修士論文

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

2017 年 2 月 3 日

指導教員　佐藤　洋一 教授

情報理工学系研究科電子情報学専攻

48146423　杉田 祐樹

内容梗概

頭部に装着したウェアラブルカメラによる撮影では，カメラを外界に設置することによって生じる行動空間の制約や，カメラを手に持つことによって生じる身体的制約を受けることなく，撮影者自身が主体的に関わる体験を撮影者の一人称視点から記録することができる．この一人称視点映像には撮影者の移動や視点の変化，頭部運動が反映されており，このような特徴を活用することによって主観視点での体験の記録に様々な付加価値を与えることができる．一方で，手術や清掃作業のように複数人が作業場所を共有し，作業者間のインタラクションが頻繁に発生するような協調作業を複数のウェアラブルカメラで記録して，人間がその映像を同時に閲覧する場合，このような一人称視点映像の特徴は各撮影者の絶対位置や相対的位置関係の把握を困難にする．

そこで，本研究では協調作業の様子を記録した複数一人称視点映像の閲覧において，各撮影者の位置に関する情報の把握を支援するインターフェースを提案する．並べて配置した複数の一人称視点映像に加えて，３次元復元した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したビューを提示することによって，撮影者の絶対位置や相対的位置関係の把握を支援する．５つのタスクによるユーザ評価実験の結果，提案ビューを閲覧することによって撮影者の絶対位置や相対的位置関係が正しくかつ容易に把握できることが示された．また複数の一人称視点映像を見比べるタスクや，特定の一人称視点映像を注視している割合が高くなるようなタスクにおいても，提案ビューが撮影者位置の把握に役立ち，かつ提案ビューの参照が映像からの情報の把握に悪影響を及ぼさないことも確認された．評価実験における被験者の注視点解析の結果からは，ウィジェットの遷移回数や一人称視点映像同士の見比べ回数，注視点移動量の減少傾向などの閲覧行動への効果も確認された．さらに，評価実験での被験者へのインタビュー結果を元に，提案インターフェースを改善するための知見も得た．

目次

図目次

表目次

第1章

序論

1.1 本研究の背景

近年，Google Glass や Apple Watch のような軽量安価なウェアラブルデバイスの開発と市場投入が活発に行われ，個人でもウェアラブルデバイスを活用して日々の体験を映像や音声, 移動履歴などの多様な形態で容易に記録することが可能となった. その中でも，ウェアラブルカメラは豊富な情報を有する映像という形態での体験の記録を可能にする. また，動画共有サービスやソーシャルネットワークの普及などの社会的背景の後押しを受けて，ウェアラブルカメラで撮影した映像を投稿して他者と共有することによって，映像記録の共有資産化にも大きく貢献している. このような共有資産としてのウェアラブルカメラの映像記録が新たにどのような付加価値をもたらしうるかを見出し，その新しい価値を効率的に活用するためにどんなシステムをどのように構築するかということが，社会において注目を集める大きな課題の一つである.

ウェアラブルカメラによる撮影では，カメラの設置場所を考慮する必要がなく，また撮影中に両手を自由に使用することができるといった利点を有する. このような利点によって，撮影時の行動可能範囲に関する空間的制約を受けないだけでなく，カメラを手に持つといった撮影行為に伴う身体的な制約を受けることもなく，撮影者自身が主体的に行動に参加している様子を記録することが可能となる. このウェアラブルカメラを頭部に装着して撮影することによって得られる一人称視点映像からは，固定カメラ映像で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が得られるため (図 1.1(a))，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，撮影者が物体の受け渡しや会話・握手といった他者とのインタラクションを行なう際には，その相手の様子が克明に記録される (図 1.1(b)). さらに，一人称視点映像の画面の動きは撮影者の頭部の動きを反映し，また撮影者が手元で作業を行なっている際には，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示唆する.

このような一人称視点映像の持つ特色は多様な分野の研究者の注目を集めており，コンピュータビジョンの技術を用いて撮影者本人の手元の自己動作を認識する研究 [1, 2, 3, 4] や，撮影者の移動経路や行動を予測する研究 [5, 6]，撮影者がインタラクションを行なっている相手の動作を認識する研究 [7] などが近年活発に行われている. また，一人称視点映像をライフログや視覚障碍者のナビゲーションに応用するアプローチ [8, 9] も開拓されている. 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，一人称視点映像を

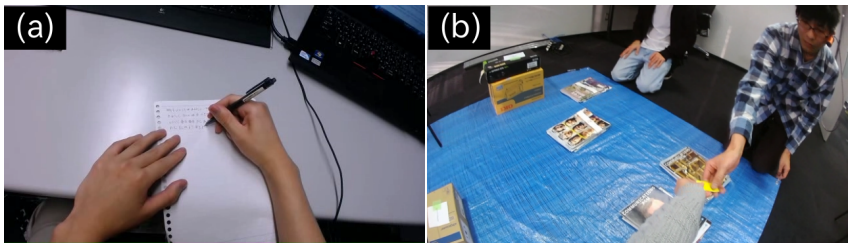

図 1.1: 一人称視点映像の例. (a) 撮影者の手元の様子や, (b) インタラクションの際の相手の様子が克明に記録されている.

用いて撮影者本人と映像に映る他の複数の作業者との間のインタラクションを解析するアプローチ [10] や，複数の作業者全員がウェアラブルカメラを装着するという状況を想定して，複数の一人称視点映像に映る重要な物体や，各撮影者間の相互作用の解析を行なうアプローチ [11, 12, 13]，ヒューマン・コンピュータインタラクション (HCI) において遠隔協調システムへと応用するアプローチ [14, 15, 16] などが多数考案されている．さらに，複数の一人称視点映像の広域監視システムへの応用に関して，日本では 2016 年に大規模な実証試験が成田国際空港や東京国際空港で実施された [1,2].

ここまで述べてきたように，ウェアラブルカメラを用いて撮影した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撮影者の頭部の動きは映像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に関する情報の把握を困難にする. このような閲覧の難しさは，複数の映像を同時に閲覧しようとする際により顕著になる. 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の各所で同時的に，ないし経時的に発生することが考えられる. この際，図 1.2 に示したように (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しているか， (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，などの時間的な変化を逐一把握することが困難となる. 一人称視点映像を共有資産として十分に活用するためには，このような閲覧の難しさを取り除くための可視化手法を備えたシステムの開発が不可欠である.

