

一人称ライフログ映像からの顔検出に基づいた 社会活動計測：当事者，二人称，他者視点による印象評価

奥野 茜¹ 角 康之¹

概要：本研究では，人と対面する日常的な社会活動への参与度の定量化と可視化を目的として，一人称ライフログ映像に映り込む顔を数えることでカメラ装着者の社会活動を計測する手法を提案する．顔の数を数えるほかに，顔の大きさと時間継続性を考慮することで，一対一で対話をしている状況と単に人とすれ違った状況を数え分けて実際の印象に近づける．本稿では，学会に参加したカメラ装着者である当事者，二人称，他者視点から，複数の一人称ライフログ映像から一様に抽出した社会活動状況に対する社会活動量への印象評価を行い，提案手法の有効性を検討する．会話や発話をしていて結果的に周囲の顔がカメラ装着者の方を向いている状況では，社会活動量が大きく評価される傾向があることを確認した．このことから，発話やジェスチャそのものを計測せずにそれらの結果として得られるカメラ装着者の方を向いた他者の顔に基づいて，人と対面する日常的な社会活動を定量化できることが示唆された．最後には，人によって社会活動量の印象が異なる状況および新しい計測手法の可能性について議論する．

Social Activity Measurement by Counting Faces Captured in Lifelog Video: Impression Evaluation by Parties, Second-Person and Others

AKANE OKUNO¹ YASUYUKI SUMI¹

1. はじめに

本研究では，人と対面する日常的な社会活動の定量化と可視化を目的として，一人称ライフログ映像に映り込む顔を数えることでカメラ装着者の社会活動を計測する手法を提案する(図1)．個々人が身に着けたカメラに映り込んだ顔の数を数えることで，日々の社会活動量を測ることができるか，というのが本研究の素朴なアイデアである．歩数計が加速度変化から歩数を数えるのと同様，本研究では対面する顔の数を数えることで「顔数計」を実現することを目指す．

歩数計は従来，単純に歩数を数えるだけのものであった．しかし近年，身体動作の揺れのパターンを見分ける技術が進み，リストバンド型の活動量計(Fitbit, SmartBand, Jawbone, Apple Watch など)は，歩行，ジョギング，乗り物による移動，睡眠などを識別できるようになり，活動記

録のライフログとして活用されるようになった．また，数万人単位のユーザのデータを集約・比較することで，個々人の運動量や睡眠量の客観視が簡便化され，そのことが，運動への動機づけを促進している．本研究も，まず対面する顔の数を数えるという単純な方法から始める．そして歩数計が身体活動の計測器として発展したように，顔検出の時間継続性に基づいた数え分けを行うことで，社会活動計測器としての可能性を探索する．



図1 一人称ライフログ映像からの顔検出に基づいたカメラ装着者の人と対面する日常的な社会活動の定量化

¹ 公立はこだて未来大学
Future University Hakodate



図 2 一人称ライフログ映像に映り込む場面の例

対面する顔の数を数える手段としては、胸につけたカメラに映り込んだ顔を顔検出によって数える方法を採用した。ウェアラブルカメラの普及により、映像を常時記録することが可能になっている。本研究では、一人称視点映像をライフログとして使用する(以下、一人称ライフログ映像と呼ぶ)。カメラを装着して行動することで、図2のように様々な場面が捉えられる。日常生活の中で、人と関わり合うときには顔と対面することが多い。例えば、人と会話をするとき、雑踏を通り抜けるときである。その他には、テレビや雑誌を見るとき、鏡を見るときにも顔と対面する。まずは、人と会話をするときや雑踏の通り抜けのように、直に他者と関わり合う社会的な活動への参加度の定量化について取り組む。

カメラ装着者の顔との対面具合を社会活動への参加度と解釈する。定量化と可視化を行う際に、顔の数を数えるだけでは、一対一の会話と一人の人とすれ違う状況、近距離での会話と遠くに人が座っている状況などを等しく扱ってしまう。そのため、顔の大きさと時間連続性を考慮することで、実際の印象に近づけることを行う。社会活動をライフログとして長期的に記録し、振り返ることを可能にすることで、社会活動における孤独感や疲労感の軽減のような社会的健康 [1], [2] に向けた行動変容につながることを期待している。

本稿では、学会に参加したカメラ装着者である当事者、二人称、他者視点から、複数の一人称ライフログ映像から一様に抽出した状況に対する社会活動量への印象評価を行い、顔検出に基づいてカメラ装着者の社会活動を計測する提案手法の有効性を検討する。最後には、人によって社会活動量の印象が異なる状況および新しい計測手法の可能性について議論する。

