Sensors Know Which Photos Are Memorable

Soyoung Kim *
Department of Industrial Design,
KAIST
Daejeon, Republic of Korea
soyoung.kim@kaist.ac.kr

Kun-Pyo Lee
Department of Industrial Design,
KAIST
Daejeon, Republic of Korea
kplee@kaist.ac.kr

Kalenzi Asio Evelyn Patra *
Knowledge Service Engineering,
KAIST
Daejeon, Republic of Korea
blessedeve@kaist.ac.kr

Aviv Segev
Knowledge Service Engineering,
KAIST
Daejeon, Republic of Korea
aviv@kaist.edu

Auk Kim
KAIST
Daejeon, Republic of Korea
kimauk@kaist.ac.kr

Uichin Lee
KAIST
Daejeon, Republic of Korea
uclee@kaist.ac.kr

* Two authors contributed equally to this work

CHI'17, Extended Abstracts, May 06-11, 2017, Denver, CO, USA
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3053198

Abstract
The goal of this study is to determine if physiological signals are salient in the detection of memorable personal photos. We begin by collecting physiological sensor data as well as memorability and emotion ratings for photos. We then build a mixed model to evaluate the predictive power of physiological variables on memorability and emotion by examining whether or not the photographer's data is useful for predicting the ratings of the photographer or the ratings of the subjects in the photos. Our results suggest that heart rate and GSR (galvanic skin response) data are the major predictors of memorability for photographers, and that the sensor signals are not particularly useful for predicting memorability ratings of subjects in the photos.

Author Keywords
Memorable photos; memorability; emotion; physiological data; wearable devices.

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
People take such a large number of photos during various activities such as tours, weddings, and birthday parties that it can become impractical to manually sort through all the

photos. There is a desire to find memorable photos that are worth sharing via social media such as Facebook and Instagram, or to automatically shuffle memorable photos to help people reminisce about important life events. Photo memorability is an important aspect of photography and social culture because it helps determine if a photo is worth sharing or displaying.

Our research investigates how to automatically detect memorable photos. There has been a significant amount of research in the field of computer vision to uncover what makes an image memorable or forgettable [2, 5, 7, 4] and to estimate the memorability of images [6]. However, prior studies primarily focused on objects in the photos such as analysis of visual cues and heat distributions for inferring memorability. Some other research on life-logging [8] uses similar devices to measure user emotions and categories pictures and videos, but does not necessary measure the memorability of these pictures or videos. Furthermore, memorability was treated as a non-personal property, meaning that photos that were not taken by the study participants were used for measuring and inferring memorability.

Due to the increasing popularity of wearable devices and their excellent portability, recent research has attempted to use wearable devices and their built in sensors for various image-related research. Our research attempts to answer the following research question: “Can wearable sensors determine which photos that a user has taken are memorable?” If they can, we would like to identify the features that are the most relevant for predicting memorable photos. Our work was inspired by a recent study by Feng et al. [9] that recommends personalized walking routes by leveraging their findings that heart rate changes when encountering interesting walking locations. Prior studies suggest a relationship between the memorability of images and emotions in the cognitive science fields [1, 10], such as arousal affecting long-term memory [14], but none of these studies have examined which physiological signals are directly related to the memorability of personal photos. Mobile devices with built in sensors were used by the people tagging the photos (e.g. TagSense [13]), but none of the mobile sensor studies examined the memorability of personal photos.

In this study, we used the Microsoft Band 2, a wearable smart-band that contains various physiological sensors including heart rate (HR) and galvanic skin response (GSR). Our study was carried out in four stages: data collection, photo rating, data merging, and data analysis. We developed a custom app to record the sensor data captured by the MS Band 2 and used a phone camera to capture photos that can be rated. The timestamp from the sensor data were then linked to the time the photo was taken. For each photo, the participants rate its memorability and their emotion at the time the photo was taken by answering a questionnaire in our custom photo-rating app. We extract the key features and analyze how rating variables are correlated to each other and to the extracted features. We then build a mixed linear (or multi-level regression) model and evaluate the predictive power of physiological variables for memorability and emotion. Our analysis examines whether or not the sensor data is useful for predicting the ratings of the photographer or the ratings of the subjects in the photos.

Our preliminary experimental results showed that heart rate and GSR data were major predictors of memorability for photographers. However, none of the sensor signals appeared to be useful for predicting the memorability ratings of the subjects in the photos. Our findings are one more step towards automatically detecting memorable photos, which will facilitate various kinds of personalized photo recommendation services.

