

グループ学習における個人視点映像を用いた注目行動の自動認識に関する基礎調査

近藤 一晃¹ 森 幹彦¹ 小泉 敬寛² 中村 裕一¹ 喜多 一³

概要：

ワークショップでは進行役であるファシリテータの役割が特に重要であり、十分な協同学習効果を得るためには適切な誘導・介入が求められる。この問題に対して我々は、情報技術を用いてファシリテーション行動の分析・評価を行い、次回以降のワークショップへフィードバックすることを目指している。本稿では重要な評価項目として、参加者のファシリテータへの注目行動を取り上げ、個人視点映像という一人称の映像記録から自動的に認識することを試みる。小規模ワークショップによる実証実験を行った結果、提案手法の性能や誤検出の原因が明らかとなり、その解決方法についても検討することができた。

1. はじめに

ものづくりやサービスなどの価値を創造する活動が広まってきている。このような活動では、学習者が社会的な状況の中で主体的な活動を通じて学習することが有効であるとされ、協同学習 [1]、すなわち複数の学習者が協同で取り組む学習形態で行われることが多い。また、協同学習の一形態として、近年ではワークショップが注目されている。ワークショップは、「参加者が主体的に議論に参加したり、言葉だけでなくからだやこころを使って体験したり、相互に刺激しあい学びあう、グループによる学びと創造の方法」[2]とされ、これはまさに参加者たちによる協同学習である。ワークショップは、様々な分野における活動で取り入れられており、価値を創造する活動の文脈でもしばしば用いられる。

ワークショップが教育手法として注目されている中、そのファシリテーション、すなわちワークショップを進行する仕事や手法の重要性が高まっている。とくに推進役となるファシリテータの役割は重要であり、情報技術による支援の必要性が指摘されているが、実際にファシリテータが使える技術が提供されているとは言い難い。近年になっ

て、低廉化したカメラや各種のセンサが提供されるようになってきている。これらに加えて、携帯機器やウェアラブル機器も様々な形態で提供されるようになってきているため、これらを利用することが可能な環境が整いつつある。機器の低廉化や入手の容易性の向上とともに、映像メディア処理技術は、ライフログ [3], [4] やユビキタスホーム [5] などの研究が活発になり技術開発も進んできている。

このような背景から、我々はワークショップの実践のニーズの立場と、映像メディア処理技術のシーズの立場から共同で情報技術によるワークショップの支援に関する研究を進めてきた。本報告では、ワークショップで行われる諸活動のうち、グループで行う活動にファシリテータが介在する場面での記録・分析に関わる問題に注目し、ファシリテータに対する注目行動を自動処認識する手法の検討のための基礎調査について報告する。

2. 情報技術を用いたファシリテーション評価の支援とその課題

グループ活動は単独で行う活動に比べて役割分担による共同作業が行われるため、学習意欲の向上、対話を通じた相互理解・気づきの創出、といったメリットを持つ。これを最大限に引き出すには、ファシリテータ（活動を円滑に進めるための誘導役）による適切な誘導・介入が必要不可欠であり、参加学習者（以降、参加者と呼ぶ）の行動分析を通じたファシリテーション評価は、活動を振り返り次回以降の活動を改善するための良い材料となる。しかしその評価は、印象的だった場面を思い出す・実施時に記録したメモを見返す・記録ビデオを逐一閲覧するなどの、手作業が

¹ 京都大学学術情報メディアセンター
Accademic Center for Computing and Media Studies, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan

² 京都大学大学院工学研究科
Graduate School of Engineering, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan

³ 京都大学国際高等教育院
Institute for Liberal Arts and Sciences, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan

多く定量的な分析に向かない方法で行われてきた。このような問題に対して我々は、種々の装置・センサを用いてグループ活動の場を記録し、それらの自動分析を通してファシリテーション評価を効率化することを目指している。