2. 関連研究

2.1 社会的状況の認識

個人および集団の社会的状況を非視覚情報から認識する技術は、これまで多くの研究でされている。例えば、加速度センサから運動、スピーカから音声、Bluetooth から人への接近、IR センサから対面対話の認識を組み合わせることで、社会的状況の様々な側面を計測し、生産性および職務満足度などの結果を予測している [3]。一方で、社会活動の重要な側面を掘り下げて状況を解釈する研究がされている。例えば、対面対話時の発話を支援するためにモバイル端末を用いて発話を計測する技術 [4] では、単純な発話情報の解釈をより深く探究することの可能性を示している。一方で、会話場を検出する技術 [5] では、ネットワークモジュールとマイクを備えたモバイル端末を用いて単純なアルゴリズムと軽量プロセスで動作することを可能としている。また、装着者自身をセンシングすることで社会的状況を認識する研究もされている。例えば、眼鏡にフォトフレクタを組み込み皮膚変形から表情を日常的に計測する技術 [6] では、入力情報を複数個用意して機械学習することで表情という複雑な状況をライフログとして記録することを可能としている。さらに、社会的状況を認識した結果をもとにフィードバックを与えて支援する研究もされている。例えば、AR グラスを用いて装着者の非言語情報を認識し、発表者の社会的インタラクションの質を向上するための研究 [7] がある。一方で、仮想空間上の長期的な大量の情報を用いた社会的状況の認識 [8] や、社会的つながりが健康的な行動に及ぼす影響の分析もされている [9]。

これらのように、社会活動には様々な側面があり、目的や応用範囲に応じて重要な情報を認識する必要がある

ことがわかる。非視覚情報から認識できることがあれば、視覚情報から新たに認識できる社会的な状況があると考えられる。近年ではモバイル端末にも高精度のカメラが搭載され、ウェアラブルカメラが人々に利用される機会が増えている。プライバシーの問題に焦点を当てた研究もされている [10]。本研究では、ウェアラブルカメラから得られる一人称視点映像に含まれる視覚情報に普遍的に存在する顔を手掛かりにして、カメラ装着者の社会的な活動への参加度を計算する。設計の簡素化やプライバシーの観点から、個人の顔を特定せず顔検出結果だけを利用する。

2.2 一人称視点映像を用いた技術

一人称視点映像を振り返る技術はこれまでに多く研究されている。例えば、カメラ装着者の移動、静止、手作業、人との会話場面を手掛かりに、関心に合わせて映像の再生速度を変更し、表示を強調して簡単に振り返る技術がある [11]。また、深層学習で一人称視点映像を解釈して場面を分類する手法も提案されている [12]。社会的な状況を認識するための技術に関しても、これまでに研究がされている。例えば、顔の位置および向きからカメラ装着者の対話相手の視線を計算し、3D 空間へのマッピングやヒートマップの作成、さらに複数人での計測からグループ内での対話相手や役割の推定をしている [13]。また、グループ会話のような複数人のカメラ装着者が対面しているときの互いの頭の動きの相関を計算することで、カメラ装着者自身の顔の位置を特定することを可能としている [14]。頭部方向の親和性を社会的インタラクションの距離と定義することで、グループ会話の識別を可能としている技術もある [15]。さらには、一人称視点映像を用いて、人の健康の支援 [16]、[17] や知覚の拡張 [18] の研究がされている。

一人称視点映像からカメラ装着者の状況を認識し、活動を支援することが可能になってきている。本研究では、カメラ装着者の社会活動への参加度を、検出された顔の個数、顔の大きさ、時間継続性から計算する。日常的に人と対面する社会活動をライフログとして長期的に記録し、振り返ることを可能にすることで、社会活動における孤独感や疲労感の軽減のような社会的健康に向けた行動変容につながることを期待している。

3. 顔検出に基づいた社会活動計測

本研究では、日常生活における顔との対面具合を社会活動への参加度と解釈する。どれくらいの量、どのような状況であったのかを、社会活動量、社会活動状況として定義し、定量化と可視化を行う。具体的には、量の算出と状況の判別を行う。顔検出から得られる顔の個数、顔の大きさ、時間継続性をもとに人数、距離の近さ、持続性を考慮することで実際の印象に近づけることを考える。設計の簡素化やプライバシーの観点から、個人の顔を特定せず顔検出結

