情報処理学会論文誌 Vol.62 No.2 607–616 (Feb. 2021)

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)]

受付日 2020年4月17日, 採録日 2020年7月7日

概要：胸に装着したカメラによる一人称ライフログ映像に映り込んだ対面者の顔の数を数えることで，カメラ装着者の対面的な社会活動量を計測する方法を提案する．社会的な場への参与の深さを測るために，検出された顔ごとの近接性（検出された顔画像の大きさ）と時間継続性（顔が検出された連続時間）の重みづけをする工夫をした．実際のライフログ映像を用いて，当事者およびその知人たちに協力してもらい，映像閲覧から読み取れる社会活動量の主観評価実験を行った．複数場面の比較による社会活動量の大小についての主観評価は，実験協力者の間で大きな偏りがないことを確認したうえで，それらの映像データに提案手法を施して算出された社会活動量の値の比較分析を行った．その結果，多くのシーンにおいて，提案手法は実験協力者の主観評価をよく再現することが確認され，単純に顔の数を数えるだけの手法よりも明らかに適切な結果を提示できることが確認できた．一方，近接した対話者とのシーンにおいては，通常の画角のカメラでは近接した対話者の顔をとらえることができず，提案手法の出力する値が主観評価を大きく下回るという問題があった．そこで，広角カメラを用いた予備検討を行い，この問題が解決できる見通しを示す．

キーワード：社会活動計測，一人称映像，ライフログ，顔検出

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

Akane Okuno[1,a)] Yasuyuki Sumi[1,b)]

Received: April 17, 2020, Accepted: July 7, 2020

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. Through experimental evaluation, many participants 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 paper briefly describes how the results can be improved by widening the camera's field of view.

Keywords: social activity measurement, first-person view video, lifelog, face detection

1. はじめに

本論文では，対面した人の数によって，日常の社会活動を計測する方法を提案する．具体的には，胸に装着したカメラに映り込んだ対面者の顔の数を数えることで，一定時間中の対面社会活動の量を計量することを試みる．

提案手法は，歩数計からの類推により発想した．歩数計は元々，単純な体の揺れから歩数を数えるだけのものであった．しかし，身体動作の識別技術の向上により，歩数計は身体活動計に発展した．歩行，ジョギング，睡眠といった身体活動の見分けが可能になり，そのことによって，

[1] 公立はこだて未来大学
Future University Hakodate, Hokkaido 041–8655, Japan
[a)] a-okuno@sumilab.org
[b)] sumi@acm.org

本論文は，Okuno, A. and Sumi, Y.: Social activity measurement by counting faces captured in first-person view lifelogging video, *Proc. 10th Augmented Human International Conference 2019*, No.19, pp.1–9, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3311823.3311846 の内容を拡張したものである．

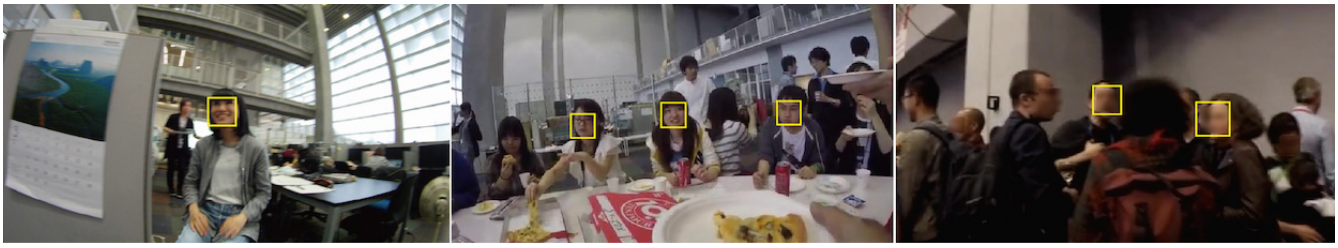


図 1 一人称ライフログ映像に対面者の顔が映り込むシーンの例
Fig. 1 Examples of scene where faces are captured in first-person view lifelogging video.

日々の運動量や睡眠量の振り返りや，ライフログのインデクス情報として活用されるようになった [8].

筆者らは同様に，対面者の顔の数を数えるだけの，いわば「顔数計」という単純な発想から検討を始めた．歩数計は，オフィスでの移動も家での家事も外での散歩も分け隔てなく，歩数の積み重ねで 1 日のおおよその身体活動量を定量化している．同様に，家族も知人も赤の他人も区別せず，他者との対面の量を機械的に数え，その積算によって一定時間中の社会活動を計量しよう，という割り切りが，我々の提案手法の特徴である．

本研究で計量を試みる対面的な**社会活動**とは，実空間における対面状況において他者と何らかの関わり合いを持つ行為全般を指す．具体的には，2 人から 10 人程度によって形成される立ち話，打ち合わせ，共同作業，共食などを想定している．計量対象者による，その社会的な場への関わりの度合い（つまり，発話したり積極的に共同作業に関与したりする度合い）を，本論文では対面的な社会活動への**参与度**と表現する．参与度は時々刻々と変化すると思われる．Goffman [7] は会話に参加している人々を，話し手，聞き手，傍参与者といった参与役割に分類し，発話交替が起きるたびに動的にそれらの役割が交代する現象を議論した．本研究ではその考え方を受け，動的に変化する参与度を，ある時点で計量対象者に向けられる他者の顔の数で表し，その時間積分を**社会活動量**とすることを提案する．社会的な場への参与は，本来はその当事者の内面から生じるものであると考えるのが自然である．それにもかかわらず，当事者の周辺にいる人の反応（すなわち，当事者に顔を向けるという行為）によって，間接的に，当事者の参与度を測ろうとするところに，本提案の面白さがあると考える．

ここまでは基本的な考えを伝えるために，説明を単純化して「対面者の顔の数を数える」と記してきたが，本論文では重要な工夫をしている．それは，対面者の近接性と時間継続性を参与度計算に考慮することである．つまり，装着カメラの映像中で検出された顔ごとに，その大きさ（物理的近さに比例する）と，検出の時間継続性（カメラ装着者へ顔を向け続けた時間に対応する）を考慮して，値の重みづけを行う．そうすることで，雑踏の中で多くの異なる

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
Fig. 2 Social activity measurement by counting faces captured in first-person view lifelogging video.