1.2　本研究の目的とアプローチ

本研究では，複数人による協調作業を記録した一人称視点映像を閲覧する際に生じる難しさ，すなわち上述のような各撮影者の絶対位置や撮影者間の相対的位置関係の把握が難しくなるという問題に着目し，その解決のために，複数の一人称視点映像の効率的な閲覧を支援するユーザインタフェースシステムの構築を目指す．具体的には，タイル状に並べて配置した複数の一人称視点映像に加えて，図 1.3 のように (A) 作業が行われている空間を 3 次元復元して提示し，(B) ウェ

[1] www.naa.jp/jp/press/pdf/20160215-keibinaa.pdf
[2] www.tiat.co.jp/pdf/2016/20161121jp.pdf

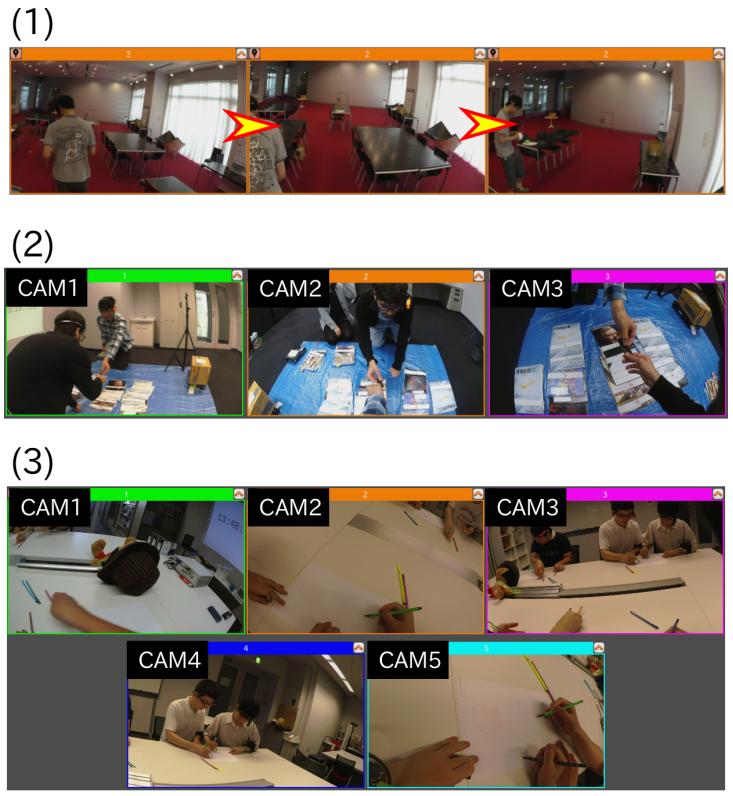

図 1.2: 一人称視点映像からの撮影者位置情報の把握 – (1) 撮影者が多くの机から構成される複雑な空間内を移動しており，移動経路の把握が難しい．(2)3 名の撮影者がビニルシートを囲んで作業を行なっているが，彼らの相対的位置関係の把握は難しい．(3)5 名の撮影者が 2 名ずつペアを組んで協力して 1 枚の紙に絵を描いているが，誰と誰がペアであるかについてグループ分けを行なうことは難しい．

アラブルカメラの位置と (C) 向き，そして (D) カメラの移動軌跡をその上に表示するという可視化手法（ワークスペースビュー）を提案する．

複数の映像を同時に閲覧するためのユーザインターフェースとして，既存のものは主に複数の固定監視カメラ映像の閲覧を想定してデザインされている [17, 18, 19]. それらのインターフェースでは複数の映像がタイル状に並べて配置されており，固定カメラの設置位置や視点方向を記した 2 次元マップ [17] や，行動空間の 3 次元モデル [18] を合わせて提示するものも考案されている．しかしながら，このようなインターフェースで複数一人称視点映像を閲覧する場合，単にタイル状に映像を配置するだけの可視化では行動空間の全体像の把握と撮影者位置の把握がともに困難となり，また，マップを表示する場合においても，行動空間の中を絶えず動き回るウェアラブルカメラの視点位置と方向を閲覧者が把握することが困難であるため，撮影者の位置を把握することは難しい．

これに対して，本研究で提案する可視化手法によって，閲覧者はカメラの位置情報を用いて撮影者の位置や撮影者同士の近接を速やかに把握し，カメラの向きを用いて一人称視点映像の視点とワークスペースビューの視点との対応関係や，撮影者の相対的位置関係や協調行動の有無，作業が行なわれている場所などに関する手掛かりを得ることができると期待される．さらに，カメラの移動軌跡を用いて撮影者の移動経路や撮影者同士の近接・協調行動に関する時間的な手掛か

りを得ることも期待される.

本研究は，複数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取り組みである．そのため，提案するワークスペースビューを閲覧することによって，協調作業における撮影者位置や行動空間への理解がどのように改善されるかを検証するとともに，ワークスペースビューを閲覧者がどのように利用して複数一人称視点映像の閲覧行動がどのように変化したか，また，ワークスペースビューを新たに追加することによって位置情報以外の一人称視点映像に関する内容理解にどのような影響を及ぼすかについて多面的に調査し，今後のインターフェースの開発者のために，本研究で用いた可視化手法をどのように改善するべきかについて明らかにすることも重要である.

本研究では 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すと同時に，ワークスペースビューの参照によって一人称視点映像自体の内容理解が妨げられないことを確認する．さらに，被験者の視線解析とインタビューや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ることを目指す.

1.3　本論文の構成

本論文は次のように構成される．まず第 1 章（本章）では本研究の背景と目的，アプローチについて述べた．第 2 章では一人称視点映像の閲覧負担をどのように緩和するかという問題について，一人称視点映像の編集という観点と複数映像を同時に閲覧するインターフェースという観点からそれぞれ関連研究を紹介する．次に, 第 3 章では提案する可視化手法のデザインとその実装方法について詳細に述べる．第 4 章では提案手法を用いたユーザー評価実験について，仮説や実験タスクの設定，実験手順，実験条件，評価方法に関して詳細に述べる．第 5 章では評価実験の結果をまとめ，得られた知見に関して議論する. 最後に第 6 章で本研究の結論と今後の課題を述べる.

