Social Activity Measurement with Face Detection Using First-Person Video as a Lifelog

Akane Okuno
Future University Hakodate
116-2 Kamedanakano-cho,
Hakodate, Hokkaido, Japan
a-okuno@sumilab.org

Yasuyuki Sumi
Future University Hakodate
116-2 Kamedanakano-cho,
Hakodate, Hokkaido, Japan
sumi@acm.org

Abstract

This paper proposes a simple method of using video as a lifelog to measure the social activity of a camera-wearer from a first-person perspective, aiming to quantify and visualize social activities that are performed unconsciously. The context of social activity is determined by the number and continuity of detected faces, whereas the amount of social activity is calculated by the number, size, and continuity. This taxonomy allows users to ascertain whether they tend to pass by people or spend time with them, and how many people there are. Our expectation is to enable users to change their behavior toward achieving social health by providing visual feedback. In this paper, we show an implementation of our social activity measurement and report on our feasibility study.

Author Keywords

Social activity measurement; first-person video; lifelog; face detection; social health; visual feedback.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

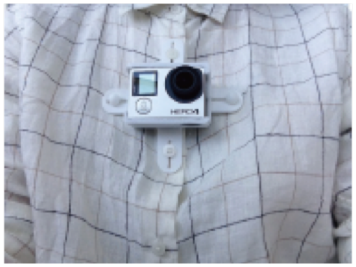


Figure 1: Lifelogging by wearing a camera on a chest.

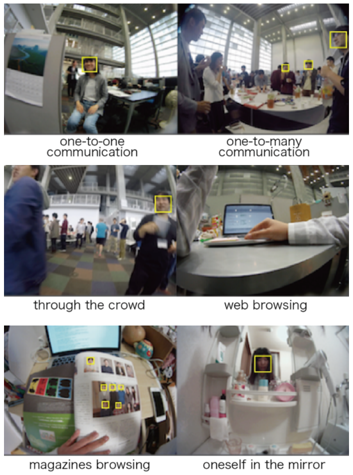


Figure 2: Examples of the scene captured using first-person lifelog video.

Introduction

This paper proposes a method of using video as a lifelog to measure the social activity of a camera-wearer from a first-person perspective, aiming to quantify and visualize social activities that are performed unconsciously. It is a simple idea of this research whether counting the number of faces in the images of the cameras worn by individuals can measure the daily social activity amount. Just as the pedometer counts the number of steps from the acceleration change, this research aims to realize the "facemeter" by counting the number of faces. In recent years, as a result of advances in technology to distinguish patterns of fluctuations in body motion, list band type activity meters (Fitbit, SmartBand, Jawbone, Apple Watch, etc.) identify walking, jogging, sleeping, etc. It has enabled people to adjust physical activity for health. Like as a pedometer developed as a measure of physical activity, we explore possibilities as a social activity measuring instrument by counting based on the time pattern of face detection.

Recently, wearable cameras have become widespread. We use the first-person video as a lifelog (first-person lifelog video). By wearing a camera (Figure 1), various scenes can be captured (Figure 2). Face-to-face communication occurs in daily life, such as when we meet and talk to people. Additionally, there are scenes of television-viewing, magazine-reading, and using social networking services as forms of face-to-face contact. We focus on scenes facing the face, among various social activities. Our expectation is improving social health [5; 11] via visual feedback of measured. In this paper, we show an implementation of our social activity measurement and report on our feasibility study.

