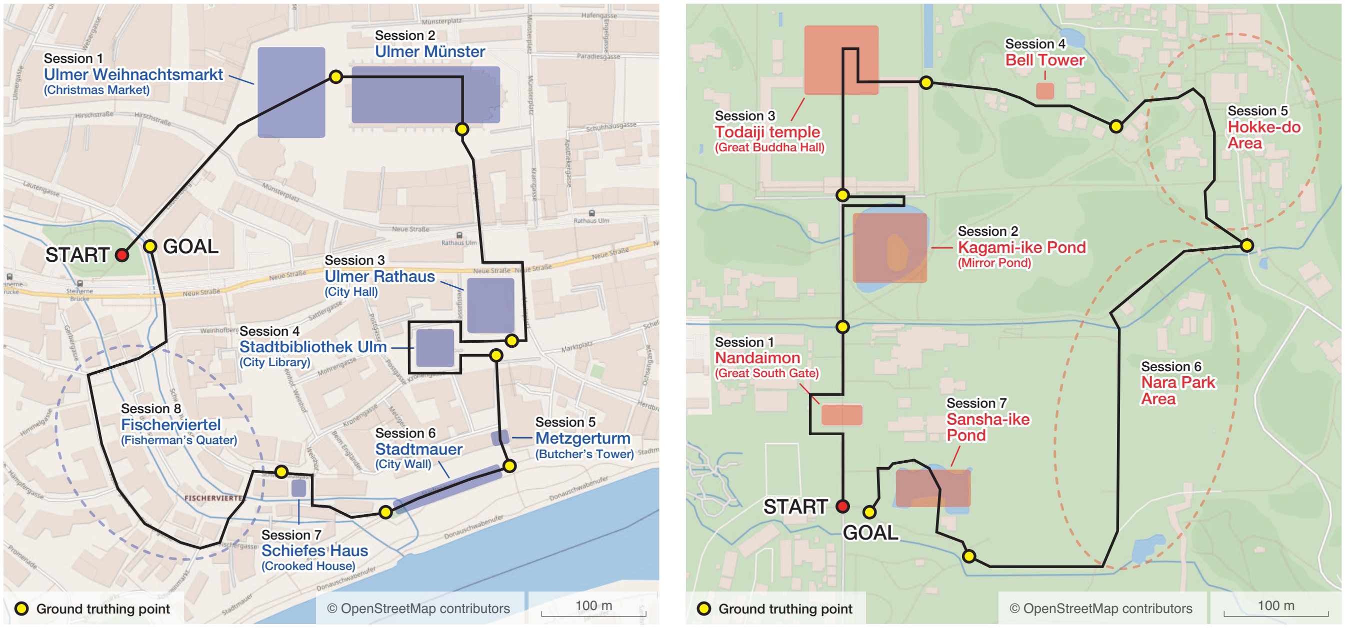
Fig. 4. Two-dimensional emotion status model. Figure taken from [14], [15]

IV. Preliminary Experiments and Results

We conducted preliminary experiments in real-world conditions to find the dependencies between features and labels. As the experimental field, we selected the two touristic areas depicted in Fig. 5, which have completely different conditions. The first one is the center of Ulm, Germany. The sights in this area include particular buildings as well as the walking routes with high touristic value (e.g., Fisherman's Quarter). As it is located in the city center, the sights are surrounded by the common city buildings and may be crowded, depending on the time. The approximate length of the route is 1.5 km divided into eight sessions. The second area is the Nara Park, the historic outskirts of Nara, Japan. The route through the area includes many temples and shrines, it is located in nature and has no distraction from the sights included in the sessions. The approximate length of the route is 2 km divided into seven sessions.

Participants were asked to follow the described routes and take as much time as they need to see the sights. During the sessions, we recorded the features described in Section III-B in real time. At the end of each session, participants rated it using emotion and satisfaction scales.

In this experiments, nine participants were asked to make a short sightseeing tour. The meta information about participants is the following: six participants are in Ulm (five male, one female) and three participants are in Nara (three male). For some participants, one or several sets of features can be missing due to technical problems.



(a) Experimental field in Ulm, Germany

(b) Experimental field in Nara, Japan

Fig. 5. Experimental field

A. *Eye gaze features*

To find dependencies between eye gaze features and labels, we use several features derived from the raw output data, extracted directly by Pupil Labs eye tracker. We used *theta* and *phi* values, which represent the normal of the pupil as 3D circle in spherical coordinates. Hence we can use only two variables to describe the position of pupils and thus eye gaze. The mean values of these variables differ across users and depend on the physical setting of a camera and eye peculiarity. Therefore, they should not be generalized to different users.

