Vol.2017-HCI-171 No.26
2017/1/24

複数一人称視点映像閲覧における行動空間とカメラ位置姿勢の3次元可視化による効果

杉田 祐樹[1,a)]　樋口 啓太[1,b)]　米谷 竜[1,c)]　佐藤 洋一[1,d)]

概要：本研究では協調作業を記録した複数一人称視点映像の閲覧において、撮影者の絶対位置や相対位置、およびそれらの時間変化の把握を支援するために撮影者の行動空間とウェアラブルカメラの位置姿勢および移動軌跡を可視化するアプローチを提案する。複数のタスクによるユーザ評価実験の結果、（1）提案ビューを閲覧することによって撮影者の絶対位置・相対位置およびその時間変化のより正しくかつ容易な把握が支援されることが示され、また（2）一人称視点映像数が増加した場合や一人称視点映像を閲覧している割合が高くなるようなタスクにおいても提案ビューの参照がパフォーマンスに悪影響を及ぼさないことも示された。その際に（3）一人称視点映像同士を見比べる回数や注視点移動総量の減少傾向等のユーザの閲覧行動の変化も観察された。

Effects of 3D Visualization of Workspace and Camera Poses in Browsing Multiple First Person Videos

Sugita Yuki[1,a)]　Higuchi Keita[1,b)]　Yonetani Ryo[1,c)]　Sato Yoichi[1,d)]

1. はじめに

近年急速に普及が進んでいるウェアラブルカメラではカメラの設置場所を考慮する必要がなく、撮影者が両手を自由に使用することができるといった利点によって、撮影行為に伴う意識的制約や空間的制約にとらわれない自然な行動の記録が可能となる。また、カメラを頭部に装着して撮影した一人称視点映像では、固定カメラ映像からは得ることのできない豊富な情報を引き出すことが可能である。例えば、一人称視点映像ではより作業空間に接近した手元の詳細な映像を得られるため、作業の細部にわたって手順を克明に記録することが可能である。また、一人称視点映像の画面の動きは撮影者の頭の動きを、映像の中心付近にある物体は撮影者の注目している可能性のある物体をそれぞれ示唆する。このような一人称視点映像の特色に着目して、コンピュータビジョンの技術を用いて撮影者が何をしているかを解こうとする研究 [1] や、ライフログや視覚障害者のナビゲーション等に応用する研究 [2], [3] が行われている。それらに加えて、一人称点映像を多人数による協調作業の記録と解析に用いようとする研究も行われ始めており、複数の映像に映る重要な物体といった集合視に着目して解析を行うアプローチ [4] や、HCI における遠隔協調システムへの応用 [5], [6] 等が多数考案されている。このように、一人称視点映像による体験の記録は共有資産としての映像記録に新たな価値をもたらす可能性を秘めている。

しかし一方で、一人称視点映像を人間が閲覧しようとする際には多大な困難が伴う。一人称視点映像に特有な断続的な頭の動きは画面の断続的なブレとして顕在化する。また、撮影者の移動に伴う視点移動や、視野の断続的な変化は撮影者の位置情報の把握を困難にする。こうした閲覧の難しさは複数の映像を同時に閲覧しようとする場合により顕著になる。例えば、複数の撮影者が広い作業空間内を移動しながら行なう引っ越しや部屋の模様替えのような協調作業の場合、複数人で物を運搬するという協調行動が空間

[1] 東京大学生産技術研究所
IIS, Meguro-ku, Tokyo 153–8505, Japan
[a)] yusugita@iis.u-tokyo.ac.jp
[b)] khiguchi@iis.u-tokyo.ac.jp
[c)] yonetani@iis.u-tokyo.ac.jp
[d)] ysato@iis.u-tokyo.ac.jp

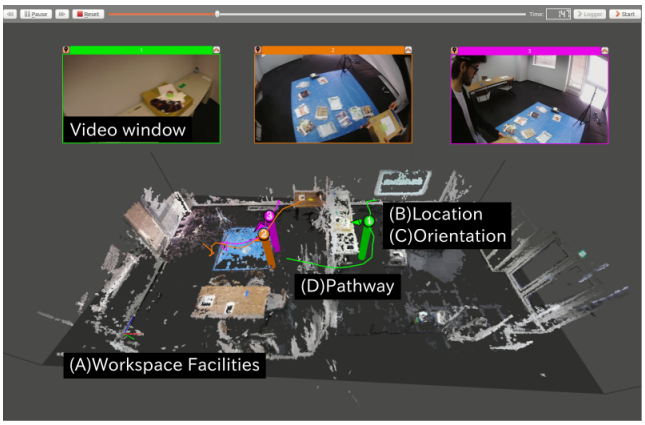

図 1: ワークスペースビューによる可視化 – (A) 作業空間の 3 次元復元, (B) 撮影者の位置, (C) 向き, (D) 移動軌跡が表示されている.

