

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

奥野 茜^{1,a)} 角 康之^{1,b)}

概要: 本研究では、無意識的に行われている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を用いて、映像中に映り込んだ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。社会活動の状況を、検出された顔の個数と連続性で決定し、社会活動の量は顔の個数、顔の大きさ、時間連続性で計算する。これにより、カメラ装着者が単に人とすれ違っただけなのか、それとも特定の人と一緒に過ごしていたのか、また、その人数が一人だったのか複数人だったのかを見分けることで、活動の種類を数え分けることができる。社会活動をライフログとして記録し視覚的フィードバックを行うことで、社会活動の充実感の向上と疲労感の低下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。本稿では、学会に参加した複数人の2日間の一人称ライフログ映像から社会活動を計測して提案手法の有効性を検討する。時間連続性を考慮した社会活動のラベル付けをすることで対面コミュニケーション場面を抽出できること、映り込んだ顔の数だけでなく対面距離と継続時間を考慮することで対面コミュニケーションの種類を見分けることが可能になること、複数人の社会活動を計測することで同一空間に参加していた参加者ごとに役割の相違が顕在化されることが示唆された。

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

AKANE OKUNO^{1,a)} YASUYUKI SUMI^{1,b)}

Abstract: In this paper, we propose 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 major aim is to enable users to change their behavior toward achieving social fulfillment and reduction of fatigue. Our focus is on the improvement of social health by providing visual feedback of measured social activities. We experimented by tracking and measuring multiple people for two days. Afterwards, we discussed the effectiveness of the proposed method with them. For the effectiveness measure, we 1) labeled social activity per the face-to-face communication continuity; 2) calculated the amount of social activity based on both the number of people and the distance and continuity, considering the quality of one-to-one communication; and 3) measured the social activities of multiple camera-wearers to ascertain the differences and similarities of the users' time, space, and role.

1. はじめに

本研究では、無意識的に行われている社会活動の定量化と可視化を目的として、一人称視点映像をライフログとし

て使用し、映像中に存在する顔からカメラ装着者の社会活動を計測する手法を提案する。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。万歩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「万顔計」を実現することを目指す。

万歩計は従来、単純に歩数を数えるだけのものであった。

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

a) a-okuno@sumilab.org

b) sumi@acm.org



図 1 一人称ライフログ映像に映り込んだシーンの例

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。そして万歩計が身体活動の計測器として発展したように、顔検出の時間パターン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求する。

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画像処理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 1 のように様々な場面がとらえられる。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、人と会ったり人と話すときである。その他には、テレビを見るとき、雑誌を読むとき、コンピュータ上で SNS を利用するとき顔と対面する。本研究では、様々な社会活動の中でも特に、顔と対面する場面に着目する。顔との対面具合を社会活動への参加度と解釈し、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。顔検出から得られる顔の個数、顔の大きさ、顔の連続性をもとに人数、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。長期的な目標としては、長期的に一人称ライフログ映

像 (図 1) から計測される社会活動の視覚的フィードバックを行い、社会活動の充実感の向上と疲労感の低下のような社会的健康 [13], [19] につながる行動変容を目指している。

本稿では、学会に参加した複数人の 2 日間における一人称ライフログ映像から社会活動を利用して、提案する手法による社会活動計測の有効性を調べる。時間継続性を考慮して社会活動のラベル付けをすることで対面コミュニケーション場面を抽出できるのか、社会活動量を人数だけでなく距離と持続性から計算することで一対一の対面コミュニケーションの質を考慮することができるのか、複数人の社会活動を計測することで時空間と役割における相違点を知ることができるのかどうかについて焦点を当てる。

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 [24]、スピーカから音声 [25]、Bluetooth から人への接近 [9]、IR センサから対面対話の認識 [6] 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 [26]。一方で、社会活動の重要な側面を単純ながらも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 [18] 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 [23] 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な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 [21] では、単純な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] やモバイル端末で社会的インタラクションの認識をし、身体活動量のモチベーションにつなげる研究 [1] 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 [27] 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 [4]。

