

一人称ライフログ映像からの顔検出に基づいた 社会活動計測と主観評価

奥野 茜^{1,a)} 角 康之^{1,b)}

概要：人との対面時の参与度を定量化して時間方向に積分すると、日々の人と直に関わり合う社会活動への参与度を可視化して振り返ることができる。本研究では、対面時の参与度を時間方向に積分した値を社会活動量と定義して計ることに取り組む。我々は既に、対面時の参与度を簡素な方法で定量化することを目的として、一人称ライフログ映像中の顔を検出してカメラ装着者の日々の社会活動を計測する手法を提案してきた。顔の個数を数え上げるだけだと雑踏での他者との遭遇や特定の人物との密な対話を同一に扱ってしまうが、距離の近さと時間連続性で重み付けをして数え分けることができる。本稿では、社会活動量が多いと感じる傾向がある状況を主観評価実験から調べ、詳細なセンシングをせずに一人称ライフログ映像からの顔検出に基づいて人との対面時の参与度を定量化することの有効性と課題について議論する。最後には、斜めや隣り合わせでの対話の計測と画角について述べる。

1. はじめに

「個々人が身に着けたカメラに映り込んだ顔を数え上げることで、日々の社会活動量を計ることができるか？」というのが本研究のアイデアである。歩数計は従来、歩数を数えるものであった。身体動作の揺れのパターンを計測する技術が進み、リストバンド型の身体活動量計（Fitbit, Jawbone など）は、歩行、ジョギングなどを凡そ認識できるようになり [1]、ライフログデバイスとして活用されるようになった。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。我々は、まず対面する顔を数え上げるという単純な方法から始め、歩数計が運動時の加速度の変化を積み重ねて日々の身体活動量を記録するのと同様、対面時の参与度の変化を積み重ねて日々の社会活動量を記録する顔数計を実現することを目指す。日々の社会活動量を記録し振り返ることで、直に人と関わることが極端に少なすぎる、あるいは多すぎるといったことに気づき、孤独感や疲労感の軽減といった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。

人との対面時の参与度を定量化して時間方向に積分すると、日々の人と直に関わり合う社会活動への参与度を可視化して振り返ることができる。本研究では、対面時の参与

度を時間方向に積分した値を社会活動量と定義して計ることに取り組む。我々は既に、人との対面時の参与度を簡素な方法で定量化することを目的として、一人称ライフログ映像中の顔を検出してカメラ装着者の日々の社会活動を計測する手法を提案してきた [2]。カメラを装着して行動することで、図 1 のように様々なシーンが捉えられる。例えば、人と会話をするとき、雑踏を通り抜けるときに人と対面する。我々は一人称視点映像を用いて活動をライフログングすることにした。

一人称ライフログ映像からの顔検出に基づいて、直に人と対面するときの参与度および社会活動量の定量化（図 2）を行う。安定した映像を記録するためにカメラは胸の部分に装着する。なお、現在は研究における分析のために映像を全て記録しているが、実際に利用するときには画像処理された結果および数値のみを記録しスマートフォンなどで振り返ることを想定している。カメラは、自然なセンシングが可能になるくらい小さくなるだろうと考えている。

本稿では、社会活動量が多いと感じる傾向がある状況を主観評価実験から調べ、詳細なセンシングをせずに一人称ライフログ映像からの顔検出に基づいて人との対面時の参与度を定量化する手法の有効性と課題について議論する。最後には、斜めや隣り合わせでの対話の計測と画角について述べる。

