

顔情報に着目した一人称画像ライフログによる社会活動計測

奥野 茜^{1,a)} 角 康之¹

概要：本研究では，一人称画像ライフログ中に含まれる顔情報を抽出して，カメラ装着者の社会活動量を計測する手法を提案する．長期的な一人称画像ライフログ中に含まれる顔情報に着目することで，カメラ装着者の複数人との関わりや，その時間，空間の継続性を測ることができ，そこから社会活動のパターンや社交性の推定ができると考える．研究交流会の際に記録されたデータを題材に予備検討を行い，一人称画像中に検出される顔の個数，大きさ，時間継続性から社会活動状況の分類（一対一会話，多人数会話，雑踏の中の移動等）や会話インタラクションへの参与レベルの変化を推定することの可能性を議論する．

Social Activity Measurement by the First-person-vision of Life-log Tagged with Face Detection

AKANE OKUNO^{1,a)} YASUYUKI SUMI¹

Abstract: In this research, we propose a method to extract the face information contained in the first-person-vision of life-log and measure the social activity amount of the camera wearer. By paying attention to the face information included in the long-term first-person-vision of life log, it is possible to measure the relationship with a plurality of people of the camera wearer, the time and the continuity of the space, and we think that we can estimate social activity patterns and sociability from there. We conducted a preliminary study using data recorded at the research exchange meeting as a subject, and discuss the possibility of estimating the change in participation level to conversation interactions and the scene classification (one-on-one conversation, multi-person conversation, crowd etc.), which from the number, size and time continuity of faces detected in the first-person-vision.

1. はじめに

私たちは，日々，様々な人と関わり合って活動をする社会的動物である．近年，メディアが高度に発達し，実世界での対面した会話に限らず，SNS，ゲームなどの仮想空間でも社会活動が行われており，複雑化している．その中，何人の人と実際に会って話したのか，それは一対一の会話だったのか，多人数での会話だったのかを把握することは，一日単位でも難しい．社会心理学の分野では，人間社会の中で起きている，人と人，あるいは集団や団体や社会と個人といった組み合わせの中で，そこにどんな法則性があるのか，どんな形で問題が生じるのか，あるいは合理性ということを考えてときに何か不都合なことは起きていな

いか，こうしたことを根気強く明らかにすることを目指している [1]．

社会活動の中で，顔は重要なコミュニケーションツールとして機能している．例えば，心の窓としての顔，コミュニケーションメディアとしての顔，存在としての顔，文化としての顔などがある [2]．そこで，我々は顔情報をもとに，当人の社会活動量を計測することを考え，社会活動量を測る歩数計を作る．

歩数計は，最初は単純に歩数を数えるだけであった．しかし，単純に歩数を数えることから，大体の活動量を知ることが可能になり，活動量を増やす動機付けも高めた．また，加速度センサからの行動判別の高度化 [3] やクラウド化によって，より詳細な行動記録や振り返りを可能にしつつある．我々は，その社会活動版をやりたいと考える．最初は，顔の個数を数えるという単純なことから始め，歩数計の発展を参考にして，以下のような研究ランドマークを

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

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

設定する。

一人称視点である画像のライフログ（以下，一人称画像ライフログ）から顔情報を抽出し，顔の個数から人数判定を行う．次に，顔の大きさをもとに，人との距離感判定を行う．顔の時間継続性から会話の持続性判定を行い，人との距離感判定と合わせて会話への参与レベル判定を行う．また，人数と会話の持続性判定をもとに当人の社会活動の状況を推定する．例えば，一対一会話なのか，雑踏の通り抜けなのかの判定を行う．さらに，長期的に社会活動量を計測することで，当人の内的状況を推定することができる．