ファシリテーションの評価は一般に参加者の心理状態・行動や場の状況を分析することで行われるが、それを推論・理解するには高レベルな認知能力が必要であり、正確に自動処理することは未だに困難である。最終的にはやはり人間が直接記録を閲覧して判断するべきと我々は考えており、その閲覧作業を簡単かつ少ない労力で行えるよう支援することが本研究の目的である。そのためには、

要件 1 場を詳細かつ網羅的に記録すること

要件 2 状況を端的に説明する特徴量やインデックスを記録から取り出すこと

要件 3 上記 2 項の結果から閲覧しやすいインターフェースを構築すること

が求められる。例えば、野外調査における事後学習支援を目的として提案されたシステム DigitalEE-III [6] では、小型カメラ・マイクとともに GPS や加速度センサを参加者に装着して記録を行い、主に加速度センサの情報を基に重要イベントを切り出して実写仮想空間上に提示している。間瀬らの研究グループは、同じく装着型のカメラによりポスター発表会場における人物間インタラクションを記録することで、コミュニケーションの時間構造の分析法や要約ビデオの自動生成手法などを提案している [7], [8], [9]。しかし、上記要件 1-3 がどれだけ満たされていれば実用上十分なのか、そのための方法にはどのようなものが考えられるのか、はこれら従来研究では明らかにされていない。そこで本報告では、ファシリテータに対する参加者の注目を具体的な評価項目の一つとして取り上げ、要件 1 で参加者視点の映像記録のみが得られており、要件 2 で単純な視覚マーカーの検出に基づいた注目行動認識を行ったときの、その推定性能と課題について検証を行う。

3. 個人視点映像を用いた注目行動認識

3.1 集散的個人視点映像によるグループ活動の記録

本提案におけるグループ活動記録は、人間が閲覧して場の状況を理解しやすく、また要件 2 で特徴量やインデックスが取り出しやすい必要がある。そのために、まず前者を十分に満たすような記録方法を設計し、その記録に基づいた処理方法が後者をどの程度満たすのかを確認する。そしてその上で、記録方法の再検討を行う。

人間が閲覧することを考えるとやはり映像が最も適した記録メディアであろう。しかし、グループ活動中の人物は自由に歩き回って行動するため、固定カメラだけでは分析・評価に関連する出来事を十分な密度で記録することが難しい。加えて野外でのグループ活動など、固定カメラを設置すること自体が現実的でない場面も多い。記録係の力

メラマンを導入する方法も考えられるが、分析に耐えうる水準の記録を得るためには高い撮影技術が必要である。そこで本研究では、参加者全員の頭部に小型カメラを装着して複数の一人称視点の映像（集散的個人視点映像と呼んでいる）を記録する方式を採用している。集散的個人視点映像は以下のような特長を持つ。

- (1) インタラクションを主観・客観を含めて多視的に記録することができる。
- (2) 個人視点映像とは、すなわち装着者が見ていたシーンであるため、撮影されている対象やカメラワークから興味や注意といった心理状態をある程度推測できる。
- (3) 参加者間のインタラクションに対して相補的に働く可能性が高い。例えば、参加者 A の個人視点映像に参加者 B の正面顔が記録されているならば、逆に参加者 B の個人視点映像にも参加者 A が記録されているはずである。

人間が閲覧する時には第一項・第二項が、計算機による自動処理では第二項・第三項が効果的に働く。

3.2 視覚マーカーの検出を通じた注目行動認識

注目行動を自動認識する初期の提案として、視野内に対象人物が現れていることという単純な指標を用いる。具体的な手続きは以下の通り。

- 注目対象として扱いたい人物（ファシリテータ）の身体に視覚マーカーを貼り付ける
- 注目行動を認識したい人物（参加者）の頭部に小型カメラを装着して個人視点映像を撮影する
- 記録された個人視点映像から視覚マーカーが検出されたとき、注目が発生したとみなす