主観評価実験から得られた結論を以下に示す。

- 発話をしている状況ほど社会活動量が多いと評価され、相手の顔がカメラ装着者の方向を向く傾向を確認し

¹ 公立はこだて未来大学

^{a)} a-okuno@sumilab.org

^{b)} sumi@acm.org

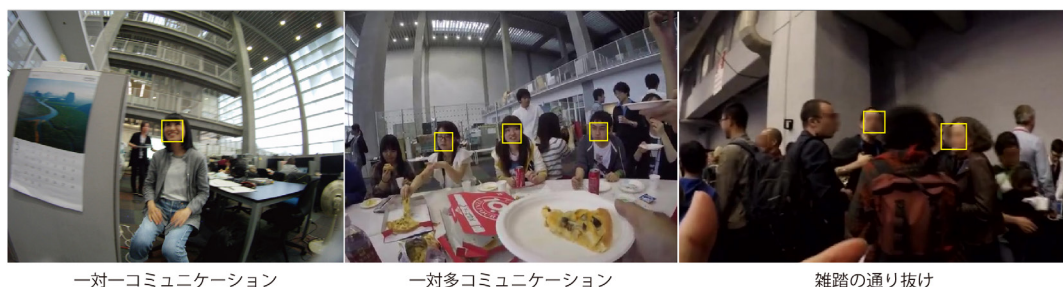


図 1 一人称ライフログ映像に対面者の顔が映り込むシーンの例



図 2 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

た、つまり発話やジェスチャそのものを計測せずに、カメラ装着者が主体的な行動をした際に向く相手の顔を検出することで社会活動量を計測できることが示唆された。

- 顔の個数のみからの計測だと主観評価と異なる状況があることを確認した。主体性を考慮して社会活動量を定量化するには距離の近さと時間継続性による重み付けが必要である。
- 斜めや隣り合わせでの対話が多く、180 度以上の画角が必要なことがわかった。200 度の半球カメラを使った結果、対面時の参加度が高い状況の計測を改善できることが示唆された。

2. 関連研究

2.1 非視覚情報から社会的状況を認識する技術

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、会話を検出する技術 [4] 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、眼鏡にフォトリ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 [5] 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。さらには、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 [6] も行われている。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。

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顔を手掛かりにして、カメラ装着者の日々の社会的な活動への参加度を計算する。

2.2 一人称視点映像から社会的状況を認識する技術

一人称視点映像から社会的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 [7]。一方で、カメラ装着者自身の行動の計測から社会的状況を認識する研究も多くされている [8]。例えば、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 [9]。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 [10]。一人称視点映像を使ってカメラ装着者の周囲および自分自身の状況を認識し、様々な種類の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、カメラ装着者の周囲に存在する人の正面を向く顔を手掛かりに、カメラ装着者自身の人との対面時の参加度を計測し社会活動量を推定することに取り組む。

3. 顔検出に基づいた社会活動計測

本研究では、対面時の参加度を時間方向に積分した値を社会活動量と定義して計ることに取り組む。顔の個数を数え上げるだけだと雑踏での他者との遭遇や特定の人物との密な対話を同一に扱ってしまうが、距離の近さと時間継続性で重み付けをして数え分けることができる。社会活動量は、フレームごとに人数、距離の近さ、時間継続性をもとに算出された値の時間積分とする。