の各所で同時的に、ないし経時的に発生する。この際、(1) 撮影者がどのような経路を通って作業空間内のどこからどこへと移動するか、 (2) 各撮影者の両隣や向かいにいる人物は誰か、そして (3) 誰と誰がグループとなって協調行動しているか等を時間的に追いかけてその変化を逐一把握することが困難となる。

本研究では、協調作業における上記のような撮影者の絶対位置や相対的位置関係を把握する難しさに着目し、その支援を行なうために**図 1** のように (A) 作業が行われている空間を 3 次元復元し、(B) 各カメラの位置と (C) 向き、そして (D) 移動軌跡をその上に表示する可視化手法（ワークスペースビュー）を提案する。これは複数の一人称視点映像における撮影者位置把握の支援という点において我々の知る限りでは初めての試みである。我々は 8 つの協調作業データセットを構築してユーザ評価実験を行ない、提案手法が協調作業における撮影者位置情報のより正しくかつ容易な把握を支援することを示し、ワークスペースビューの参照によって一人称視点映像の閲覧が妨げられないことを確認した。さらにユーザの視線解析とアンケートの分析を通して閲覧行動の変化を明らかにし、現在の可視化手法が抱える課題についての知見を得た。

2. 関連研究

2.1 一人称視点映像自体の編集

一人称視点映像には撮影者の断続的な頭の動きによって多大な画面の動きやブレが表れている。この頭の動きを滑らかにしようとする手法が [7], [8] である。また、効率的にフレームをサブサンプリングするアプローチが [9] らによって考案されている。これらの手法は一人称視点映像を早送り再生しても観賞に耐えることを可能にする。一方で、複数一人称視点映像から重要な要素だけを抽出して映像自体を要約して短縮する手法も考案されている。これらの手法では要約映像に用いるショットを選択するために、重要な人物や物体 [10]、複数の映像間でのシーンの共起関係 [11]、そして集合視に着目する手法 [12] などが取られる。本研究ではこれらの研究とは異なり、複数一人称視点映像を編集せずに閲覧する場合において、映像のみからでは把握の難しい撮影者の位置情報を可視化するという立場を取る。

2.2 映像閲覧インターフェース

複数の映像を同時に再生するユーザーインターフェースとしては、固定監視カメラを想定したものが古くから研究され実用的な場面での導入も活発に行われている。並列提示された映像に加えて、作業空間のマップを提示する手法も数多く開発され、3D モデルを利用したマップ [13]、複数の魚眼レンズを元に構築したマップ [14] 等が利用されている。またマップにシーン解析の結果を可視化する手法 [15] も考案されている。本研究では、撮影者と共に移動するウェアラブルカメラで記録した映像の閲覧支援のために作業空間を俯瞰する視点からの 3 次元マップを提示し、それが積極的に利用されるかどうかを検証する。複数一人称視点映像の閲覧に関して、[16] は複数一人称視点映像を HMD 上に並べて表示し、共同作業者の視線も提示してリアルタイムでの没入型協調作業支援システムを開発した。また [17] では複数人のカメラ位置と姿勢を復元し、共同注視領域を算出して可視化を行なった。本研究ではこれらの手法で用いられているような可視化手法 (映像の並列提示や位置姿勢の提示) が複数一人称視点映像の撮影者位置を把握する上でどのように役立つかを綿密なユーザー評価実験に基づいて検証する。

3. 閲覧システムのデザイン

本研究では、広い空間内で撮影者の移動や協調動作が多数発生するような協調作業について、その映像記録を第 3 者があとで見返すような閲覧スタイルを想定する。複数映像を同時に提示するもっとも単純な手法は映像をタイル状に並べるような可視化であり、本研究ではこれをベースラインと呼ぶこととする。しかしながら一人称視点映像のように撮影者が空間内を絶えず移動する場合には、視野の断続的変化によって作業空間の情報が断片的にしか得られず、また撮影者同士の近接や見ている方向の一致に気づきづらいため (1) 各撮影者の絶対位置や (2) 相対的位置関係、(3) 協調行動しているグループ等を逐一把握してその時間変化を追うことは容易ではない。

このような問題に対処するため、我々はタイル状に配置した複数一人称視点映像に加えて 3 次元復元した作業空間を表示し、その上に各カメラに関する位置情報（位置・向き・移動軌跡）を可視化したビュー（ワークスペースビュー）を提示する。3 次元復元による作業空間の可視化は作業空間の全体像に対する速やかな理解を促し、カメラ位置の可視化は撮影者の絶対位置や相対的位置関係、近接の速やかな発見を促すと期待される。また、カメラの向き

を併用することで、一人称視点映像の撮影視点とワークスペースビューの視点との速やかな対応や近接する撮影者同士の関係性や共通視を把握する手掛かりが得られ、さらに、カメラの移動軌跡の可視化によって、撮影者の移動経路だけでなく撮影者同士の近接を前もって発見する時間的な手掛かりが得られると期待される。

4. 閲覧システムの実装

4.1 作業空間とカメラ位置情報の復元

作業空間の復元には事前に撮影した作業空間の映像を用いて VisualSFM [*1] による３次元復元のフレームワーク (Structure from Motion と Bundle Adjustment による粗い形状復元および Patch-based Multiview Stereo による密な形状復元) を適用する。復元された作業空間の点群には Point Cloud Library [*2] を用いたノイズ除去とデータサイズの圧縮を適用する。カメラの情報（位置・向き・移動軌跡）の復元には、まず各カメラで撮影した映像のフレームに対して作業空間の撮影に用いたフレームおよび他のカメラのフレームとの特徴点マッチングを行なう。次にこれらのマッチング情報を元に VisualSFM のフレームワーク (Strucure from Motion と Bundle Adjustment) を適用してカメラの位置と姿勢の復元を行なう。

