Social Activity Measurement by Counting Faces Captured in First-Person View Lifelogging Video

Akane Okuno
Future University Hakodate
Hokkaido, Japan
a-okuno@sumilab.org

Yasuyuki Sumi
Future University Hakodate
Hokkaido, Japan
sumi@acm.org

Figure 1: Examples of the scene captured using first-person view lifelogging video.

ABSTRACT

This paper proposes a method to measure the daily face-to-face social activity of a camera wearer by detecting faces captured in first-person view lifelogging videos. This study was inspired by pedometers used to estimate the amount of physical activity by counting the number of steps detected by accelerometers, which is effective for reflecting individual health and facilitating behavior change. We investigated whether we can estimate the amount of social activity by counting the number of faces captured in the first-person view videos like a pedometer. Our system counts not only the number of faces but also weighs in the numbers according to the size of the face (corresponding to a face's closeness) and the amount of time it was shown in the video. By doing so, we confirmed that we can measure the amount of social activity based on the quality of each interaction. For example, if we simply count the number of faces, we overestimate social activities while passing through a crowd of people. Our system, on the other hand, gives a higher score to a social actitivity even when speaking with a single person for a long time, which was also positively evaluated by experiment participants who viewed the lifelogging videos. Through evaluation experiments, many evaluators evaluated the social activity high when the camera wearer speaks. An interesting feature of the proposed system is that it can correctly evaluate such scenes higher as the camera wearer actively engages in conversations with others, even though the system does not measure the camera wearer's utterances. This is because the conversation partners tend to turn their faces towards to the camera wearer, and that increases the number of detected faces as a result. However, the present system fails to correctly estimate the depth of social activity compared to what the camera wearer recalls especially when the conversation partners are standing out of the camera's field of view. The paper briefly descibes how the results can be improved by widening the camera's field of view.

AH2019, March 11–12, 2019, Reims, France

ACM ISBN 978-1-4503-6547-5/19/03.
https://doi.org/10.1145/3311823.3311846

CCS CONCEPTS

• **Human-centered computing** → *Human computer interaction (HCI)*; *Ubiquitous and mobile computing*; *Visualization*; Empirical studies in HCI.

KEYWORDS

Social activity measurement, first-person view video, lifelogging, face detection, quantified self, social health

ACM Reference Format:
Akane Okuno and Yasuyuki Sumi. 2019. Social Activity Measurement by Counting Faces Captured in First-Person View Lifelogging Video. In *Augmented Human International Conference 2019 (AH2019), March 11–12, 2019, Reims, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3311823.3311846

1 INTRODUCTION

The question of this research is, "Is it possible to measure the face-to-face engagement level, that is, the amount of daily social activity, with a simple method?" This paper proposes a method to measure the social activity of engagement with other people by counting faces captured in a lifelogging video (Figure 1).

Originally, the pedometer was an instrument to count steps. In recent years, as a result of advances in technology to recognition patterns of fluctuations in body motion, many wristband-type activity meters (Fitbit, Jawbone, etc.) are able to identify walking, jogging, sleeping, etc. [7]. By aggregating and comparing the data of tens of thousands of users, the objective vision of individual exercise and sleep quantity is simplified, which promotes motivation for exercise.

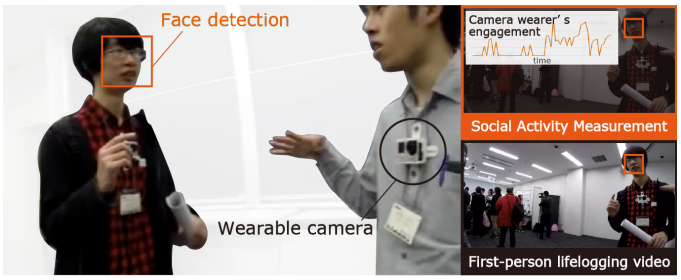

Figure 2: Social activity measurement by counting faces captured in first-person view lifelogging video

First of all, we count the number of faces. We aim to realize a face-meter that keeps track of changes in face-to-face engagement based on the time pattern and that records daily social activity. This is an analogy with the pedometer that accumulates changes in acceleration during exercise and records daily physical activity.

We quantify face-to-face engagement with people and integrate over time. This will make it possible to visualize and review the level of engagement in daily face-to-face social activity, of which it is difficult to be aware. In this study, the value obtained by integrating the face-to-face engagement level over time is defined as the amount of social activity, which we measure. By wearing a camera and acting, various scenes are captured, as shown in Figure 1. Face-to-face communication occurs often in daily life, such as when we meet and talk to people. We use the first-person view video as the lifelogging sensor device (first-person view lifelogging video).

We propose a method to measure the daily social activity of a camera wearer by detecting the face captured in the first-person view lifelogging video. This is done to quantify the face-to-face engagement level using a simple method (Figure 2). To record a stable image, we attach the camera to the chest. In this research, to discuss the requirements of system realization (camera view angle, evaluation formula) and subjective evaluation of users, we record video as part of the research process. The recorded data are used after the camera wearer has confirmed whether there are data to be considered. When this system is used outside the experiment, only the image-processing results and numerical values will be recorded. Users will look back on their own social activity on a smartphone or similar device. We also expect the camera to be small enough to allow for natural sensing.