2. 関連研究

2.1 ライフログ中の顔情報に着目した研究

捧らは，主観写真ライフログを効果的に活用するために，時間，空間，人間での画像検索を可能とした．顔認識結果から，当人から見た人間関係を抽出し，人間関係を反映した検索および閲覧を可能としている [4]．高松らは，日常生活の中でも人間同士の直接面会に着目し，ライフログレコーダによって取得される画像の中から人の顔画像を抽出し，個人毎に分類することによって過去の面会履歴の検索を可能としている [5]．一方，福本らは，顔情報の中でも微笑みと笑いに着目し，顔認識ではなく，フォトリフレクタを用いて日常的に笑顔を検知し，動画に記録された楽しい出来事や面白い出来事にタグ付けをしている [6]．本研究では，主観的ではなく，客観的である一人称画像ライフログに含まれる顔情報に着目する．そして，顔情報の中でも，顔の有無・顔の大きさ・顔の時間継続性に着目し，人数判定・人との距離感判定・会話の持続性判定・会話への参与レベル判定を行い，当人の社会活動量を計測する．

2.2 一人称画像，一人称映像中の情報に着目した研究

Ortis らは，一人称画像ライフログに含まれる家の中における日常生活の場所を，CNN をもとに得られる特徴から分類し，家の中における日常生活の場所と時間の使われ方を可視化している [7]．樋口らは，一人称映像を用いて，一対一の対面会話における話者間の社会的地位を推定している [8]．ソーシャルインタラクションのさらなる理解を目的とし，頭部の動き特徴のみから，マルチクラス識別器を用いて社会的地位を推定している．本研究では，一人称画像ライフログに含まれる顔情報を，顔認識と物体追跡の技術を用いて，何人の人と実際に会って話したのか，それは一対一の会話だったのか，複数数での会話だったのかについて，いつ，どのくらいの量だったのかを推定する．

3. 一人称画像ライフログ

本研究では，一人称画像ライフログに含まれる顔情報を抽出して，社会活動量を計測する．一人称画像ライフログ



図 1 一人称画像ライフログ

Fig. 1 First-person-vision of Life-log.



図 2 カメラの装着

Fig. 2 Wearing camera.

には，図 1 のような画像が映り込む．そこで，顔の有無・顔の大きさ・顔の時間継続性に着目し，人数判定・人との距離感判定・会話の持続性判定・会話への参与レベル判定を行い，当人の社会活動量を計測する．複数人の関わり合い方，時空間の使われ方に焦点を当て，当人の社会活動を可視化する．図 2 のように，カメラを胸部にクリップで付け，装着して撮影する．アクションカメラである GoPro を使用する．

4. 顔情報を用いた社会活動計測

4.1 システム概要

顔の有無・顔の大きさ・顔の時間継続性に着目し，人数判定・人との距離感判定・会話の持続性判定・会話への参与レベル判定を行い，当人の社会活動量を計測する．複数人の関わり合い方，時空間の使われ方に焦点を当て，当人の社会活動を可視化する．人物判定，モバイル環境下での利用を展望として考えているため，深層学習を用いた顔認識オープンソースである OpenFace[9] を用いた．OpenFace では，顔認識と物体追跡に，dlib ライブラリが用いられている．顔認識には，hog 特徴と SVM で既に学習されているモデルが用いられている．物体追跡には，ロバストな視覚追跡のための正確なスケール推定が用いられている [10]．

4.2 人数判定

顔認識をされた顔の個数を，人数とする．

4.3 人との距離感判定

顔認識をされた顔の大きさを画像の解像度で割り，百を

掛け百分率としたものを人との距離感とする．人数を n とする．

$$g(n) = \frac{\text{顔の大きさ}}{\text{解像度}} \times 100 \quad (1)$$

4.4 会話の持続性判定

顔認識と物体追跡をされた顔に ID を割り当てる．この場合，同一人物との対面関係が，ある程度継続しており，会話インタラクションが成立していると解釈する．

4.5 会話への参与レベル判定

会話の持続性判定をもとに，人との距離感を，会話への参与レベルとして換算する．会話の持続性の継続時間を j とする．会話の持続性がない場合は， j を 0 とし，会話への参与レベルは 1 とする．持続性がある場合は，フレームごとに j は 1 増加する．

$$h(j) = 1 + \frac{j}{10} \quad (j > 0) \quad (2)$$

4.6 社会活動量の計測

式 (1)，(2) をもとに社会活動量を計測する．式 (3) は，会話の持続性がある場合である． i をフレーム番号とし， a を一画像中の人数の合計とし， n を人数とする．例えば， i が 100， a が 3， n が 3 のとき，100 フレーム目における 3 人のうち 3 人目の計算を行い，人との距離感を測定し，会話への参与レベルの測定値を与える．式 (4) は，会話の持続性がない場合である．