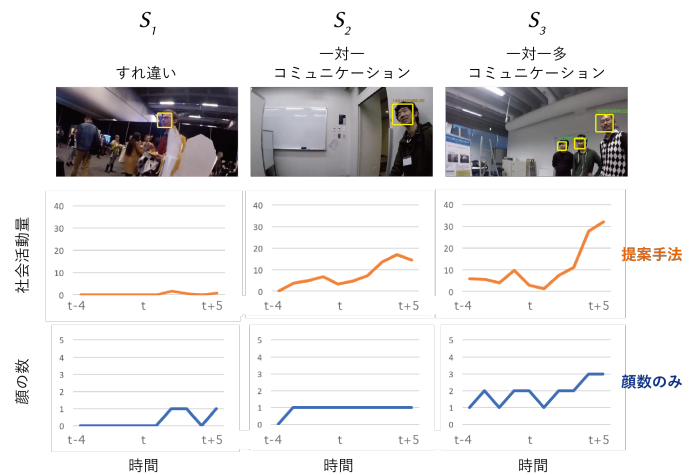


図 3 顔検出に基づいた社会活動計測の例

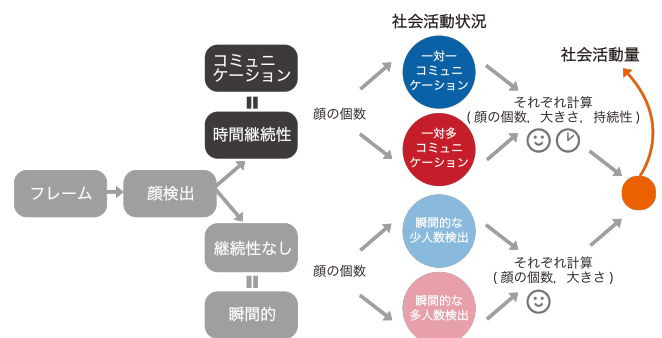


図 4 顔検出に基づいた社会活動計測の流れ

果だけを利用する。顔検出には、カーネギーメロン大学が開発した OpenFace [19] を用いた。中で使われている dlib ライブラリのフェイストラッキングは、フレーム間で同一人物の顔と推定されたものを追跡する機能がある [20]。そこで本研究では、フェイストラッキングが同一人物の顔を連続検出した場合にその相手との持続的なインタラクションと解釈することとする。これらの技術を用いて得た数値をもとに、社会活動量の算出と社会活動状況の判別を行う。

図 3 は社会活動計測の結果の例である。単純に顔の数を数えるだけだと、雑踏での他人とのすれ違いも特定の人物との密な対話も同一に扱ってしまうため、相手との距離の近さ (顔の大きさ) と持続性 (時間継続性) を考慮する。社会活動量は、フレームごとに人数、距離の近さ、持続性をもとに算出された値の時間積分とする。その際に、人数と持続性をもとに 4 種類の社会活動状況を判別し数え分けて値を求める (図 4)。これにより、図 6 の $t+1$ フレーム目のように複数の状況が混在している場面でも印象に近い社会活動量を求めることができると考える。また、状況ごとに数え分けることで、日常生活で利用する際には各々の社会活動量の目標値を設定して行動することができると考える。また、量の算出に加えて、4 種類の社会活動状況のラベル付けを行う。これにより、長時間の計測をした際には、カ

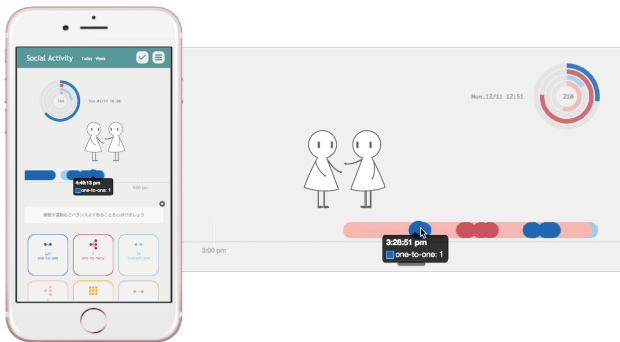


図5 顔検出に基づいた社会活動量と社会活動状況のラベル付けの応用例 (Apple Watch のアクティビティゲージ, Sony SmartBand のアニメーションを参考にしたプロトタイプ)