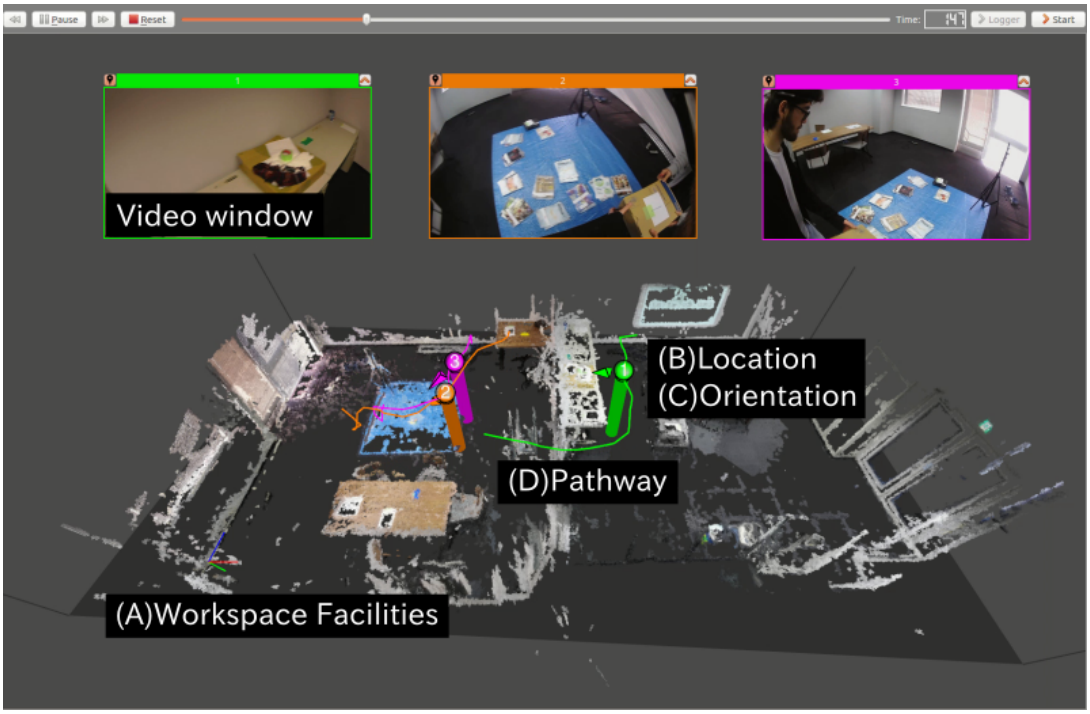

図 1.3: ワークスペースビューによる可視化 – (A) 作業空間の 3 次元復元，(B) 撮影者の位置，(C) 向き，(D) 移動軌跡が表示されている．

第2章

関連研究

本章では，複数一人称視点映像の閲覧支援に関して，一人称視点映像自体に編集を加えて閲覧の負担を軽減するという観点，および複数の映像を同時に閲覧するためのインターフェースという観点からそれぞれ関連研究を紹介する.

2.1 一人称視点映像の編集

一人称視点映像には，撮影者の断続的な頭の動きが多大な映像の動きやブレとして表れている. この映像の動きはしばしば映像中の物体や，映像中で何が行われているかに関する詳細な把握を困難にするため，この動きを滑らかにしたりブレを除去することが一人称視点映像閲覧の負担を軽減するアプローチの1つとして考えられる. 一方で，ウェアラブルカメラで日常生活を長時間記録するような場合，一人称視点映像には撮影者の長時間の移動や休息などといった部分が含まれ，1日の行動変化のログを取ったり，特定の行動記録だけを探すような閲覧目的においては，このような部分は相対的に重要度が低く冗長となる. また，映像が複数となった場合には，協調作業における共同作業場所や共同注視物体などの重要部分は作業に参加している複数の撮影者の映像に映っているため，この共同注視物体の時間変化を把握するという目的においては，協調作業の様子を最も鮮明に映している撮影者の映像以外の重要度は相対的に低くなると考えられる. このように，それぞれの閲覧目的に沿って映像自体の冗長部分を除去したり，複数映像の取捨選択を行なって，1つの短い要約映像として抽出するというアプローチも閲覧負担を軽減するためには重要である.

2.1.1 カメラ運動のスムージング

一人称視点映像の動きは，カメラ運動を補正することによって取り除くことができる. ウェアラブルカメラでは，カメラの微細な運動を物理的に除去する機構を有するスタビライザーにカメラをマウントすることによって，撮影時に画面の微細な動きを抑制することができ，たとえば二輪車等の乗り物に乗っている際にも地面の微細な振動を抑制した安定的な映像を得ることができる. 一方で，すでに撮影された映像の補正はコンピュータビジョンの技術を用いて行われる. たとえば，[20] の手法では2次元のカメラ運動とピクセル単位の移動情報を用いて補正を行なっている. 隣接フレーム間の疎な画像特徴点の対応を用いて画像平面間の2次元の射影変換を計算して [21]，各ピクセルに対してこの射影変換を行なった後に隣接フレームにおけるピクセル移動 (オプティカルフロー) を算出することによって密な移動情報を得る. このピクセル単位のフローのうち移動物

体境界付近に該当する時空間的に不連続なフローを除去してからカメラ運動とピクセル移動の補正を行なうことで，スムーズかつ歪みのない映像を得ることができる．また，[22] では 3 次元のカメラ運動（位置と姿勢）を復元して，それらを元にスムーズな仮想的カメラパスを構築し，パス上の各仮想カメラから見えるべき映像を，そのカメラに近い位置と姿勢を持つ複数のフレームを用いてレンダリングする Hyperlapse と呼ばれる手法を考案した. この手法では，カメラの微細な動きだけでなくより周期の長いカメラ運動をスムージングできるため，一人称視点映像を早送り再生した場合でも観賞に耐える映像を提供することが可能である．他のスムーズな早送り映像を提供する手法としては，一人称視点映像を適応的にサブサンプリングする EgoSampling が [23] によって考案されている. この手法では，密なオプティカルフローを用いてサンプルフレーム間の見た目と速度の連続性，および移動方向 (動きの拡大点: FOE) の連続性を担保しながらフレームを選択する．また [24] では，EgoSampling を類似した方向をとらえた複数一人称視点映像の場合に拡張して，スムーズな早送りパノラマ映像を合成した．このように，カメラの断続的な運動が加わる一人称視点映像では，固定カメラ映像とは異なり短い周期での微細な運動だけでなく，映像を早送りした際に顕在化する長い周期の運動にも対処するアプローチが閲覧負担を軽減する上で重要である.

2.1.2　一人称視点映像の要約

一人称視点映像の要約では，映像を複数のショットと呼ばれる部分区間に分割し，その中から出力する要約映像に用いるショットを様々な基準を用いて選択する．単一の一人称視点映像の要約手法として，複数のショットに登場している物体を介したショットとショットの接続性（ストーリー）に着目したもの [25] と重要な物体や人物に着目したもの [26] を紹介する．[25] の手法では，オプティカルフローを特徴として各フレームを移動中・頭部運動中・静止状態の 3 クラスに識別し，その結果を元にショット分割を行なっている．ショットの選択には，画面のブレの程度や物体がどのように映っているかを元に算出したショットの重要度，外見ベースの画像特徴を元に算出したショット間のシーンの多様性，そして，各ショットにおける各物体の出現頻度を元に算出したショット間の接続性 (ストーリー) を基準として用いている．この手法では，各物体と各ショットをノードとしてそれらの間に双方向結合をもつグラフを構築した際に，ランダムウォークによってあるショットから別のショットへと物体を介して接続する確率をショット間の接続性（ストーリー）として定義し，これを重要視している．また，[26] の手法では，各フレームのカラーヒストグラムを元にショット分割を行ない，各ショットが 1 つのイベントを表しているとみなしてその中で重要な人物や物体の映っている代表フレームを抽出する. この手法では，各フレームを super pixel と呼ばれる領域に分割し，各領域について，手や画面中心との距離やその領域が前後のフレームにどのくらいの頻度で登場するか (一人称特徴)，その領域の外見がどの程度物体らしさ・顔らしさを持ち，領域の周囲と比べて動きがどの程度異なっているか (物体特徴)，領域の大きさやバウンディングボックスの形状 (領域特徴) といった特徴を用いて回帰することによって各フレームの重要度を算出している.