4.2 閲覧システムの可視化

閲覧システムの可視化は図 1 に示したように行なう。3 次元復元した作業空間の点群は真上から約 45 度の俯瞰視点から見下ろしたものを表示する。真上からの視点でなくこのような斜め上からの視点を用いる理由は、作業空間の情報を３次元で提供するとともに撮影者の頭部という低い視点で撮影された一人称視点映像とのスムーズな視点の対応を考慮してのものである。本研究では画面に表示された作業空間の点群をワークスペースビューと呼ぶこととする。ワークスペースビューには (A) 作業空間の点群のほか、(B) カメラ位置を丸印で、カメラの作業空間床面からの高さを円柱でそれぞれ表示し、(C) カメラの向きを矢印で、そして (D) 移動軌跡を線で表示する。移動軌跡は現在位置の前後約 5 秒分の長さを表示する。これらの各カメラの可視化と各撮影者の一人称視点映像との対応は色と撮影者ＩＤが記されたラベルによって行なう。例えば図 1 ではワークスペースビューに緑色で表示されたカメラ位置と向きおよび移動軌跡は緑色の窓枠を持つ一人称視点映像に対応している。

5. ユーザ評価実験

提案手法が協調作業を記録した複数一人称視点映像の閲覧において撮影者位置情報の効果的な把握を支援することを示すために、我々は 2 つの仮説を構築し、それぞれに対してタスク群（タスク群 1・タスク群 2）を設定してユーザー評価実験を行なう。タスク群 1(タスク 1-3) ではワークスペースビューを閲覧することによって一人称視点映像のみからでは困難である撮影者位置情報の把握がベースラインよりも正しくかつ容易に達成できることを確認する。一方、タスク群 2（タスク 4, 5）ではワークスペースビューを補助的に閲覧するような場面を想定して、提示する一人称視点映像の数を変化させたり (タスク 4)、様々な役割推定に取り組んだりしてもらい (タスク 5)、そのような場合でもワークスペースビューの参照が各タスクの正答率や難易度に悪影響を及ぼさないかどうかを多角的に検証する。ユーザ評価実験ではユーザの視線計測やアンケート、インタビューも実施して閲覧行動の変化や提案手法へのフィードバック等の知見を得る。

仮説 1: ワークスペースビューの閲覧によって、ユーザは撮影者の位置情報をより正しくかつ容易に把握することができる

ワークスペースビューを閲覧することよって撮影者の絶対位置の変化、相対的位置関係、およびグループといった撮影者位置情報の把握がより正しくかつ容易に行われるかどうかをタスク 1-3 を通して確認する。

タスク 1: 撮影者の絶対位置変化の把握

複数人が作業空間に存在する状況下で特定の１名の撮影者について、実験用映像の再生時間内における移動経路を白地図に記入してもらう。実験用映像は 12 秒程度の長さで 1 名分の一人称視点映像だけが提示される。ユーザはカメラの現在位置と移動方向を示唆するカメラの向き、そして前後約５秒の移動軌跡を活用して移動経路をより正しくかつ容易に把握することができると期待される。

タスク 2: 撮影者の相対的位置関係を把握

実験用映像の再生が終了した時点での特定の 3 名の撮影者について、相対的位置関係を白地図に記入してもらう（1 名分が既に記入済)。実験用映像は 30 秒程度の長さで 3 名分の一人称視点映像が提示される。実験用映像には相対的位置関係が途中で変化するかどうかのバリエーションを設ける。ユーザはカメラの現在位置と向きを活用して相対的位置関係をより正しくかつ容易に把握することができると期待される。

タスク 3: 同一グループに属する撮影者の把握

5 名の撮影者が複数のグループに分かれて協調作業（紙への描画や箱の運搬）を行なっている実験用映像を閲覧し、その中の１つのグループについて全メンバーを白地図に記入してもらう。実験用映像は 30 秒程度の長さで 5 名分の一人称視点映像が提示される。実験用映像にはグループが当初から形成されているか、それとも途中で形成されるかのバリエーションを設ける。ユーザはカメラの現在位置を

[*1] http://ccwu.me/vsfm/
[*2] http://pointclouds.org/

用いてグループ候補となる近接する撮影者の一団を速やかに発見し、向きと移動軌跡を用いて実際に協調行動をとっているかどうかを速やかに判断してグループをより正しくかつ容易に把握することができると期待される。

仮説 2: 一人称視点映像を閲覧する割合が高くなるような場合でもワークスペースビューの参照が一人称視点映像の閲覧を妨げない

撮影者が何をしているかを把握する場合には一人称視点映像を閲覧している時間割合がより高くなると考えられ、ワークスペースビューは位置情報参照のための補助的な利用が想定される。そのような場合でも撮影者の位置情報をより正しくかつ容易に把握することができると同時に、ワークスペースビューの参照によって新たに発生する一人称視点映像とワークスペースビューとの頻繁な見比べ行動が一人称視点映像の内容理解自体に悪影響を及ぼさないことを期待する。提示する一人称視点映像の数を変化させたり様々な役割推定タスクに取り組んでもらい (タスク 4, タスク 5) 多角的に検証する。

タスク 4: 提示する一人称視点映像数を変化させた場合の移動経路の把握

3 名または 5 名の撮影者が協力して物体を運搬している実験用映像を閲覧し、その中で 2 回目に発生した運搬の経路を白地図に記入してもらう。提示する一人称視点映像数は 1・3・5 と変化させる（FPV1, 3, 5)。FPV1 の場合は 1 名の撮影者による運搬を把握し、FPV3 または FPV5 の場合はその中の 2 名の撮影者による協調的運搬を把握する。実験用映像はそれぞれ 40 秒程度の長さである。閲覧すべき一人称視点映像数が増加する場合、複数の一人称視点映像を見比べる回数が増加して処理すべき情報量が増大することが予想される。その中で撮影者の位置情報把握のために行われる見比べはワークスペースビューの活用によってその回数を大きく減らすことができると期待される。しかし一方で、提案手法ではワークスペースビューというウィジェットが追加されることによる新たな見比べ行動が発生する。複数のカメラのワークスペースビュー上での表示位置が断続的に変化する場合には固定位置に提示される一人称視点映像との間のスムーズな視線移動に困難が生じる可能性もある。そのような場合にもワークスペースビューによる位置情報の迅速な把握の効果が新たに発生する負担を補って余りあるほど大きく、タスク正答率と主観的難易度に悪影響を及ぼさないと期待する。

