Towards Real-Time Contextual Touristic Emotion and Satisfaction Estimation with Wearable Devices

Dmitrii Fedotov*†, Yuki Matsuda‡§¶, Yuta Takahashi‡, Yutaka Arakawa‡‖, Keiichi Yasumoto‡¶, Wolfgang Minker*
* *Ulm University*, Ulm, Germany,
Email: {dmitrii.fedotov, wolfgang.minker}@uni-ulm.de.
† *ITMO University*, Saint Petersburg, Russia.
‡ *Nara Institute of Science and Technology*, 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*, Tokyo, Japan.
¶ *RIKEN Center for Advanced Intelligence Project AIP*, Tokyo, Japan.
‖ *JST Presto*, Tokyo, Japan.

***Abstract*—Following the technical progress and growing touristic market, demand on guidance systems is constantly increasing. Current systems are not personalized, they usually provide only a general information on sightseeing spot and do not concern about the tourist's perception of it. To design more adjustable and context-aware system, we focus on collecting and estimating emotions and satisfaction level, those tourists experience during the sightseeing tour. We reducing changes in their behaviour by collecting two types of information: conscious (short videos with impressions) and unconscious (behavioural pattern recorded with wearable devices) continuously during the whole tour. We have conducted experiments and collected initial data to build the prototype system. For each sight of the tour, participants provided an emotion and satisfaction labels. We use them to train unimodal neural network based models, fuse them together and get the final prediction for each recording. As tourist himself is the only source of labels for such system, we introduce an approach of post-experimental label correction, based on paired comparison. Such system built together allows us to use different modalities or their combination to perform real-time tourist emotion recognition and satisfaction estimation in-the-wild, bringing touristic guidance systems to the new level.**

***Index Terms*—ubiquitous computing, emotion recognition, satisfaction estimation, contextual modelling, wearable computing, smart tourism**

I. Introduction and Related Work

Over the past decades, the level of internationalization has increased dramatically, which results in growing tourism market. According to Statista [1], total contribution of travel and tourism to the global economy has strong positive trend and reached 8.27 trillion U.S. dollars in 2017. Following this tendency, demand on touristic guidance system is rising yearly. Wide spread of such devices as smartphones allowed to have a personal guide in one's pocket, but are they truly personal? Traditional guidance system provides some general information on particular area or sight, historical background, recent or upcoming events, etc., no personalized information is usually included and no user-adaptability is possible.

Some consumer touristic services, such as TripAdvisor [2], are allowing users to give reviews for sights using 5-star rating system. This creates a human opinion basis for the guidance system, but the reviews in such systems are usually biased and it is critically important to keep users motivated, especially for middle-rating reviews. Moreover, touristic impressions are always subjective and cannot rely on any facts and characteristics, in contrast to products reviews.

Emotions and satisfaction level are personal for each tourist and in addition to data-related problems in the area of general emotion recognition, such as expression of emotions and their annotation, touristic emotion recognition introduces the challenge of data labelling. It is common for emotion recognition area to annotate the data by several experts to reduce the subjective fluctuations, which is not the case in touristic domain. The label can be measured only by the tourist himself and not by any third-party person.

A lot of research has been done in area of emotion recognition and satisfaction estimation [3], [4]. Most popular modalities used in this context are audio and visual, although some researchers employed physiological features as well [5], [6]. Even though such system perform relatively good, they are often working only with acted data, collected in laboratory conditions, and are having troubles with in-the-wild data [7].

II. Proposed Approach

In this paper, we introduce a system for collecting real-time touristic data and performing an emotion recognition and satisfaction level estimation, based on it. We collect data during a sightseeing tour divided by several sessions. Each session includes one sight.

A. Technical Aspects

To build a truly personal touristic guidance system, working in real-time, we designed the technical setup, that allows to collect the information in a background mode. To avoid

This research was funded by Ministry of Science and Higher Education of the Russian Federation (grant number 2.12795.2018/12.2), Japan Society for the Promotion of Science (JSPS) KAKENHI (grant number 16H01721) and German Academic Exchange Service (DAAD)

978-1-5386-9151-9/19/$31.00 ©2019 IEEE

changes in tourist's behaviour due to the knowledge, that his actions are being recorded, we use wearable devices, collecting the data insensibly (see Fig. 1).

To get comprehensive information about the behaviour of the tourist, we use the following devices:

- Pupil Eye Tracker [8] with two infra red global shutter eye cameras. Each camera produces video of an eyeball with a resolution of 200x200 pixels, 120 fps. Based on videos from both eyeballs, it determines the gaze direction of a person, providing an information about coordinates and system confidence.
- Sensor board SenStick [9], mounted on an ear of the eye tracker device. It provides a wide spectrum of an environmental information, as well as accelerometer and gyroscope data required for determining the head and body movement. Combination of eye gaze and head movement provides a clearer picture of visual behaviour of a tourist.
- Empatica E4 wristband [10]. It is equipped with PPG sensor, allowing to measure the heart rate; EDA sensor, estimating electrical response of the skin; skin temperature sensor; and accelerometer.