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上手に認識する必要があることが分かる。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たり、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 [14], [17]。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オブジェクトを手掛かりにカメラ装着者の社会的な状況を認識することを考える。本研究の焦点は、無意識的に行われている社会活動の定量化と可視化をするために、一人称視点映像でカメラ装着者の周囲に存在する顔を検出して、人との関わり合いの質と量を計測しライフログとして記録することである。

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、ディープラーニング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 [5], [30]。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 [10]。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 [31]。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 [2]。さらには、一人称視点映像を用いて、人の健康の支援 [12], [20], [29] や知覚の拡張 [15], [16] 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いることが分かる。本研究では、人との関わり合いの質と量を、検出された顔の個数から人数、顔の大きさから距離の近さ、顔の連続性から持続性を計算して計測する。これは、長期的にカメラ装着者の社会活動をライフログとして記録し、社会的健康につながるフィードバックをすることを目指している。

3. 一人称ライフログ映像

図 1 のように、一人称ライフログ映像にはカメラ装着者の周囲に存在する人の顔が写り込む。今回は研究交流会においての社会活動計測を行うが、本来は社会活動が行われるフィールドは限定しないことを想定している。図 2 のように、カメラを装着して一人称ライフログ映像を取得する。装着位置は、ライフログとして記録する際には、胸の部分が安定する場所として適していると考えた。



図 2 胸にカメラを装着して一人称ライフログ映像を取得

4. 顔検出に基づいた社会活動計測

日常生活における顔との対面具合を社会活動への参与度と解釈する。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、量の算出と状況のラベル付けを行う。顔検出から得られる顔の個数、顔の大きさ、顔の連続性をもとに人数、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。そのため、同一人物であるかどうかといった顔の識別や表情、口の開閉の識別は行っていない。顔検出には、カーネギーメロン大学が開発した OpenFace[3] 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 [8]。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) と時間の継続性 (顔の連続性) を考慮する。