一方で，複数の一人称視点映像の場合には，多くの撮影者の映像に映る物体や人物の重要度が高くなると考えられる．複数の映像間でのシーンの共起関係に着目した手法としては [27] のアプローチが挙げられる. この手法では，大きな色変化を検出したフレームでショット分割を行ない，N 本の映像の各ショットのうち類似した物体や人物を映しているものが同一グループに組み入れられるようにショットのクラスタリングを行なう. この際，各ショットをノードに，各ショットペ

アの特徴ベクトルの類似度をエッジの結合の重みとしてそれぞれ配置した 2 部グラフを構築して，一定以上の類似度を持つエッジの中から重みが最大となるようなエッジを選択しつつ 2 部グラフの切断を行なうという問題に帰着させた. 要約映像に用いるショットは，切断によって得られた各クラスタの各ショットの中から重みスコアの高いものを順番に採用している．また，[28] のアプローチでは，カメラの位置姿勢を 3 次元復元して，それらの情報を用いて複数のカメラの視野範囲や撮影者の視線モデルの重なりとしての共同注視点を算出して，ショット選択の基準の一つとして用いている. この手法では，あらかじめ分割されているショットの中から要約映像に採用するショットを選択するのではなく，各時点での各カメラのフレームを評価しながら要約映像に採用するカメラを切り替えるタイミングを選択する，という形でショットの分割と選択が一意的に行われる. カメラの切り替えの選択は，各カメラの各フレームを全てノードとして縦横に配置したトレリスと呼ばれる構造を持つグラフを作成して，最もコストの低くなるような経路を選択する．この際，カメラの切り替えでは視点が 180 度変わらないようにする，共同注視点に近いカメラを選択するなどといった映像編集現場での知見に基づいたコストを設定している.

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．公共施設などにおける映像監視を行なう部署では，複数の固定監視カメラ映像がタイル状に並列配置され，警備員がその映像を注視するといった光景が一般的である．また，並列配置された固定監視カメラの映像に加えて，2 次元マップや 3 次元モデルを用いて作業空間の全体像を提示したり，マップやモデル上にカメラの設置位置や視点方向を表示することによって，作業空間やカメラ位置の把握を支援する手法も考案されている [17, 29, 19, 18].

一方で，複数の一人称視点映像については，その可視化手法を明示的に評価した事例は我々の知る限りではまだない. そのため，本章では複数一人称視点映像を活用している様々なシステムにおいて，どのような可視化手法が採用されているかに関して幾つかの事例を紹介する.

2.2.1　作業空間と各固定カメラの位置把握を支援するインターフェース

複数の固定監視カメラ映像の閲覧において，作業空間とカメラ位置の把握を支援するインターフェースとして作業空間の 2 次元マップと 3 次元モデルを用いるものを紹介する.
2 次元マップを利用するインターフェースでは，あらかじめ構築されたフロアの白地図にカメラ位置と視点方向が図示されるものが一般的である. このような 2 次元マップを用いたインターフェースとして，フロアマップ上に固定カメラの位置や方向，カバーする視野範囲等を表示して，固定監視カメラをどのように配置するべきかについて，設置計画段階から監視閲覧業務に至るまでを一体的に支援するものが考案されている [1][2].

作業空間の 3 次元モデルを利用するインターフェースでは，あらかじめ構築された 3 次元モデルと設置されたカメラの位置姿勢との位置合わせを行なって，3 次元モデル上に映像を投影したり，映像のシーン解析結果を 3 次元モデル上に反映させるといった可視化が試みられている [29, 19, 18]. 例えば，[29] ではあらかじめ作成した 3 次元モデル上に複数のカメラ映像をリアルタイムで投影し，任意の仮想視点から 3 次元モデルを見ることが可能な閲覧システムを構築した. 多数のカメラ

[1] http://www.tienda24hs.com/HD-IP-camera-video-analysis
[2] https://www.security-camera-warehouse.com/security-camera-software/

を用いて算出した各映像の各ピクセルの奥行き情報を用いることで，各仮想視点からモデルを閲覧する際に，各映像の投影部分のうち遮蔽によって見えないはずの部分を算出して，可視化しないようにするといった工夫を行なっている．また，[19] では，2 次元フロアマップを元にして CAD を用いてあらかじめ作成したフロアの 3 次元モデルの上面に，複数の魚眼カメラの映像を投影して可視化している．この可視化手法では，各部屋の天井に設置された魚眼カメラ映像がマップ状に連結されており，映像内容と各カメラの位置関係，そしてフロアのレイアウトをリアルタイムで同時に把握することが可能である．さらに，[18] では，3 次元モデル上に各固定監視カメラの視野範囲をハイライト表示することで閲覧者が 3 次元モデル上のどの部分に注目するべきかを示唆しているほか，複数の人物の移動軌跡を可視化するなどといった機能も提供している．

一方で，[17] のように，2 次元フロアマップと 3 次元モデルの両方による空間把握の機能を提供しているインターフェースも存在する．このインターフェースでは作業空間を真上から俯瞰する 2 次元フロアマップ上に各カメラの位置と視点方向を可視化したビューと，フロアマップから書き起こした 3 次元モデル上に，選択したカメラの映像を重畳表示するビューを提供している．また，監視カメラの設置計画に関する支援ツールのほか，設置したカメラと 3 次元モデルとのキャリブレーションを容易に行なうことのできるツールも提供しており，映像に映っている構造物と 3 次元モデルの外枠線とを手動で視覚的に位置合わせすることによって，カメラの位置姿勢を容易に求めることができる．さらに，複数の映像にまたがって映っている人物をトラッキングする解析機能も有しており，各映像中にその人物のバウンディングボックスを表示するとともに，その人物の位置と移動軌跡がフロアマップ上にも可視化されている．

2.2.2　複数一人称視点映像を用いたシステムにおける可視化の事例

[30] では，各撮影者のカメラの位置と姿勢，および作業空間の点群を 3 次元復元し，共同注視領域を算出して可視化している．この手法で推定された共同注視領域は，先述した複数一人称視点映像の要約手法においても活用されている．