All of the devices listed above, perform the measurements continuously in real-time and the data can be merged together to provide more reliable information about tourist's behaviour.

At the development stage and later at adjustment stage, the system associates collected raw sensor data with labels, provided for each session by user after visiting it (see Fig. 2). User selects the satisfaction level between 0 (fully unsatisfied) and 6 (fully satisfied), as well as one of the emotions from the following list: excited, happy/pleased, calm/relaxed, neutral, sleepy/tired, bored/depressed, disappointed, distressed/frustrated, afraid/alarmed.

Along with the label, user records a short video of himself giving a feedback to recently visited spot. The videos are used later to extract audio-visual features and use them as additional modalities.

Taking both sources of data into account, the system records tourist's behaviour in unconscious form during the session and his impression in conscious form right after sightseeing session is finished.

B. Implementation of Recognition System

When enough data is collected, estimation model is being trained, that may be refined as soon as new data is available. We use raw data to extract high-level features, that can describe the human behaviour without unnecessary noise. We use body movement data to determine pace rate of the user [11]; head movement data to count the number of head turns; eye gaze data to count the number of looks up/down/left/right [12]; audio signal to extract low-level descriptors [13]; and video frames to detect action units [14].

We base our recognition system on neural networks, as they are powerful, flexible and easy to refine with a new data. We train subsystems for each modality and then use meta-classifier to fuse the results and get final prediction for either emotion or satisfaction level. As the dimensionality for audio and visual feature sets is relatively high and these kind of data can easily be generalized, we use additional databases of emotionally rich interactions to train the contextual emotion recognition subsystem [15] in a cross-corpus fashion [16].

C. Post-experiment corrections

As we mentioned above, when it comes to touristic domain, labels reach an extreme level of subjectivity and there is no opportunity to engage a third person to correct them. Labels play the most significant role in building an estimation model and a little amount of noise and discrepancy may bring a lot of chaos into the system.

During the sightseeing tour, user unconsciously utilizes the information about sights, appeared in previous sessions, comparing the current sight to them. It leads to biased labels assignation, when early sessions have higher chance to be labelled positively, than late sessions, as no competition precede them.

To cope with it, we develop a method of post-experiment label correction. It should be performed immediately after the sightseeing tour, while the tourist's impressions are still fresh. The user assigns the satisfaction level for each sight, having the whole tour picture in mind, similar to the procedure during the session. This way user may concentrate one more time on each session, ranging them anew. After this user fills in a paired comparison table, comparing each sight to each another and deciding, which of two is better.

Having paired comparison, we can perform cross-analysis. First, we calculate the trust level for each participant as the percentage of cases, in which session preferred in paired comparison has equal or higher satisfaction value, than the second one. That is, if the ratings from post-experiment survey and paired comparison are in perfect consensus, the trust level will be equal to 1.0; if ratings have discrepancy in 10% of cases, the trust level will be equal to 0.9; etc. Secondly, we calculate the post-survey correction vector $\{c_i\}$ as difference between labels given during the session and in post-survey with respect to past experience:

$$c_i = (l\_orig_i - l\_post_i) * \frac{N-i}{N},$$

where l_orig_i is original label value, provided during the session i; l_post_i is label for session i, assigned in post-experiment survey; N - total number of sessions. We can later correct labels by the corresponding c_i value. Such correction strategy allows to rely on post-experiment survey more for the first sessions, than for the last.

III. Feasibility and Future Perspectives

During our recent experiments, we collected initial amount of data in two different touristic areas in Nara, Japan and Ulm, Germany. Given the features and methods described above, we were able to build the described system and achieve moderate performance up to 0.48 of unweighted average recall score (UAR) for three-class emotion classification problem and

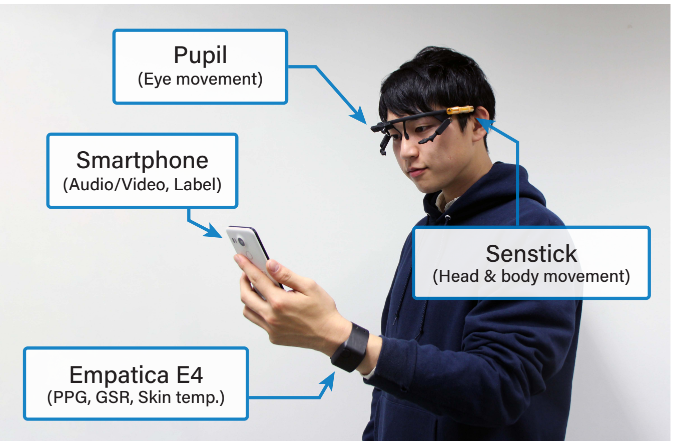

Fig. 1. Devices for data collection during sightseeing: Pupil [8], SenStick [9], Empatica E4 [10], and smartphone.