視覚マーカーおよびその検出には ARToolKit [13] を用いた。この手法では画像処理による矩形領域の検出とパターン認識に基づいたマーカー識別を行い、その時のマーカー・カメラ間の相対的な位置と姿勢も同時に推定する。そのため、それぞれの人物に異なる視覚マーカーを用いれば、誰が映っているのかがその距離や方向とともに取り出すことができる。

注目行動の自動認識には、固定カメラの映像から人物の顔位置と顔向きを推定する方法が古くから用いられてきた。それらの手法では、人物の位置を固定とみなしていたり [10], [11]、人物が存在する空間に制限があったりするため [12]、グループ活動のように人物が自由に行動する場には不向きである。一方、視覚マーカーと頭部可視光カメラの組み合わせの代わりに、赤外線 LED の発光パターンを赤外線カメラで認識する仕組みを頭部装着型のセンサとして用いる手法も提案されている [8]。ただし、この方法では視野内に対象マーカーが存在していたことしか分からず、注目行動を認識する上で重要な追加の情報、例えば、対象人物がどちらを向いていたのか、近かったのか遠かったの



図 1 実験の様子（左）と記録機器の外観（右）。

か、などが得られない。加えてシーンの状況を記録するための可視光カメラを追加で装着する必要がある。それに対して、視覚マーカーを利用した本提案は人物の向きや距離を同時に知ることができ、頭部に装着するカメラも1台で済む。

4. 実験

4.1 目的と方法

本稿における注目行動の認識方法は視覚マーカーの検出に基づいた単純なものであるため、おそらくそのままでは実用に耐えうる認識結果は得られない。そこで、注目行動をどれだけ正しく取り出せるのかを実験を通して確認し、その結果から提案手法に足りない要素を分析、さらにそれを補う手法について考察する。

実験では「トイブロックを用いて題目に沿った作品を作る」という仮の小規模ワークショップを開催して記録し、その時に発生した注目行動を分析した。被験者は、ファシリテータ1名とグループ構成者（参加者）4名で、全ての被験者の頭部および胸部に視覚マーカーを貼り付けるとともに小型カメラを装着して個人視点映像を記録した（図1）。実施時間はおよそ30分で、個人視点映像の解像度およびフレームレートは1920×1080 pixel, 30 fpsである。

ファシリテータに対する参加者の注目行動が分析対象であるため、参加者の個人視点映像からファシリテータの視覚マーカーを検出を行った。ただし、真の注目対象は当事者にしかわからず、さらに無意識下での注目行動も多いため、それらを正確に逐一記録することは難しい。本実験では、実験に同席していた著者らが分析対象者の個人視点映像を視聴して注目行動の有無を判定し、それをもって注目の真値とした。

4.2 結果と分析・考察

ファシリテータの頭部・胸部に貼り付けた視覚マーカーのどちらかが検出されたときを注目状態とみなした。実験結果の一部を図2に示す。図の縦軸はマーカーの検出結果（0=未検出/1=検出）、横軸は時刻（画像フレーム単位）を示している。おおよそ4500 frameを境に、前半はファシリテータによる説明時間、後半は自由作成時間である。

被験者A, Bともに説明時間にはファシリテータを頻繁に注目する一方で、自由作成時間ではほとんど手元を見な

がら作業を行っており、その傾向が検出結果にも現れていた。検出結果に移動平均フィルタなどをかければ、時間方向に粗く評価したときの注目度は十分に表すことができる。ただし、誘目動作や問いかけ等の特定行動に対する反応など、より時間方向に詳細な注目行動もファシリテーション評価には重要である。そこで以降の考察では検出結果をフレーム単位で検証する。

マーカーの検出結果が注目行動と一致しない理由には、(1) 画像中にマーカーが存在するにも関わらず自動検出に失敗している、(2) マーカーが視野内にあっても注目しているとは限らない、の大きく2つの要因が考えられる。そこでまず、マーカーの検出失敗がどのようなときに現れ、それが注目行動の解釈にどのような影響を与えるのかを確認する。続いてそれら検出失敗が修正されたと仮定した上で、検出結果のパターンと実際の注目の開始・継続・終了との対応を検証する。