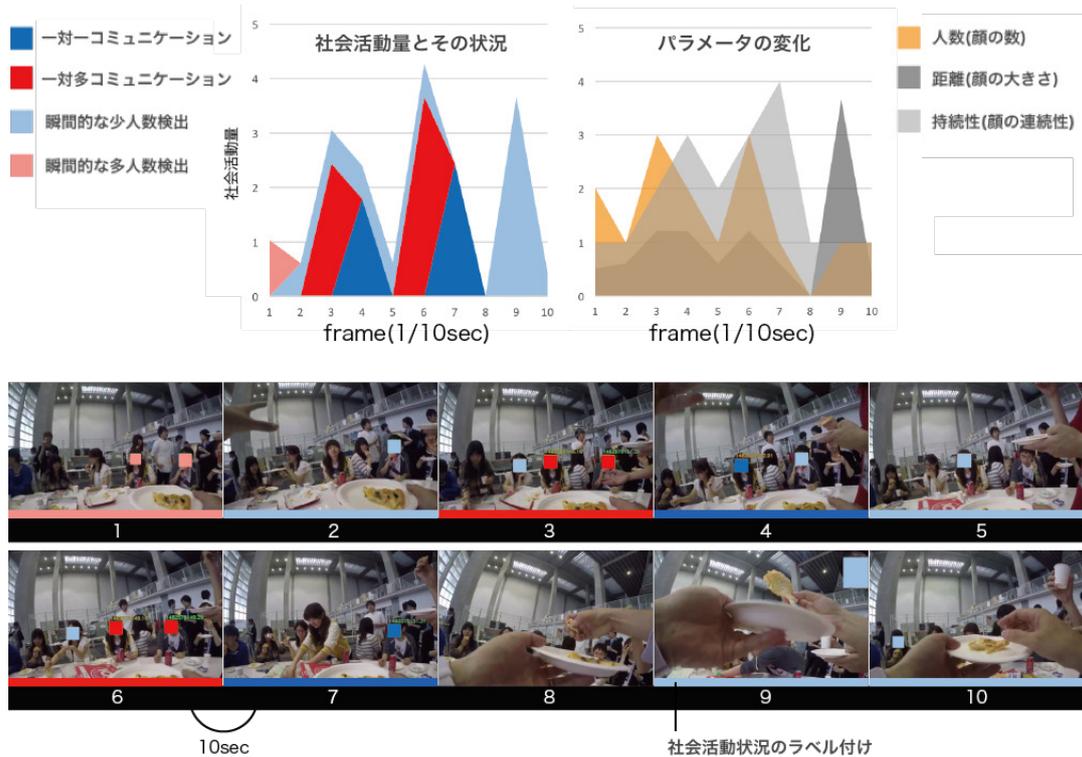


図 3 顔検出に基づいた社会活動計測：量の算出と状況の判別

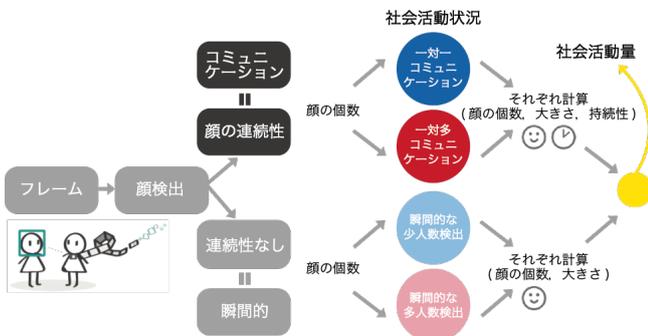


図 4 顔検出に基づいた社会活動計測の流れ

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3の3、4、6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。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4.1節、社会活動状況の判別の詳細については4.2節で述べる。

4.1 社会活動量の計算

本研究では、社会活動量を、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計算する。その際、映り込んだ顔ごとに、そ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 $S(t)$ は次の式で計算する(図5参照)。

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

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



図 5 検出された顔ごとの大きさと持続性の計算

つまり、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。そしてそれを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 w 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、今回利用している OpenFace が、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、時間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 からスタートする。

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を表す。具体的には次の式で計算する。

$$D_i = \frac{\text{検出顔 } i \text{ の幅} \cdot \text{検出顔 } i \text{ の高さ}}{\text{画面解像度}} \cdot 100 \quad (2)$$

4.2 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。

- (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 6 のような 4 種類の社会活動状況が判別される。社会活動状況を決定する際に、人との距離を考慮していないのは、距離が遠くても人と一緒に過ごしている場面は社会活動への参加度が高く対面コミュニケーションを行っている状況に近いと考えたからである。

5. 実験

5.1 目的

提案手法の動作確認と、計測された社会活動量の印象評価を行うために、学会参加の際に記録されたライフログ映像に提案手法を適用してみた。具体的には、インタラクション 2017 に参加した筆者らの研究グループの学生 8 名が、インタラクティブ発表セッションにおいて収録した一人称ライフログ映像を利用した。学会のインタラクティブ発表セッションでは、多くの人とすれ違いつつ、いくつかの発表ブースでは特定の相手と緊密な対面コミュニケーションが発生すると考えられ、様々な種類の社会活動が計測されると期待される。人数だけで社会活動を計測すると常に人に囲まれる発表者のほうが聴衆者より社会活動は濃密で数値が高い結果が得られると考えるが、人数、距離の近さ、持続性を考慮することで一対一コミュニケーションを考慮したような結果が得られるのかを、複数人の発表者と聴衆者の実際の状況を比較して調べる。本研究の提案手法が、社会活動の定量化と可視化をすることに有効であるのかについて議論する。

5.2 カメラ装着者

研究交流会に参加した複数人の 2 日間における一人称ライフログ映像から社会活動を計測した。カメラ装着者は、発表者あるいは聴衆者として参加した学生 8 名 (A, B, C, D, E, F, G, H) とした。1 日目の発表者は D, 2 日目の発表者は A, H であり、それ以外が聴衆者であった。年齢は 21 から 23 歳で、男性 7 名、女性 1 名である。各日、約 2 時間であった。