Eye gaze data was analyzed using two methodologies: (1) mean, minimum and maximum values for *theta* and *phi* were calculated for each participant; four levels (20, 40, 60, 80%) were set for the range [mean, min(max)] as shown at Fig. 6 and then used to count percentage of time outside each level per session (feature F1 in Table I); (2) standard deviation of *theta* and *phi* was calculated for small window of recorded data and the values corresponding to the same session were averaged (feature F2 in Table I). The following window sizes were used: 1, 5, 10, 20, 60, 120, 200, 300 seconds with the offset of $\frac{1}{3}$ of the window size.

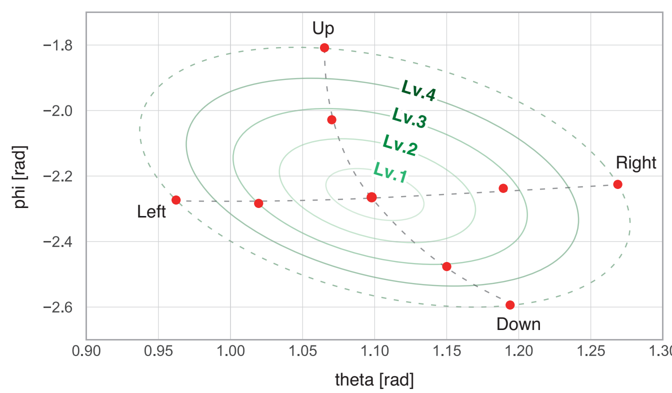


Fig. 6. Eye gaze in spherical coordinates

B. *Head movement features*

To collect head movement, we used sensor data of accelerometer and gyroscope from SenStick [13]. Both sensors have three axes: X-axis, Y-axis and Z-axis. In the experiment, the sampling frequency of both sensors is set to 50Hz to detect even slight head motion. Accelerometer and gyroscope ranges were set to 2G and 250dps respectively.

From sensor data, we can get two types of features: head tilt and state. Head tilt feature is the number of head tilt such as up/down or left/right and it can be detected from gyroscope due to its rotation motion. First, we calculated average μ and standard deviation σ of gyroscope values of each axis i and set the upper and lower thresholds with following equations:

$$\psi_{upper,i} = \mu_i + 2\sigma_i \tag{1}$$

$$\psi_{lower,i} = \mu_i - 2\sigma_i \tag{2}$$

Then, we checked windows with the size of one second and counted the number of gyroscope value of which axis is above $\psi_{upper,i}$ or below $\psi_{lower,i}$ as a head tilt. In our condition, Y-axis indicates looking up/down motion and Z-axis indicates looking left/right motion.

Head state feature is the time while keeping head to look up, down or front. It can be easily distinguished from an accelerometer. Features of looking left/right can be extracted

TABLE I
RESULT OF CORRELATION ANALYSIS BETWEEN SATISFACTION LEVEL AND FEATURES

Category	Feature ID	Feature	Ulm, Germany		Nara, Japan		All	
			R	p-value	R	p-value	R	p-value
Eye gaze[1]	F1	Lv.1	0.294	0.077	0.171	0.558	0.315	0.025
		Lv.2	0.344	0.037	0.243	0.403	0.284	0.043
		Lv.3	**0.378**	**0.021**	0.086	0.771	**0.384**	**0.005**
		Lv.4	0.206	0.221	0.116	0.894	0.228	0.107
	F2	theta	**0.376 (60 sec)**	**0.020**	**-0.584 (10 sec)**	**0.028**	0.259 (10 sec)	0.064
		phi	0.290 (10 sec)	0.077	-0.258 (5 sec)	0.372	0.235 (10 sec)	0.094
Head movement[2]	F3	left/right	**0.470**	**0.007**	0.159	0.491	0.098	0.486
		up/down	**-0.471**	**0.007**	**-0.503**	**0.020**	**-0.528**	**0.000**

[1] 38 sessions, 5 participants (Germany) and 14 sessions, 3 participants (Japan) [2] 32 sessions, 4 participants (Germany) and 21 sessions, 3 participants (Japan)

in the case of head tilt, whereas it is difficult to extract it in the case of the head state. To distinguish head state, we used the axis parallel to the face direction (Y-axis). Similarly to head tilt feature process, first, we checked the one second windows and calculated the average of values on Y-axis. Then we used the values of Y-axis to detect looking up (above $\psi_{upper,Y}$), down (below $\psi_{lower,Y}$) or front (none of the conditions) actions. These thresholds are experimentally determined as follows: $\psi_{upper,Y} = 0.4$ and $\psi_{lower,Y} = -0.4$. The threshold values indicate the head tilts about 24 degrees.

C. Results of preliminary experiments

At first, we analyzed the relationship between tourist satisfaction level and features by using Pearson correlation coefficient. Table I shows the results of correlation analysis. The significant results are highlighted with bold font. The number in brackets for feature F2 indicates window size, which provides the corresponding result, as the best among other options. Due to the technical issues as detailed in next section, we could not use several sets of features for some participants.