$$f(i) = \sum_{n=1}^a (g(n)) \times h(j) \quad (3)$$

$$f(i) = \sum_{n=1}^b (g(n)) \quad (4)$$

4.7 社会活動状況の推定

人数判定と会話の持続性判定をもとに，以下の計 4 つの状況を定義した．

- (1) 一対一インタラクション：1 つだけ ID 付きの顔検出
- (2) 多人数インタラクション：2 つ以上 ID 付きの顔検出
- (3) 少人数検出：1 つだけ ID 無しの顔検出
- (4) 多人数検出：2 つ以上 ID 無しの顔検出

上記の (1)，(2) の状態を，会話状態であると解釈する．以下に示す図 3 の状態遷移図のように，社会活動状況の状態は遷移する．

5. ポスターセッション時の社会活動計測

5.1 概要

予備検討として，ポスターセッション時の社会活動計測を行った．比較的短い時間に会話の出入りが多く，様々な種類の社会活動が計測できると考えたからである．約 1 時

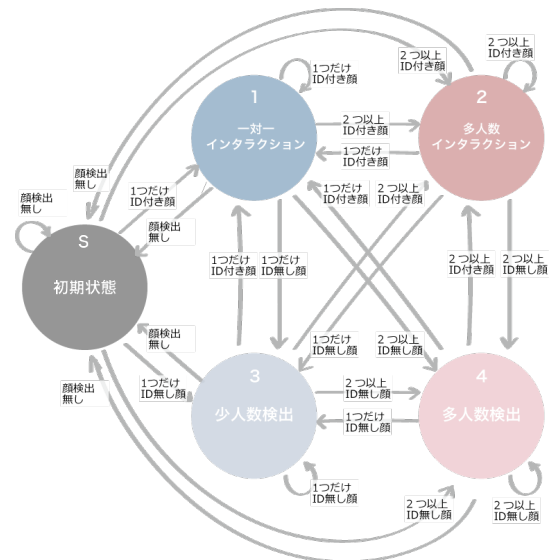


図 3 社会活動状況の状態遷移図

Fig. 3 State transition diagram of social activity situation.

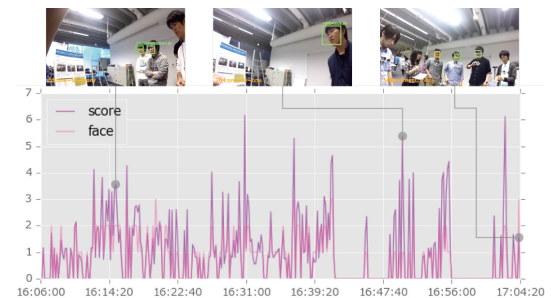


図 4 会話への参与レベル判定を反映した計測値と顔の個数

Fig. 4 Measured value reflecting participation level judgment in conversation and number of faces.

間のセッションが 3 回あった．以下に，発表をしていた人の一人称画像ライフログに含まれる顔情報をもとに，社会活動量を計測した方法と結果について述べる．

5.2 会話への参与レベル判定を反映した計測値

58 分 20 秒の動画から，10 秒おきの連番画像を 350 枚作成した．図 4 の score は，式 (3)，(4) を合わせた計測値である．face は一画像中における人数（顔認識をされた顔の個数）の合計である．

図 4 において，16 時 15 分頃，16 時 50 分頃，17 時 4 分頃の計測結果を取り上げた．単純に人数を数えた場合では，17 時 4 分頃が最も高い社会活動量となっているが，会話への参与レベルを与えた場合では，16 時 50 分頃が最も高い社会活動量となっていた．会話への参与レベルを与えると，人との距離感の小さい持続した一対一の会話のほうが，人との距離感が大きい持続した一対多の会話より社会活動量は高くなっていた．