タスク 5: 協調作業における撮影者の役割把握

3 名または 5 名の撮影者が協調作業を行なっている実験用映像を閲覧し、特定の役割を担っている撮影者を把握して白地図に記入してもらう。実験用映像はそれぞれ 50 秒程度の長さである。本人の一人称視点映像のみから役割推定ができる場合 (ROLE1: 雑誌整理におけるタグ配布者)、手元が映っておらず本人以外の映像から役割推定を行なう必要がある場合 (ROLE2: 掃除中にゴミ袋を縛っている人物)、そして同時に複数の撮影者の役割推定を行なう場合 (ROLE3: ブロック組み立て作業における 2 名の「何もしない」作業者) の 3 種類を行なう。ROLE2 では同時に操作対象物体の位置も記入してもらう。提示する一人称視点映像数はそれぞれ 3 名・5 名・5 名分である。本タスクでは撮影者の移動が比較的少なく相対的位置関係があまり変化しないようなシーンを用いる。役割推定の最中でも映像中の人物が誰でどこにいるかについてワークスペースビューの頻繁な参照による迅速な把握が行われ、タスク正答率と主観的難易度に悪影響を及ぼさないことが期待される。

5.1 データセットの構築

本研究では 3 名または 5 名で行われる協調作業を頭に装着したウェアラブルカメラで記録して 8 種類のデータセット (1-8) を構築した。各データセットで行われる協調作業は (1) 箱の梱包と運搬、(2) ポスターの見回り、(3) 雑誌と箱の集配と整理、(4) 机を囲んだ乾杯、(5)2 人 1 組での描画、(6) 箱の運搬、(7) 部屋の掃除、(8) ブロックの組み立てである。各データセットはそれぞれ 30-630 秒程度の長さがあり、異なる 5 つの場所のうち 1 ヶ所以上で収録された。ユーザ評価実験における各タスクではこれら 8 種類のデータセットから互いに重複しないように多数の部分データを切り出して実験用映像として使用する。実験用映像はワークスペースビューや複数の一人称視点映像を同時に再生している様子を画面キャプチャして 1 つの映像として切り出される。作業空間と撮影者のカメラ位置および姿勢の復元には GPU が搭載されたマルチコア環境のデスクトップ PC を用いた。作業空間の 3 次元復元には 165-289 分、カメラ位置姿勢の復元には 88-2576 分の時間を要した。また、複数の撮影者の映像はあらかじめ手動によるフレームの時間同期が行われている。

5.2 実験の詳細

5.2.1 実験手順

ユーザ評価実験では、各タスクに関して被験者全員にベースライン（ワークスペースビューなし）と提案手法（ワークスペースビューあり）の両方を閲覧してもらった。各タスクの開始前には本番タスクとは異なる実験用映像を使用した練習問題に取り組んでもらった。被験者はタスク群 1(タスク 1-3) またはタスク群 2(タスク 4-5) のどちらか一方に取り組み、タスク 1,2,3 または 4,5 にはこの順序で取り組んだ。各タスクではまずベースライン・提案手法のうちどちらか一方について連続して 2 つ（タスク 1-3）または 3 つ（タスク 4-5）の映像を閲覧し、手法を入れ替えてもう一方を同様に閲覧した。各タスクの解答終了後には難易度に関してアンケートに回答してもらった。全タスクの

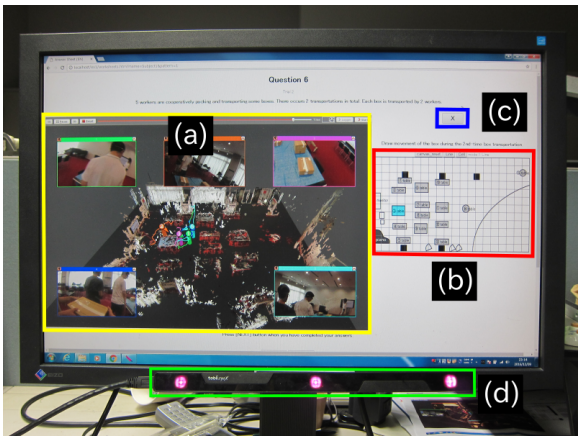

図 2: 実験用インターフェース – 24 インチモニター上に (a) 実験用映像と (b) 解答入力用キャンバスが表示されている．被験者は (c) ボタンで映像を一度だけ再生できる．閲覧中の被験者の視線は (d) アイトラッカで記録する．