メラ装着者が人とすれ違っていたのか、人と一緒に過ごしていたのか、独りで過ごしていたのかを知ることができると考える。

社会活動量の計算の詳細については 3.1 節, 社会活動状況の判別の詳細については 3.2 節で述べる。

3.1 社会活動量の計算

本研究では, 一人称ライフログ映像内に映り込んだ顔の数の時間積分で社会活動量を計算する。その際には, 映り込んだ顔ごとに顔の大きさと時間持続性で重み付けすることで, より近くに対面し, かつ, 持続的なインタラクションを重視することとする。具体的には, ある時刻 t の社会活動量 S は次の式で計算する (図 6 参照)。

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

$$\left(\begin{array}{l} i: \text{検出された顔の識別番号,} \\ T_i(t): \text{持続性 (同一顔の検出継続フレーム数),} \\ D_i(t): \text{顔の大きさ (画面全体に占める顔の面積),} \\ m: \text{時刻 } t \text{ までの計測フレーム数 (経過時間),} \\ n: \text{時刻 } t \text{ までの累計人数 (顔の個数)} \end{array} \right)$$

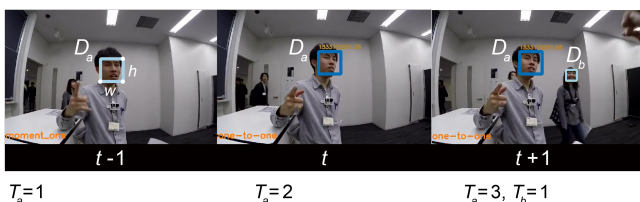


図6 検出された顔ごとの大きさと持続性の計算

1 フレームごとに, 検出された顔ごとの大きさとその時点での持続性の積を求め, それらを累積する。そして時間積分することで, 例えば, 1 日全体の社会活動量を測ったり, ある特定のシーンを時間的に切り出して, そのシーン

の社会活動量を評価することが可能になる。

検出された顔の識別番号 i は, 新たな顔を検出する度に新しく発行する ID を利用する。具体的には, あるフレームで新たに検出された顔は, それぞれ異なる ID が発行される。一方, 直前のフレームで検出された顔と同一人物と判定された顔には同じ ID が付与される。ただし, 2 フレーム以上の未検出フレームが間に割り込んだ際は, 例え同一人物の顔でも別の新しい ID が発行される。この性質を利用して, 我々は, 同一 ID が連続フレームで検出された場合には, その ID の T_i をカウントアップしていき, 持続性として利用することとした。なお, すべての T_i は 1 から始まる。

顔の大きさ D_i は, 撮影画面全体に占めるその顔の大きさを表す。具体的には次の式で計算する。

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

$$\left(\begin{array}{l} w_i: \text{検出顔 } i \text{ の幅,} \\ h_i: \text{検出顔 } i \text{ の高さ,} \\ R: \text{画面解像度 (pixel).} \end{array} \right)$$

3.2 社会活動状況の判別

人数と持続性をもとに, 以下の 4 つの状況を定義した。

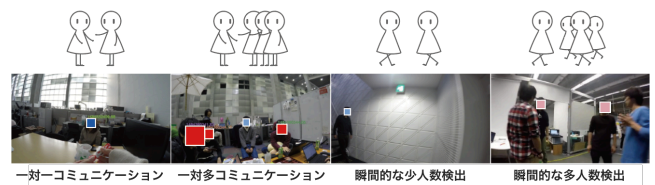


図7 社会活動状況のラベル付け定義。

- (1) 一対一コミュニケーション: 連続検出された顔が 1 つだけある状態
- (2) 一対多コミュニケーション: 連続検出された顔が 2 つ以上ある状態
- (3) 瞬間的な少人数検出: 新たに 1 つだけの顔が検出された状態
- (4) 瞬間的な多人数検出: 新たに 2 つ以上顔が検出された状態

社会活動状況の判別は, 社会活動量の数え分けと, 社会活動状況のラベル付けのために行う。上記の (1), (2) の状態を, 参加度が高く対面コミュニケーションが成立している状況であると解釈する。初期状態と顔検出なしの場面は独りであるとする。フェイストラッキングがされた顔に ID を割り当て, 持続性があると判定する。持続性と顔検出から得られた人数に基づいて社会活動状況の状態は遷移する。

ラベル付けにおいて, 1 つのフレームに連続検出された顔

表 1 印象評価の参加者

	評価 1	評価 2
当事者	P8	P1
二人称	P1	P8
他者	P2,P3,P4,P5,P6,P7	P2,P3,P4,P5,P6,P7

と新規検出された顔が混在した際には、対面コミュニケーションが成立している状況 (1) または (2) であると解釈する。参加度が高い状況のほうがカメラ装着者の社会活動状況に関わっているはずであると考えたからである。図 7 のような 4 種類の社会活動状況が判別される。