会話への参与レベルを与えることで，単純に顔の個数を足す場合よりも現実の感覚に近い結果が得られることができると考えるが，人数，人との距離感，会話の持続性のそ

それぞれの比重については、今後、検討する必要があると考える。

5.3 人数と会話の持続性判定による社会活動状況の推定

10秒おきに計測し、人数と会話の持続性から、当人の社会活動状況を推定した(図5)。

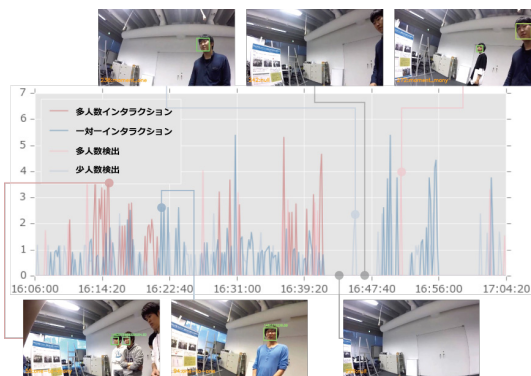


図5 人数と会話の持続性判定による社会活動状況の推定

Fig. 5 Estimate the situation of social activities by judging sustainability of conversation and number of people.

図5において6つの計測結果を取り上げた。16時14分頃では、多人数で会話していたことが読み取れ、16時21分頃では、一対一で会話をしてきたことが読み取れた。16時43分頃では、誰とも会話していないことが読み取れた。16時45分頃に一人と会っていたことが読み取れたが、16時46分頃では相手の顔が見切れており、状況は計測できていなかった。16時52分頃では、会話をしてきた人と近くに寄って来た人が写り込んでいた。

計測結果と実際の状況は概ね合致していたが、相手の顔を見ずに会話を行っている場合は、相手の顔が画面の端に見切れていることが多く、計測することができていなかったと考える。

5.4 計測する時間の粒度による差

10秒おきに計測した場合と、1秒おきに計測した場合の社会活動の状況推定結果を比較した(図6, 図7)。1秒おきに計測した場合のほうが、会話をしている状況の推定の割合が多かった。また、一対多の会話の推定の割合が増加した。

16時55分頃において、当人は2人と会話をしてきたが、10秒おきの計測では一対一の会話と判定しており、1秒おきの計測では多人数での会話と判定していた。10秒の間に、会話相手あるいは当人が移動すると、会話の持続性がないと判定されるため、1, 2分ほどの短い時間の一対多の会話を計測するためには、1秒おきに計測する必要があると考える。

撮影終了直前に、当人は歩いて人とすれ違っていたが、その場合には顔は追跡されていなかった。したがって、当

人の雑踏の通り抜けと会話を区別することができると思う。しかし、10秒おきの計測では、短時間の歩行中の画像は1フレームほどしかなく、雑踏の通り抜けと判定するほどの計測はできていなかったと考える。

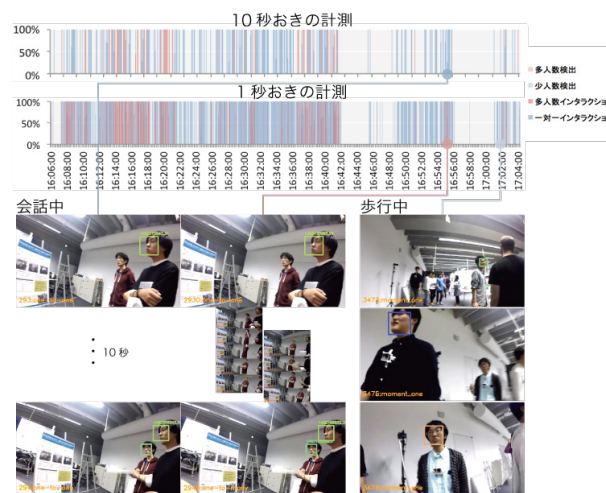


図6 10秒おき, 1秒おきにする計測の場合

Fig. 6 In the case of measuring every 1 second, and 10 seconds.

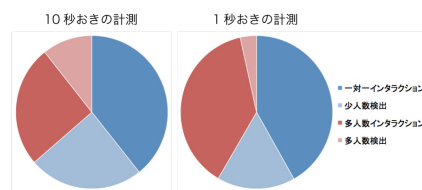


図7 社会活動状況の計測値の割合

Fig. 7 Percentage of measured values for each situation of social activities.