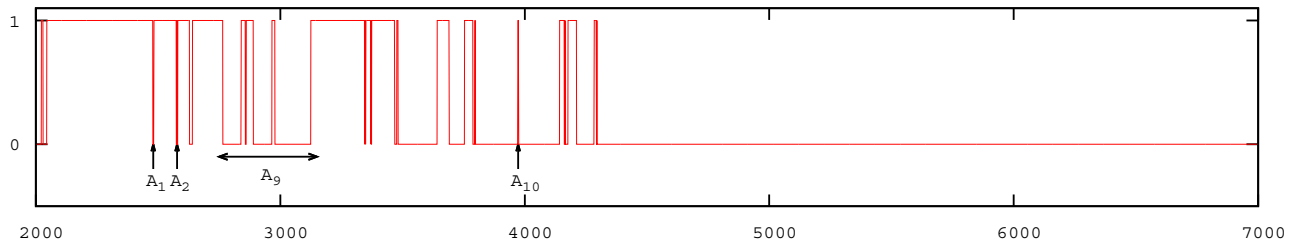
(1) マーカーの検出失敗

人物行動の特性を考慮すると、0.1秒などのごく短時間中に何度も注目対象を変えることは考えにくい。そこで、検出結果が突発的に谷になっている時刻 A_1, A_2 （図2を参照）や激しく振動している区間 B_1, B_3 などを対象に、実際の注目行動との関係を分析した。その結果、注目対象が高速に切り替わったわけではなく、ほとんどの場合でマーカーの検出に失敗していることが確認された（図3）。視覚マーカーの自動検出に失敗する要因には、モーションブラー*1などにより画像がぼけている・隠れや視野の限界が原因でマーカーの一部分が欠けている・マーカーが斜めであったり小さく映っている、などがある。カメラ装着者の頭部やファシリテータの動きにより映像内に上記悪条件が発生すると誤検出を起こしてしまうとみられる。

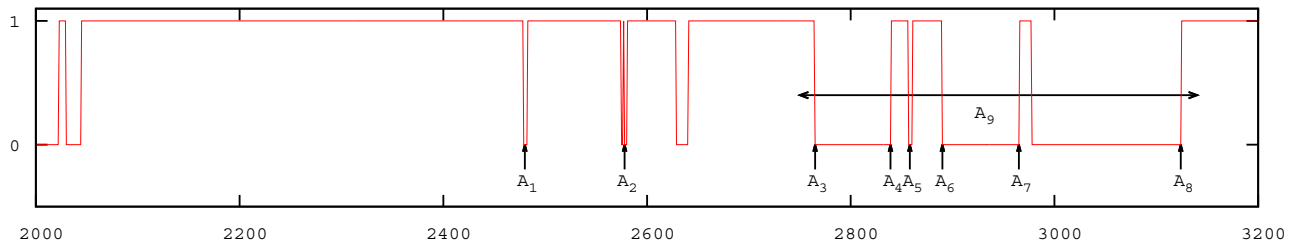
上記以外の区間でも同様の原因による検出失敗が見られたが、その種類によっては注目行動に与える解釈が異なることも判明した。例えば、時刻 A_6, A_7 では激しいモーションブラーが原因で検出に失敗しているが、そのほとんどは注目対象に視線を移すための頭部運動が原因であった。この場合、マーカーが視野内に現れている時間はごく短いため、検出に失敗する区間も同様に短く、それ故に注目の開始・終了タイミングは比較的正確に捉えられていた。それに対し、図3(b)-(d)のようにマーカーが隠れている・視野の端で切れている・斜めに映っているといった注目行動と無関係な原因では、検出失敗の区間が長くなることが多く、マーカー出現・消失のタイミングを大きく間違える傾向があった。