5.3 1 名の社会活動計測の結果

計測は、10 秒おきに行なった。図 7 は 1 日目に発表者として参加していた D から得られた結果である。ラベル付けされている箇所と実際の場面を見比べてみると、一対一対面コミュニケーションをしている場面と、一対多で少し距離をおいて対面コミュニケーションをしている場面が区別されていた。また、通りすがりの 1 名が一瞬検出されていた。また、社会活動量と顔の個数のグラフを見比べてみると、一対一コミュニケーションの質が抽出されていたのが分かった。距離の近さと持続性の観点から結果を見てみると、距離は常に幅広く変動していたが、持続性は一定の場面のみで変動していた。これらのことから、発表者 D に関しては、持続性を考慮して社会活動のラベル付けをすることで対面コミュニケーション場面を抽出できていると考える。また、社会活動量は、距離の近さと持続性の両方を互いに補うように機能しており、提案手法が意図しているような結果が得られていたと考える。



図 6 社会活動状況の判別

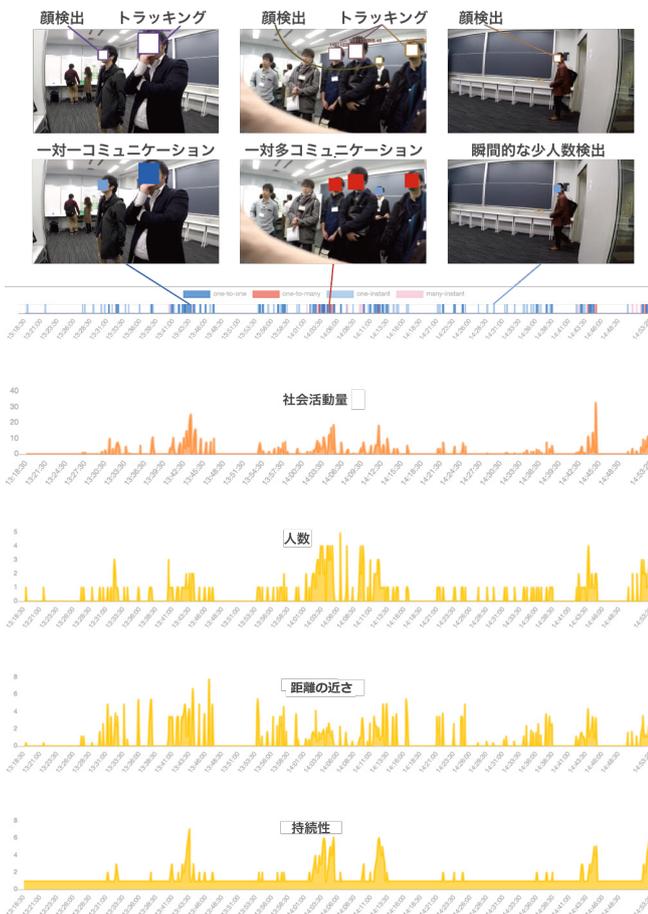


図 7 発表者 D の社会活動計測の結果 (学会参加 1 日目)

次に、複数の同じラベル付けがされている箇所と実際の場面を見比べて、同じような社会活動状況が抽出できているかを調べた(図 8)。一対多コミュニケーションを一対一コミュニケーションと判別している場面があったが、大体のフレームの前後では正しく判別できていたため、概ね一致していたと考える。瞬間的な検出に関しては、持続した人との関わり合いが始まる前の場面を少数人、多人数とも判別することができていた。瞬間的な少数人数検出では、人とすれ違う場面についても判別できていたと考える。瞬間的な多人数検出では、コンピュータを見た直後の場面が判別されており、顔を持たないモノを介した間接的な人との関わり合いの直後も瞬間的な多人数検出と判別される可能性があると考え。また、顔検出がされない独りの場面では、人がいない状況、コンピュータを見ている状況、顔

が見切れている状況が判別されていた。カメラの画角が広すぎないことで、聴衆者から少し視線を逸らしているような場面とも捉えられ、それぞれ同じ社会活動状況を判別することができていると考える。

今回は、社会活動計測をした結果と実際の状況を見比べて提案手法の妥当性を確認した。今後は、社会活動計測の利用者であるカメラ装着者本人に、得られた社会活動量と社会活動状況のラベル付けが実際の印象に近いかどうかを点数で定量的、書き起こしで定性的に評価してもらうことで妥当性を調べることを考える。