顔が入れ代わり立ち代わり検出されたとしても，その時間帯の社会活動の参与度を過大に加点することを避け，逆に，少人数だとしても同一人物と対面で対話しているような状況の参与度を重視して加点することが可能となる．

以下，本論文では，これまでに筆者らが蓄積してきたライフログデータを用いて議論を進める．筆者らはこれまで折りに触れ，小型カメラを胸に装着し，研究室内の日常活動や学会参加の一人称映像（以下，一人称ライフログ映像と呼ぶ）を記録してきた．**図 1** に示すように，一人称ライフログ映像には様々な社会活動のシーンが記録される．図に示したように，映像内で同時に検出された顔の数とその時系列パターンから，各シーンに対応する社会活動の種類やカメラ装着者の参与度について，大まかな傾向を推定できないか，というのが本研究の問いである．

図 2 に提案システムの全体像を示す．システム利用者は各々が胸にカメラを装着し，一人称ライフログ映像を記録する．画像処理によって映像中に映り込んだ対面者の顔を検出し，それに基づいて時々刻々の参与度が計算され，その時間積分によってカメラ装着者の社会活動量が計算される．その結果は，社会活動の量や，1 日に占める社会活動の種類ごとの割合をグラフ化してユーザに提示される [19].

以下，実際のライフログデータを用いて，社会活動量に関する人の主観評価実験を行い，提案手法によって計算された計量値がどの程度，当事者の主観評価を再現できるかを示す．

2. 関連研究

2.1 各種センサ情報による社会的状況の認識技術

個人および集団の社会的状況を認識する技術は，これまで多くの研究で提案されてきた．たとえば，加速度センサによる運動，スピーカによる音声，Bluetooth による他者との近接，IR センサによる対面者の認識といった複数のセンサ情報を統合することで社会的状況を計測し，組織の生産性や職務満足度の改善が試みられている [3], [12], [20].

Sumi ら [13] は，一人称映像記録用のヘッドセットに組み込んだ赤外線 ID トラッカによって注視対象を自動判別することで，多人数インタラクションにおける相互注視，共同注視を特定し，協調体験データの自動インデキシングを行うシステムを提案した．また，各ユーザが置かれている環境音の近さから会話場を検出する技術を提案し [11]，ライフログデータの閲覧支援に応用した [16].

装着者の表情変化に着目してライフログデータにインデクスをつける試みもなされている．Fukumoto ら [6] は一人称映像記録のためのカメラを装着した眼鏡にフォトリフレクタを組み込み，装着者自身の笑った時間帯を特定してライフログ閲覧に応用した．

本論文では，他の特殊なセンサ類は利用せずに，自らの社会活動量を測りたいユーザ自身が身に着けたウェアラブルカメラだけを用いる．そして，その一人称視点映像に映りこむ対面者の顔を手がかりにして，カメラ装着者の日々の社会的な活動へのかかわりの度合いを計量する方法を提案する．

2.2 一人称視点映像による社会的状況の認識技術

一人称視点映像に対する画像処理によってカメラ装着者の社会的状況を認識する試みも多くなされてきた [14]. Fathi ら [5] は，映像中の顔の抽出や人物特定だけでなく，カメラ装着者に対する相対的な対面者の顔の向きや立ち位置まで特定し，その時間的変化パターンから社会的インタラクションの状況を推測することまで試みている．Alletto ら [1] は，一人称映像に映り込んだ人々の頭部の位置と向きから，会話グループの特定を試みた．Yonetani ら [18] は，将来的に多くの人が各々ライフログカメラを装着している未来を想像し，他者の映像に映った自分自身を特定する手法を提案した．

本研究もこれらと同様に，一人称映像に映りこんだ他者の顔を画像処理によって抽出する．しかし，ここで紹介した先行研究は，相対的な立ち位置の関係を推定するといった先進的な技術を必要とする．それに対して本論文で提案する手法は，映像内の顔の抽出のみを行い，個人情報を必要としない．そのような頑健かつ個人情報を利用しない簡易的な方法で，いかにカメラ装着者の社会活動を計量できるかを見極めることが，本論文の目的である．

3. 顔検出に基づいた社会活動計測

本研究では，他者と関わる場において，一人称映像への対面者の顔の映り込みからカメラ装着者のその場への参与度を推定し，その時間積分から社会活動を計量しようとするものである．しかし，ただ顔の個数を数えるだけでは，雑踏で大勢の他人とのすれ違い状況を過大評価してしまうであろうし，逆に，特定の人物との密な対話状況を過小評価してしまうと思われる．そこで，本論文では，検出された顔の大きさから対面者の物理的近さを読み取り，同一顔を連続して検出した継続時間を対面インタラクションの密度ととらえる．そして，それらを重みづけしながら抽出された顔の数を数えあげることにより，社会活動量に対する人の印象に近づけることを試みる．

具体例を**図 3** に示す．この例では，2 人の対面者の顔が検出されている．検出されたそれぞれの顔は，前後の複数の時刻で検出される範囲において同一の識別番号が割り当てられる．各時刻ごとに，それぞれの大きさ（図中の D）と連続抽出された経過時間（図中の T）が計算される．検出された顔ごとにかけ合わせて，時間方向に累積したものを社会活動量とする．

ある時刻 t の社会活動量 S は式 (1)，(2) で計算する．式 (2) の顔の大きさ D_i は，撮影画面全体に占めるその顔の面積の割合を表す．新しく検出された顔には新しい識別番号 i が割り当てられる．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識別番号が引き継がれる．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，たとえ同一人物の顔でも別の新しい識別番号が発行される．同一の識別番号に対応する顔が連続で検出されると，その T_i をカウントアップしていき，検出の継続時間とする．計量対象となる時間（つまり，$t:1\sim m$）について積算したものを社会活動量 S とする．