(2) マーカーの存在と注目行動の有無との関係

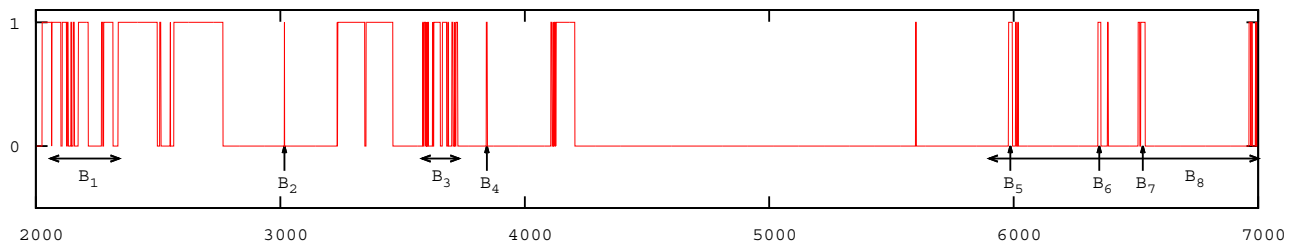
*1 カメラで撮影するとき、対象やカメラ自身が動いている場合に生じるぶれのこと。



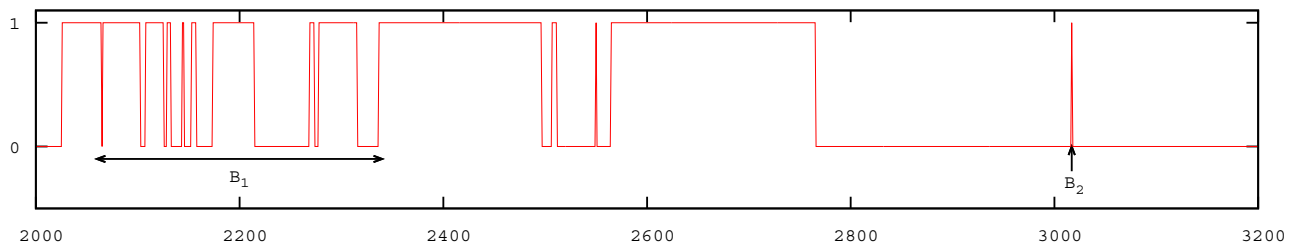
(a-1) 被験者 A. 時刻 2000 – 7000 frame.



(a-2) 被験者 A. 時刻 2000 – 3200 frame (a-1 の一部を拡大したもの).



(b-1) 被験者 B. 時刻 2000 – 7000 frame.



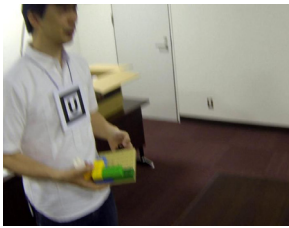
(b-2) 被験者 B. 時刻 2000 – 3200 frame (b-1 の一部を拡大したもの).

図 2 被験者二名におけるマーカー検出結果の抜粋．縦軸は検出結果（0=未検出/1=検出），横軸は画像フレーム単位の時刻を示す（1 フレーム = およそ 1/30 秒）．図中のラベル A_n, B_n は本文中で参照される時刻や区間を表す．

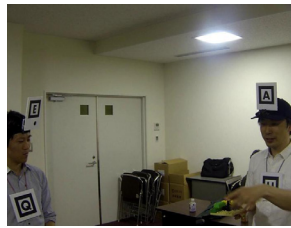
上述で触れた検出失敗が修正されたと仮定した状態で、マーカーが検出された区間と実際の注目行動との対応を分析した。その結果、説明時間中では、視野内にマーカーが存在している = ファシリテータに注目している、の仮定がほぼ成立していた。しかし、自由作成時間中では、他の参加者の様子・自身が行っている手元作業・完成品のサンプルといった同一視野内の他の対象に注目していた場合も多く、マーカーが視野内に存在していることと注目行動との相関性は、状況依存・被験者の心理状態依存であることが再確認された。