5.4 複数人の社会活動計測の結果

図 9 が発表者として参加していた 8 名から得られた社会活動計測の結果である。1 日目の平均値は 321, 2 日目の平均値は 261, 両日の平均値は 291 となった。1 日目は発表者 D より聴衆者 H のほうが数値が高い結果が得られた。また、2 日目は発表者 A が最も高く、発表者 H は平均値より低かった。他の聴衆者として参加した人々は、合計値にあまり差は見られないが、ラベル付けされる時間帯や社会活動状況の比率と数値の差に特徴が見られた。また、2 日目の A と H は同じ発表者として同じ空間にいたが、結果が異なっていた。これらことから、複数人の社会活動を計測することで、時空間と役割において似ている部分と異なっている部分を知ることができると考える。

社会活動量の数値が高くても、対面コミュニケーションの相手が必ずしも多いとは限らない結果となっていた。図 9 の 1 日目の聴衆者 H と 2 日目の発表者 A のラベルを見比べてみると、対面する頻度が少なくても濃密なコミュニケーションをしている可能性がある場合と、対面する頻度が多くても濃密なコミュニケーションをしている可能性がある場合が考えられる。提案手法が意図していることは、1 日目の発表者 D がした聴衆者に対する濃密な対面コミュニケーションと同じくらいかそれ以上、聴衆者 H が一対一の対面コミュニケーションをしていたことが数値として現れることである。そこで、次に、参加者 A と H の実際の場面と比べながら、距離の近さと持続性のパラメータの働きが上手くいっているのかを調べる。

図 10 は参加者 A と H の実際の社会活動状況である。1 日目は、2 名とも聴衆者として参加したが、得られた結果

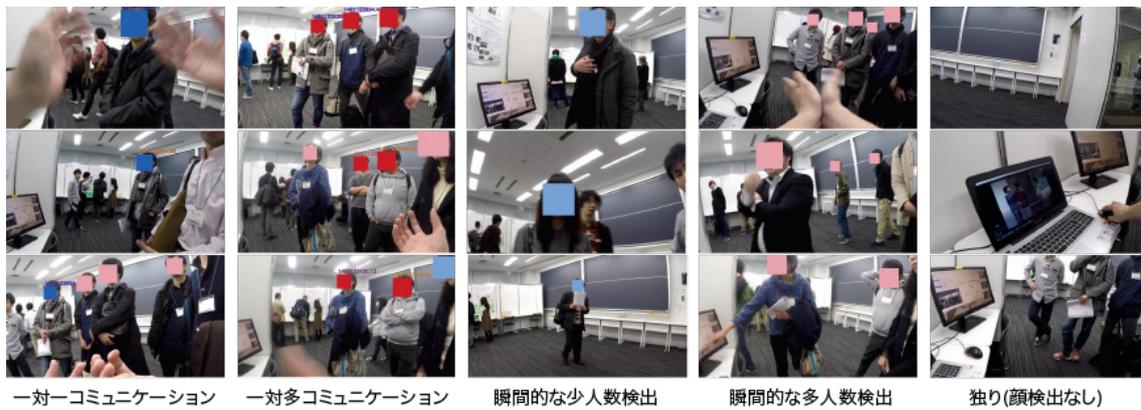


図 8 発表者 D の社会活動の状況判別 (学会参加 1 日目)