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

i：検出された顔の識別番号
$T_i(t)$：継続時間（同一顔の検出継続フレーム数）
$D_i(t)$：大きさ（画面に占める顔の面積の割合）
m：時刻 t までの経過時間（フレーム数）
n：時刻 t までの累計人数（顔の個数）

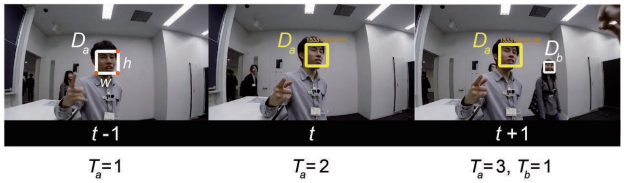


図 3 検出された顔ごとの大きさと時間継続性の計算
Fig. 3 Calculation of size and continuity of each detected face.

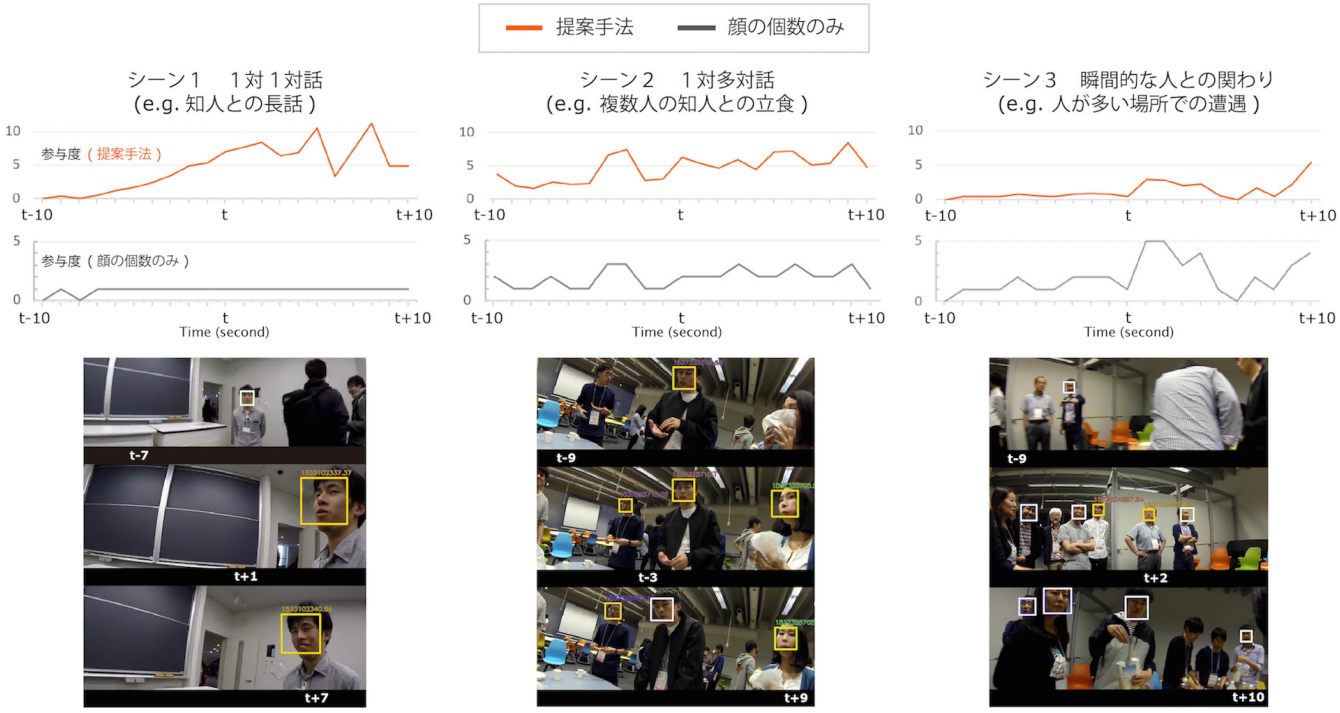


図 4 参与度計算の例：1 対 1 対話，1 対多対話，瞬間的な関わり

Fig. 4 Participation degrees of example scenes: one-to-one, one-to-many and instant communication.

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

$$\left(\begin{array}{l} w_i：\text{検出顔 } i \text{ の幅} \\ h_i：\text{検出顔 } i \text{ の高さ} \\ R：\text{画面サイズ} \end{array} \right)$$

図 4 の具体例を見ながら，提案手法の効果を説明する．3 種類のシーンについて，各フレームごとの参与度の変化をグラフにした[*1]．カメラ装着者は，シーン 1 では知人と 1 対 1 で長話をしており，シーン 2 では複数人が集まった場所で立食をしており，シーン 3 では大勢の人々が集まっている場所を渡り歩いている．

提案手法による計算結果とあわせて，検出された顔の個数をそのまま表示したグラフを示す．図の下部には，各シーンの代表フレームにおける顔検出結果を載せた．白い枠線は新規に検出された顔であることを示し，連続して検出された顔は黄色い枠線で囲んだ．

顔の個数を単純に参与度として使う場合，カメラ視野内で検出された顔の数がそのまま使われるので，大勢が密集している場所に身を投じている場合に値が大きくなる．なので，シーン 2 とシーン 3 に比べてシーン 1 の参与度は低く評価される．一方，シーン 2 とシーン 3 の違いは顕著には現れない．

それに対して，提案手法では対面者との距離と対面継続時間を考慮しているため，シーン 1 のように対面者が 1 人の場合でも継続的な対話をしているシーンの参与度は徐々に高くなる．一方，シーン 3 のように各時刻で大勢の人と対面していたとしても，対面時間の継続性や対面者との距離が遠いままの場合は，提案手法では参与度は低い値となる．また，同様に複数人と対面している場合，シーン 2 のように同一の人々としばらく対面時間が続く場合は参与度に寄与する．