In the future, we will aim to provide feedback that leads to behavioral changes to improve social health [9], such as loneliness and fatigue in social activities. By counting according to the situation, users can set the target amount of social activity when using the system in daily life. Figure 3 is our prototype application. This taxonomy allows users to ascertain whether they tend to pass by people or spend time with them, and how many people they interact with. In addition, by recording and recalling daily social activity, we think that users might be able to perceive that there were too few or too many face-to-face engagements with people. The prototype UI is based on an activity gauge, like that of the Apple Watch, and an animation of reviewing the activity, like that of SmartBand.

We examined the case in which the amount of social activity tends to feel large by subjective evaluation experiment. In this

Figure 3: Amount of face-to-face social activity can be visualized similarly to the amount of physical activity on a smart phone.

paper, we discuss the effectiveness and limitation of quantifying face-to-face engagement level based on face detection in terms of the inclusiveness of multiple active behavioral characteristics without detailed sensing. Finally, we improved the measurement of dialogue at diagonal or side-by-side positions by using a hemispherical camera. The conclusions obtained from subjective evaluation experiments are shown below.

- In the situation in which the user spoke and/or engaged, the amount of social activity was evaluated as high, and we confirmed the tendency of the other person facing the camera wearer. In other words, it was suggested that social activity can be measured by detecting the face of the partner when the camera wearer performs an active behavior, without measuring the utterance or the gesture itself.
- It was shown that there were scenes that were different from the subjective evaluation when the amount of social activity was calculated by only counting the number of faces. To quantify the amount of social activity, considering the active behavior, it is necessary to weight by proximity and time continuity.
- A few diagonal and side-by-side dialogues were found, and it was determined that an angle of view of 180° or more was necessary. As a result of using a 200° hemispherical camera, it was suggested that measurement of situations with a high level of face-to-face engagement can be improved.

2 RELATED WORK

2.1 Recognition of Social Contexts

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 [17], a voice with a speaker [16], distance with a Bluetooth link [5], and face-to-face detection with an IR sensor [3], various aspects of social context have been measured. The results include predictions about productivity and job satisfaction [15]. Alternatively, technologies that interpret a talking field [14] enabled researchers to use a simple algorithm and a lightweight, networked mobile terminal equipped with a microphone to work in crowds. There are also studies that recognized social context by sensing the wearers themselves. For example, a technique of incorporating a photo reflector into eyeglasses and measuring facial expressions from skin deformation [13] enabled researchers to record complex, daily facial expressions in lifelogs using machine learning. Moreover, social context recognition has been studied using a large amount of long-term data from virtual space [19].

There are many aspects of social activities that require us to well-recognize important information, according to the purpose and scope of its application. We assess the face-to-face social context of a camera wearer using faces that are cues from the first-person view lifelogging video.

2.2 Technique with First-Person View Video

Many techniques have been studied for analyzing first-person view video. There have also been studies on techniques for recognizing social contexts. 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, thus creating a heat map. This can be used to estimate the partner's profile and role in a group setting [6]. A few studies recognize egocentric social situations by measuring the behavior of the camera wearer using the first-person view video[21]. 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 [23]. There is another technique that uses the affinity of head direction to assess the social factors in a group conversation [1].

It is useful to understand the social context of the camera wearer from first-person view video. With our approach, we measure the level of engagement with others by calculating the number of people with detected frontal faces, distances, and the time-weight of the continuity to quantify the daily face-to-face social activity with a simple method.

2.3 Utilization of Captured Experiential Data

Research that extends meta-recognition by combining egocentric and objective information using first-person view video and/or metadata has been studied. For example, there is research that supports memory recall that is difficult for people with memory impairment [8] and research that supports control of everyday eating habits [18]. Additionally, the influence of egocentric and/or objective information on metacognition has been studied empirically. For example, although vision information promotes recall of detailed memory, location information is reported to support inferential processes [11]. Furthermore, because metacognition perceptually changes over time, information on meta-viewpoints is reported to be useful for reviewing experiences [20]. On the other hand, there is research to expand self-perception by using the first-person view video of others in parallel [12].

We quantify face-to-face engagement with people using a first-person view lifelogging video. Our motivation is that this will enable users to visualize and reflect on the level of engagement in daily face-to-face social activity that it is difficult to be aware of.

3 SOCIAL ACTIVITY MEASUREMENT

In this study, the value obtained by integrating the face-to-face engagement level over time is defined as the amount of social activity, which we measure. We show the example result (Figure 4) obtained from social activity measurement using our proposed method compared with a result using a method that only counts the number of faces. If we only count the number of faces, we treat encounters with other people in crowds and close dialogue with specific persons in the same way; thus, we propose counting them separately with distance and time continuity. The amount of social activity shall be the time integral of the value calculated on the basis of the number of people, the proximity, and continuity for each frame. The proposed method and method of counting the number of faces measured the amount of social activity every second.

For example, the amount of social activity is calculated for a situation in which one person talks to a specific person at a short distance (Figure 4: S_1). Additionally, the amount of social activity is calculated in consideration of the situation where three people are talking while maintaining the distance (S_2). Furthermore, the amount of social activity is calculated considering the instantaneous involvement with people in the crowd (S_3).