終了後には実験全体を振り返ってワークスペースビューによる可視化が理解しやすかったかどうかについてアンケートを実施し、その後インタビューを実施した。

5.2.2 実験条件

ユーザ評価実験はコンピュータビジョン分野の大学院生や博士研究者 14 名を集めて実施した。被験者のうち 8 名がタスク群 1 に取り組み、8 名がタスク群 2 に取り組んだ。各タスクではベースラインまたは提案手法に取り組む順序と提示する実験用映像の種類の各組み合わせに対して 2 名ずつを割り当てて、カウンタバランスを取った。さらにタスク 4 とタスク 5 では 3 つの映像の提示順序をランダムに変化させた。実験は**図 2** に示すような環境下で実施した。実験用映像は解像度 1260x840 で 24 インチモニター（1920x1200）上に提示され、再生はモニター上のボタンによって行なわれる。被験者には再生ボタンを押す前に解答すべき内容を把握してもらい、また白地図を見て作業空間の大まかな把握を行なってもらった。各映像の再生は 1 度のみとし、巻き戻し・一時停止等の操作は一切できない。各映像の再生開始前には 3 秒のカウントダウン映像が挿入される。映像は再生終了後には速やかに非表示となる。各タスクに関する指示や解答入力も同一のモニター上にて行なう。解答入力はモニター上の 600x360 のキャンバスに対してマウスで行なう。キャンバスには 20 ピクセル毎に罫線が引かれ、机やドア、本棚等のラベルを付与した白地図が表示されている。また、被験者の閲覧行動を観察するために実験と並行して Tobii EyeX Controller を用いて被験者の視線を約 60Hz で記録した。実験終了後にはインタビューを実施し、その様子を録音した。

5.3 評価方法

客観評価

以下に示した採点基準に沿って手動で採点し、正答率を算出する。

タスク 1: 出発点・到達点・途中経路のそれぞれについて各 1 点の 3 点満点とする。出発点と到達点は白地図上のマス目に基づいて 1 マス分の誤差を許容する。途中経路については通過すべき全ての構造物（机・椅子等）の間を過不足なく通過していた場合にのみ得点を与える。

タスク 2: 白地図に既に記入されている 1 名から見た他の 2 名の相対位置について 8 方位に基づいて各 1 点を与え 2 点満点とする。

タスク 3: グループの全てのメンバーを正しく特定できた場合に 1 点、グループの絶対位置に 1 点を与え 2 点満点とする。グループの絶対位置は白地図上に記入された全てのメンバーから最も近い作業場所（机・椅子等）とする。

タスク 4: タスク 1 と同様の基準を用いる。

タスク 5: ROLE1 では役割を担う撮影者を正しく特定できた場合に 1 点、撮影者の絶対位置に 1 マスの誤差を許容して 1 点を与え 2 点満点とする。ROLE2 では ROLE1 に加えて操作物体の相対位置が正しい場合に 1 点を与え 3 点満点とする。ROLE3 では 2 名について各 1 点、2 名から最も近い作業場所について 1 点を与え 3 点満点とする。

主観評価

主観難易度を難しいを 1、易しいを 7 とした 7 段階で回答してもらった。

6. 実験結果

タスク 1-3

各タスクの正答率と主観難易度をそれぞれ**図 3**(A, D) に示す。Wilcoxon の符号順位検定の結果、タスク 1 とタスク 2 において提案手法では有意に正答率が高くなることが確認された（それぞれ $p < .05$, $p < .001$）。主観難易度では、タスク 1・2・3 の全てにおいて提案手法では有意に易化することが確認された (それぞれ $p < .01$, $p < .001$, $p < .01$)。またタスク 1-3 の提案手法において、実験用映像上で視線が観測された総時間に占めるワークスペースビューの割合はそれぞれ $66.3 \pm 15.4\%$, $57.4 \pm 13.8\%$, $43.5 \pm 13.8\%$ であった。

タスク 4

提示する一人称視点映像の数を 1・3・5 と変化させた場合（FPV1, FPV3, FPV5）の正答率および主観難易度をそれぞれ図 3(B, E) に示す。タスク正答率では有意差は確認されず、また FPV1 での提案手法とベースラインの双方、FPV3 での提案手法において全被験者が満点となった。主観難易度では、FPV5 の場合に提案手法において有意に易化することが確認された ($p < .05$)。提案手法において、FPV1, FPV3, FPV5 のそれぞれについてワークスペースビューを閲覧していた割合は $38.1 \pm 7.4\%$, $36.0 \pm 7.7\%$, $36.3 \pm 20.4\%$ であった。また被験者の注視点をもとに算出

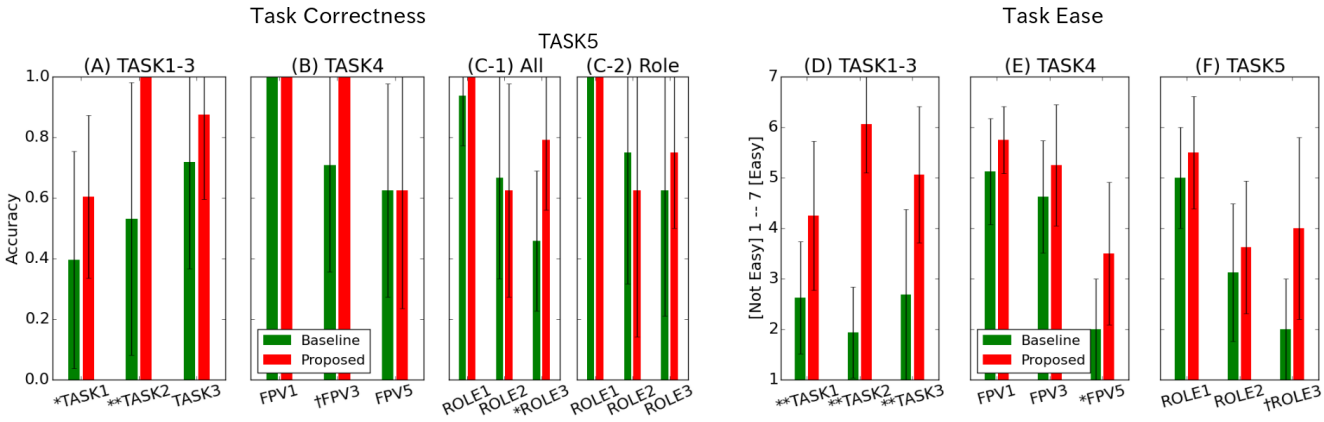

図 3: (A-C) タスク正答率, (D-F) タスク主観難易度 – (†, ∗, ∗∗) はそれぞれ有意水準 ($p < .1, .05, .01$) を示す.