顔検出には，カーネギーメロン大学で開発された OpenFace [2] を用いた．OpenFace 内で使われている Dlib ライブラリには，フレーム間で同一人物と推定された顔を追跡する機能がある [4]．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．なお Dlib は，同一顔の未検出が 1 フレームだけ割り込んだ場合は，同一顔として追跡を続けるので，本研究でもそのまま採用した．

また Dlib は，真横を向く顔や後頭部は検出せずに，正面を向いた顔のみを検出する．このことも本研究には都合が良いので，そのまま用いた．なぜなら，対面者がカメラ装着者の方に顔を向けているということが，カメラ装着者がその社会的な場に一定の度合いで参与していることを表していると考えるからである．逆に，真横を向いた顔でも検出してしまう頑健性の高すぎる顔検出システムは，筆者らの考える参与度計算にはふさわしくない．

[*1] ここでは 1 秒ごとに 1 フレームとして顔検出の画像処理を施したので，合計 20 フレーム，約 20 秒間に相当する．

4. 映像閲覧から得られる社会活動量の主観評価実験

本論文の目的は，提案手法によって提示される社会活動量が，人の印象をどの程度再現できるかを見極めることである．そこで，ライフログの当事者とその知人を実験参加者とし，ライフログから抜粋した複数の映像を閲覧してもらい，それらのシーンにおけるカメラ装着者の社会活動量についての印象を答えてもらった．

主観評価に参加した実験参加者は 8 人である．ライフログ映像から抜粋された 1 分間ずつの 10 本のシーン映像を閲覧してもらい，それらを，カメラ装着者の社会活動量が多いと感じる順番で並べてもらった．8 人分のデータを集計して主観評価の傾向を確認した．そのうえで，10 本のビデオに提案手法を施して社会活動量を計算し，上記の主観評価との比較を行った．

4.1 実験に用いた一人称ライフログ映像

2017 年 3 月に開催された学会「インタラクション 2017」に参加した際に記録された一人称ライフログ映像を評価実験に用いた．実験参加者の 1 人である P1 がデモおよびポスター発表のあるインタラクティブセッションを見学して回った約 1 時間半の一人称ライフログ映像を用いた．カメラは身体装着が容易な GoPro HERO4 を用い，解像度 1,280 × 720，超広角設定（垂直画角 69.5 度，水平画角 118.2 度）で記録した．また，カメラを胸に装着することで，ぶれの影響が少なく安定した映像を記録し [17]，超広角設定により周囲の人が写るように考慮した．なお，記録の際には映像として記録したが，後述するとおり，1 秒ごとに画像を切り出して，画像処理および社会活動量の計算を施した．

学会参加のときのライフログ映像を用いた理由は，様々な種類の社会活動，すなわち，大勢での発表見学，少人数での会話，個人的なデモ体験などが含まれるからである．その中には，カメラ装着者自身が主体的に発話者になっている場面や，大勢のうちの 1 人として聞き手になっている場面も含まれる．また，映像中には，カメラ装着者の知人と他人が適度に登場するため，評価素材として適していると考える．

表 1 は実験に用いた映像のリストである．約 1 時間半の一人称ライフログ映像から，各 1 分間の様々な種類の社会活動に相当する場面を抽出した．

4.2 実験参加者

実験参加者はすべて筆者の研究グループに所属する大学生である．そのうちの 5 人（P1 から P5）は，実験対象となる映像が記録された学会に参加していた．その中の 1 人である P1 が装着していた一人称ライフログ映像を評価実験に用いた．P2 は評価用映像の中に数回登場し，P1 と会話している場面が含まれる．別の 3 人（P3 から P5）は学会には参加していたが評価用映像の中には登場せず，残りの 3 人（P6 から P8）は学会に参加していない．まとめると，実験参加者の内訳は**表 2** のようになる．

表 2 に整理したように，実験参加者はライフログ映像をそれぞれ一人称，二人称，三人称の視点で閲覧・評価することになる．実験参加者はすべて同じ研究グループに所属しているので，映像に出てくる人々が知人なのかどうかの見分けができ，また，映像から聞こえてくる声がカメラ装着者のものなのかどうかも容易に聞き分けられる．したがって，ここで行う評価実験は，ライフログの当人および身近な人々の印象を測るものであり，本論文はそういった当事者たちの印象を提案手法がどの程度再現可能かを確かめることになる．

4.3 社会活動量の大きさへの印象評価

8 人の実験協力者各々に，表 1 に示した 10 本の映像を視聴してもらい，そこから読み取れるカメラ装着者の社会活動量の大きさの順序を回答してもらった．順序を回答してもらったのは，社会活動量の大きさの絶対値を回答することは困難であると考えたからである．具体的には，10 個の映像を視聴してもらい，それらのシーンの社会活動量の大きい順に記号 $>$ と $=$ を使って並べ直してもらった．あわせて，各自の判断基準を自由形式で記述してもらった．

表 1 P1 がデモ見学中の一人称ライフログ映像から抽出した 10 個の映像

Table 1 Ten extracted scenes from P1's ego-centric video during demo tour.

	映像の内容
A	1 人で廊下を歩いて移動
B	雑踏の中を移動して発表者がいる場所に移動
C	発表者と対話をした後に雑踏の中を移動
D	展示デモを体験しながら発表者と対話
E	人物 P2 と 1 対 1 で対話
F	人物 P2 と他の人物との複数人で対話
G	遠くから発表者の話を聞く
H	人物 P2 と遭遇して短い対話
I	発表者と他の見学者の話を背後から聞く
J	多くの聴衆と一緒に遠くから発表者の話を聞く

表 2 印象評価実験の参加者

Table 2 Participants in the experiment for subjective evaluation.