図 3 は、顔の個数を数え上げる方法と、顔の大きさと時間継続性を考慮する提案手法で対面時の参加度を定量化し、社会活動量を計算した結果の例である。例えば、図 3

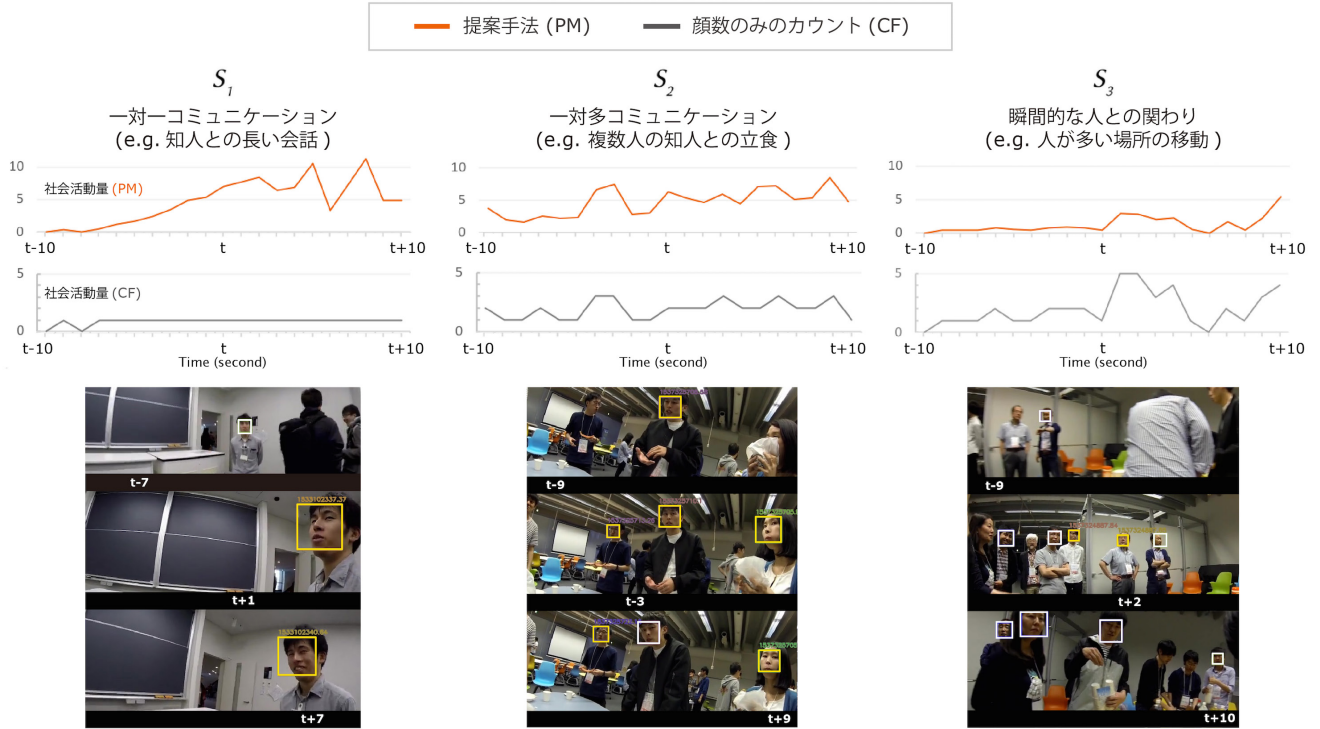


図 3 社会活動計測の結果の例：一対一対話，一対多対話，瞬間的な関わり

の S_1 ような一対一で特定の人物と近い距離で会話をしている状況， S_2 のような多人数で立食時に会話をしている状況， S_3 のような人混みの中で瞬間的に人と関わり合う状況を考慮して社会活動量を計算することができる．顔の個数を数え上げる方法では，約 20 秒間の社会活動量の累計は $S_1 < S_2 \approx S_3$ の順番になるが，提案手法では $S_3 < S_2 \approx S_1$ となる．図 3 の S_2 : $t+9$ フレーム目のように，会話相手がカメラ装着者の方を向き続けていない複数の状況が混在しているシーンを考慮する．これにより印象に近い社会活動量を求めることができるのではないかと考え計算式を作成した．

顔検出には，カーネギーメロン大学が開発した OpenFace[11] を用いた．中で使われている dlib ライブラリでは，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 [12]．そこで本研究では，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にした．真横を向く顔や後頭部は検出せずに，正面を向く顔のみを検出する．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している．

具体的には，ある時刻 t の社会活動量 S は式 (1)，(2) で計算する．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である．これらの内容の詳細については [2] の第 4 章で述べている．

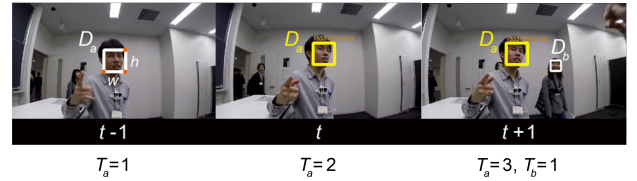


図 4 検出された顔ごとの大きさと時間継続性の計算

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

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

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

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

4. 主観評価実験

社会活動量が多いと感じる傾向がある状況を主観評価実験から調べ，詳細なセンシングをせずに一人称ライブログ映像からの顔検出に基づいて人との対面時の参与度を定量

化することの有効性と課題について議論する。

以下の2つの疑問について調べた。

- どのような状況に対して社会活動量が多いと感じる傾向があるのか？
- 個々人が身に着けたカメラに映り込んだ顔を数えることで、日々の社会活動量を計ることができるか？

我々は社会活動の量の多さという印象を定量的に評価するために、8名に対して1分間の一人称視点映像の視聴および並び替え作業を教示した。

使用した映像は、学会に参加した際に記録した一人称ライフログ映像から1分間の活動を一樣に抽出した映像である。最後には、複数の評価者によって定量化された社会活動量の多さへの印象と、提案手法および顔の個数を数える方法で定量化された社会活動量を比較した。