した注視点移動量とウィジェットの遷移回数を**表 1** に示す。注視区間の検出には、[18] らのウェーブレット解析ベースの手法を用いて約 100ms 以上の停留区間を抽出した。

タスク 5

ROLE1-3 におけるタスク正答率と主観難易度を図 3(C-1, F) に示す。正答率では ROLE3 の場合に提案手法において有意に高くなったが ($p < .1$)、その他では有意差は確認されなかった。また ROLE1-3 のそれぞれについて、役割担当者を特定する項目のみの正答率を図 3(C-2) に示す。すべての場合において有意差は確認されなかった。タスク難易度では ROLE3 の場合に有意に易化したが ($p < .05$)、その他では有意差は確認されなかった。提案手法において ROLE1-3 のそれぞれの場合についてワークスペースビューを閲覧していた割合は $29.5 \pm 14.7\%$, $29.9 \pm 8.9\%$, $26.7 \pm 12.8\%$ であった。また注視点をもとに算出した注視点移動量とウィジェットの遷移回数を表 1 に示す。

6.1 実験後アンケートの結果

実験終了後アンケートでは、(Q1) ワークスペースビューによる可視化を容易に理解して利用できたかどうか、(Q2) ワークスペースビューにおいて作業空間の 3D モデルを閲覧する視点は適切だったかを最も不適切を 1 とした 1-7 の 7 段階で尋ねた。Q1 では 6.3 ± 0.8、Q2 では 6.4 ± 0.7 という結果を得た。

6.2 インタビュー結果の集約

実験終了後のインタビューでは、被験者 14 名に次の 5 つの質問 (質問 2 と 4 は 8 名) に回答してもらった。

1 ワークスペースビューにおける可視化 3 要素はどのような場面で役に立ったか？

カメラ位置: 誰がどこにいるのかの把握 (7 名)、移動軌跡の把握 (3 名)、空間の把握（人の位置を起点にして作業空間の背景や物体の位置が分かった）(2 名)、一緒に行動している撮影者の把握 (2 名)。

カメラの向き: 意識的には使用しなかった (6 名)、全く使用しなかった (3 名)、撮影者の移動方向の把握 (3 名)、一人称視点映像との視点の対応 (映像でこちらを向いている人物が誰なのかを向きから判断)(3 名)、作業場所の把握 (2 名)、複数人による協調の有無の判断 (1 名)、位置ズレの補正 (1 名)。使用すればよかった (2 名) や向き表示がなければ難易度が上昇しただろう (1 名) と回答した被験者も見られた。

カメラの移動軌跡: 撮影者や物体の移動経路を把握 (5 名)、先読み (撮影者位置やペア形成など)(4 名)、過去位置の確認 (1 名)、位置が時々ずれるので使用しなかった (1 名)、位置だけで十分だったので使用しなかった (1 名)。

2 役割推定ではどのようにワークスペースビューを利用したか？

複数の一人称視点映像を見比べてその役割を担当している撮影者の外見を決定した後に、ワークスペースビューを用いてそれが誰であるか (ID) を位置とともに決定した (3 名)。上記の回答をした被験者のうち 2 名は位置の解決が終わらないうちに映像が終了してしまい時間が足りなかったと述べた。映像がどの向きから撮影されたかを把握して映像に映っている人物が誰であるかを把握しながら役割担当者を決定した (2 名)。映像に映っている人物が誰でどこにいるのか分かりづらかった (2 名)、撮影者がもっと動いていれば軌跡を利用した (1 名) という意見もあった。

3 ワークスペースビューによる可視化はどのような場面で役に立たなかったか／あることによって却って混乱したか？

役に立たなかった場面: 撮影者が動いていない (位置のみ役に立った)(4 名)、手元での作業内容を把握する (1 名)、手元が見えない (位置のみ役に立った)(1 名)、運搬 (1 名)。また、移動軌跡のうち後方の軌跡が不要 (1 名) あるいは前方の軌跡が不要 (1 名) いう意見もあった。

却って混乱招いた場面: 特になし (6 名)、撮影者位置推定の誤り (3 名)、向き推定の誤り (1 名)、撮影者が動かない (1 名)、関係のない人物が映り込んでいる (1 名)。

	TASK4						TASK5					
	FPV1		FPV3		FPV5		ROLE1		ROLE2		ROLE3	
	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed	Baseline	Proposed
(1)Fixation movements	$119_{\pm 36}$ *	$227_{\pm 59}$	$534_{\pm 128}$	$482_{\pm 94}$	$629_{\pm 120}$	$577_{\pm 36}$	$260_{\pm 68}$	$299_{\pm 54}$	$357_{\pm 95}$	$342_{\pm 109}$	$370_{\pm 81}$	$358_{\pm 102}$
(2)Widget transitions	$0.0_{\pm 0.0}$ *	$1.1_{\pm 0.3}$	$1.5_{\pm 0.4}$	$1.7_{\pm 0.2}$	$1.5_{\pm 0.4}$	$1.6_{\pm 0.3}$	$0.6_{\pm 0.2}$ *	$1.3_{\pm 0.3}$	$1.1_{\pm 0.3}$	$1.3_{\pm 0.4}$	$1.1_{\pm 0.4}$ †	$1.5_{\pm 0.5}$
(FPV-FPV transtitions)	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$1.5_{\pm 0.4}$ *	$0.7_{\pm 0.2}$	$1.5_{\pm 0.4}$ *	$0.8_{\pm 0.4}$	$0.6_{\pm 0.2}$ †	$0.4_{\pm 0.2}$	$1.1_{\pm 0.3}$ *	$0.8_{\pm 0.2}$	$1.1_{\pm 0.4}$	$0.8_{\pm 0.4}$