また，[31] では，同一のパフォーマーについて複数人がそれぞれの手持ちカメラで撮影して得られた複数の映像記録を，後から効率的に見返すための閲覧システムを構築した．このシステムでは背景空間とパフォーマーを別々にモデル化している．背景空間は事前に 3 次元復元され，また各カメラの位置姿勢の変化も事前に計算されている．パフォーマーは空間上のビルボードとしてモデル化されており，閲覧者が一つのカメラから別のカメラの映像へと切り替えることを選択した際に，快適な仮想カメラの経路を推定して，その経路上の各仮想カメラから見えるべきパフォーマーの映像をビルボードに合成して，さらに背景空間とも合成した映像を用いて補間することによって，スムーズで自然な視点移動を実現している．また，パフォーマーを取り囲む全てのカメラを，背景空間の 3 次元点群とともに俯瞰的に見るビューも提供している．

さらに，複数一人称視点映像は遠隔協調作業支援システムへの応用が検討されている．例えば [32] では，互いに遠隔した位置にいる複数の作業者の一人称視点映像を，リアルタイムで各ヘッドマウントディスプレイ上に並べて配置し，遠隔作業者の視線も表示することで，複数人が見ている映像と視線を共有する没入型のテレプレゼンスシステムを開発した．また，[33] では，複数の作業者がウェアラブルカメラを装着して作業の様子を撮影し，作業空間を複数台の距離画像センサで撮影しているという状況において，遠隔地にいる閲覧者がリアルタイムで各一人称視点からの映像に没入し，また各一人称視点映像間を，作業空間の点群や俯瞰視点映像を経由しつつスムー

ズに行き来することのできる Jack-in-Space と呼ばれるテレプレゼンスシステムを開発した.

2.3　関連研究のまとめと本研究の立ち位置

これまでに紹介した関連研究についてまとめ，本研究の立ち位置を明確にする．2.1 節では，単一あるいは複数の一人称視点映像の閲覧負担を，カメラ運動のスムージングや要約といった映像自体の編集によって軽減するアプローチを紹介した．補間や適応サンプリングに基づくカメラ運動のスムージング [22, 23] は早送り再生での映像の安定化を目的とし，微細な画面のブレの除去 [20] は映像の視認性の向上を目的とした処理である (たとえば看板文字を判別できるようにする，など). 本研究では，複数人が作業場所を共有して頻繁にインタラクションを行なっているような場面における各撮影者の位置情報把握の支援を目的としている．そのような場面での撮影者位置把握の困難の要因は，画面の微細なブレではなく撮影者の位置移動そのものや，大きな頭部運動（俯く, 横を向く，など）であると考えられる．また，そのような場面における映像の詳細な内容理解では，不連続な頭部運動や位置変化そのものが，撮影者間のインタラクションを把握する上で重要な手掛かりであると考えられる (並んで歩行中に会話のために突然横を向く，机越しに物を手渡すために突然身を乗り出す，など). そのため，本研究ではカメラ運動のスムージングによる閲覧支援ではなく，複数の映像をそのまま閲覧する場合において各撮影者の位置を把握する難しさに対処する．また，本研究では，協調作業において複数人が共通して注目している物体や共同作業場所だけでなく，それぞれの作業者がどんな役割を担っているか，また，誰と誰が協調して作業を行なっているかなどといった各作業者に焦点を当てた内容理解が伴うような場面での閲覧支援を目的としている．そのため，本研究では各映像の個人性を損なわないような形での閲覧支援を重視し，複数の一人称視点映像を要約編集は行わずにそのまま提供する.

一方で，2.2 節では，複数の映像を閲覧するためのインターフェースという観点から関連研究を紹介した．本研究が想定する複数一人称視点映像を編集せずに同時に閲覧するスタイルは，既に豊富な研究事例の存在する固定監視カメラ映像の閲覧との共通点があるため，固定監視カメラの閲覧システムから代表的な研究事例を幾つか抜粋して紹介した．[29, 19, 18] のような 3 次元モデルによる作業空間の提示は，固定カメラ映像に比べてより作業空間の把握が困難になると考えられる複数の一人称視点映像の場合においても有効であると期待される．また，これらの手法では映像中の人物のトラッキング結果を表示しており，監視システムでは人物の可視化が重要視されていることが窺える. 本研究においても作業者に焦点を当て，その位置情報を可視化することによって，どの作業者がどの場所で誰と一緒に行動しているか，という観点から協調作業の理解を促進することを目指す.

また，複数一人称視点映像の可視化事例に関しても紹介した.[32] では複数の一人称視点映像が HMD 上に同時に提示されている．この事例では没入型の協調作業支援を想定しており，HMD の装着者は自身の一人称視点映像に基づいて自己位置を容易に把握することができる一方で，作業に参加していない第 3 者がこれらの映像を閲覧する場合には全員の位置把握が困難になると考えられる．[30, 31, 33] では，本研究と同様に作業空間と各カメラの位置姿勢を 3 次元復元して可視化している. しかしながら，[30] は共同注視領域の算出過程での可視化であり，[31] は同一の対象を撮影した複数の映像を閲覧するシステムである. また，[33] では遠隔地にいる閲覧者が一度に閲覧する映像は一つであることを前提としている. そのため，複数の作業者（撮影者）に焦点を当てて，その位置関係や作業を行なっている人物，あるいはグループを把握するような，本研究が想

定している場面において，これらの研究で用いられているような可視化手法がどのような効果をもたらすかをユーザ評価実験によって明らかにすることが，本研究の目的の一つである．

第3章

複数一人称視点映像閲覧システム

本章では，本研究で提案する，複数一人称視点映像における撮影者の位置情報把握を支援する閲覧インターフェースについて，まず一人称視点映像をタイル状に配置したベースラインインターフェースとその問題点について述べる．次に，ベースラインで浮き彫りとなった問題点から，撮影者の位置情報把握を支援するためのインターフェースに必要となる機能を洗い出してデザインコンセプトを述べ，最後に閲覧システムの実装の詳細について述べる．閲覧システムの実装の詳細は，複数のウェアラブルカメラの位置推定手法や作業空間の復元手法に関する項目と，可視化手法に関する項目について，それぞれ別の節で述べる．