Experimental Framework

Capturing photos and physiological signals

The Nexus 5 [3] with a 8-megapixel rear camera and a 1.3-megapixel front camera was used to capture the photos. To capture physiological signals, we used the Microsoft Band 2 [12], a second-generation smart-band with smart watch features. Its sensors include a heart rate monitor, 3-axis accelerometer, UV sensor, gyroscope, GPS, microphone, ambient light sensor, GSR sensor, capacitive sensor, and barometer. Readings from each sensor were captured for each participant at the time a photo was taken, and a timestamp was recorded to keep track of what data belongs to each photo. The smart band and the phone were synchronized for accurate data tracking. Note that we developed a custom sensor logging application to record all data captured by the MS Band 2. This application was installed on the Nexus 5 smartphone and synced with the MS Band user profile created on each phone, then synced with the MS Band 2. A GPS logging app [11] was also installed on the phones and used to log the GPS location data of the participants.

Measuring memorability and emotions

For each photo, we asked the participants to rate the memorability of the photo and their emotion at the time the photo was taken. To rate memorability, the participants were provided with a scale of 1−7 (least memorable:1 − most memorable:7), corresponding to low-high memorability. For emotion measurement, we used the self-assessment manikin and semantic differential (SAM) questionnaire. SAM is an emotion assessment tool that uses graphic scales (-4 − +4) depicting cartooned emotions of pleasure, arousal, and dominance. It provides a clear and intuitive method for describing emotions related to the photos. As shown in Figure 1, we developed the photo rating app in the form of a questionnaire, where participants select a photo they

Figure 1: Photo-rating application screenshot

took (and later a photo that was taken of them), and rate it according to their feelings. In Figure 1, the SAM scale is located on the left with three subscales for pleasure, arousal, and dominance , and the memorability scale is located on the bottom right.

Data pre-processing

The Microsoft Band 2 generates a significant amount of stream sensor data. In our data collection app, we used SQLite for data storage. We first decided to eliminate the variables with missing data and those that were not fully captured (UV, Skin temperature, and Pedometer). We also removed the variables whose values did not vary significantly. The barometer measures atmospheric pressure, and its value was consistent because photo-taking activities in our experiment occurred at a consistent elevation. Similarly, the values from the motion sensors (accelerometer, gyroscope, and manometer) did not vary significantly, thus the results were not good predictors of memorability and emotions. During our pilot and main studies (campus tour and Arboretum visiting), our participants taking a photo or being photographed tended to hold their poses to prevent blurry photos while taking photos, resulting in almost static values from the motion sensor. While not taking photos, participants walked around at a similar pace, resulting in non-significant variation for the values. Note that motion sensor data could be more helpful and meaningful for other situation, such as during visits to a sporting event or theme park, which require various motion patterns at different paces. We smoothed the time-series sensor data using an exponential moving average to filter out high frequency noise. For data pre-processing, we used MySQL workbench, and exported the resulting data for further analysis.

Feature generation and regression analysis
Feature generation was performed using a windowing approach. We used a window size of 5 seconds with a 40% overlap. First, we calculated the average, minimum, maximum, median, and standard deviation of all kept variables. These variables were then windowed and organized into a .csv file for each participant. The rating data was separated into two categories: photographer focused and subject focused data. Photographer focused means that the sensor data of the photographer is analyzed with the ratings from the photographer; these could be selfies or photos of others. Subject focus means that the sensor data of the photographer is analyzed with ratings from the subject in the photo. Subject focus was used to determine if a photographer's biological signals could be used to infer the memorability and emotions of subjects in the photos. To perform this analysis, we merged the corresponding sensor data and ratings from separate files. Initially, we analyze how the rating variables are correlated with each other and how they are correlated with the extracted features. We then build a mixed linear (or multi-level regression) model using R and use it to evaluate the predictive power of physiological variables for memorability and emotion.

Pilot User Study
The goal of the pilot study was to test the feasibility of our initial experimental plans. We recruited four participants who were selected based on existing social ties. The participants were asked to sign consent forms. First, we asked the participants to download the required apps on their smartphones and to pair their phones with the MS Band 2. They were then taught how to use the app in conjunction with the MS Band 2. We asked the participants to move freely around campus in groups and take as many photos as possible. Finally, they returned to the briefing room for interviews and to perform photo rating (using printed answer sheets).

The pilot study helped us to improve the experimental procedure. We found that using participants' smart phones delayed the entire process with delay proportional to group size; the pilot study took approximately two hours. One participant in a group did not own an Android phone (in this case, an Android phone was provided). Additionally, the time on the MS Bands was not automatically synchronized. To address these concerns in our main study, we decided to provide users with pre-configured Android smartphones that were paired with the MS Band with correct time synchronization. Another issue was that during our interview and rating session the participants felt tired and were less willing to rate or sort photos. This prompted us to design the custom rating app discussed above, enabling participants to quickly rate photos.