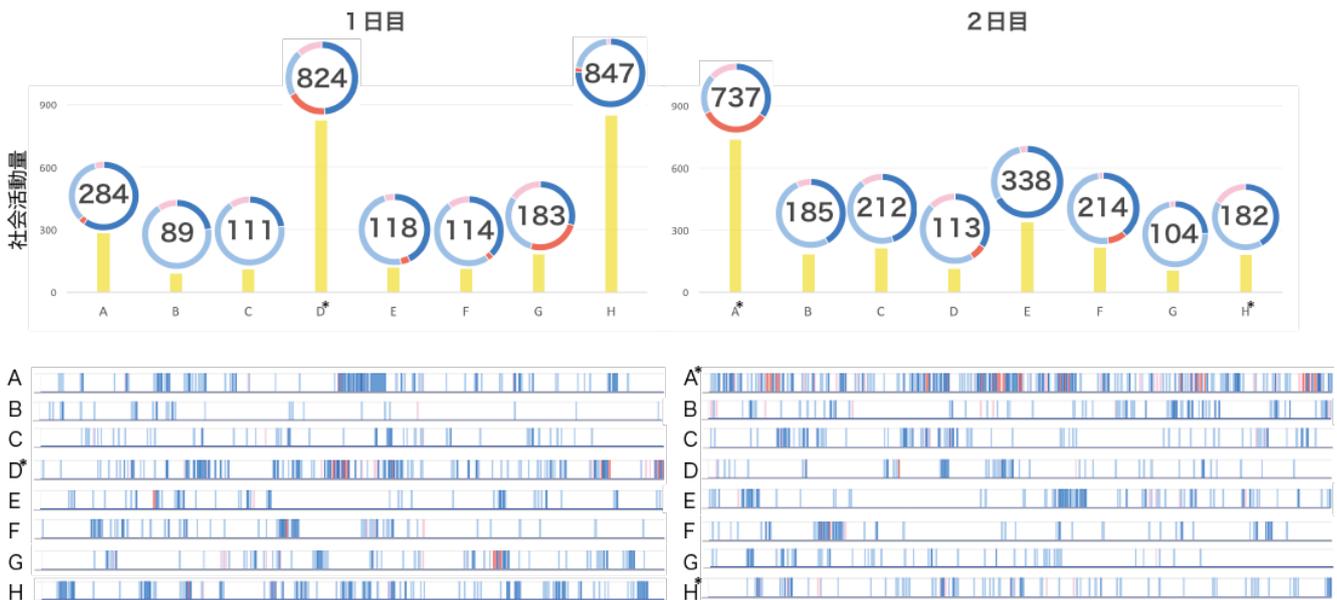


図 9 社会活動量と状況判別結果のラベル付け：1 日目発表者 D*, 2 日目発表者 A*, H*

は大きく異なっていた。2 日目は、2 名とも同じ場所で隣り合わせで発表者として参加したが、得られた社会活動状況は異なっていた。図 11 は持続性と距離の近さのパラメータの値の頻度を示している。一見、聴衆者 A は一対一で対面コミュニケーションを行っている時間が多く感じられるが、社会活動量は高くない。図 10 から対話相手との距離が遠かったことが分かった。聴衆者 H は一対一で対面コミュニケーションを行うときの距離が近かった。図 11 を見ると、持続性は聴衆者 A と H で違いはないが、距離の値は H のほうが接近性のある頻度が高かった。2 日目は、発表者 A は一対多でコミュニケーションを行っている時間があつたが、H にはなかつた。図 10 から、発表者 A は常に人通りがある状況で対面コミュニケーションをしていたことが分かる。発表者 H は、主に HMD の装着や PC の作業をする発表者としての役割をしており、対面コミュニケーションをしていない時間が多かつた。その一方で、発表者 A とは異なり個別で聴衆者に対応している場面があ

り、距離が近い一対一の対面コミュニケーションをしていた。図 11 を見ると、持続性は発表者 A のほうが高く、H は低かつた。距離の値は、どちらも接近性が低い。A に関しては頻度が多く、H に関しては母数に対する大きさが目立っていた。この結果は、本人たちの印象と合っていた。

これらのことから、社会活動量を人数だけでなく距離の近さと持続性から計算することで、一対一の対面コミュニケーションの質を考慮することができると考える。一方で、常に人通りがある場所で社会活動をしていると、常に顔を浴びることになり、対面コミュニケーションと無関係に濃密な社会活動と解釈される。このような一対多コミュニケーションの社会活動状況と、社会活動量の値が高い状況は必ずしも質が高いという訳ではなく、社会活動への参加度が高い状況であつたという解釈が提案手法の限界であるとする。改善策として、顔の向きを判定を加えることで、一対多の対面コミュニケーションの質を考慮することができると考える。



図 10 参加者 A, H の一人称ライフログ映像のシーン例 (1 日目: 見学者, 2 日目: 発表者)