Interestingly, we confirmed that some of the features correlate to tourist satisfaction level regardless of sightseeing area. The most correlated features are: F1–Lv.3 ($R = 0.384$, $p < 0.05$) and F3–up/down ($R = -0.528$, $p < 0.05$).

In contrast, we found that some features are strongly affected by differences of sightseeing areas. For example, F2–theta represents the standard deviation of horizontal eye movements and it has the positive correlation with satisfaction level in the case of city center sightseeing in Ulm, Germany ($R = 0.376$, $p < 0.05$), but the negative correlation in the case of historic outskirts sightseeing in Nara, Japan ($R = -0.584$, $p < 0.05$). F3–left/right in Germany has the correlation with satisfaction level ($R = 0.470$, $p < 0.05$), however this correlation cannot be observed when we use features of both area ($R = 0.098$, $p > 0.05$).

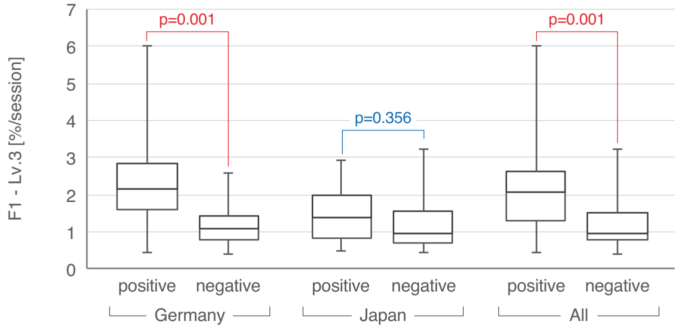
We also analyzed the relationship between tourist emotions and features. Since we did not have much data, we categorized all emotion labels into 2 emotion groups: positive and negative (in the context of touristic application). According to Fig. 4, the emotion labels 0–2 belong to positive and emotion label 3–8 belong to negative emotion group. Fig. 7 shows the distribution of feature values among emotion groups for different sightseeing areas. The statistical significance was calculated using Student's t-test analysis at the 10% confidence level.

We found significant differences in feature values between two emotion groups (indicated as red color at Fig. 7). In most of the cases, we can use the derived features to distinguish emotion group of each session. However, we haven't found differences in data of Japanese sessions (Fig. 7–(a) and (c)). The most probable reason is the lack of data; therefore, we should collect more data and confirm the relationship in the future research. Also, we found the same effect of the F2–theta on the emotional label as on satisfaction level. The dissimilarity in the behavior during the sightseeing session in different areas makes it necessary to distinguish the area type when eye gaze features are used.

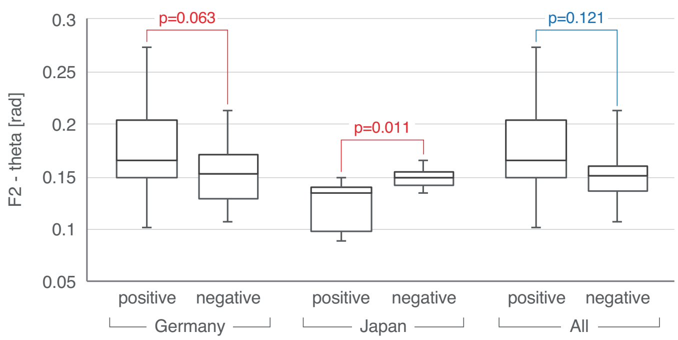
To understand the emotional state of tourist more precise, we need to recognize emotions in the positive group. Thus, we investigated the differences between emotions which belong to the positive group. Fig. 8 shows F1–Lv.3 (eye gaze) and F3–up/down (head movement) features for each emotion in the positive group. We found significant differences between exciting (0), happy (1) and calm (2) indicated as red color. This result indicates the potential possibility that eye gaze and head movement features can be used for emotion recognition during the sightseeing.

V. CONCLUSION

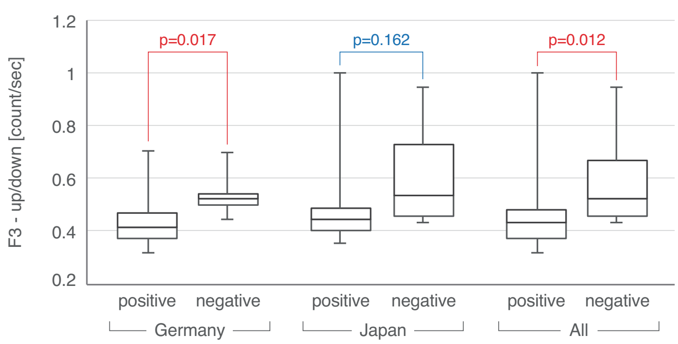
Following the increasing usage of smart devices, the intelligent tourist guidance is required as well. To design more enjoyable tourist routes, we must consider emotions and satisfaction level in a viewpoint of tourists. In this study, we have developed a methodology to estimate emotions and satisfaction level based on continuous observation of the tourist behavior with wearable devices during the sightseeing. Preliminary experiments have proved the feasibility of such research as