Main Study
Procedure
Eleven participants (three males and eight females) in four groups were engaged in the main study. The participants were compensated with 10 USD for their time and effort. We ensured that these participant groups were formed based on existing social ties so that they can interact naturally with each other. The study began with the participants being given an overview of the purpose of study. They were asked to sign consent forms due to the necessity of collect-

ing their physiological data. We provided a Nexus 5 phone and a MS Band 2 to each participant and explained how to use the devices. The participants were asked to walk around a park in Daejeon, South Korea (called Hanbat Arboretum) for approximately thirty minutes and to take at least 15 photos. After taking each photo, the participants immediately rated the photo using the photo-rating app that incorporated the SAM model and photo memorability. The rating was performed in two ways, as a photographer and as the subject of the photo. Each method yielded different results, which were stored in a .csv file for each participant.

The eleven participants captured 304 photos, with approximately 10,000 instances of sensor data for 12 general variables for each participant. From the 12 variables, we selected 6: ambient light, distance walked, GSR, heart rate, temperature, and calories burned. As illustrated earlier, gyroscope, accelerometer, and altimeter readings are related to the movement of the participants, but are not related to the prediction of participant memorability or emotion ratings because participants do not typically move while taking photos. Barometer readings were the same everywhere in the park and UV and skin temperature were not properly captured during the experiment.

As mentioned in the experimental framework section, we merge the log data set from the MS Band with the participants' ratings on photos based on pleasure, arousal, dominance, and memorability before performing analysis. After pre-processing, we performed a Pearson correlation using SPSS (Statistical Package for the Social Sciences) for finding correlated variables among 34 different variables. Four variables came from the photo-rating app and the other 30 came from the MS band. They are the min, max, average, standard deviation, and median of ambient light, distance walked, GSR, heart rate, temperature, and calories burned.

Next, multi-level regression analysis was performed in R to investigate which variables are useful for predicting memorability, pleasure, arousal, and dominance. Because there were 11 participants with differing ranges of sensor data, we considered each participant as a different level of data. The regression analysis was performed four times with four different dependent variables and 30 points of sensor data, which were used as independent variables.

Note that we consider two different focuses for analysis: (1) photographer focus: photographers rate the photos that they took (selfies or photos of others), and (2) participant focus: participants rate photos of themselves taken by others.

Results

Photographer focus: Pearson correlation analysis showed that memorability and three of the subscales from SAM were highly correlated (memorability-pleasure: r =-$.505$, memorability-arousal: r =-$.539$, memorability-dominance: $r = .368$) with an alpha level of 0.05. Additionally, GSR and calories burned also had a strong correlation with memorability. From the multi-level regression analysis, we discovered that the minimum value and standard deviation of heart rate can predict memorability as well ($\beta = 9.051, p < .001$ and $\beta = 17.622, p < .001$ respectively). Furthermore, the standard deviation of heart rate is an important predictor for the pleasure and arousal dimensions (β =-$13.296, p < .05$ and β =-$14.777, p < .001$). The minimum value for heart rate was also significant for predicting arousal (β =-$7.481, p < .05$). Finally, the maximum value and standard deviation of GSR and the maximum value for calories burned were useful for predicting dominance, which was significantly different from the results of previous regression analysis ($\beta = 0.00026$, β =-0.00061, and β =-56.524 for all $p < .05$).

Participant focus: Similar analysis was performed with a participant focus. Pearson correlation analysis showed that memorability and three of the subscales of SAM were even more highly correlated than in the above analysis (memorability-pleasure: r =-$.507$, memorability-arousal: r =-$.623$, memorability-dominance: $r = .389$) with an alpha level of 0.05. Additionally, heart rate and calories burned had a strong relationship with memorability. From the multi-level regression, however, we could not find any significant factors for predicting memorability, pleasure, or arousal. The maximum value for heart rate proved to be an important factor for predicting dominance (β =-$7.563, p < .05$). We suspected that lack of relationships for many variables may be due to regression analysis being performed on the pooled data of all participants. We then performed multiple regressions with each individual's dataset, but again, we were unable to find any significant variables. Therefore, we concluded that sensor data is not useful for predicting photo ratings in a participant focused scenario.

Discussion

Our experimental results showed that heart rate and GSR are the major predictors for the memorability ratings of photographers. Because memorability was correlated with other semantic dimensions of photos, we assume that emotional responses affect memorability. Furthermore, biological data at the time of photo capture affected memorability ratings, indicating that there is a good possibility of detecting memorable photos using physiological signals. Photos taken by the participants themselves have greater potential for photo memorability prediction when compared to photos taken of the participants by others.

The results of our analysis showed that the standard deviation of heart rate and GSR could reliably identify memorable photos from the photographer focus, whereas there were no significant data points from the participant focus. This indicates that the feelings of the person who takes the photos, rather than the subject of the photo, is a more important factor for photo memorability. This is likely because the person who takes the photo concentrates more on the scenery and subject while looking for the perfect moment, whereas the subject only cares about how they feel and look in the photo. Additionally, photographers consider their own opinion and viewpoint when taking memorable photos, and this affects physiological data such as heart rate and GSR.