With the method of counting faces, the cumulative amount of social activity for approximately 20 s is in the order of $S_1 < S_2 \approx S_3$, but with the proposed method, it is $S_3 < S_2 \approx S_1$. Additionally, our method is calculated considering scenes in which there are situations when conversation partner does or does not keep facing the camera wearer, as in frame $t + 9$ of Figure 4.

3.1 Requirements

For face detection, we use OpenFace developed by Carnegie Mellon University [2]. Face tracking of the dlib library [4] in OpenFace tracks what was estimated as the same person's face between frames (Figure 5). In the case of measurement every 10 s, the time continuity $T_i(t)$ increases by 1. We do not detect the sideways face and occipital area but detect frontal faces and measure face-to-face engagement. In terms of design simplicity and privacy consideration, we think it is useful to use the face-detection result without identifying the individual's face.

3.2 Implementation

Specifically, the amount of social activity S of a certain time t is calculated by Formula (1). The closeness $D_i(t)$ represents the size

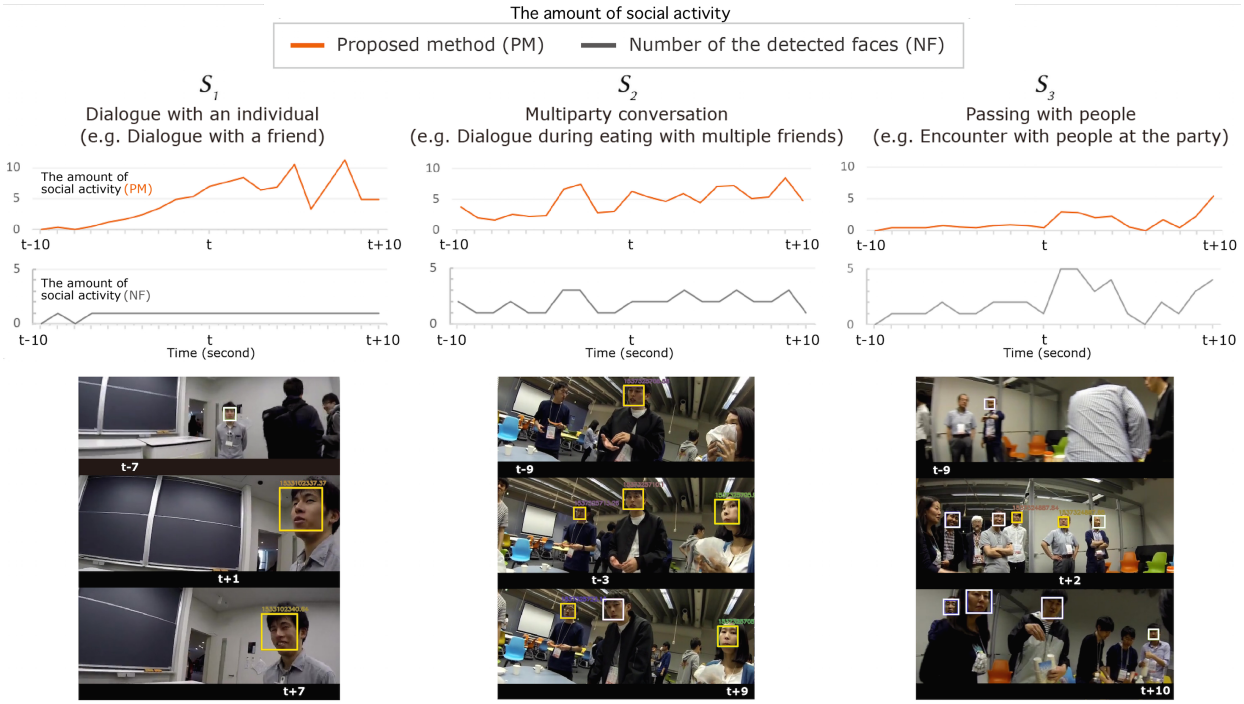

Figure 4: Examples of measurement: Dialogue with an individual, Multiparty conversation, and passing with people

of the face occupying the entire screen. Specifically, it is calculated by Formula (2).

That is, for each frame, the product of the size of each detected face and the continuity at that time is obtained and accumulated. By time-integrating this information, for example, it is possible to measure the amount of social activity during the whole day, extract a specific scene in time, and evaluate the amount of social activity of the scene.

For the detected face identification (ID) number i, the newly issued ID is used every time OpenFace detects a new face. Specifically, different IDs are issued for newly detected faces in a certain frame. However, the same ID is given to the face determined to be the same person as the face detected in the immediately preceding frame. However, when two or more undetected frames intervene, another new ID is issued, even for the same person's face. Utilizing this property, we decided to increment $T_i(t)$ for that ID if the same ID is detected in consecutive frames, and we use this value as time continuity. $T_i(t)$ always starts from 1.