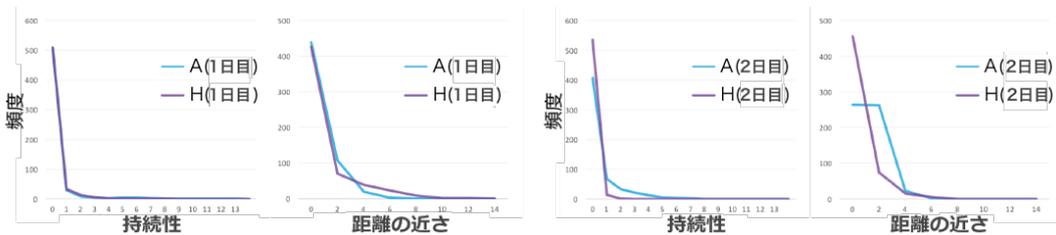


図 11 持続性と対面距離の値の出現頻度

5.5 同一シーンへの参与度の違いの顕在化

同一の時空間を共有している場合でも、各自のその場での役割や積極性の差から、シーンへの参与度は異なる。今回提案した手法は、そういった各自の参与度の違いを定量的に表現することができるのではないかと考える。



図 12 学会シーンで見られた立ち位置の非対称性



図 13 会食シーンで見られた立ち位置の非対称性

例えば、図 12 は、学会 1 日目の聴衆者 A と発表者 D が対面している状況である。D は発表者なので、多くの聴衆者に向かい話しかけており、つまり、一対多で対面コミュニケーションをしている。一方で、聴衆者 A は一瞬だけの対面コミュニケーションと判定されていた。この例は、学会での発表者と聴衆者という役割の違いを顕在化した例と言える。

図 13 は日常シーンの例である。具体的には、同じ食事会に参加していた 2 人の同一シーンの一人称画像を表す。座敷の中心に座っていた参加者のライフログ映像 (左) からは多くの同席者との継続的な対面コミュニケーション

が検出されているが、座敷の角に座った参加者のライフログ映像 (右) からは、個別の同席者との断続的なコミュニケーションが多く検出された。このことは、会食における座った位置によって、無意識のうちに対面する顔の量に大きな差があり、このことが無意識のうちに会話インタラクションの種類に影響を与え、ひいては社会活動量に差をもたらすことを顕在化する可能性を示しているのではないだろうか。

これまでに複数人の一人称視点映像から社会的な状況を認識する研究はされている [2], [10], [31]。複数人で人との関わり合いの質と量の計測を行うことで、このようなコミュニケーションへの参与の非対称性を検出できる可能性がある。カメラ装着者の内的な状況を計測しようとする、多対多、一対多の対面コミュニケーションでは状況の感じ方は異なるはずである。例えば、充実感や疲労感が関係していると考えられる。しかし、これには様々な要因が複雑に絡んでいて難しい。そのため、身体活動量計と合わせて社会活動を計測し、人との関わり合いがどのように変化するかの変容の観察を通して、これらの状況を認識することをしていきたいと考える。

6. おわりに

日々無意識のうちに行っている社会活動量の可視化を目的として、一人称ライフログ映像を用いて、映像中に存在する顔からカメラ装着者の社会活動を計測する手法を提案した。日常での運用テストや学会参加者の複数人の一人称ライフログ映像に提案手法を適用した結果、顔検出から得られる人数、距離の近さ、持続性に基づいて社会活動を定量化することで、実際の印象に近い結果を得ることができた。提案手法は個人の顔を特定せず、顔検出結果だけを利

用しているため、設計の簡素化やプライバシーの観点から実用化しやすく、一人称ライフログ映像から社会活動を記録・解釈することに適していると考えられる。

今後は、社会活動量の可視化による行動変容について調べたい。つまり、長期的に蓄えられた一人称ライフログ映像から社会活動の視覚的フィードバックを行い、社会活動の充実感の向上と疲労感の低下のような社会的健康につながる行動変容支援を目指す。今回はその第一歩だと考える。社会活動と睡眠との関係 [22] などの調査にも役立てると考える。また、本手法は簡易的に複数人の社会的インタラクションを計測する手法として優れていると考えるので、社会的インタラクション研修のためのマルチモーダルコーパスの作成 [28] にも活用したい。