	参加者
カメラ装着者本人	P1
映像中の対話者	P2
映像中の非対話者（会議参加）	P3，P4，P5
映像中の非対話者（会議不参加）	P6，P7，P8

表 3 に実験協力者 8 人の主観評価の結果を示す．8 人の主観評価の傾向と分散を見やすくするために，これらの順序情報を数値に変換した．具体的には，10 個の映像の評価が高い順に $10, 9, 8, \ldots, 1$ の点数を与えた．ただし，＝の記号が使われた場合には，その範囲の平均点を与えた．

上記の手順で数値化した結果を**図 5** にグラフ化した．その際，各映像ごとに 8 人から与えられた数値の中央値を求め，その順序で図中に表示した．中央値が同じ場合には平均値で順序を決定した．なお，図中には各実験協力者のデータから得られた値をプロットし，値の分散を大まかに見るために箱ひげ図を表示している．

この結果から以下のことが読み取れる．

表 3 社会活動量の大きさへの印象によって映像が並べ替えられた結果

Table 3 Subjective orders of social activity amount among the ten scenes.

	社会活動量の大きさに関する印象順序
P1	F > E = D > H = C > B > J = G = I > A
P2	F > E > H > D = C > J > B = G = I > A
P3	F > E = H = C > D = G > J > I > B > A
P4	F > E > D = H > C > J = G = B > I > A
P5	F > E > D > H > C = J > I > G > B > A
P6	F = E = D > H > C > J = B = G = I = A
P7	D = C > F > B > E = H > J = G = I > A
P8	D > F > E > H = C > B = G > J = I > A

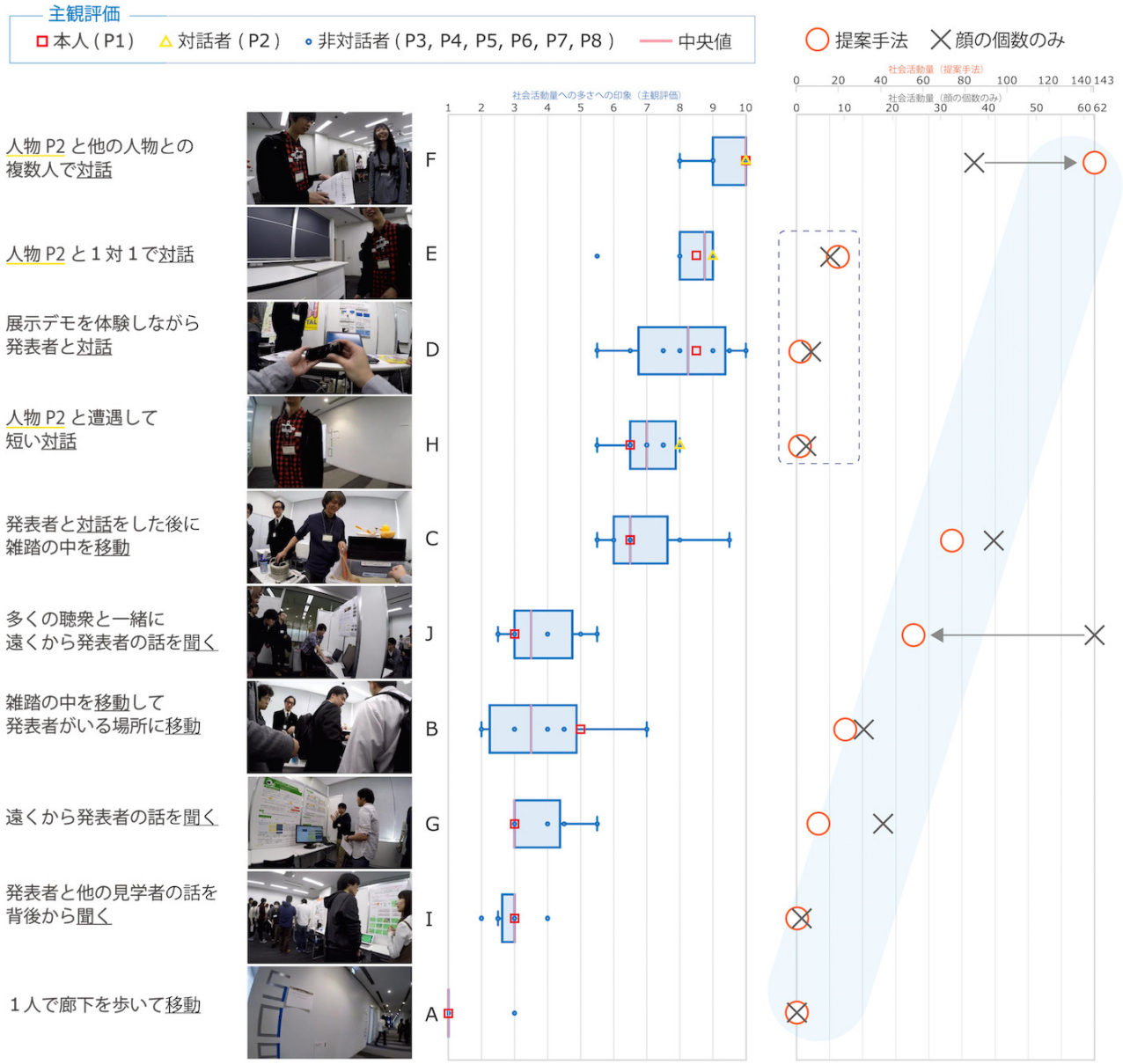


図 5 本人，対話者，非対話者による社会活動量の主観評価と，提案手法および顔数のみから計算された社会活動量の比較

Fig. 5 Comparison between human impression on social activity amount and calculated results by the proposed methods.