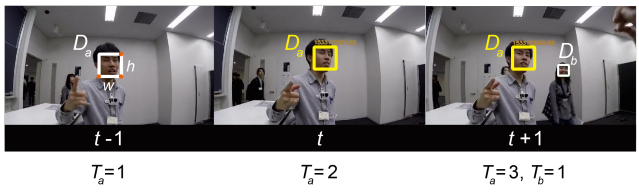

Figure 5: Calculation of size and continuity of each face

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

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

$$D_i = \frac{w_i \cdot h_i}{R} \cdot 100 \tag{2}$$

w_i : The width of the detected face,
h_i : The height of the detected face,
R : Screen resolution.
The units are pixels.

4 SUBJECTIVE EVALUATION EXPERIMENT

We examined the scene in which the user tends to feel that the amount of social activity is large by subjective evaluation experiment. The results of the subjective evaluation experiment will reveal two questions.

- What kind of scene tends to feel as though the amount of social activity is large?
- Is it possible to measure the face-to-face engagement level, that is, the daily amount of social activity, with a simple method?

In order to quantitatively evaluate the subjective amount of social activity, we instructed 8 persons (Table 2) to browse and rearrange the various first-person view videos (Table 1). Each of video contents is a uniformly extracted 1-min activity from the first-person view lifelogging video that is recorded by a visitor who participated in the demonstration and poster session at a conference. Finally, we compared the quantified subjective evaluation using an ordinal scale that is obtained from 8 evaluators and the amount of social activity quantified by the proposed method and by the method of only counting the number of faces.

4.1 Data Collection

From the 8 first-person view lifelogging videos recorded during the demonstration and poster session, we chose Person P8's video of the person who had conversations with other participants in the same place. We uniformly extracted 10 videos that contain 1 minute of various social activities from the approximately 1.5-h first-person view lifelogging video (Table 1).

Table 1: The extracted 10 videos from the 1.5-h video

	Video contents
A	Walking alone in the hallway
B	Walking in the crowd and moving toward the presenter
C	Talking to the presenter and walking in the crowd
D	Talking to the presenter while experiencing the exhibit
E	Talking to the Person P1
F	Group conversation with the Person P1 and another person
G	Hearing a presenter's talk from afar
H	Encounter with Person P1 and having a short conversation
I	Hearing a conversation behind the presenter and visitor
J	Hearing a presenter's talk from afar with many visitors

4.2 Experiment Participant

The evaluators were a total of 8 persons, including a camera wearer, a dialogue partner, and 6 third parties (Table 2). Furthermore, the subjective evaluation was compared from the following three viewpoints. This is to consider the influence of the difference in experience on the evaluation.

- Person's viewpoint as a camera wearer
- Person's viewpoint as a dialogue partner in videos
- Third-party viewpoint that did not appear in videos

Table 2: Participants in subjective evaluation experiment

	Participants
Camera wearer	P8
Dialogue partner	P1
Third person	P2,P3,P4,P5,P6,P7

4.3 Experimental Procedure

We uniformly extracted 10 videos that contains 1 minute of various social activities from the approximately 1.5-h first-person view lifelogging video (Table 1). Additionally, the 8 evaluators (Table 2) were instructed to watch all 10 videos and rearrange them using two symbols, < and =, to sort the videos in ascending order of the amount of social activity. We also instructed them to describe the memo of 0 to 100 and the judgment criteria so as not to mistake the sorting order. The time required for evaluation was approximately 20 min. These tasks were conducted online by all evaluators. Then, we quantified the subjective evaluation of the amount of social activity using an ordinal distance in which the videos were rearranged. Finally, we compared the quantified subjective evaluation using an ordinal scale that was obtained from multiple evaluators with the amounts of social activity quantified by the proposed method and by the method only counting the number of faces. The proposed method and the method counting the number of faces measured the amount of social activity every second.

4.4 Results of the Experiment

We show the experimental results obtained from 10 videos rearranged in ascending order of quantified subjective evaluation in Figure 6. The results were rearranged in order of median of quantified subjective evaluation of the amount of social activity. When the same value was obtained, the videos were sorted alphabetically. We also plotted the amount of social activity quantified by the proposed method (PM) and that quantified by counting the number of faces (NF). Figure 6 shows the order of the obtained results. Because each detailed scale is different, the calculated amount of social activity is plotted in accordance with the maximum value.

In scenes where the amount of social activity is small (A, I) and those where it is large (E, F), the subjective evaluations were generally consistent among the evaluators (Figure 6). There were also scenes (G, J, C, H) in which evaluations tended to be scattered slightly, and scenes (B, D) in which evaluations were varied greatly among evaluators. The outline of the obtained results is shown below.

- In the scene in which the camera wearer spoke and/or engaged, the amount of social activity was evaluated much, and we confirmed the tendency of the other's face facing the camera wearer (Figure 6: C, H, D, E, F).
- There were scenes that were different from the subjective evaluation when the amount of social activity was calculated by only counting the number of faces (Figure 6: J, F).
- There were scenes in which the social activity quantified by both methods was in a different order from that of the subjective evaluation (Figure 6: H, D, E). The face was not completely captured when the dialogue partner of the camera wearer was too close or the standing position became oblique.

The detailed results of the subjective evaluation experiment and the scene are described in Sections 4.3.1 and 4.3.2.