ファシリテータへの注目行動を正しく表していた例とし

て区間 A_9 を挙げる。当初ファシリテータに注目していたが、指差し動作に反応して離れた場所にあるトイブロックの山に注目対象を移した（時刻 A_3 および図 4 (a)-(d)）。その後、ファシリテータによる説明が続いていたために一度そちらに注目を戻し（時刻 A_4 および図 4 (d)-(f)）、再度ブロックの山を注目する行動を繰り返したのちに、ファシリテータへの注目に落ちていた（時刻 A_8 ）。時刻 A_5 は微小な動きによるモーションブラーが原因の検出ミスであり、これが正しく修正されたなら、区間 A_9 では正しく注目の開始・終了を表せていることになる。また、区間 B_2 では「動作や発話に反応してファシリテータに注目したがファ



(a) モーションブラー



(b) 隠れ (胴体部マーカー)

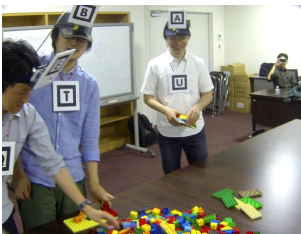


(c) 欠け (頭部マーカー)



(d) 斜め方向 (頭部マーカー)

図 3 種々の原因による視覚マーカーの検出失敗例。()内は検出に失敗しているマーカー部位を示す。



時刻 B_5



時刻 B_6

図 5 区間 B_8 で観測された非注目行動。

シリテータは他の場所を指し示していたのでそちらに注目を映した」、区間 B_4 では「一度注目したがさほど興味が引かれなかったので自身の作業に戻った」という注目行動が現れていた。注目行動を正しく表せていたこれらの区間に共通の性質として、注目対象を移すための頭部回転運動が含まれていたことが挙げられる。

マーカーが検出されているにもかかわらず、実際には他の参加者の様子に注目していた例には区間 B_8 が挙げられる。その時刻では2名の参加者が互いに相談しながら作品を作っており、被験者 B はその様子に興味を惹かれて注目していた。このとき、視野内に偶発的にファシリテータが入り込んでくるのが何度かあり、それが区間 B_8 中の時刻 B_5 , B_6 , B_7 で検出されている(図5)。また、区間 A_{10} では被験者の頭部がほぼ静止しているところにファシリテータが視野内に入り込んできたためにマーカーが検出されていた。受動的にファシリテータが出現した形であるため、やはりこちらも注目が開始したとは考えにくい。

4.3 提案手法の課題と解決策

これらの分析により、個人視点映像から視覚マーカーを検出する手法で注目行動を自動認識するには、いくつかの課題があることが明らかとなった。課題を整理したものと

それに対する解決法を以下に述べる。

マーカーの検出失敗

安定した注目行動の認識には、カメラ装着者の頭部の揺れや撮影対象の動きによって生じるモーションブラー・マーカーの隠れ・欠け・向きなどが原因による検出失敗を解決しなければならない。本提案で用いたマーカー検出のアルゴリズムは一枚の静止画を入力として用いており、画質が悪い場合は検出失敗が避けられない。このような問題には時間方向への連続性に基づいたトラッキングとの併用が必要である。すなわち、信頼度の高い検出結果を初期値として時間前後方向に探索を行うことでマーカーもしくは対象物体を追跡する。検出を妨げる要因は上記のように複数存在するため、各々に適したトラッキング処理を用いて安定した追跡を目指すべきである。

徐々に悪い条件に変化したことによる検出失敗はトラッキングによりある程度解決できる一方で、マーカーが撮影されている一連の区間において全く検出されない場合もある。そのようなケースに対応するためには、視覚マーカーだけでなく一般的な人物検出アルゴリズムを組み合わせる必要があると考える。