3.1　ベースラインシステムとその問題点

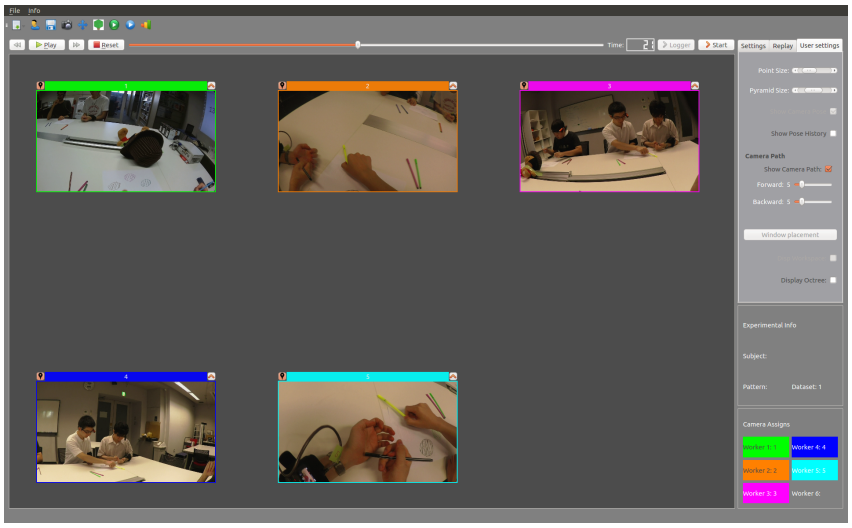

図 3.1: ベースラインシステム．複数の一人称視点映像がタイル状に配置されている．

複数映像を同時に提示するもっとも単純な手法は，映像をタイル状に配置す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 (図 3.1)．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，各カメラの作業空間における絶対位置が変化する．また，撮影者が方向を変えたり物体に接近することによってカメラの視点方向や視野範囲が断続的に変化するため，作業空間の情報は断片的にしか得られない．

このため，閲覧者がドアや机がどこにあるかといった作業空間のレイアウトを理解しながら，同時に各カメラの作業空間における絶対位置を把握することは困難であり，カメラを装着している各撮影者の絶対位置を把握することは難しい. 例えば図 1.2(1) のように複数の撮影者が作業空間内を移動している場面において，撮影者は頻繁に首を左右に振る，俯くなどといった行動を取りながら移動しているため，それぞれの撮影者について，作業空間内のどの場所からどの場所へと，どの経路を通って移動しているかといった，絶対位置の時間的な変化を把握することは難しい.

また，各カメラの位置と視点方向を把握することが困難であるため，複数の撮影者の相対的位置関係を把握することは難しい. 例えば図 1.2(2) のように 3 名の撮影者がブルーシートの周囲に座っている場面において，ある撮影者 1 名の一人称視点映像のみでは隣や向かい側に誰が座っているかを把握することができないため，撮影者 3 名がどのようなフォーメーションを形成しているかを理解するためには 3 つの映像を見比べる必要がある. 閲覧者は映像を見比べながらシーンの背景や各撮影者の顔や身体，衣服などの対応を取って，フォーメーションを推測すると考えられる. そのような状況で，例えば閲覧者が注視している映像の撮影者が手元の作業に没頭するために下を向いた場合には，向かい側に座っている他の撮影者の顔や身体がカメラの視野範囲から外れてしまい，撮影者のフォーメーションを瞬時に把握することを一層困難にする.

さらに，各カメラの近接の有無や視点方向が一致しているかどうかを把握することが困難であるため，協調行動しているグループの存在を逐一把握してその時間変化を追うことは難しい. 例えば図 1.2(3) のように 5 名の撮影者が机を囲んで座り，2 名, 2 名, 1 名の 3 グループに分かれて各グループで 1 枚の紙を共有して絵を描いている場面において，撮影者は下を向いて紙に視線を落としているため周囲の状況はカメラの視野範囲にはほとんど映っていない. また，紙に絵を描くという同一の協調行動が同時に複数発生している. そのような場合に，5 つの一人称視点映像を見比べて手の動きや衣服，あるいはペンの対応を取りながら，5 名を 3 つのグループに分けて，各グループに誰が属しているかを把握することは困難である.

3.2　閲覧システムのデザイン

上述のような問題に対処するために，本研究では 3.2.1 節で述べるような閲覧スタイルを想定して，3.2.2 節で述べるような機能を有する閲覧システムを構築する.

3.2.1　本研究で想定する閲覧スタイルと閲覧支援の内容

本研究では引っ越し作業や清掃作業，手術のように複数人が作業空間を共有して行なうような，また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，その複数の映像記録を，第 3 者があとで同時に再生して見返すような閲覧スタイルを想定する. また閲覧支援では，各撮影者が作業空間内のどこにいて，他の撮影者とどのような位置関係にあって，そして誰と協調行動しているかといった撮影者自身の情報の把握を支援することを目指す.

3.2.2　閲覧システムに実装する機能

- 3 次元での作業空間の把握を支援する機能
  一人称視点映像は撮影者の頭部という視点から撮影されているため，撮影者の位置とともに，一人称視点映像に映っている撮影者の付近に存在する物体やシーン背景の構造物が作業空間

のどの部分に該当するかを容易に把握できるようにするためには，作業空間を真上から見た 2 次元のフロアマップではなく，カメラの視点方向に近い角度で作業空間を側面ないし斜め上から閲覧するような 3 次元モデルの提示が効果的であると考えられる. この際，作業空間内の様々な場所にいる他の撮影者の位置も同時に把握できるようにするためには，各カメラの視点方向に近い角度で，かつ作業空間全体を見渡すような高い視点から 3 次元モデルを提示することが効果的だと考えられる. そのような作業空間の 3 次元モデルによって閲覧者は各一人称視点映像の背景が 3 次元作業空間内のどの部分に対応しているかを容易に把握し，映像に映る人物や物体の絶対位置に関する理解が促進されることが期待される.

また，[17, 29, 19, 18, 31] の事例のように，監視システムや映像をあとで見返すような閲覧スタイルを想定する場合，作業空間や撮影者，物体の 3 次元の情報を保存して再利用できるようにすることが重要であり，事前に構築した静的な 3 次元モデルが使用される. 本研究でもこれらの事例に倣って，再利用性を担保して事前に作成した 3 次元モデルに対して様々な要素を可視化する. さらに，本研究では，[19, 18] のようなフロアマップから書き起こした床と壁のみから構成されたシンプルな 3 次元モデルではなく，一人称視点映像に映るシーン背景との容易な対応を取ることができるように，作業空間のテクスチャを詳細に再現したモデルを構築する.

- カメラの絶対位置を表示する機能

カメラ位置の可視化によって，閲覧者が作業空間における撮影者の位置を速やかに把握することが可能となり，各時点における撮影者の絶対位置や相対的位置関係，そして撮影者同士の近接の速やかな発見が促されると期待される. また本研究では，協調作業において各撮影者がどこに位置し，誰と協調行動しているかといった撮影者自身の情報把握を支援することを目的としているため，各カメラがどの撮影者を示しているかを色とラベルによって明確に区別できるようにする. また，3 次元モデルを斜め上から俯瞰しているため，各カメラと地面との距離感に誤解を与えないようにカメラ位置の地面からの高さを可視化する.

- カメラの視点方向を表示する機能

カメラの視点方向の可視化によって，一人称視点映像が撮影された視点と作業空間の 3 次元モデルを閲覧している視点との速やかな対応を取ることが可能となるほか，近接する撮影者同士の関係性や共通視の位置を速やかに把握する手掛かりとなることが期待される.