6. 約4時間の個人間における計測値の推移

上記のポスターセッション時の予備検討とは別に、ポスターセッションと懇親会に渡る約4時間の一人称画像ライフログから社会活動量を計測した。対象は、教員1名(Participant1), コンサルタント職2名(Participant2, 3), 学生1名(Participant4)の計4名とした。Session1はポスターセッション前半, Session2はポスターセッション後半, Session3は懇親会である。10秒おきに計測し、計測値は、4.6で述べた式で算出した。

各セッションごとに、社会活動の計測値の累計を出してグラフ化した(図8)。scoreが計測値であり、faceは人数(顔認識された個数)である。社会活動の計測値の推移は、個人間で異なることが分かった。

各セッションごとに、社会活動状況の割合を出してグラフ化した(図9)。社会活動状況の割合の推移は、個人間で異なることが分かった。ポスターセッション(Session1, 2)では、一対一の会話がが多く、懇親会(Session3)では、複

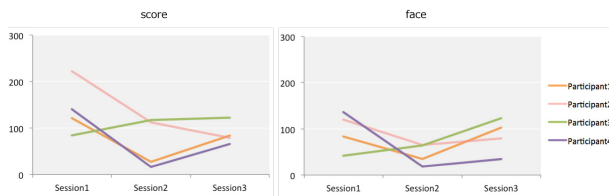


図 8 個人間の計測値と人数の推移

Fig. 8 Changes in cumulative measured values and number of people among individuals.

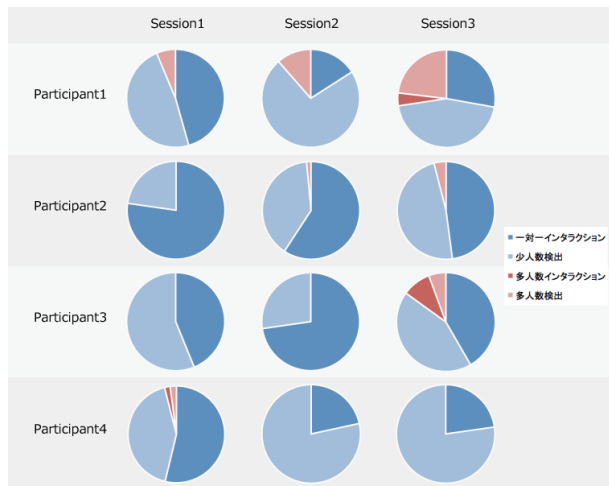


図 9 個人間の社会活動状況の割合の推移

Fig. 9 Changes in proportion of social activity situation among individuals.

数人と会う人、一対一の会話をする人など個人間で異なっていた。懇親会 (Session3) の同時刻における個人間の社会活動状況の違いについて、図 10 を示す。より長期間の渡る複数人の計測値の推移を比較することで、似た傾向の時間帯などが発見できると考える。今回の一人称画像ライフログの一部が、逆光の影響を受けており、そのような環境が与える影響についても検討する必要があると考える。

7. 課題

今回は、社会活動状況の推定からラベルデータを作成し、iCourpusStudio[11] を用いて、計測値と実際の様子の整合性を確認した。現状では、ラベル付けされたフレームを目視で確認しており、正解データがないため、今後は手作業でラベル付けを行うことを考える。また、iCourpusStudio では、マルチモーダルデータを扱うことができるため、音環境比較による会話場の検出 [12] から得られるラベルデータとともに参照することもできると考える。また、今回の予備検討において、会話をしていない人の小さな顔も ID 付けされていたり、会話をしている人の大きな顔が ID 付けされていなかったため、人との距離感を社会活動状況の推定に与えることを考える。