4. 当事者および他者による印象評価

顔検出に基づいて社会活動を計測する提案手法の有効性を検討するために、学会に参加したカメラ装着者である当事者および他者に、複数の一人称ライフログ映像から一様に抽出した社会活動状況に対する社会活動量への印象評価を行った。

4.1 印象評価の参加者

学会に参加した際に記録した 8 名の一人称ライフログ映像のうち、デモセッションに参加して見てまわっていたカメラ装着者の中で、互いに居合わせていた時間帯が複数あった 2 名のデータを使用した。これは、

- カメラ装着者である当事者視点
- 互いに居合わせていた二人称視点
- 互いに居合わせていなかった他者視点

の 3 つの視点から社会活動量への印象を比較するためである。評価者は、1 セットにつき 1 名の当事者、1 名の二人称視点、6 名の他者視点の合計 8 名とした (表 1)。6 名の他者視点の中には、当日に近くにいたが偶然居合わせていなかった人と、その場に全く居合わせていなかった人の両方が含まれている。

4.2 印象評価の手順

実験者が、様々な 1 分間の社会活動状況の映像を約 1.5 時間の一人称ライフログ映像から一様に 10 個抽出した。次に、評価者には 10 個の映像を全て視聴し社会活動量が小さい順に「<, =」の 2 つの記号を使って並び替えることを教示した。その際に、0 から 100 の数値と判断基準について記述することも合わせて教示した。実験者が、順位付けされた値を用いて社会活動量の正解データを作成した。最後に提案手法と顔数のみで 1 秒おきに計算した場合と比較した。これを 2 セット行った。

4.3 印象評価の結果

1 セット目 (評価 1) の社会活動量が小さい順に並び替えられた 10 個の映像から得た結果を図 8、2 セット目 (評価

2) を図 9 に示す。また、並び替え結果から得られた正解データの中央値と、提案手法および顔数のみの計算から得られる社会活動量を比較した結果を図 10 に示す。図 10 の横軸の値は、正解データの中央値の合計が 55 であるため、範囲がなく計算される社会活動量と顔数のみの値については正解データの全体の幅に合わせている。

評価 1 と評価 2 の両方で、社会活動量が小さい場面 (P8-A, P8-I, P8-G, P8-J, P1-B, P1-F, P1-C) や大きい場面 (P8-C, P8-H, P8-E, P8-F, P1-A, P1-D, P1-I, P1-E) では評価者間で印象が概ね合致していた (図 8, 図 9)。当事者、二人称視点の印象においては外れ値が生じることがなかった。並び替えの判断基準で多く見られた記述は、「会話をしているか、発話をしているかどうか」であった。会話および発話をしている場面の実際の様子を確認してみると、会話相手の顔がカメラ装着者の方向を向いていた。

また、評価者間で印象が分散している場面 (P8-B, P8-D, P1-J) があった (図 8, 図 9)。その場面の実際の様子を確認してみると、デモの体験をしている後ろ姿の人の間を歩き最後に発表者の前でデモを体験する様子 (P8-B)、会話および発話をしながらデモを体験しているが、会話相手の顔がカメラの画角に収まっていなく見切れている様子 (P8-D)、近い距離で発表者から説明を受けているが、会話相手の顔がカメラの画角に収まっていなく見切れている様子 (P1-J) となっていた。

正解と提案手法および顔数のみの計算から得られる社会活動量を比較した結果から (図 10)、提案手法では評価 1 と評価 2 の両方で正の相関 ($r=0.6$) が認められた。一方で顔数のみの計算では評価 1 において、ほとんど相関が見られなかった ($r=0.1$)。顔が見切れていた場面 P8-D, P1-J のほかにも正解と異なっている場面があった (P8-H, P8-E, P1-H)。実際の様子を確認してみると、場面 P8-E では一対一で会話をしていたが、立っている場所が正面ではなく斜めでありカメラの画角から外れることが数回あった。また、場面 P8-H, P1-H では会話相手の立ち位置に加え偶発的な短い会話となっており、顔を捉えることができていなかった。