- カメラの移動軌跡を表示する機能

カメラの移動軌跡の可視化によって，閲覧者は撮影者の移動経路を速やかに把握することが可能であるほか，撮影者同士の近接を事前に発見する手掛かりとなるなど，撮影者の位置情報に関する時間的な理解を促進することが期待される.

本研究では図 1.3 のようにタイル状に並べて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のモデルを表示し，その上に各カメラに関する情報（位置・向き・移動軌跡）を可視化したビューを提示する. 本研究ではこのビューを**ワークスペースビュー**と呼ぶこととする.

3.3　閲覧システムの実装

本研究で提案する複数一人称視点映像の閲覧システムについて，作業空間の 3 次元復元手法 (3.3.1 節)，ウェアラブルカメラの位置姿勢の復元手法 (3.3.2 節)，それらの可視化手法 (3.3.3 節)，そして閲覧インターフェース (3.3.4 節) のそれぞれに関して実装の詳細を述べる.

3.3.1　作業空間の 3 次元復元

協調作業が行われるに先立って，本研究では協調作業の撮影に使用するものと同様のウェアラブルカメラを用いて事前に作業空間を撮影し，その映像を元にして作業空間の 3 次元モデルを構築する. 空間の 3 次元モデルを作成する方法は多数存在し，点群やメッシュモデル [1]，フロアマップから書き起こした壁面構造にテクスチャを貼り付けるもの [29] など様々な形態でモデルを得ることができる. また，本研究のように静的なモデルを事前に作成するアプローチだけでなく，Microsoft Kinect などの距離画像センサを用いてリアルタイムで作業空間の動的な点群を得るものも存在するが，カメラ位置の把握にはモーションキャプチャや加速度センサといった追加のデバイスが必要となる [33].

本研究では，追加のデバイスが不要である点や，カメラの位置姿勢推定を画像特徴のマッチング情報を用いて容易に行なうことができる点を重視して，ウェアラブルカメラで作業空間を撮影して得られた多数のフレームから Structure-from-Motion (SfM) ベースの手法を用いて 3 次元モデルを構築する.Structure-from-Motion とは，複数の画像の特徴点のマッチング情報を用いて，特徴点群の 3 次元位置 (structure) と各画像を撮影した際の各カメラの位置姿勢 (motion) を同時に復元する手法である [34].

図 3.2 に本研究での 3 次元モデル構築の流れを図示する.3 次元モデル構築の流れとしては，まず前処理として映像からブレの小さなフレームを抽出し，レンズの歪みを補正する. 次に抽出されたフレーム群を用いて作業空間の形状復元を行なう. 形状復元は粗い形状復元と密な形状復元の 2 段階で行なう. 粗い形状復元では Structure-from-Motion (SfM) を用いて，各フレームを撮影した際のカメラの位置姿勢とフレーム中の疎な特徴点の 3 次元位置を同時に推定し，作業空間の粗い点群を作成する. 一方，密な形状復元では，作業空間の粗い点群に対して Multi-View-Stereo (MVS) を用いて作業空間の密な点群を得る．Multi-View-Stereo とは，同一の物体や構造物を写した画像が多数あって，各画像を撮影したカメラの位置関係が既知の場合に，それらの画像中の密な点対応の情報を用いて各画像の密な奥行きマップを作成し，対象物の詳細な形状復元を行なう手法である [35]．この形状復元で得られた密な 3 次元点群は多くのノイズを含むため，後処理としてノイズ除去を実施する. さらに本研究では広い空間を多数のフレームを用いて復元したため得られた点群のサイズが大きく，データサイズの圧縮を行なった. 以下では各段階の処理について詳細に述べる.

1. ブレの小さなフレームの抽出とレンズ歪みの補正

3 次元モデル構築の前処理では，作業空間をスキャンした映像について Kanade-Lucas-Tomasi (KLT) 法 [36] を用いた隣接フレーム間の特徴点のトラッキングを行なってオプティカルフローを計算し，一定フレーム毎にそれらの平均フローのサイズがもっとも小さいものを抽出して，形状復元に用いるフレームとする. ただし，抽出されたフレームのうちフローサイズが 5 ピクセルを超

[1] http://www.meshlab.net/

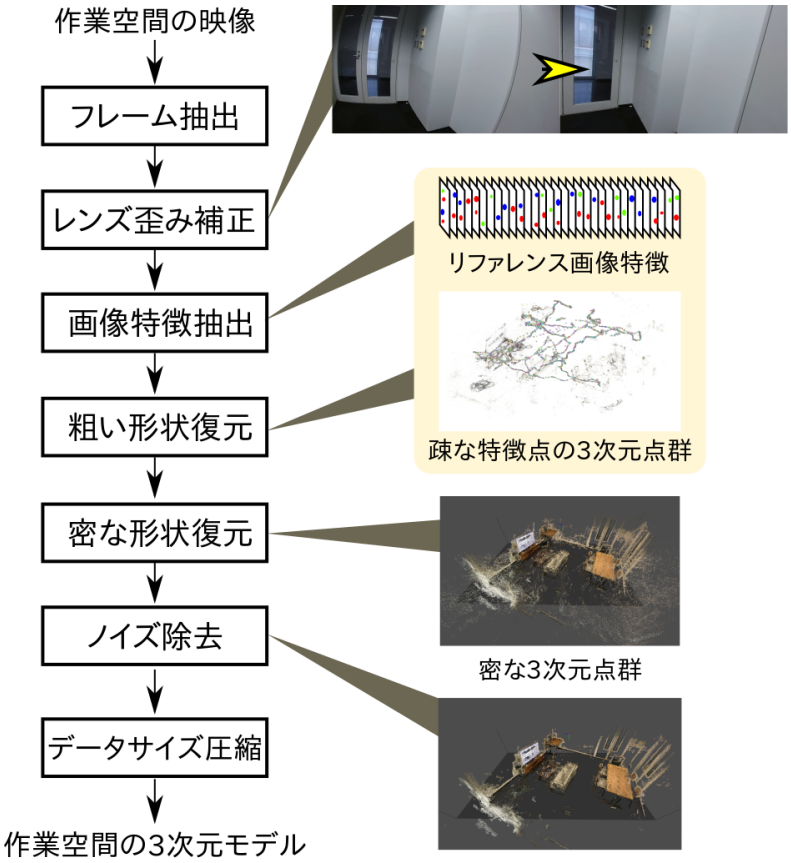

図 3.2: 3 次元モデル構築の流れ．前処理として映像からブレの小さなフレームを抽出してレンズ歪みを補正する．形状復元は Strucure-from-Motion 法によって画像特徴点から粗い 3 次元点群を復元し，Multi-View-Stereo 法によって密な点群を復元するという 2 段階で行われる．最後に後処理としてノイズ除去とデータサイズの圧縮が適用される．