注目の開始や終了には高速な頭部運動が伴われていることが多いため、それを認識することができればより安定した注目行動の認識が可能となる。そのような運動下ではそもそも隣接画像間に共通の領域がほとんどない上に激しいモーションブラーが現れるため、映像に基づいた頭部運動の認識は困難と予想される。我々は、より直接に運動を計測できる機器が追加が必要と考え、無線加速度センサとの同時計測・分析を現在導入中である。

マーカーは検出されたが注目はしていない場合

該当する状況の多くでファシリテータよりも他の対象に注目していたことから、それらも注目候補として検出することで誤認識を減らすことができる。すなわち、ファシリテータへの注目行動だけでなく、グループ活動におけるインタラクションを一般的に扱う必要があると分かった。本提案では用いていないが、マーカーの距離・向き・視野内での位置・検出された時刻前後における行動などの特徴も誤認識を避けるために利用可能である。ただし、マーカーの検出と同様にあくまで「注目しているときに起こりやすい現象」であるため、注目行動の有無ではなく確率や度合いとして取り扱う仕組みも求められる。

5. まとめ

本報告では、協同的な学習としてのワークショップを念頭に置き、グループ活動におけるファシリテーション評価の支援に求められる記録方法・自動処理方法について基礎的な提案および調査を行った。具体的には、ファシリテータ

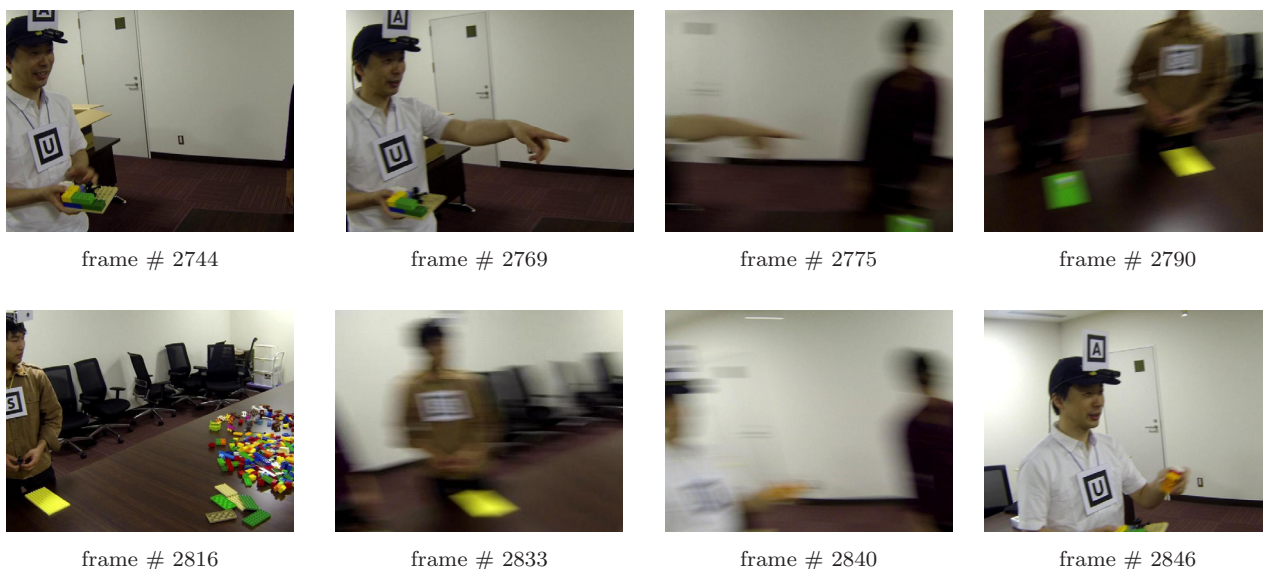


図 4 時刻 A_3, A_4 付近で観測された注目行動 . 左から右, 上段から下段へと時刻が進んでいる .