4.4 印象評価の考察

多くの評価者が社会活動量を大きく評価する傾向がある場面は、会話に参加して発話をしている場面であった。また、会話および発話をしている場面の実際の様子を確認してみると、会話相手の顔が結果的にカメラ装着者の方向を向いていたことから、周囲の顔がカメラ装着者の方を向いている状況では、社会活動量が大きく評価される傾向があった。これらのことから、発話やジェスチャーそのものを計測せずにそれらの結果として得られるカメラ装着者の方を向いた他者の顔に基づいて、人と対面する日常的な社会活動を定量化できることが示唆された。

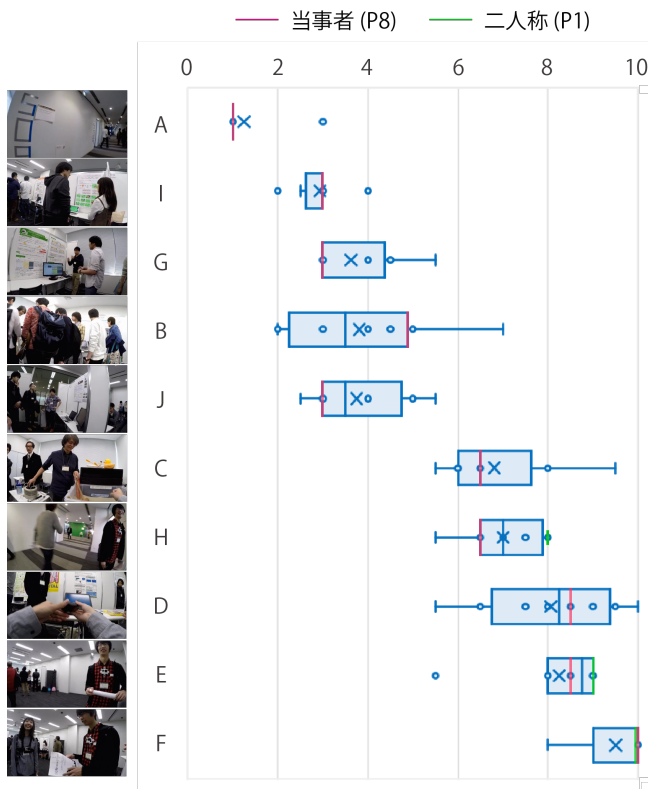


図 8 評価 1：当事者 (P8) の一人称ライフログ映像から一様に抽出した社会活動状況に対する社会活動量への印象評価

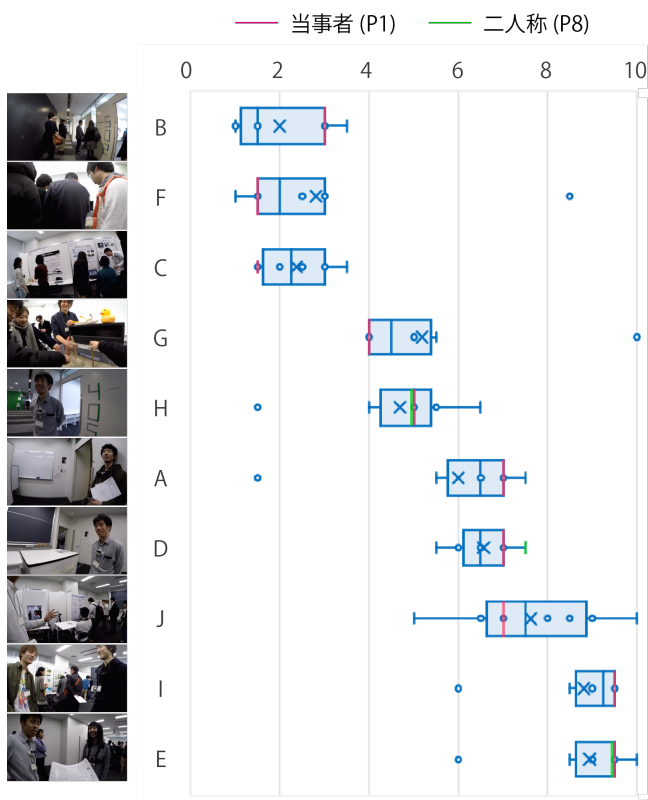


図 9 評価 2：当事者 (P1) の一人称ライフログ映像から一様に抽出した社会活動状況に対する社会活動量への印象評価

一方で、評価者間で印象が分散している場面にあった対面していない人との関わりの定量化については、提案手法では難しく限界となっている (P8-B)。また、場面 P8-B の前半に見学者の後ろ姿、後半に発表者の顔が映っており状況が混ざっているため、評価者によって違いが生じた可能性がある。当事者 P8 は社会活動量を他の人よりも大きく評価していることから、人と対面した状況のほうが印象に残る傾向があるとも考えられる。