4.1 実験参加者

2017年3月に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参加者P8の映像を使用した。以下の3つの視点から社会活動量の多さへの印象を比較する。

- カメラ装着者である本人視点
- 互いに居合わせていた対話者視点
- 互いに居合わせていなかった第三者視点

評価者は、1名の本人、1名の対話者、6名の第三者視点の合計8名とした。6名の第三者視点の中には、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 主観評価実験の手順

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樣に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少ない順に「<、=」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、並び替え順を間違えないように0から100の数値のメモと判断基準について記述することも合わせて教示した。我々は、映像が社会活動量が少ない順に並び替えられた順序の距離を用いて社会活動量の多さへの印象である主観評価を定量化した。最後に、定量化された社会活動量の多さに関する印象と顔検出に基づいた評価の順序により整合性を比較した。

4.3 主観評価実験の結果

社会活動量が少ない順に並び替えられた10個の映像から得た結果を図5に示す。結果は定量化された印象の中央値順に並び替えた。同列のときはアルファベット順とした。また、並び替え結果から定量化された社会活動量への多さの印象(SE)と、提案手法(PM)および顔数のみの計

算(CF)から得られる社会活動量をプロットした。なお、順序を示すために尺度が異なる各々の値を最大値に合わせてプロットしている。社会活動量が少ないシーン(A, I)や多いシーン(E, F)では評価者間で印象が概ね合致していた(図5)。少し評価が分散している傾向があるシーン(G, J, C, H)、評価者間で印象が大きく分散しているシーン(B, D)もあった。

会話や発話をしているほど社会活動量への印象は高く評価されていた。会話および発話をしているシーンC, H, D, E, Fの実際の様子を確認してみると、会話相手の顔や体がカメラ装着者の方向を向いていた。

提案手法で定量化した社会活動量はシーンF, Jを主観評価と同じ順序になる値であったが、顔の個数を数える方法で定量化した社会活動量はシーンF, Jを主観評価と異なる順序になる値であった。

また、シーンH, D, Eでは両方の方法で定量化した社会活動量が印象に近くなかった。映像を確認するとカメラ装着者の会話相手が近づきすぎたり立ち位置が斜めになったときに顔が見切れていた。

5. 主観評価の考察

5.1 社会活動量が多いと評価されたシーン

カメラ装着者が他の人と会話をしているシーン(図5: C, H, D, E, F)を、社会活動量の多いシーンと評価する。特に、カメラ装着者自身が発話をしていると社会活動量の多いシーンと評価される。ここで面白いのは、我々の提案手法では「発話量」そのものは計測していないことである。それに関わらず、会話シーンにおいて顔を検出する提案手法で高いスコアを提示することができる。それは、カメラ装着者が発話をする際には、周りの人がカメラ装着者の方に対面する傾向が高まり、結果的に顔が検出されるからである。つまり、カメラ装着者が会話に参加して発話していると結果的に周囲の顔がカメラ装着者の方向を向く状況を利用することができる。したがって、発話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と考えられる。

5.2 社会活動量の多さにつながる対面シーン

対面している人の人数はあまり社会活動量の評価に依存せず、同一人物との対面時間の継続性と近さの方が社会活動量の多さの評価につながる。単純な顔の個数のカウントだけでなく、対面時間の継続性を考慮したスコアリングがうまく機能している(図5: F)。顔の個数のカウントだけでは、遠くから話を聞いているシーンで主観評価の結果よりも高くスコアリングしてしまう(図5: J)。したがって、直に人と関わり合う社会活動量を計測する際には、距離の近さと時間継続性を考慮することが重要である。