- 対話，聞く，移動という行為の順序で社会活動量の大きさが評価された．
- 対話中に本人（P1）が多く発話している場面の社会活動量が高く評価された．
- 会話場面では，会話相手の人数が多い方がより高く評価された．
- 大勢と対面していても本人（P1）が聞き手になっている場合は，社会活動量は低く評価された．
- 誰とも会わない場面は，全員が最も低く評価した．
- 本人（P1）の印象と他者（P1 以外）の印象には大きな違いはなかった．

評価基準に関する自由記述には，実験参加者が「本人（P1）が発話しているかどうかを重視した」と書いており，このことは上記の結果とよく符合する．

なお，学会に参加していなかった P7 と P8 の 2 人だけ，上位に位置づけた場面が他の実験協力者と異なっていた．現場での体験の共有の有無が社会活動量に関する印象に影響する可能性が示唆され興味深いが，今回の限られたデータからは有意なことはいえない．より多くのデータによる分析と議論は今後の課題としたい．

以上のとおり，一部の場面については実験協力者の評価の分散が大きいものもあったが，全体としてはおおむね，複数の場面間の社会活動量の大小について，複数評価者の印象が一致していることが確認された．また，社会的場面における当事者の参与度が高いほど，つまり，発話量が多かったり積極的にその場の活動に関わっているほど，社会活動量が高く評価されており，それはまさに筆者らが定量化を目指す社会活動量の定義そのものである．

4.4 提案手法によって計算された社会活動量と当事者らの印象の比較分析

ここまでで，筆者らが定量化を目指している社会活動量の定義が，当事者とその友人たちによる印象とよく対応していることが確認された．いよいよ本論文の本題に入る．つまり，本論文で提案した方法，つまり，その近さと対面時間を考慮したうえで対面者の顔の数を数えるという単純な方法が，当事者やその周辺にいる人たちの社会活動量に関する印象をどの程度再現できるかを見てみる．

図 5 の主観評価データの右に，各々の映像データに提案手法を施して得られた値を表示した．比較対象として，近接性や時間継続性は考慮せずに各時刻に検出された顔の個数を加算するだけの方法で得られた値も示す．なお，顔検出の画像処理は 1 秒ごとに適用した．

この図は，主観評価，提案手法，比較手法の 3 つの方法で定量化された値の変化の傾向を比較検討するために，1 つの図の上に可視化したものである．それぞれの尺度は異なるものなので，それらの絶対値を比較することには意味がなく，ここでは，それぞれの変化傾向の対応を見ることとする．

まずいえることは，提案手法（図中の丸印）の並びは，一部のデータを除いて，右上がり，つまり主観評価の並びの傾向を再現していることが見てとれる．特に，場面 J と F に注目してほしい．場面 J は，多くの他者と対面しているため，単純に顔数を数えるだけだと社会活動量が過大評価されてしまう．事実，場面 J のバツ印は全映像中で最も高い値をとっている．しかし実際には，カメラ装着者である P1 は集団の比較的遠くから発表者の話を受動的に聞いているだけであった．つまり，筆者らによる定義では社会活動量は低くなるべきであり，図に見られるように，提案手法が提示した値は低めに抑えられている．

一方，場面 F は，見学中に出会った P2 と立ち話を始め，そこに他の知人 2 人が加わってきて話を続けるシーンである．つまり，対面者は 3 人だけであり，顔数を数えるだけの方法では値はそれほど大きくならない．しかし，この場面ではカメラ装着者である P1 が主に発話し続けており，その話を聞くために 3 人が近寄ってきている場面であり，筆者が定義するところの当事者の参与度は高い．その結果，場面を通した当事者の社会活動量は大きな値になることが期待され，事実，提案手法は 10 個の場面中で最も大きな値を提示した．このように，全体の傾向としては，筆者らの提案する社会活動量の計量手法は，うまく当事者およびその周辺の人々の印象を再現しているといえよう．

一方，大きくこの傾向から外れた場面がある．場面 E，D，H である．図中で破線で囲ったとおり，この 3 つの場面については，提案手法と比較手法の両方について，期待される値に比べて大幅に小さい値を示した．その理由は，図のサムネイルを見ると分かるように，対面者の顔が映像視野の外に出てしまっているために，顔検出がほとんどなされていないからである．場面 E と H ではいずれも，親しい友人 P2 と立ち話をしており，親しいゆえに立ち位置の近い時間が続き，胸に装着したカメラの超広角設定では顔をとらえることが難しかった．また，場面 D では，P1 本人はデモ機器を触りながらデモ発表者と会話をしている．P1 は一定の発話量を保っており，また，デモ機器を操作していることを考えると，筆者らの定義による参与度は高い値をとることが期待される状況であった．しかし，この場面では P1 はしゃがみ込んでデモ体験を行っており，対話相手である発表者は頭上から覗き込むように対話に参加していた．したがって，提案手法から得られる社会活動量を当事者らの印象に近づけるためには，より画角が大きいカメラが必要であることが分かった．

図 5 を見ることで大まかな傾向を確認できたと考えるが，実際にどれくらい当事者らの印象を反映しているか定量的に確認する．筆者らの提案する社会活動量の計量手法は，単純に顔数を数える計量方法と比べてどれくらい印象との誤差があるのかを調べた．各社会活動量を 1 から 10

のスケールに変換し，場面 E，D，H 以外の 7 つの場面に対して，当事者らの印象との平均絶対誤差（MAE：Mean Absolute Error）を計算した．得られた結果は，当事者らの印象と提案手法から得られる社会活動量の MAE は 0.89，当事者らの印象と単純に顔数を数えて得られる社会活動量の MAE は 1.93 であった．したがって，近接性と時間継続性を考慮することで当事者らの印象に近づけることができたと考える．

一方，対話者の顔がカメラに映らなければ提案手法は無力である．この実験を含む長期的運用を通して学んだことであるが，人は誰かと親しく会話をしているときほど立ち位置が互いに近くなる．また，少人数（2 人ないし 3 人）の立ち話では互いの正面に対面して立つことは稀であり，ある程度角度を保ちながら立つことが多い．したがって，対話相手が当事者の方を向いていたとしても，その顔を継続的にとらえることが難しい．そのため，筆者らが高い値を期待するような社会的状況であるときに，むしろカメラが対話相手の顔をとらえることができず，結果的に社会活動量の出力値が期待よりも小さくなってしまうことがある．このことは，現時点での提案手法の大きな弱点である．しかし，この弱点は技術的に十分改善可能であると考える．つまり，カメラ装着者の前面全体を半球状にとらえることができるくらいの画角の大きなカメラを使えばよいと思われる．次章ではその予備検討について述べる．