外れ値は他者視点で生じ、その判断基準は、「会話をしている時間のみを見た」「敬語を使う必要のある相手の話を聞いたり話したりすることで疲れるかどうか」であった。他者視点での外れ値の判断基準から、会話の時間のみだと多くの人の印象とずれることが考えられる。一方で疲れを基準にすると会話場面であっても相手の属性で変わってしまい多くの人の印象とずれることが考えられる。これらのように人によって社会活動量の大小を感じる背景が異なる場合には、提案手法の限界があることがわかった。

会話相手の顔がカメラの画角に収まっていなく見切れている様子 (P8-H, P8-D, P8-E, P1-H, P1-J) に関しては、カメラの画角を広角にすることで改善できると考える。ただ、立ち位置が斜めである会話のほか、真横の会話でカメラ装着者の方を向いている顔を捉えるためには、180 度以上の画角のカメラが必要であると考えられる。また、偶発的な短い会話での顔を捉えるためには計測する時間の粒度を細かくすることで改善できる可能性があると考えられる。

正解と提案手法および顔数のみの計算から得られる社会活動量を比較した結果 (図 10) から、評価 1 の評価結果の顔数のみでは、遠くで発表を聞いて他の見学者も遠くにいる場面が最も社会活動量が大きいという解釈になり、正解の並び替えの順 $P8-J < C < F$ とは逆の順 $P8-F < C < J$ となっているためほとんど相関が見られなかったと考える。評価 2 の評価結果では、提案手法と顔数のみからの計算結果が似ていた。顔が検出された場面とできなかった場面の差は提案手法の方が大きくなっているため顔数のみのほうが正の相関が強い結果になっていたと考えられる。これらのことから、カメラの画角と計測する時間の粒度を改善することで、提案手法は実際の印象により近づくことが示唆された。

5. おわりに

人と対面する日常的な社会活動への参加度の定量化と可視化を目的として、一人称ライフログ映像に映り込む顔を数えることでカメラ装着者の社会活動を計測する手法を提案した。学会に参加したカメラ装着者である当事者、二人称、他者視点から、複数の一人称ライフログ映像から一様に抽出した社会活動状況に対する社会活動量への印象評価を行った結果、会話や発話をしていて結果的に周囲の顔がカメラ装着者の方を向いている状況では、社会活動量が

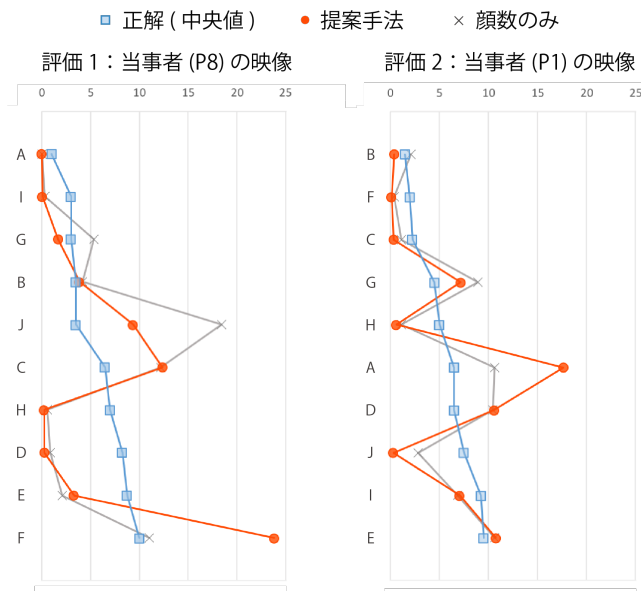


図 10 正解データの中央値と、提案手法および顔数のみの計算から得られる社会活動量を比較した結果 (正解との相関係数 (評価 1, 評価 2), 提案手法: 0.6,0.6, 顔数のみ:0.1,0.7)

大きく評価される傾向があることを確認した。このことから、発話やジェスチャそのものを計測せずにそれらの結果として得られるカメラ装着者の方を向いた他者の顔に基づいて、人と対面する日常的な社会活動を定量化できることが示唆された。また、後ろ向きの人との関わりの計算といった提案手法の限界とカメラの画角、計測時間の粒度を改善するといった課題が明らかになった。

今後は、長期的に蓄えられた一人称ライブログ映像から社会活動の視覚的フィードバックを行い、社会活動における孤独感や疲労感の軽減のような社会的健康につながる行動変容の支援を目指す。今回はその第一歩だと考える。

参考文献

[1] House, J. S., Landis, K. R. and Umberson, D.: Social relationships and health, *Science*, Vol. 241, No. 4865, pp. 540–545 (1988).