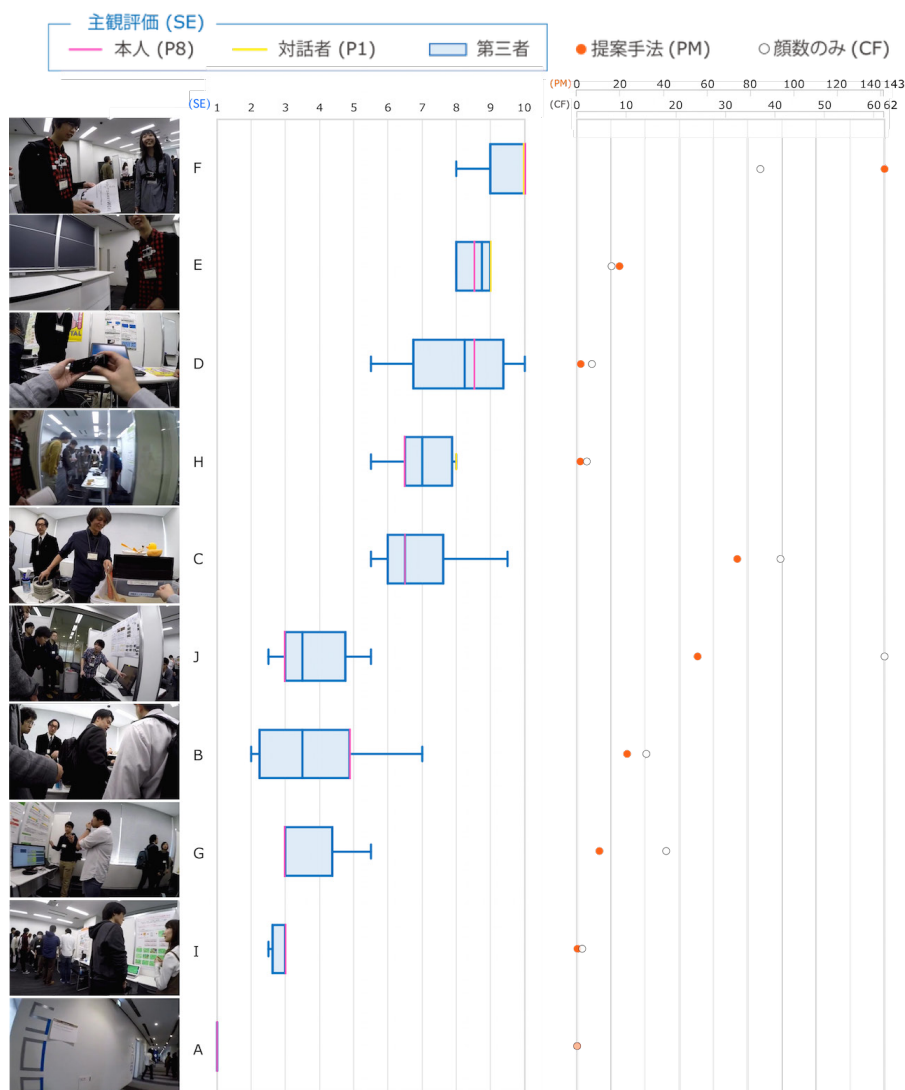


図 5 主観評価実験の結果：社会活動量の多さへの印象に関する主観評価 (SE) と提案手法 (PM) および顔数のみの計算 (CF) から得られる社会活動量の順序による整合性の比較。

5.3 評価者間の印象の一致

主観評価の結果は、評価する人がカメラ装着者本人か、それとも対話者と第三者かによってあまり影響を受けない。本人と対話者にとっては、出来事の体験直後ではなく約1年半前に自身が体験した様子からの社会活動量の多さを判断することになり、会話の中身や感じたことといったエピソード的な事実について触れることがなかったと考えられる [13]。しかし、第三者にとっては初めての出来事に対する印象の判断にも関わらず、本人と対話者と同様に、会話への参加や発話の有無から判断する傾向があった。つまり社会活動量の多さという印象は、エピソードにはあまり影響を受けないことが考えられる。

5.4 シーンの種類による印象の一致と不一致

シーンの種類によって、主観評価の分散が異なる。例えば、特定の人との長話シーン (図 5: F, E) は高いスコア、廊下を移動するシーン A は低いスコアであり、誰もが共通

の判定をする。一方で、大勢のグループに bystander 的な立ち位置で参加した後に発表者の目の前で展示を体験しているシーン B、特定の人と会話しながら展示を体験しているシーン D は評価の分散が大きい。これらのようなシーンの印象の個人差の配慮については提案手法の限界である。