(a) Relationship between emotion groups (eye gaze feature F1 - Lv.3)



(b) Relationship between emotion groups (eye gaze feature F2 - theta)



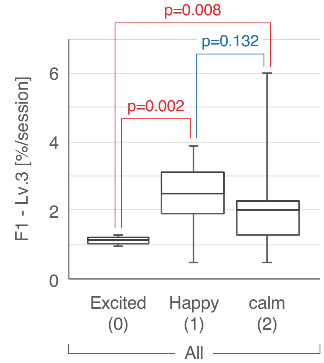
(c) Relationship between emotion groups (head movement feature F3 - up/down)

Fig. 7. Relationship between two emotion groups

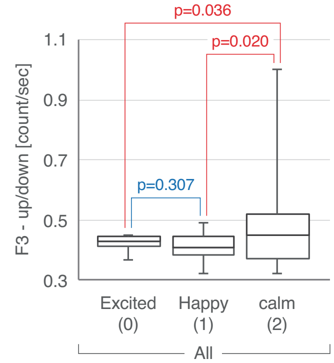
well as revealed the challenges, that have to be faced while solving a task of tourist satisfaction estimation. As a result, we have found a moderate correlation between physiological features and emotion and satisfaction level of a tourist. This correlation may be positive as well as negative, depending on the type of sightseeing area. In our future research, we will conduct more experiments to collect additional data and will build a prototype of tourist satisfaction estimation system based on machine learning techniques.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers 16J09670 and 16H01721.



(a) Eye gaze feature F1 - Lv.3



(b) Head movement feature F3 - up/down

Fig. 8. Relationship between emotions in the positive group

References

[1] E. Balandina, S. Balandin, Y. Koucheryavy, and D. Mouromtsev, "IoT use cases in healthcare and tourism," in *2015 IEEE 17th Conference on Business Informatics*, vol. 2, July 2015, pp. 37–44.

[2] S. Morishita, S. Maenaka, N. Daichi, M. Tamai, K. Yasumoto, T. Fukukura, and K. Sato, "Sakurasensor: Quasi-realtime cherry-lined roads detection through participatory video sensing by cars," *Proc. UBICOMP '15*, pp. 695–705, 2015.

[3] J. Alegre and J. Garau, "Tourist satisfaction and dissatisfaction," *Annals of Tourism Research*, vol. 37, no. 1, pp. 52 – 73, 2010.

[4] C.-F. Chen and F.-S. Chen, "Experience quality, perceived value, satisfaction and behavioral intentions for heritage tourists," *Tourism Management*, vol. 31, no. 1, pp. 29 – 35, 2010.

[5] B. Resch, A. Summa, G. Sagl, P. Zeile, and J.-P. Exner, "Urban emotions – geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data," 11 2014, pp. 199–212.

[6] W. Y. Quck, D. Y. Huang, W. Lin, H. Li, and M. Dong, "Mobile acoustic emotion recognition," in *2016 IEEE Region 10 Conference (TENCON)*, Nov 2016, pp. 170–174.

[7] S. Shapsough, A. Hesham, Y. Elkhorazaty, I. A. Zualkernan, and F. Aloul, "Emotion recognition using mobile phones," in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Sept 2016, pp. 1–6.

[8] F. Ringeval, F. Eyben, E. Kroupi, A. Yuce, J. P. Thiran, T. Ebrahimi, D. Lalanne, and B. Schuller, "Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data," *Pattern Recognition Letters*, vol. 66, pp. 22 – 30, 2015, pattern Recognition in Human Computer Interaction.

[9] T. W. AlHanai and M. M. Ghassemi, "Predicting latent narrative mood using audio and physiologic data." in *AAAI*, 2017, pp. 948–954.

[10] 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.

[11] W. L. Zheng, B. N. Dong, and B. L. Lu, "Multimodal emotion recognition using eeg and eye tracking data," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2014, pp. 5040–5043.

[12] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ser. UbiComp '14 Adjunct, 2014, pp. 1151–1160.

[13] Y. Nakamura, Y. Arakawa, T. Kanehira, M. Fujiwara, and K. Yasumoto, "Senstick: Comprehensive sensing platform with an ultra tiny all-in-one sensor board for IoT research," *Journal of Sensors*, vol. 2017, 2017.

[14] J. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[15] G. Paltoglou and M. Thelwall, "Seeing stars of valence and arousal in blog posts," *IEEE Transactions on Affective Computing*, vol. 4, pp. 116–123, Jan. 2013.