5. 広角カメラによる改善の検討

通常カメラでは近くや斜め前に立っている対話相手の顔をとらえられないという問題に対処するために，現在は広角カメラを用いて「顔数計」の試作に取り組んでいる．具体的には，画角が 200 度のカメラモジュールを使って小型マイコンデバイス Raspberry Pi を用いた独自のカメラモジュールを試作している．

図 6 に，同じ場面を同時撮影した通常カメラ（GoPro HERO4）の画像と新しく試作している広角カメラによる画像を示す．この例は，参照物を挟んで斜め横に立っている他者と対話している，典型的な社会活動の場面である．しかし，これまでに試用してきた通常カメラでは，たとえ広角モードで撮影したとしても対話者の顔は映っていない．それに対して，身体前面全体をとらえる広角カメラであれば同じ状況でも対話者をとらえることができ，画像処理によって顔抽出も問題なく成功していることが確認できる．

図 7 は，上述した画角の広いカメラモジュールを日常運用した中で得られた例である．食事やデスクワーク中に隣に座った人と話をしたり，立っている人と座っている人が混ざっていたりするような会話状況では，これまでの通常画角のカメラによるライフログでは対面者の顔が映っていないことが多かった．それに対して，画角を広げた試作システムによるライフログでは，そういった場面でも対話者

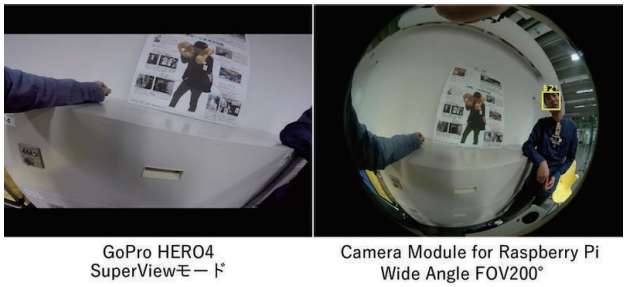


図 6 カメラ画角を広げることによる対話者の捕捉の改善
Fig. 6 Improvement of capturing conversational partners by extending FOV of camera.

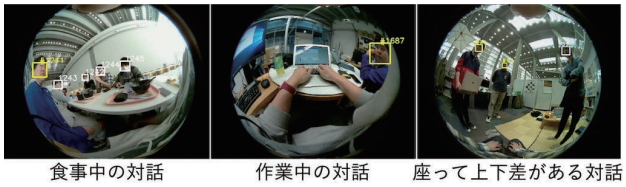


図 7 広角カメラによって非正面の対話相手の捕捉が可能になった例
Fig. 7 Successful examples of capturing conversational partners staying side by side by using wide-angle camera.

の顔をとらえることができ，図に示したように，継続的な顔検出が成功していることが分かる[*2]．

ただし，広角カメラで撮影された画像は周辺視野が歪むため，顔検出の画像処理を施す際に，事前に歪み補正を行ったり，頑健な顔検出アルゴリズムを適用したりするなどの工夫が必要である．また，このような魚眼画像では，中心視野は周辺視野よりも相対的に小さく映るので，対面者との近接性に相当する顔面積の大きさの計算を再検討する必要がある．現在，これらの試行錯誤に取り組んでいる．

6. おわりに

胸に装着したカメラによる一人称ライフログ映像に映り込んだ対面者の顔の数を数えることで，カメラ装着者の社会活動量を計測する方法を提案した．社会的な場への参与の深さを測るために，検出された顔ごとの近接性（検出された顔画像の大きさ）と時間継続性（顔が検出された連続時間）の重みづけをする工夫をした．

実際のライフログ映像を用い，当事者およびその知人たちに協力してもらい，映像閲覧から読み取れる社会活動量の主観評価実験を行った．複数場面の比較による社会活動量の大小についての主観評価は，実験協力者の間で大きな偏りがないことを確認したうえで，それらの映像データに提案手法を施して算出された社会活動量の値の比較分析を行った．その結果，提案手法は実験協力者の主観評価をよく再現することが確認され，単純に顔の数を数えるだけの手法よりも明らかに適切な結果を提示できることが確認で

[*2] 顔の黄色い囲みは，その顔が複数フレームで連続抽出されていることを示している．

きた．

一方，通常画角のカメラでは近接した対話者の顔をとらえることができず，提案手法の出力する値が主観評価を大きく下回るという問題があった．そこで，広角カメラを用いた予備検討を行い，この問題が解決できる見通しを示した．

当事者とその周辺の知人による主観評価実験を通して，社会活動への参与度の高さには，単純な対面人数だけでなく，当人の発話量の多さや社会的場に対する積極的な関わりが強く影響することが確認された．しかし提案手法は，音声処理による発話量を測ったり，映り込んだ自らの手のジェスチャを認識したり，頭部方向の変化を測ったりすることもない．ただ対面する人の顔を手がかりにして，間接的に，当人の社会的参与度を測っている．

このような単純な方法が，当事者らの社会活動量に関する印象を再現できることは驚きである．しかし考えてみると，カメラ装着者が発話や手作業によって，目の前の社会的な場に積極的に参与している場合，周辺の人々の顔はそのカメラ装着者を注視する傾向があり，その結果，提案手法によって計算される値は大きくなるであろう．したがって，一人称映像中の対面者の顔検出から間接的に，当人の社会活動量を測るという方法は理にかなっていると考える．

近い将来，装着デバイス内で即座に顔検出を行い，それらの時系列データのみを用いて社会活動の計量と，ユーザへの可視化フィードバックが可能となろう．提案手法は個人を識別する情報を排除した顔検出情報のみを用いており，かつ，社会的参与度の計量を目指しているにもかかわらず音声データなども利用しないという意味で，プライバシ配慮への有効性は高いと考えている．