Related Work

Techniques for recognizing the social context of individuals and groups from nonvisual information have been studied by many researchers. For example, by combining exercise with an acceleration sensor, a voice with a speaker, distance with a Bluetooth link and face-to-face detection with an IR sensor, various aspects of the social context have been measured [9]. The productivity and job satisfaction were predicted by measuring. Additionally, studies have been conducted to simply interpret the important aspects of the social activities themselves. For example, technologies that interpret a talking field [8] enabled researchers to use a simple algorithm and a lightweight, networked mobile terminal equipped with a microphone to work in crowds. On the other hand, many techniques have been studied for analyzing first-person video. For example, calculating the line-of-sight of another person's face from the position and orientation of the camera-wearer allows software to create a 3D mapping of the environment, creating a heat map. This can be used to estimate the partner's profile and role in a group setting [3]. Additionally, multiple camera-wearers' scenes can be analyzed to correlate head movement and faces during a group conversation. Thus, it is also possible to derive the position and orientation of the face of the camera-wearer [12]. Furthermore, studies on human health [4; 7; 10] and perception extension [6] have been conducted using wearable cameras.

There are many aspects of social activities that require us to well-recognize important information, according to the purpose and scope of its application. If there is something of value that can be recognized from non-visual cues, we believe that there is a social context that can be recognized anew from visual information. It

$$S = \sum_{t=1}^{m} \sum_{i=1}^{n} T_i(t) \cdot D_i(t) \qquad (1)$$

i : The identification number of the detected face,
$T_i(t)$: Continuity (same face detection frame, starts from 1),
$D_i(t)$: Closeness (the area of the face occupying the whole),
m : The number of measurement frames (elapsed time) up to the time t,
n : The cumulative number of people (number of faces) up to the time t.

$$D_i = \frac{w_i \cdot h_i}{R} \cdot 100 \qquad (2)$$

w_i : The width of detected face i,
h_i : The height of detected face i,
R : Screen resolution. Unit: pixel.

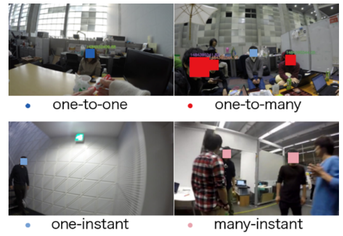


Figure 4: Determination of social activity situations for labeling.

1. one-to-one : Only one face is tracked with an ID.
2. one-to-many : Two or more faces are tracked with IDs.
3. one-instant : Only one face is detected with no ID.
4. many-instant : Two or more faces are detected with no ID.

is useful to understand the context of the camera-wearer from first-person video and to support their activities as necessary. With our approach, we measure the degree of face-to-face relationship with others by counting based on the time pattern of face detection. Our expectation is to provide feedback that leads to greater social health.

Social Activity Measurement

We interpret the face-to-face relationships in daily life as participation in social activities. Also, we define quantities and visualization of how much and what kind of situations were defined as social activity volume and social activity situation. Considering approaching the actual impression by considering the number of people, distance nearness, continuity based on the number of faces obtained from face detection, face size, and continuity of faces. Therefore, identification of faces such as whether they are the same person, expressions, identification of opening and closing of the mouth are not done. For face detection, we use OpenFace [1]. Face tracking of the dlib library used in there has the function of tracking what was estimated as the same person's face between frames [2]. Therefore, in this research, if face tracking detects continuous faces of the same person, it interprets it as sustained interaction with that person.

Based on the numerical values obtained using these technologies, we calculate the amount of social activity and discriminate the social activity situation. Based on the number of people and time continuity, we discriminate four types of social activity situations and count them separately (Figure 3). If we simply count the number of faces, we will treat the same in a crowd with another person as well as a close dialogue with a specific person. For this reason, we use the closeness of the distance and continuity of the time. The amount of social activity S shall be the time integral of the value calculated on the basis of the number of people, the closeness of distance, and continuity for each frame (Formula 1). For the detected face identification number i, the newly issued ID is used every time OpenFace being used this time detects a new face. Specifically, different IDs are issued for newly detected faces in a certain frame. Utilizing this property, we decided to increment $T_i(t)$ for that ID if the same ID is detected in consecutive frames and use it as time continuity. The closeness D_i represents the size of the face occupying the entire screen (Formula 2).