We think it would be interesting to examine how the photo subjects' sensor data is related to their ratings, which will be included our future work. We also did not consider people's subjective tastes and perspectives on photos. Because memorability is subjective, getting group viewpoints is important, but our focus in this paper was on individual users (particularly photographers). While our work is limited to quantitative analysis, interview-based studies would likely yield significant insight into how memorability and emotion are related to physiological signals. We collected biological data using a smart band, but the data was sometimes noisy or incomplete; adopting more accurate sensors or advanced calibration techniques would aid further research. Furthermore, the experiment only lasted for approximately 30 minutes. To generalize our findings would require large-scale, naturalistic user studies such as a group of users visiting a theme park and staying there for at least 5 hours.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. B0101-15-1272, Development of Device Collaborative Giga-Level Smart Cloudlet Technology). U. Lee is the corresponding author.

References

[1] Larry Cahill and James L. McGaugh. 1995. A Novel Demonstration of Enhanced Memory Associated with Emotional Arousal. *Consciousness and Cognition* 4, 4 (1995), 410–421. DOI:http://dx.doi.org/10.1006/ccog.1995.1048

[2] Rachit Dubey, Joshua Peterson, Aditya Khosla, Ming-Hsuan Yang, and Bernard Ghanem. 2015. What Makes an Object Memorable?. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) (ICCV '15)*. IEEE Computer Society, Washington, DC, USA, 1089–1097. DOI:http://dx.doi.org/10.1109/ICCV.2015.130

[3] Google. 2016. Nexus. (2016). Retrieved December 23, 2016 from https://www.google.com/intl/ko_kr/nexus/5x/.

[4] Phillip Isola, Devi Parikh, Antonio Torralba, and Aude Oliva. 2011. Understanding the Intrinsic Memorability of Images. In *Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS'11)*. Curran Associates Inc., USA, 2429–2437. http://dl.acm.org/citation.cfm?id=2986459.2986730

[5] Aditya Khosla, Wilma A. Bainbridge, Antonio Torralba, and Aude Oliva. 2013. Modifying the Memorability of Face Photographs. In *Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV '13)*. IEEE Computer Society, Washington, DC, USA, 3200–3207. DOI:http://dx.doi.org/10.1109/ICCV.2013.397

[6] Aditya Khosla, Akhil S. Raju, Antonio Torralba, and Aude Oliva. 2015. Understanding and Predicting Image Memorability at a Large Scale. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) (ICCV '15)*. IEEE Computer Society, Washington, DC, USA, 2390–2398. DOI:http://dx.doi.org/10.1109/ICCV.2015.275

[7] Aditya Khosla, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2012. Memorability of Image Regions. In *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS'12)*. Curran Associates Inc., USA, 296–304. http://dl.acm.org/citation.cfm?id=2999134.2999168

[8] Su Gon Kim, Seong Hoon Moon, and Da Young Ju. 2013. Memorable Moment Extraction using the Pupil Dilation Analysis. *The Japanese Journal of Ergonomics* 49, Supplement (2013), S559–S561. DOI:http://dx.doi.org/10.5100/jje.49.S559

[9] Feng Liang, Masashi Nakatani, Kai Kunze, and Kouta Minamizawa. 2016. Personalized Record of the City Wander with a Wearable Device: A Pilot Study. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 141–144. DOI:http://dx.doi.org/10.1145/2968219.2971462

[10] James L. McGaugh. 2006. Make mild moments memorable: add a little arousal. *Trends in Cognitive Sciences* 10, 8 (2006), 345 – 347. DOI:http://dx.doi.org/10.1016/j.tics.2006.06.001

[11] Mendhak. 2016. GPS Logger. (2016). Retrieved December 23, 2016 from https://play.google.com/store/apps/details?id=com.mendhak.gpslogger.

[12] Microsoft. 2016. Microsoft Band. (2016). Retrieved December 23, 2016 from https://www.microsoft.com/ko-kr/store/p/microsoft-band/9wzdncrfjbcx.

[13] Chuan Qin, Xuan Bao, Romit Roy Choudhury, and Srihari Nelakuditi. 2011. TagSense: A Smartphone-based Approach to Automatic Image Tagging. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys '11)*. ACM, New York, NY, USA, 1–14. DOI: http://dx.doi.org/10.1145/1999995.1999997

[14] Corina Sas, Tomasz Fratczak, Matthew Rees, Hans Gellersen, Vaiva Kalnikaite, Alina Coman, and Kristina Höök. 2013. AffectCam: Arousal- Augmented Sensecam for Richer Recall of Episodic Memories. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 1041–1046. DOI: http://dx.doi.org/10.1145/2468356.2468542