参考文献

- [1] Aharony, N., Pan, W., Ip, C., Khayal, I. and Pentland, A.: The Social fMRI: Measuring, Understanding, and Designing Social Mechanisms in the Real World, *Proceedings of the 13th International Conference on Ubiquitous Computing*, pp.445–454, ACM (2011).
- [2] 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).
- [3] 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).
- [4] Centola, D.: The spread of behavior in an online social network experiment, *Science*, vol.329, pp.1194–1197. American Association for the Advancement of Science (2010).
- [5] 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*, pp.75–82 (2015).
- [6] Choudhury, T. and Pentland, A.: Sensing and modeling human networks using the sociometer, *Proceedings of the 7th IEEE International Symposium on Wearable Computers*, pp.216–222 (2003).
- [7] Damian, I., Tan, C.S.S., Baur, T., Schöning, J., Luyten, K. and André, E.: Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp.565–574 (2015).
- [8] 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).
- [9] Eagle, N. and Pentland, A.: Eigenbehaviors: Identifying structure in routine, *Behavioral Ecology and Sociobiology*, vol.63, pp.1057–1066, Springer (2009).
- [10] 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).
- [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*, pp.6536–6546, ACM (2017).
- [12] 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).
- [13] 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).
- [14] Hoyle, R., Templeman, R., Armes, S., Anthony, D., Crandall, D. and Kapadia, A.: Sensitive Lifelogs: A Privacy Analysis of Photos from Wearable Cameras, *Proceedings of the 33rd CHI Conference on Human Factors in Computing Systems*, pp.1645–1648, ACM (2015).
- [15] 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).
- [16] Kasahara, S. and Rekimoto, J.: JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera for Remote Collaboration, *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*, pp.217–225, ACM (2015).
- [17] Korayem, M., Templeman, R., Chen, D., Crandall, D. and Kapadia, A.: Enhancing Lifelogging Privacy by Detecting Screens, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp.4309–4314, ACM (2016).
- [18] Lee, Y., Min, C., Hwang, C., Lee, J., Hwang, I., Ju, Y., Yoo, C., Moon, M., Lee, U. and Song, J.: SocioPhone: 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*, pp.375–388, ACM (2013).
- [19] Li, T.M.H. and Wong, P.W.C.: Youth social withdrawal behavior (hikikomori): A systematic review of qualitative and quantitative studies, *Australian & New Zealand Journal of Psychiatry*, vol.49, pp.595–609, SAGE Publications (2015).
- [20] 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).
- [21] 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*, pp.317–326, ACM (2016).
- [22] Moturu, S.T., Khayal, I., Aharony, N., Pan, W. and Pentland, A.: Using social sensing to understand the links between sleep, mood, and sociability, *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, pp.208–214 (2011).
- [23] Nakakura, T., Sumi, Y. and Nishida, T.: Neary: Conversational field detection based on situated sound similarity, *IEICE Transactions on Information & Systems*,

E94-D(6):1164–1172 (2011).

- [24] Olguin, D.O. and Pentland, A.: Human activity recognition: Accuracy across common locations for wearable sensors, *Proceedings of 2006 10th IEEE International Symposium on Wearable Computers*, Montreux, Switzerland, pp.11–14 (2006).
- [25] Olguin, D.O., Paradiso, J.A. and Pentland, A.: Wearable communicator badge: Designing a new platform for revealing organizational dynamics, *Proceedings of the 10th International Symposium on Wearable Computers*, Student Colloquium, pp.4–6 (2006).
- [26] 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).
- [27] Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M.M., Larsen, J. and Lehmann, S.: Measuring large-scale social networks with high resolution, *PloS one*, vol.9, pp.e95978. Public Library of Science (2014).
- [28] Sumi, Y., Yano, M. and Nishida, T.: Analysis environment of conversational structure with nonverbal multimodal data, *12th International Conference on Multimodal Interfaces and 7th Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI 2010)*, 4 pages, ACM (2010).
- [29] Williams, A.D.: Harnessing the Quantified Self Movement for Optimal Mental Health and Wellbeing, *Proceedings of the First Workshop on Lifelogging Tools and Applications*, pp.37–37, ACM (2016).
- [30] Yao, T., Mei, T. and Rui, Y.: Highlight detection with pairwise deep ranking for first-person video summarization, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.982–990 (2016).
- [31] 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).