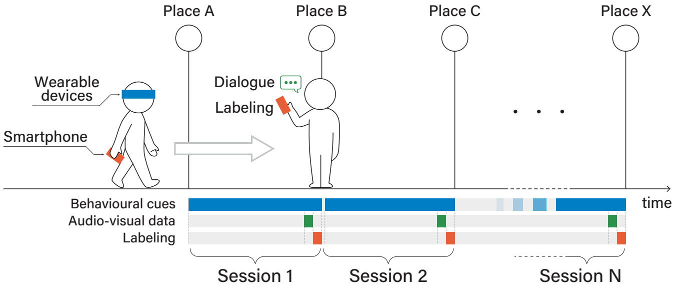

Fig. 2. Workflow to estimate tourist emotional status and satisfaction level.

1.11 of mean absolute error (MAE) for satisfaction estimation, proven the feasibility of such approach [17].

The main factor of success in the research with high label subjectivity is the diversity of data, which can be achieved by increasing its amount, keeping it balanced in such parameters as age, gender and nationality of participants. The data used in the system differs by its generalizing ability, therefore modalities can not be developed equally. Some data format is easy to collect, e.g. for improving audio-visual performance of the system, short videos from social networks, such as Instargam Stories, can be utilized, with labels taken from textual descriptions. It is harder to collect behavioural data, but with the spread of wearable devices, such as fitness-trackers, some features (pace rate, heart beat) will be easier to collect. Some are still left behind, e.g. at the moment it is hard to imaging a simple data collection method for eye gaze and head movement.

IV. Conclusions

To design more personalized touristic guidance systems, contextual information on user's emotional state and mood should be taken into an account. In this paper we suggested the way to do it, using wearable devices and smartphones. Utilizing several different modalities and covering conscious and unconscious user behaviour, we created a pipeline of real-time tourist emotion- and satisfaction estimation system. We have also considered labelling issues, while dealing with touristic data and suggested a way to overcome them with post-experiment surveys and labels corrections. Usage of proposed approaches will help to create more user-adjustable and context-aware guidance systems and bring them to a new level.

References

[1] "Direct and total contribution of travel and tourism to the global economy from 2006 to 2017 (in trillion u.s. dollars)," https://www.statista.com/statistics/233223/travel-and-tourism–total-economic-contribution-worldwide, (accessed: 26 Nov. 2018).

[2] "Tripadvisor," http://www.tripadvisor.com/, (accessed: 26 Nov. 2018).

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

[4] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," *Procedia Computer Science*, vol. 108, pp. 1175–1184, 2017.

[5] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1301–1309, 2017.

[6] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, "Introducing the recola multimodal corpus of remote collaborative and affective interactions," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–8.

[7] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon, "From individual to group-level emotion recognition: Emotiw 5.0," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ACM, 2017, pp. 524–528.

[8] 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.

[9] 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.

[10] Empatica Inc., "Empatica E4," https://www.empatica.com/research/e4/, (accessed: 26 Oct. 2018).

[11] H. Ying, C. Silex, A. Schnitzer, S. Leonhardt, and M. Schiek, "Automatic step detection in the accelerometer signal," in *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007)*, 2007, pp. 80–85.

[12] D. Fedotov, Y. Matsuda, Y. Takahashi, Y. Arakawa, K. Yasumoto, and W. Minker, "Towards estimating emotions and satisfaction level of tourist based on eye gaze and head movement," in *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2018, pp. 399–404.

[13] F. Eyben, M. Wöllmer, and B. W. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM International Conference on Multimedia*, ser. MM '10. ACM, 2010, pp. 1459–1462.

[14] T. Baltrušaitis, P. Robinson, and L. P. Morency, "Openface: An open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2016, pp. 1–10.

[15] D. Fedotov, D. Ivanko, M. Sidorov, and W. Minker, "Contextual Dependencies in Time-Continuous Multidimensional Affect Recognition," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), 2018, pp. 1220–1224.

[16] H. Kaya, D. Fedotov, A. Yeşilkanat, O. Verkholyak, Y. Zhang, and A. Karpov, "Lstm based cross-corpus and cross-task acoustic emotion recognition," *Proc. Interspeech 2018*, pp. 521–525, 2018.

[17] Y. Matsuda, D. Fedotov, Y. Takahashi, Y. Arakawa, K. Yasumoto, and W. Minker, "Emotour: Estimating emotion and satisfaction of users based on behavioral cues and audiovisual data," *Sensors*, vol. 18, no. 11, p. 3978, 2018.