本論文で行った評価実験は，ライフログを行った当人とその周辺の知人による印象と提案手法の比較検討を意図したものなので，実験参加者の属性と人数が限られたものになった．それでも，ライフログ当人と周辺の印象に大きな差がないことが確認でき，提案手法がそれらの印象を再現できることを示すことができた．

なお，今回は学会参加時の大勢での見学，少人数での会話，個人的なデモ体験という社会活動を扱った．発話，ジェスチャ，人との距離感の調節ができる社会的な場であり，日々の生活の中で共通している部分が多いであろう．また，個人で制御することが難しいような社会活動もあるだろう．たとえば，満員電車や飲食店での他者との相席などである．現状では，会話をせずに人と対面し続ける場面でも，提案手法による社会活動量は大きな値をとる．そのような身体動作の自由度が低い社会活動に対する当人らの印象との比較も興味深い．今後の課題は，多様な日常生活を対象にし，異なる社会的立場や年齢の人々を対象とした評価実験である．

提案手法の応用として，日々の社会活動による孤独感や疲労感に関わる社会的健康 [9] の計量と軽減に興味がある．たとえば，ひきこもりの若者 [15] やうつ病の高齢者 [10] の社会活動量を長期的に計測し，計測結果をフィードバックすることのよる当人の行動変容や周辺の介護者の支援の可能性を探りたい．

謝辞 本研究の一部は 2018 年度未踏 IT 人材発掘・育成事業の支援を受けた．

参考文献

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

[2] Amos, B., Ludwiczuk, B. and Satyanarayanan, M.: OpenFace: A general-purpose face recognition library with mobile applications, Technical Report, CMU-CS-16-118, CMU School of Computer Science (2016).

[3] Choudhury, T. and Pentland, A.: Sensing and modeling human networks using the sociometer, *Proc. 7th IEEE International Symposium on Wearable Computers, ISWC '03*, pp.216–222, IEEE Computer Society (online), available from ⟨http://dl.acm.org/citation.cfm?id=946249.946901⟩ (2003).

[4] Danelljan, M., Häger, G. and Khan, F. and Felsberg, M.: Accurate scale estimation for robust visual tracking, *British Machine Vision Conference*, pp.1–11 (2014).

[5] 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 (online), DOI: 10.1109/CVPR.2012.6247805 (2012).

[6] Fukumoto, K., Terada, T. and Tsukamoto, M.: A smile/laughter recognition mechanism for smile-based life logging, *Proc. 4th Augmented Human International Conference*, *AH2013*, pp.213–220, ACM (online), DOI: 10.1145/2459236.2459273 (2013).

[7] Goffman, E.: *Forms of talk*, University of Pennsylvania Press (1981).

[8] Guo, F., Li, Y., Kankanhalli, M.S. and Brown, M.S.: An evaluation of wearable activity monitoring devices, *Proc. 1st ACM International Workshop on Personal Data Meets Distributed Multimedia*, *PDM '13*, pp.31–34, ACM (online), DOI: 10.1145/2509352.2512882 (2013).

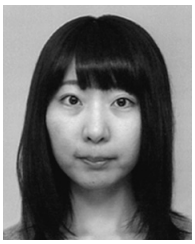
[9] House, J.S., Landis, K.R. and Umberson, D.: Social relationships and health, *Science*, Vol.241, No.4865, pp.540–545 (1988).

[10] Kaji, T., Mishima, K., Kitamura, S., Enomoto, M., Nagase, Y., Li, L., Kaneita, Y., Ohida, T., Nishikawa, T. and Uchiyama, M.: 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*, Vol.64, No.4, pp.426–434 (2010).

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

[12] Olguín, D., Waber, B.N., Kim, T., Mohan, A., Ara, K. and Pentland, A.: Sensible organizations: Technology and methodology for automatically measuring organiza-

tional behavior, *IEEE Trans. Systems, Man, and Cybernetics, Part B* (*Cybernetics*), Vol.39, No.1, pp.43–55 (2009).
[13] Sumi, Y., Ito, S., Matsuguchi, T., Fels, S., Iwasawa, S., Mase, K., Kogure, K. and Hagita, N.: Collaborative capturing, interpreting, and sharing of experiences, *Personal and Ubiquious Computing*, Vol.11, No.4, pp.265–271 (2007).
[14] Tadesse, G.A. and Cavallaro, A.: Visual features for ego-centric activity recognition: A survey, *Proc. 4th ACM Workshop on Wearable Systems and Applications, WearSys '18*, pp.48–53, ACM (online), DOI: 10.1145/3211960.3211978 (2018).
[15] 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).
[16] Toyama, K. and Sumi, Y.: Quick browsing of shared experience videos based on conversational field detection, *9th International Conference on Mobile Computing, Applications, and Services* (*MobiCASE 2018*), pp.40–55 (2018).
[17] Wolf, K., Abdelrahman, Y., Schmid, D., Dingler, T. and Schmidt, A.: Effects of camera position and media type on lifelogging images, *Proc. 14th International Conference on Mobile and Ubiquitous Multimedia*, pp.234–244, Association for Computing Machinery (online), DOI: 10.1145/2836041.2836065 (2015).
[18] Yonetani, R., Kitani, K.M. and Sato, Y.: Ego-surfing first person videos, *2015 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp.5445–5454 (online), DOI: 10.1109/CVPR.2015.7299183 (2015).
[19] 奥野 茜，角 康之：1 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ョン 2018，pp.173–182 (2018).
[20] 早川 幹，大久保教夫，脇坂義博：ビジネス顕微鏡：実用的人間行動計測システムの開発，電子情報通信学会論文誌，Vol.J96-D, No.10, pp.2359–2370 (2013).



奥野 茜 （学生会員）

2017 年公立はこだて未来大学卒業，2019 年同大学大学院博士前期課程修了，現在，同大学院博士後期課程在学中．

角 康之 （正会員）

1990 年早稲田大学理工学部卒業，1995 年東京大学大学院修了後，ATR 主任研究員，京都大学准教授を経て，2011 年より公立はこだて未来大学教授．博士（工学）．本会フェロー．