4.4.1 Scene in which the Evaluation was Consistent among the Evaluators. In scenes where the amount of social activity is small (A, I) or large (E, F), the subjective evaluation generally were consistent among the evaluators (Figure 6). There were also scenes (G, J, C, H) in which evaluations tended to be scattered somewhat.

When the camera wearer participated in the conversation and spoke, the amount of social activity tended to be evaluated as large

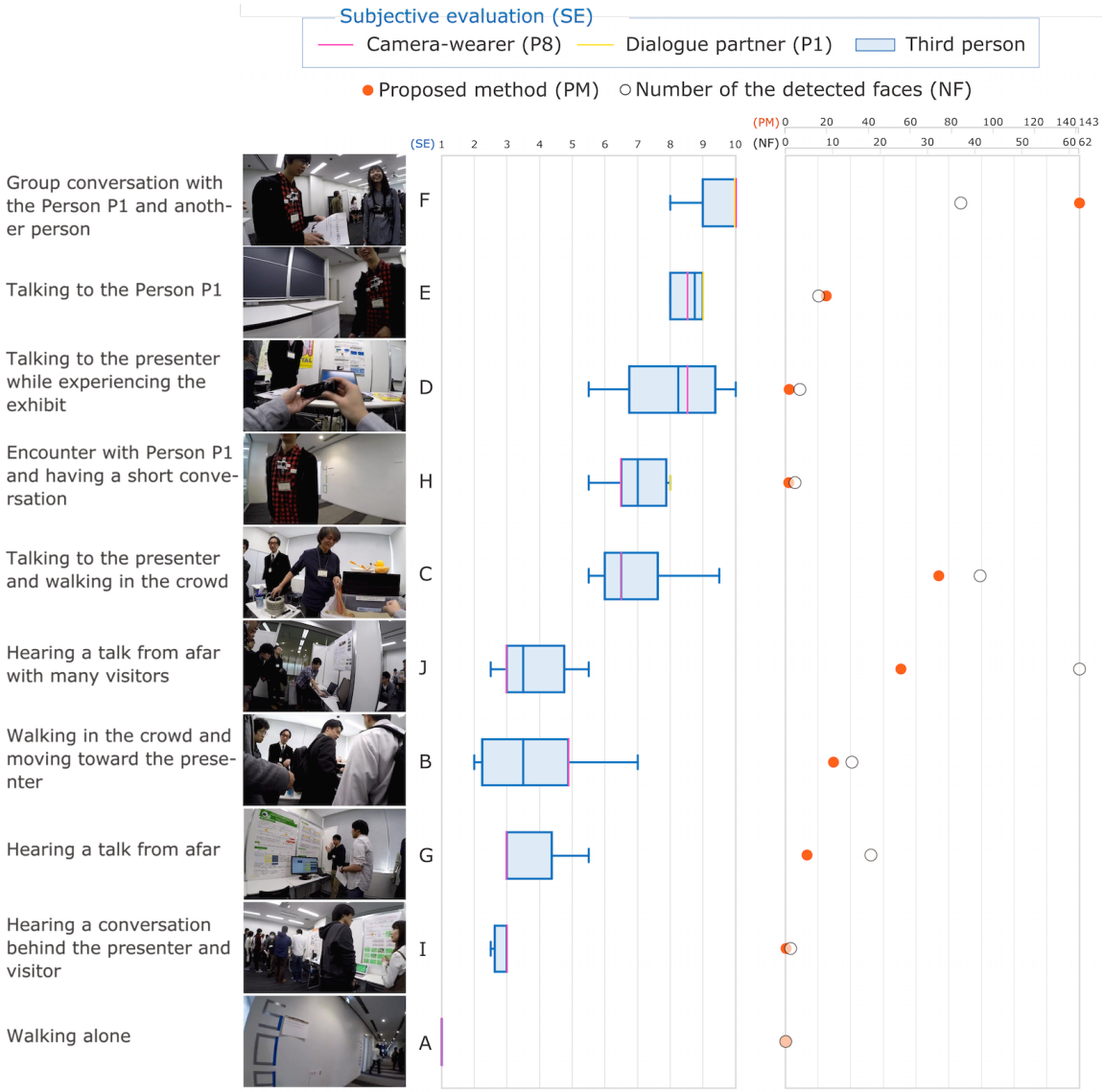

Figure 6: Subjective evaluation results that were quantified by rearranging 10 videos (SE). A comparison of SE and the amount of social activity by proposed method (PM) and with counting the number of faces (NF).

by evaluators. The description that was often seen in the sorting criteria was "*Whether camera wearer is speaking and/or participating in a conversation.*" We confirmed the tendency of the conversation partner facing the camera wearer when the camera wearer was speaking (Figure 6: C, H, D, E, F).

Additionally, the amount of social activity of Scenes J and F quantified by the proposed method was in the same order as in the subjective evaluation. However, the amount of social activity quantified by counting the number of faces was different from the order of subjective evaluation.

On the other hand, in Scenes H and E, the social activity quantified by both methods was in a different order from the subjective evaluation. The face was not completely captured when the dialogue partner of the camera wearer was too close or the standing position became oblique.

The contents of each scene that was consistently evaluated are described below (Scenes A, I, E, F).