In addition to calculating the social activity amount, we label four social activity situations. As shown in Figure 4, the state of social activity context transition is based on the result of face detection. Based on the number of people and the continuity, we define the four contexts. Conditions 1 and 2, left column, are interpreted as contexts where engagement is high and face-to-face communication is well-established.

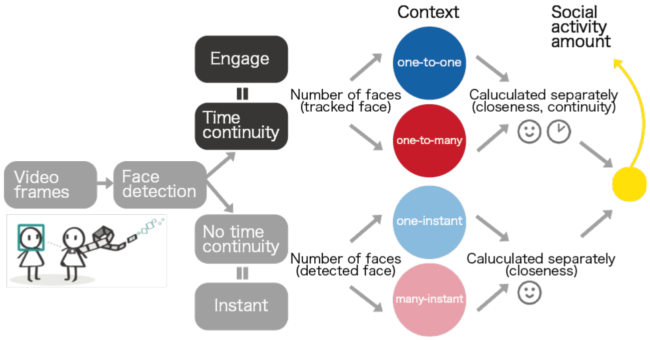
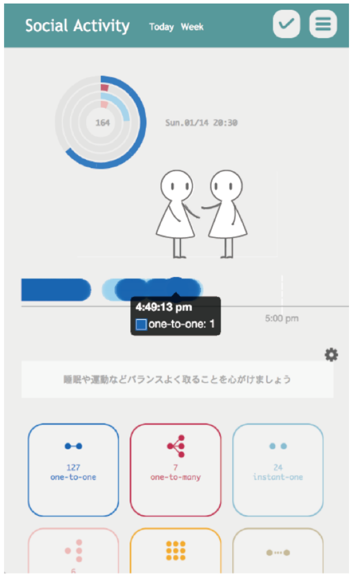


Figure 3: Flow of social activity measurement using face detection.



Social Activity Today Week

164

Sun.01/14 20:30

4:49:13 pm

🟦 one-to-one: 1

5:00 pm

睡眠や運動などバランスよく取ることを心がけましょう

127 one-to-one

7 one-to-many

24 instant-one

Figure 5: Prototype calculating the social activity amount, labeling four social activity situations. The UI is based on the gauge of activity amount like Apple Watch, and animation of looking back at the activity like SmartBand.

We expect to be able to obtain social activity quantities that are similar to impressions even in situations where multiple situations are mixed. Additionally, by counting according to the situation, when using it in daily life, users can set the target value of each social activity amount. Figure 5 is our prototype. This taxonomy allows users to ascertain whether they tend to pass by people or spend time with them, and how many people there are.

Feasibility Study

We conducted a feasibility study of social activity measurement using our proposed method. The aim of the feasibility study was to learn whether the method using simple face detection was effective for measuring the quality and quantity of communication from a lifelog recording. We attempted to measure various types of social activities during the conference where are several face-to-face communications in a relatively short period. Sessions lasted approximately 2 h per day. Measurement was carried out every 10 seconds. In this paper, we show the result obtained from a person D, who participated as a presenter on Day 1 (Figure 6).

When comparing the labeled parts of the actual scene, we identified one-to-one and one-to-many. Also, passersby were detected instantly. Moreover, when comparing graphs of social activity amounts and numbers of faces, we extracted the quality of one-on-one communication. From these facts, good results were obtained, considering the continuity of one-to-one communication. In the future, we will evaluate whether the labeling of the obtained social activity amount and social activity situation is close to the actual impression to the camera-wearer who is the user of the social activity measurement through long-term daily life.