今後は、一人称画像ライフログを数日間に渡って作成し、社会活動量を計測することを考える。実世界上で行われる

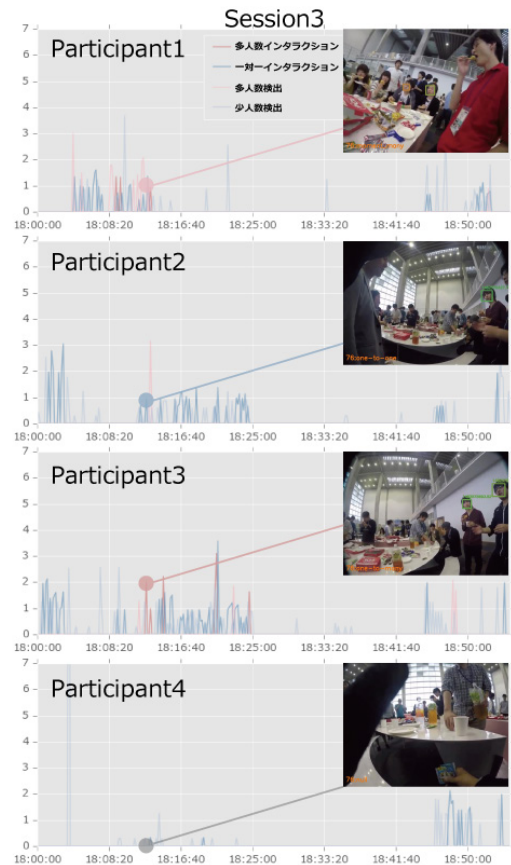


図 10 個人間の同時刻における社会活動状況の違い

Fig. 10 Difference in social activity situation between individuals at the same time.

会話場面に映り込む顔情報の他に、鏡や紙面、テレビや PC などの画面上に映り込む顔情報をもとに当人の社会活動状況の推定を行い、実世界と仮想空間の長期間における社会活動計測から当人の内的状況を推定することを目標としていく。内的状況の推定の評価のために、心拍などの情報を合わせて取得することを考える。

参考文献

- [1] 池田謙一, 唐沢穰, 工藤恵理子, 村本由紀子: 社会心理学, 有斐閣 (2010).
- [2] 原島博: 顔学への招待, 岩波書店 (1999).
- [3] 兵頭靖得, 小橋武弘, 山中泰介: IoT 行動センシングを用いた作業分類技術, 東芝レビュー Vol.71 No.5(2016).
- [4] 捧隆二, 佃洗撰, 中村聡史, 田中克己: 時間・空間・人物情報に基づくインタラクションによるライフログ画像の探索手法の提案, DEIM Forum 2012 D9-4(2012).
- [5] 高松創介, 西村邦裕, 谷川智洋, 廣瀬通孝: 顔情報を用いたライフログ利用に関する研究, 電子情報通信学会技術研究報告 MVE マルチメディア・仮想環境基礎 109(75) pp.73-78(2009).
- [6] 福本くらら, 寺田努, 塚本昌彦: ライフログにおける自動タグ付けのための笑顔認識機構の設計と実装, 情報処理学会研究報告 Vol.2013-UBI-37 No.18 pp.1-8(2013).
- [7] A.Ortis,G.M.Farinella,V.D'Amico,L.Addresso,G.Torrisi,and S.Battiatto: Organizing Egocentric Videos for Daily Living Monitoring, ACM Press pp.45-54(2016).
- [8] 樋口未来, 木谷クリス真実, 佐藤洋一: 一人称視点映像を

用いた対面会話における話者間の社会的地位の推定 (人間と機械の協働過程から生まれる知のコンピューティング技術, 人間と機械の協働過程から生まれる知のコンピューティング技術, 及び一般), 電子情報通信学会技術研究報告. CNR, クラウドネットワークロボット Vol.114 No.351 pp.43-48(2014).

- [9] B.Amos,B.Ludwiczuk,M.Satyanarayanan: 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).
- [10] M.Danelljan,G.Hager,F.S.Khan,M.Felsberg: Accurate scale estimation for robust visual tracking , British Machine Vision Conference, Nottingham, September 1-5(2014).
- [11] 矢野正治, 中田篤志, 福間良平, 角康之, 西田豊明, et al: 非言語マルチモーダルデータを用いた会話構造の分析のための環境構築, 情報処理学会研究報告 コピキタスコンピューティングシステム (2009).
- [12] 遠山魁, 角康之: 音環境比較による会話場の検出と実世界接地, 情報処理学会インタラクシオン (2016).