5.5 提案手法の改良につながるカメラの画角

会話相手がカメラ視界から外れるシーン (図 5: H, D, E) では提案手法のスコアが、主観評価に比べてかなり低い値になった。近距離での斜めや隣り合わせ、上下差のある対話で相手の顔が見切れてしまったことが原因である。

カメラの画角を広げれば改善すると考え、200 度の半球カメラを使い計測した (図 6)。立ち位置が斜めの一対一対話での相手の顔を検出し計測することができることを確認した。また、食事や作業中の対話や座って上下差のある対話中の斜めや隣り合わせの相手の顔を検出することができた (図 7)。以上から、半球カメラで立ち位置が正面では

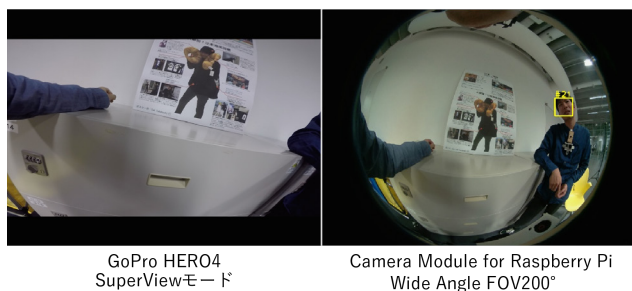


図 6 半球カメラを使った社会活動計測の検討



図 7 斜めや隣り合わせでの対話の計測の例

ない対面時の参加度も計測することができると考える。ただ、魚眼の歪みがあるため頑健な顔検出器を使用する必要がある。また、顔が映り込む場所によって顔の大きさが変わるため、歪みを除去するか、あるいは立ち位置によって顔の大きさの重み付けを再考することを考えている。

6. おわりに

人との対面時の参加度を簡素な方法で定量化することを目的として、一人称ライフログ映像中の顔を検出してカメラ装着者の日々の社会活動を計測する手法を提案した。社会活動量に関する主観評価実験の結果から、発話やジェスチャを計測せずに、カメラ装着者が主体的な行動をした際に向く相手の顔を検出することで社会活動量を計測できることが示唆された。加えて、直に人と関わり合う社会活動量を主体性を考慮して計測するには、距離の近さと時間継続性を考慮することが重要であることがわかった。

一方、斜めや隣り合わせでの対話が多く、180度以上の画角が必要ことがわかった。200度の半球カメラを使った結果、対面時の参加度が高い状況の計測を改善できることが示唆された。しかし、後ろ向きの人との関わり合いの計測、社会活動量の感じ方の個人差があるシーンを考慮した計測については提案手法の限界として残る。

今後の展望として、日々の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への行動変容の支援につながるフィードバックを目指している。我々は、複雑な社会的関係とは別の視点として、人と対面する社会活動の量は若者のひきこもりや高齢者のうつ病の傾向がある当人、家族の人、周囲の人を支援する手掛かりの1つになると考えている。今回はその第一歩だと考える。

参考文献

- [1] Guo, F., Li, Y., Kankanhalli, M. S. and Brown, M. S.: An Evaluation of Wearable Activity Monitoring Devices, *Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia*, PDM '13, New York, NY, USA, ACM, pp. 31–34 (online), DOI: 10.1145/2509352.2512882 (2013).
- [2] 奥野 茜, 角 康之: 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, *情報処理学会インタラクシオン 2018* (2018).
- [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] 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).
- [5] 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).
- [6] 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).
- [7] 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).
- [8] Tadesse, G. A. and Cavallaro, A.: Visual Features for Ego-centric Activity Recognition: A Survey, *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications, WearSys '18*, New York, NY, USA, ACM, pp. 48–53 (online), DOI: 10.1145/3211960.3211978 (2018).
- [9] 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).
- [10] 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).
- [11] Amos, B., Ludwiczuk, B. and Satyanarayanan, M.: OpenFace: A general-purpose face recognition library with mobile applications, *CMU School of Computer Science* (2016).
- [12] 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).
- [13] Sumi, Y., Suwa, M. and Hanaue, K.: Effects of Viewing Multiple Viewpoint Videos on Metacognition of Collaborative Experiences, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, New York, NY, USA, ACM, pp. 648:1–648:13 (online), DOI: 10.1145/3173574.3174222 (2018).