Scene A

The camera wearer went down the stairs. He passed a few people in the hallway. He did not talk to anyone.

Scene I

The camera wearer listened to the talks behind the conversation between the presenter and the visitor. After looking around several times, he moved and read the posters. He did not participate in the conversation directly.

Scene E
The camera wearer and Person P1 had a one-on-one conversation. They talked with each other with speech and gestures. On the way, the distance between them became closer or the standing position was oblique. There was a scene in which the face of the conversation partner was not within the viewing angle of the camera and could not be captured properly.

Scene F
The camera wearer and Person P1 had a one-on-one conversation. Then, another person joined the conversation. They talked with each other with speech and gestures. Three people saw the booklet held by Person P1 and approached. There were several scenes in which the face of Person P1 was partially out of the view angle of the camera.

The contents of each scene that tended to be scattered somewhat in evaluation are described below (Scenes G, J, C, H).

Scene G
The camera wearer turned in the direction of the presenter from afar and heard a talk. The presenter was facing the direction of the visitor and the camera wearer. In the first half, there were many visitors next door. In the second half, he looked at his surroundings. He did not have direct conversation or speak.

Scene J
The camera wearer turned in the direction of the presenter from afar and heard a talk. The presenter was facing the direction of the visitor and the camera wearer. The camera wearer was in a position facing the other visitors. He did not have direct conversation or speak.

Scene C
The camera wearer was talking to the presenter while experiencing the demonstration. They talked to each other. He did not have direct conversation, but another presenter and visitor were nearby. When he moved, he passed many other visitors.

Scene H
In the first half, the camera wearer moved after listening to a talk from afar. In the second half, he met Person P1 and talked at a short distance and with an oblique standing position. The conversation was short, approximately 20 s. There was a scene in which the face of Person P1 was not within the viewing angle of the camera and was not captured precisely.

4.4.2 Scenes in which Subjective Evaluation Did Not Match among Evaluators. There were scenes (B, D) where the subjective evaluations were dispersed among the evaluators (Figure 6). Thus, there were individual differences in the quantitative impression on several social activities.

When we confirmed the actual scene, in Scene B, the two situations were mixed in the first half and the second half. Additionally, in Scene D, the camera wearer talked to the presenter for the whole time while experiencing the demonstration. The face of the conversation partner was not within the angle of view of the camera. However, the body was facing the front when they talked.

The contents of each scene for which evaluations did not match among evaluators are described below (Scenes B, D).

Scene B
In the first half, the camera wearer walked in the crowd and moved toward the presenter. In the second half, he experienced the demonstration in front of a presenter. He touched the exhibition along with the presenter and visitor. The presenter talked to the other visitor next to the camera wearer. The camera wearer did not talk with them.

Scene D
The camera wearer talked to the presenter for the whole time while experiencing the demonstration. A conversation partner was standing in front of the seated camera wearer. The face of the conversation partner was not within the angle of view of the camera and could not be seen. However, the conversation partner's body was facing forward when they talked.

5 DISCUSSION

5.1 Scenes in which the Amount of Social Activity Is Large

In the scene in which the camera wearer spoke and/or engaged, the amount of social activity was evaluated as high, and we confirmed the tendency of the other person facing the camera wearer (Figure 6: C, H, D, E, F). Especially when the camera wearer himself speaks, it is evaluated as a scene with a large amount of social activity. The interesting thing here is that our proposed method does not measure "utterance amount" itself. Nonetheless, it is possible to present a high score by the proposed method of detecting faces in the conversation scene. This is because when the camera wearer speaks, the tendency of the surrounding people to face the camera wearer increases, and as a result, faces are detected. Thus, it is suggested that social activity can be measured by detecting the face of the partner when the camera wearer performs an active behavior, without measuring utterances or gestures themselves.

5.2 Active Face-to-Face Engagement

The number of people who face each other does not depend much on the evaluation of the amount of social activity, and the continuity and closeness with the same person leads to higher evaluation. The scoring process works well by not only the number of faces, but also proximity and time continuity (Figure 6: F). With a method of only counting the number of faces, we scored higher than the result of subjective evaluation in the scene of hearing the talks from afar (Figure 6: J). In order to quantify the amount of social activity, considering the active behavior, it is necessary to weight by proximity and time continuity.

5.3 Consistency of Evaluation by Evaluator

The result of subjective evaluation was not significantly affected by whether the evaluator was the camera wearer, dialogue partner, or third party. In other words, the evaluation of the amount of social activity is not considered to be affected much by episodes.