表 1: 注視点解析の結果 – (1) 実験用映像上で観測された注視点の単位時間あたりの移動量 [pixels/sec.] と (2) 単位時間あたりのウィジェット遷移回数 [counts/sec.] を示す. (†, ∗, ∗∗) はそれぞれ有意水準 ($p < .1, .05, .01$) を示す.

4 提示する一人称視点映像数が増加した場合に、どのような困難がどのように変化したか？

ワークスペースビューなし: 見るべき映像数が増加して処理量が増大 (3 名)、FPV3 がちょうど良い (2 名)、心理的圧迫感が増大 (1 名)、FPV1 は情報量が少なくて逆に難しい (1 名)。

ワークスペースビューあり: FPV5 では何を見るべきか分からず混乱した (3 名)、見るべき映像数が増加して処理量が増大 (1 名)、物体の移動を追う際に別の場所で行われている作業を同時に把握するのが困難だった (1 名)、ワークスペースビューからは一人称視点映像に映る人物が誰であるかを把握できなかった (1 名)、視点対応の困難が増大した (1 名)、一人称視点映像とワークスペースビュー間の視線移動が大変だった (1 名)、撮影者位置表示に対応する一人称視点映像を探す困難が増大した (1 名)。

5 提案手法への改善要望点

位置・向き・移動軌跡の推定精度向上 (4 名)、撮影者位置表示と一人称視点映像を近づけてほしい (4 名)、グループになっている撮影者の映像は互いに近づけてほしい (1 名)、撮影者位置表示から一人称視点映像へのリードが必要 (1 名)、3D モデルの粗さの改善 (1 名)、矢印を意識的に見なくても体の向きが分かる可視化 (1 名)、映像が多い際に何を見るべきかの示唆 (1 名)、映像が多い際の心理的圧迫感の解消 (1 名)、一人称視点映像のラベルが見づらい (1 名)、映像内での人や物体のローカルな位置把握とワークスペースビュー上でのグローバルな位置把握が二度手間 (1 名)、オブジェクトを表示してほしい (1 名)。

7. 実験結果に対する考察

仮説 1: ワークスペースビューの閲覧によって、撮影者の位置情報をより正しくかつ容易に把握することができる – 支持された

タスク群 1 の全てのタスクにおいてタスクの正答率が上昇し、主観難易度が有意に易化した。特に相対位置の把握で大幅な改善が見られた。以上の点から仮説 1 は支持された。

仮説 2: 一人称視点映像を見ている割合が高くなるような場合でも、ワークスペースビューの参照が一人称視点映像の閲覧を妨げない – 支持された

タスク群 2 ではタスク 5 の ROLE2 以外の全ての場合において正答率の悪化傾向は見られなかった。またタスク 5 で役割担当者を決定する項目のみに着目した場合でも ROLE2 以外では正答率の悪化傾向は見られなかった。ROLE2 では役割担当者決定の正答率において有意ではない悪化傾向が見られた。このタスクでは役割担当者本人以外の映像を長い時間閲覧する必要がある。カメラの向き表示を使用しなかった、または使用しても映像中に映る人物との対応付けに役に立たなかった場合に、一人称視点映像のみを見比べて役割担当者の外見を決定してから、ワークスペースビューを参照して該当する人物の ID と場所を決定する、という方針を取った被験者では「誰」であるかを把握する作業が二度手間となって処理時間が不足してしまったと考えられる。実際に正答率が悪化した 3 名の被験者は上記のような閲覧スタイルを取っている。一方で主観難易度では全てのタスクにおいて易化傾向が観察されている。以上の点から仮説 2 は支持された。

注視に基づく閲覧行動の変化

注視に基づく閲覧行動の解析の結果、提案手法では FPV1 と ROLE1 以外では注視点移動量の減少傾向が観察された。さらに、提案手法ではベースラインに比べてウィジェット数が増加したものの遷移回数は FPV2、 FPV3 と ROLE2 では線形増加の範囲以下に収まっており、さらに一人称視点映像同士を見比べる回数が ROLE3 以外では有意に減少している。これらの観測結果は位置決定に関する一人称視点映像の閲覧負荷をワークスペースビューが効果的に軽減している可能性があることを示し、特に多数の一人称視点映像を閲覧する場合に効果的であることを示唆する。

ワークスペースビューの使用方法に関する知見

カメラ位置表示は撮影者位置の決定のほか撮影者位置を起点とした空間の把握等で役に立った。カメラ向き表示は作業場所、協調作業の有無、移動方向、視点の対応等の様々な用途に用いられた一方で、使用方法を見出すことができず意識的に使用しなかった被験者が多数であった。カメラ