ション評価のための重要な項目として注目行動を取り上げ, 参加者の個人視点映像と視覚マーカーの検出の組み合わせた自動認識手法について検討した. トイブロックを用いた創造活動をグループで実行する実験を行った結果, 注目行動を正しく認識できる状況・できない状況について詳細な分析を行うことができた. さらに誤検出を減らすため方法としてトラッキング・人物検出手法・加速度センサ情報を併用すること, 注目行動を確率的なモデルで扱うことが必要であると確認された.

本提案で用いた集散的個人視点映像は, フィールド調査や議論シーンの記録や学習評価といったようにワークショップ全体を通じた記録メディアとして利用できる. ワークショップに限らずグループ活動において利用できるため, 自然学習・施設見学といった他の体験型の学習形態での適用も検討していきたい.

謝辞 本研究は, JSPS 科研費 24680078, 24700894 の助成を受けた.

参考文献

- [1] Johnson, D.W., Johnson, R.T. and Smith, K.A., *Active Learning: Cooperation in the College Classroom*, Interaction Book Co., 1991.
- [2] 中野 民夫, ワークショップ: 新しい学びと創造の場, 岩波書店, 2001.
- [3] Aizawa, K., Kawasaki, S., Ishikawa, T. and Yamasaki, T., *Capture and retrieval of life log*, in Proc. of Int. Conf. on Artificial Reality and Telexistence (ICAT), pp. 49-55, 2004.
- [4] Prananto, B. H., Kim, I. and Kim, H., *Multi-level Experience Retrieval for the Personal Lifelog Media System*, in Proc. of Third Int. IEEE Conf. on Signal-Image Technologies and Internet-Based System (SITIS), 2007.
- [5] De Silva, G. C., Oh, B., Yamasaki, T. and Aizawa, K., *Experience Retrieval in a Ubiquitous Home*, in Proc. of ACM workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE), 2005.
- [6] 岡田 昌也, 鳥山 朋二, 多田 昌裕, 角 康之, 間瀬 健二, 小暮 潔, 荻田 紀博, 実世界重要体験の抽出・再現に基づく事後学習支援手法の提案, 電子情報通信学会論文誌, Vol. J91-D, No. 1, pp. 65-77, 2008.
- [7] 熊谷 賢, 中原 淳, 角 康之, 間瀬 健二, 体験要約のためのビデオ自動編集手法, 人工知能学会全国大会論文集, No. 18, 1B1-02, 2004.
- [8] 間瀬 健二, 荻田 紀博, 角 康之, 小暮 潔, 片桐 恭弘, 伊藤 禎宣, 岩澤 昭一郎, 鳥山 朋二, 土川 仁, インタラクションに基づく体験共有コミュニケーション, 情報処理学会論文誌. コンピュータビジョンとイメージメディア, vol. 48, No. SIG-1, pp. 53-64, 2007.
- [9] 福間 良平, 角 康之, 西田 豊明, 人のインタラクションに関するマルチモーダルデータからの時間構造発見, 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, UBI2009-23, pp. 1-8, 2009.
- [10] 大塚 和弘, 荒木 章子, 石塚 健太郎, 藤本 雅晴, 大和 淳司, 多人数会話シーン分析に向けた実時間マルチモーダルシステムの構築, 電子情報通信学会研究会報告, PRMU2008-119, pp. 55-62, 2008.
- [11] 石塚 健太郎, 荒木 章子, 大塚 和弘, 藤本 雅晴, 中谷 智広, 音響情報と映像情報の統合による多人数会話における話者決定技術, 電子情報通信学会研究会報告, NLC2008-28, pp. 25-32, 2008.
- [12] 吉本 廣雅, 中村 裕一, ポスターセッションの分析のための不特定複数人物の頭部形状と姿勢のオンライン自動推定, HCG シンポジウム 2012, pp.344-349, くまもと森都心プラザ, Dec., 2012.
- [13] Kato, H., and Billingham, M., *Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System*, In Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99), 1999.