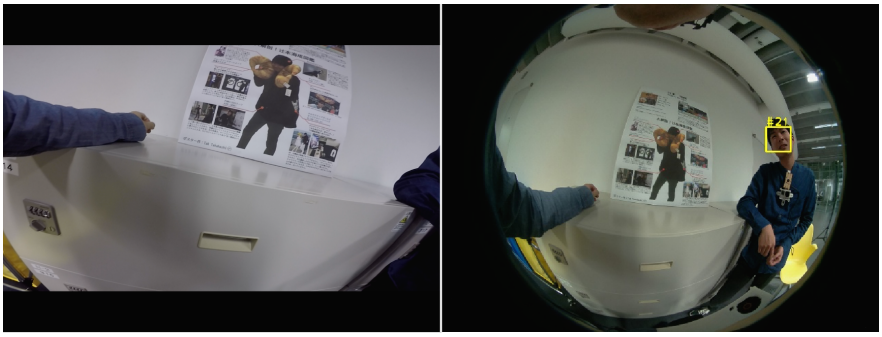

Figure 7: Angle of view for measuring the face-to-face engagement when the standing position is not in front

5.4 Consistency of Evaluation by Type of Scene

The variance of subjective evaluation varied depending on the type of scene. For example, scenes of long talk with a specific person has a high score (Figure 6: E, F), a scene moving in a hallway has a low score, and everyone makes a common judgment (Figure 6: A). Additionally, the evaluation of Scene B, experiencing demonstrations in front of the presenter after participating in a large group as a bystander, and that of Scene D, experiencing demonstration systems while talking with specific people, were quite different. Consideration of individual differences in the impression of such a scene is the limit of the proposed method.

5.5 Field of View of First-Person View Camera

In the scenes where the dialogue parties are out of camera view, the score of the proposed method differs considerably from the subjective evaluation (Figure 6: H, D, E). The reason is that the dialogue partner's face has not been captured by dialogue at diagonal or adjacent positions at a short distance, or at a position with high vertical difference.

We thought the problem would be improved if the view angle of the camera were expanded. For this reason, we measured social activity using a 200 ° hemispheric camera (Figure 7). It was confirmed that the dialogue partner's face could be detected and measured in a one-on-one dialogue with a diagonal standing position. Furthermore, within the university campus, we were able to detect side-by-side faces during meals and work, as well as detect faces during dialogue at a position with a vertical difference (Figure 8). Thus, we think that the proposed method can also measure the face-to-face engagement level when the standing position is not in front.

However, there is fish-eye distortion in this case, so it is necessary to use a robust face detector. In addition, because the size of the face varies depending on where the face is captured by the distortion, we are reconsidering the weighting of the face size by removing the distortion or based on the standing position. There is a limit, as it is not possible to measure the engagement with the person behind.

6 CONCLUSION

We propose a method to measure the daily face-to-face social activity of the camera wearer by detecting faces captured in the first-person view lifelogging video for the purpose of quantifying the face-to-face engagement degree with a simple method.

From the result of the subjective evaluation experiment, it was suggested that social activity can be measured by detecting the face of the partner when the camera wearer performs an active behavior without measuring the speech or gestures themselves. Additionally, to quantify the amount of social activity considering the active behavior, it is necessary to weight by distance shortness and time continuity.

A few diagonal and side-by-side dialogues were found, and it was found that the view angle of 180 ° or more was necessary. As a result of using a 200 ° hemispherical camera, it was suggested that the measurement of situations with a high level of face-to-face engagement can be improved.

Measurement of engagement with people facing away and measurement considering individual impression differences in a few scenes remain the limit of the proposed method.

In the future, we aim to provide feedback that leads to behavioral changes leading to social health, such as reducing loneliness and fatigue in social activities [9]. As a different viewpoint from complex social relationships, we think that the amount of social activity facing people is one of the clues to support young people's social withdrawal [22] and elderly depression [10]. We believe that our proposed method is the first step toward changes in social behavior.

ACKNOWLEDGMENTS

This study was partially supported by Exploratory IT Human Resources Project (MITOU Program) of Information-technology Promotion Agency, Japan (IPA).

REFERENCES

[1] Stefano Alletto, Giuseppe Serra, Simone Calderara, Francesco Solera, and Rita Cucchiara. 2014. From ego to nos-vision: Detecting social relationships in first-person views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 580–585.

[2] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. 2016. OpenFace: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science* (2016).

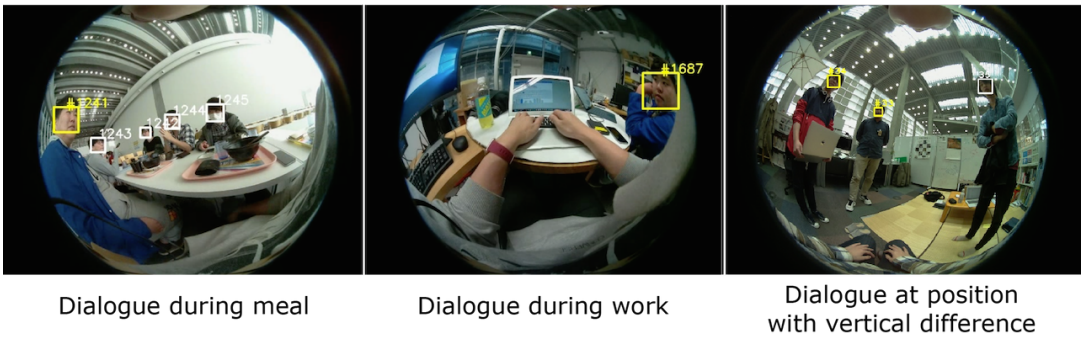

Figure 8: Measurement of dialogue at diagonal or side-by-side position

[3] Tanzeem Choudhury and Alex Pentland. 2003. Sensing and Modeling Human Networks Using the Sociometer. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers (ISWC '03)*. IEEE Computer Society, Washington, DC, USA, 216–. http://dl.acm.org/citation.cfm?id=946249.946901

[4] Martin Danelljan, Gustav Häger, Fahad Khan, and Michael Felsberg. 2014. Accurate scale estimation for robust visual tracking. In *British Machine Vision Conference, Nottingham, September 1-5, 2014*. BMVA Press.