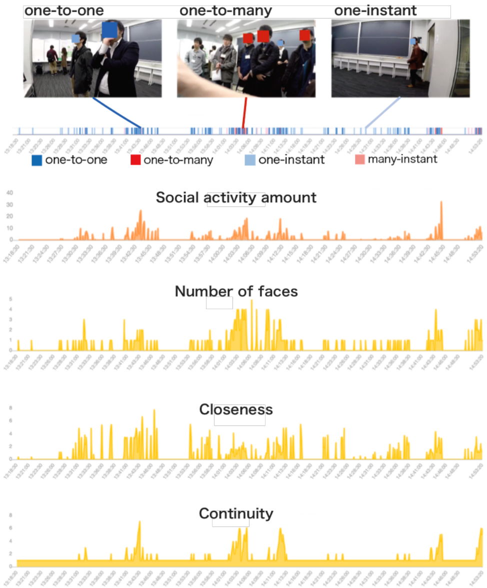


Figure 6: Result of Presenter D on Day 1.

Conclusions

To quantify and visualize social activities performed unconsciously, we proposed a method to measure the social activities using first-person lifelog video. Results of our feasibility study, we believe that the method can be used to numerically express relationships while accounting for quality and quantity although simple to perform. A simple method might be helpful for creating a corpus of multimodal data or identifying the relationship between social activities and sleep. We believe that our proposed method is the first step toward lifelogging for social health, especially during face-to-face communications. In the future, we will investigate the behavioral change by visualizing the social activity.

References

1. Amos, B., Ludwiczuk, B. and Satyanarayanan, M.: Openface: A general-purpose face recognition library with mobile applications, *Technical report, Carnegie Mellon University-CS-16-118*. Carnegie Mellon University School of Computer Science (2016).

2. Danelljan, M., Hager, G., Khan, F.S. and Felsberg, M.: Accurate scale estimation for robust visual tracking, *British Machine Vision Conference, Nottingham*, BMVA Press (2014).

3. Fathi, A., Hodgins, J.K. and Rehg, J.M.: Social interactions: A first-person perspective, *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1226-1233 (2012).

4. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N. and Wood, K.: SenseCam: A Retrospective Memory Aid, *Proceedings of the 8th International Conference on Ubiquitous Computing*, pp.177-193, Springer-Verlag (2006).

5. House, J.S., Landis, K.R. and Umberson, D.: Social relationships and health, *Science*, vol.241, pp.540-540, The American Association for the Advancement of Science (1988).

6. Kasahara, S., Ando, M., Suganuma, K. and Rekimoto, J.: Parallel Eyes: Exploring Human Capability and Behaviors with Paralleled First Person View Sharing, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp.1561-1572, ACM (2016).

7. Michael, S.L., Suneeta, G., Jacqueline, C., Melody, O., Hannah, B., Simon, J.M., Paul, K., Charlie, F., Aiden, D. and Jacqueline K.: Measuring time spent outdoors using a wearable camera and GPS, *In Proceedings of the 4th International SenseCam & Pervasive Imaging Conference (SenseCam '13)*, pp.1-7, ACM (2013).

8. Nakakura, T., Sumi, Y. and Nishida, T.: Neary: Conversational field detection based on situated sound similarity, *12th International Conference on Multimodal Interfaces and 7th Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI 2010).* E94-D (6):1164-1172 (2011).

9. Olguin, D.O., Waber, B.N., Kim, T., Mohan, A., Ara, K. and Pentland, A.: Sensible organizations: Technology and methodology for automatically measuring organizational behavior, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol.39, pp.43-55 (2009).

10. O'Loughlin, G., Cullen, S.J., McGoldrick, A., O'Connor, S., Blain, R., O'Malley, S. and Warrington, G.D.: Using a wearable camera to increase the accuracy of dietary analysis, *American journal of preventive medicine*, vol.44, pp.297-301, Elsevier (2013).

11. Teo, A.R. and Gaw, A.C.: Hikikomori, A Japanese Culture-Bound Syndrome of Social Withdrawal? A Proposal for DSM-V, *The Journal of nervous and mental disease*. 198(6):444-449 (2010).

12. Yonetani, R., Kitani, K.M. and Sato, Y.: Ego-surfing first-person videos, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.5445-5454 (2015).