の移動軌跡は経路把握だけでなくペア形成や位置の先読み、動きを用いた人物との紐付け等の使用方法が見られた。

可視化手法の改善すべき点

最も多かった改善要望はカメラ位置姿勢の復元精度と一人称視点映像の提示位置に関する項目であった。現在のカメラ位置姿勢の復元では多くのエラーフレームを含み、復元に成功した前後のフレームによる単純線形補間を行なっている。Structure from Motion ベースの手法では [12] で用いられているようなエピポーラ幾何による補間やモーションモデル等を取り入れて、より厳密に補間していく必要があると考えられる。映像の提示位置に関しては、カメラ位置表示との距離やグループを形成している撮影者の映像同士の距離を近づけて視線移動を軽減する必要があり、ワークスペースビュー内に映像を配置するほか、リードや色対応の改善によるスムーズな視線移動の促進等の対策を検討していく必要がある。またワークスペースビューと一人称視点映像との視点の対応がスムーズでなかった点も挙げられる。このことはカメラ向き表示の利用度が低かった点とも密接に関連していると考えられ、向きの意識的な利用の拡大を促す可視化や、あるいはユーザの視線計測を用いて一人称視点映像を閲覧している間にワークスペースビューを回転させて映像との視点を一致させるといった、無意識的に視点の対応をサポートする対策が効果的に働く可能性がある。また本研究では映像の事後閲覧を想定しているものの、反復・一時停止を認めない条件で実験を実施して位置把握への効果を確認した。SLAM やセンサフュージョン等を活用した高速な位置姿勢復元手法を用いることでリアルタイムシステムへの応用にも検討の余地がある。

8. おわりに

本研究では、協調作業を記録した複数一人称視点映像の閲覧において、撮影者の絶対位置や相対位置、およびそれらの時間変化の把握を支援するために、行動空間とウェアラブルカメラの位置姿勢および移動軌跡を可視化したワークスペースビューを提案した。ユーザ評価実験の結果、提案手法の閲覧がより正しくかつ容易な位置把握を支援し、一人称視点映像の内容理解を妨げないことを確認した。またユーザの視線解析の結果、一人称視点映像数が多い場合に、注視点移動量の減少傾向や映像同士の見比べ回数の減少等の閲覧行動の変化も観察された。今後は一人称視点映像とワークスペースビューとの配置関係や視点の対応に関する可視化の改善、ユーザーインターフェースとして実用的な場面でのテスト等を検討していきたい。

謝辞 本研究は JST CREST の助成を受けて行われた。

参考文献

[1] Pirsiavash, H. and Ramanan, D.: Detecting Activities of Daily Living in First-person Camera Views, *CVPR* (2012).
[2] Ishiguro, Y., Mujibiya, A., Miyaki, T. and Rekimoto, J.: Aided Eyes: Eye Activity Sensing for Daily Life, *AH* (2010).
[3] Leung, T.-S. and Medioni, G.: Visual Navigation aid for the blind in dynamic environments, *CVPRW* (2014).
[4] Kera, H., Yonetani, R., Higuchi, K. and Sato, Y.: Discovering Objects of Joint Attention via First-Person Sensing, *CVPRW* (2016).
[5] Higuch, K., Yonetani, R. and Sato, Y.: Can Eye Help You?: Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks, *CHI* (2016).
[6] Kasahara, S. and Rekimoto, J.: Jackin: Integrating first-person view with out-of-body vision generation for human-human augmentation, *AH* (2014).
[7] Kopf, J., Cohen, M. F. and Szeliski, R.: First-person hyper-lapse videos, *ACM TOG*, Vol. 33, No. 4, p. 78 (2014).
[8] Liu, S., Yuan, L., Tan, P. and Sun, J.: Steadyflow: Spatially smooth optical flow for video stabilization, *CVPR* (2014).
[9] Poleg, Y., Halperin, T., Arora, C. and Peleg, S.: EgoSampling: Fast-Forward and Stereo for Egocentric Videos, *CVPR* (2015).
[10] Ghosh, J., Lee, Y. J. and Grauman, K.: Discovering important people and objects for egocentric video summarization, *CVPR* (2012).
[11] Chu, W.-S., Song, Y. and Jaimes, A.: Video co-summarization: Video summarization by visual co-occurrence, *CVPR* (2015).
[12] Arev, I., Park, H. S., Sheikh, Y., Hodgins, J. and Shamir, A.: Automatic editing of footage from multiple social cameras, *ACM TOG*, Vol. 33, No. 4, p. 81 (2014).
[13] Rieffel, E. G., Girgensohn, A., Kimber, D., Chen, T. and Liu, Q.: Geometric tools for multicamera surveillance systems, *ICDSC* (2007).
[14] DeCamp, P., Shaw, G., Kubat, R. and Roy, D.: An immersive system for browsing and visualizing surveillance video, *ACM Multimedia* (2010).
[15] Roth, P. M., Settgast, V., Widhalm, P., Lancelle, M., Birchbauer, J., Brandle, N., Havemann, S. and Bischof, H.: Next-generation 3D visualization for visual surveillance, *AVSS* (2011).
[16] Kasahara, S., Ando, M., Suganuma, K. and Rekimoto, J.: Parallel Eyes: Exploring Human Capability and Behaviors with Paralleled First Person View Sharing, *CHI* (2016).
[17] Park, H. S., Jain, E. and Sheikh, Y.: 3d social saliency from head-mounted cameras, *NIPS* (2012).
[18] Bulling, A., Ward, J. A., Gellersen, H. and Trster, G.: Eye Movement Analysis for Activity Recognition Using Electrooculography, *IEEE TPAMI*, Vol. 33, No. 4, pp. 741–753 (2011).