[5] Nathan Eagle and Alex Sandy Pentland. 2009. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology* 63, 7 (2009), 1057–1066.

[6] Alircza Fathi, Jessica K Hodgins, and James M Rehg. 2012. Social interactions: A first-person perspective. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 1226–1233.

[7] Fangfang Guo, Yu Li, Mohan S. Kankanhalli, and Michael S. Brown. 2013. An Evaluation of Wearable Activity Monitoring Devices. In *Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia (PDM '13)*. ACM, New York, NY, USA, 31–34. https://doi.org/10.1145/2509352.2512882

[8] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Woodberry. 2006. SenseCam: A Retrospective Memory Aid. https://www.microsoft.com/en-us/research/publication/sensecam-a-retrospective-memory-aid/, In Proceedings of the 8th International Conference of Ubiquitous Computing (UbiComp 2006). 177–193.

[9] James S House, Karl R Landis, and Debra Umberson. 1988. Social relationships and health. *Science* 241, 4865 (1988), 540–545.

[10] Tatsuhiko Kaji, Kazuo Mishima, Shingo Kitamura, Minori Enomoto, Yukihiro Nagase, Lan Li, Yoshitaka Kaneita, Takashi Ohida, Toru Nishikawa, and Makoto Uchiyama. 2010. Relationship between late-life depression and life stressors: Large-scale cross-sectional study of a representative sample of the Japanese general population. *Psychiatry and clinical neurosciences* 64, 4 (2010), 426–434.

[11] Vaiva Kalnikaite, Abigail Sellen, Steve Whittaker, and David Kirk. 2010. Now Let Me See Where I Was: Understanding How Lifelogs Mediate Memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 2045–2054. https://doi.org/10.1145/1753326.1753638

[12] Shunichi Kasahara, Mitsuhito Ando, Kiyoshi Suganuma, and Jun Rekimoto. 2016. Parallel Eyes: Exploring Human Capability and Behaviors with Paralleled First Person View Sharing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1561–1572. https://doi.org/10.1145/2858036.2858495

[13] Katsutoshi Masai, Yuta Sugiura, Masa Ogata, Kai Kunze, Masahiko Inami, and Maki Sugimoto. 2016. Facial Expression Recognition in Daily Life by Embedded Photo Reflective Sensors on Smart Eyewear. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. ACM, New York, NY, USA, 317–326. https://doi.org/10.1145/2856767.2856770

[14] Toshiya Nakakura, Yasuyuki Sumi, and Toyoaki Nishida. 2011. Neary: Conversational field detection based on situated sound similarity. *IEICE Transactions on Information and Systems* 94, 6 (2011), 1164–1172.

[15] Daniel Olguín, Benjamin N Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex Pentland. 2009. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39, 1 (2009), 43–55.

[16] D Olguin Olguin, Joseph A Paradiso, and Alex Pentland. 2006. Wearable communicator badge: Designing a new platform for revealing organizational dynamics. In *Proceedings of the 10th international symposium on wearable computers (student colloquium)*. 4–6.

[17] Daniel Olguin Olgun and Alex Sandy Pentland. 2006. Human activity recognition: Accuracy across common locations for wearable sensors. In *Proceedings of 2006 10th IEEE international symposium on wearable computers, Montreux, Switzerland*. Citeseer, 11–14.

[18] Gillian O'Loughlin, Sarah Jane Cullen, Adrian McGoldrick, Siobhan O'Connor, Richard Blain, Shane O'Malley, and Giles D Warrington. 2013. Using a wearable camera to increase the accuracy of dietary analysis. *American Journal of Preventive Medicine* 44, 3 (2013), 297–301.

[19] Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann. 2014. Measuring large-scale social networks with high resolution. *PloS one* 9, 4 (2014), e95978.

[20] Yasuyuki Sumi, Masaki Suwa, and Koichi Hanaue. 2018. Effects of Viewing Multiple Viewpoint Videos on Metacognition of Collaborative Experiences. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 648, 13 pages. https://doi.org/10.1145/3173574.3174222

[21] Girmaw Abebe Tadesse and Andrea Cavallaro. 2018. Visual Features for Egocentric Activity Recognition: A Survey. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications (WearSys '18)*. ACM, New York, NY, USA, 48–53. https://doi.org/10.1145/3211960.3211978

[22] Alan Robert Teo and Albert C Gaw. 2010. Hikikomori, A Japanese Culture-Bound Syndrome of Social Withdrawal? A Proposal for DSM-V. *The Journal of Nervous and Mental Disease* 198, 6 (2010), 444.

[23] Ryo Yonetani, Kris M Kitani, and Yoichi Sato. 2015. Ego-surfing first person videos. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 5445–5454.