[2] 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*, Vol. 198, No. 6, p. 444 (2010).

[3] Olguín, D., 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, No. 1, pp. 43–55 (2009).

[4] Lee, Y., Min, C., Hwang, C., Lee, J., Hwang, I., Ju, Y., Yoo, C., Moon, M., Lee, U. and Song, J.: Socio-Phone: Everyday Face-to-face Interaction Monitoring Platform Using Multi-phone Sensor Fusion, *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys '13*, New York, NY, USA, ACM, pp. 375–388 (online), DOI: 10.1145/2462456.2465426 (2013).

[5] Nakakura, T., Sumi, Y. and Nishida, T.: Neary: Conversational field detection based on situated sound similarity, *IEICE Transactions on Information and Systems*, Vol. 94, No. 6, pp. 1164–1172 (2011).

[6] Masai, K., Sugiura, Y., Ogata, M., Kunze, K., Inami, M. and Sugimoto, M.: Facial Expression Recognition in Daily Life by Embedded Photo Reflective Sensors on Smart Eyewear, *Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI '16*, New York, NY, USA, ACM, pp. 317–326 (online), DOI: 10.1145/2856767.2856770 (2016).

[7] Damian, I., Tan, C. S. S., Baur, T., Schöning, J., Luyten, K. and André, E.: Augmenting Social Interactions: Real-time Behavioural Feedback Using Social Signal Processing Techniques, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, New York, NY, USA, ACM, pp. 565–574 (online), DOI: 10.1145/2702123.2702314 (2015).

[8] Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E. and Lehmann, S.: Measuring large-scale social networks with high resolution, *PloS one*, Vol. 9, No. 4, p. e95978 (2014).

[9] Centola, D.: The spread of behavior in an online social network experiment, *science*, Vol. 329, No. 5996, pp. 1194–1197 (2010).

[10] Hoyle, R., Templeman, R., Anthony, D., Crandall, D. and Kapadia, A.: Sensitive Lifelogs: A Privacy Analysis of Photos from Wearable Cameras, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, New York, NY, USA, ACM, pp. 1645–1648 (online), DOI: 10.1145/2702123.2702183 (2015).

[11] Higuchi, K., Yonetani, R. and Sato, Y.: EgoScanning: Quickly Scanning First-Person Videos with Egocentric Elastic Timelines, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*, New York, NY, USA, ACM, pp. 6536–6546 (online), DOI: 10.1145/3025453.3025821 (2017).

[12] Castro, D., Hickson, S., Bettadapura, V., Thomaz, E., Abowd, G., Christensen, H. and Essa, I.: Predicting Daily Activities from Egocentric Images Using Deep Learning, *Proceedings of the 2015 ACM International Symposium on Wearable Computers, ISWC '15*, New York, NY, USA, ACM, pp. 75–82 (online), DOI: 10.1145/2802083.2808398 (2015).

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

[14] Yonetani, R., Kitani, K. M. and Sato, Y.: Ego-surfing first person videos, *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, IEEE, pp. 5445–5454 (2015).

[15] Alletto, S., Serra, G., Calderara, S., Solera, F. and Cucchiara, R.: From ego to nos-vision: Detecting social relationships in first-person views, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 580–585 (2014).

[16] Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N. and Woodberry, K.: SenseCam: A Retrospective Memory Aid, *Proceedings of the 8th International Conference of Ubiquitous Computing (UbiComp 2006)*, Springer Verlag, pp. 177–193 (online), available from (<https://www.microsoft.com/en-us/research/publication/sensecam-a-retrospective-memory-aid/>) (2006).

- [17] 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, No. 3, pp. 297–301 (2013).
- [18] 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*, CHI '16, New York, NY, USA, ACM, pp. 1561–1572 (online), DOI: 10.1145/2858036.2858495 (2016).
- [19] Amos, B., Ludwiczuk, B. and Satyanarayanan, M.: OpenFace: A general-purpose face recognition library with mobile applications, *CMU School of Computer Science* (2016).
- [20] Danelljan, M., Häger, G., Khan, F. and Felsberg, M.: Accurate scale estimation for robust visual tracking, *British Machine Vision Conference, Nottingham, September 1-5, 2014*, BMVA Press (2014).