えるものは見た目でブレがあったため除外した．このような処理は，ブレのなるべく少ないフレームを選択して復元に用いることによって画像特徴点の正しい対応を多数得ることを目的としている．また，本研究における作業空間の復元では，3000〜5000 枚のフレームを用いると見た目で十分な品質が確保されたため，作業空間を 60fps で 5 分程度スキャンした映像に対して 5 フレーム (約 80ms) 毎に形状復元に用いるフレームを抽出した．これらの抽出したフレームに対しては，後段の形状復元での推定パラメータ数を減らすために，カメラのキャリブレーションによって事前に得た歪み係数を用いて歪み補正を行なう．これらの処理は OpenCV ライブラリ [2] を用いて実装した．

2. 形状復元

前述の処理で得られたフレーム群 (リファレンス画像) を用いて，粗い形状復元と密な形状復元の 2 段階の処理を行なって作業空間の 3 次元モデルを構築する．以下に示す処理の実装には，全て VisualSFM [3] が提供している機能を用いる．

[2] http://opencv.org/
[3] http://ccwu.me/vsfm/

粗い形状復元

粗い形状復元では，各リファレンス画像から SIFT[37] 特徴点を抽出し，特徴点の 3 次元位置と各画像を撮影した際のカメラの位置姿勢を同時に復元する．このような問題は Structure-from-Motion (SfM) と呼ばれている. 以下では本研究で用いた incremental SfM[38] について簡単に紹介する．この手法では，まず 2 枚の初期画像ペアを選び，特徴マッチングを行なって得られた 2 枚の画像間の特徴点位置の対応情報を用いて，その画像ペアを撮影した際のカメラ位置姿勢の相対運動を 3 次元復元する. この際に一方のカメラを基準世界座標として用いることで 2 つの初期カメラの 3 次元位置と姿勢を決定し，さらに三角測量を用いて初期画像ペアの特徴点の 3 次元位置を得る. 次に，初期復元した特徴点群を映している画像を 1 枚追加し，この画像を撮影した際のカメラの位置と姿勢を計算する.3 次元点群と 2 次元特徴点の対応関係からカメラ位置姿勢を推定する問題は Perspective-n-Point (PnP) 問題と呼ばれている. この新たに推定したカメラ位置姿勢を用いることで，新たな 3 次元点群を特徴点の三角測量によって追加することができ，画像を逐次的に追加していくことによって，全てのカメラ位置姿勢と特徴点の 3 次元点群を得ることができる．また，上述の手順の中で，PnP 問題で推定したカメラ位置姿勢や三角測量で計算した特徴点の 3 次元位置の情報は，各 3 次元特徴点を各画像に再投影した際の二乗誤差の総和を最小化するように同時に最適化する処理が行われる. この処理は Bundle Adjustment[34] と呼ばれ，初期画像ペアの復元や一定数の画像を追加する度に実施される.

一方，本研究における作業空間の 3 次元点群の復元では数千枚のリファレンス画像を取り扱っており，1 枚の画像からは数千個の特徴点が抽出される.incremental SfM の計算量は $O(n^4)$ になることが知られており [39]，画像ペア間の特徴マッチングや Bundle Adjustment を合わせたコストは膨大となる．VisualSFM では，画像特徴点のスケールの大きい順にマッチングを実施し，マッチング率の低い不良ペアの処理を早期に打ち切る [40]，あるいは GPU やマルチコア環境を利用して，特徴抽出や特徴マッチング，Bundle Adjustment における並列処理等を行なう機能 [39] を実装し，画像枚数に対して線形時間での粗い形状復元を実現している.

密な形状復元

密な形状復元では，粗い形状復元で得られた疎な 3 次元特徴点群と各画像を撮影した際のカメラ位置姿勢を元に，点を追加して密な 3 次元点群を作成する. このような問題は Multi-View-Stereo (MVS) と呼ばれている. 以下では本研究で用いた Patch-based Multiview Stereo[41] について簡単に紹介する．この手法では，まず各画像から Harris オペレータや Difference-of-Gaussian オペレータを用いて SIFT よりも密にコーナー特徴点を抽出する．画像ペア間の特徴のマッチングは，エピポーラ幾何と呼ばれる，2 つのカメラ中心間を結ぶ基線と，1 つの 3 次元点からそれぞれの画像を通って各カメラ中心へと向かう直線とで構成される平面の性質を用いて高速に行われる．この特徴マッチングの結果得られた画像特徴点の対応情報を元に，三角測量によって特徴点が 3 次元復元される. この際，特徴点は 3 次元位置と方向を持つ一定サイズのパッチとして復元され，一方の画像から切り出したテクスチャが与えられる. パッチの位置と方向は，パッチを見ることのできる全ての画像にそのパッチを投影した際の正規化相互相関がもっとも高くなるように最適化する処理が行われる.

上述の手順で特徴点対応から初期 3 次元パッチ群を作成したら，次にパッチ群を拡張する処理が行なわれる. 各画像を細かいセルに分割し，各パッチの推定に用いた特徴点が含まれているセルに隣接しているセルに着目して，そのセルが他のパッチの推定にまだ用いられていない場合に，

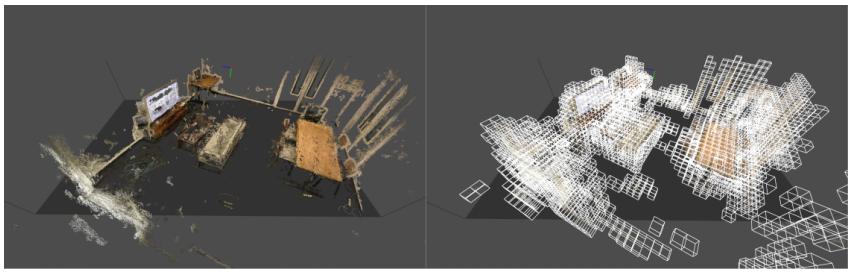

図 3.3: octree によるデータサイズの圧縮. octree では点群全体を内包する立方体を根として, 点が密にあるところに対して再帰的に 8 分割された立方体を当てはめることによって表現する.

そのセルを用いて新たなパッチが作成される. 新たなパッチは，隣接するセルに含まれている特徴点と同一の画像群から見ることができるという仮定の下で，パッチを各画像に投影した際の正規化相互相関が最も高くなるように位置と方向が最適化される．この際，新たなパッチが既存のパッチを遮蔽していたり，また反対に既存のパッチに遮蔽されたりすることのないように不適切なパッチは除去される．このような処理を繰り返すことによって，密な 3 次元パッチ群を得ることができる. 密な 3 次元パッチ群からは詳細な表面の形状を復元することができるが，本研究では，以上の手順で得られた密な 3 次元パッチの中心点群のみを以後の段階で用いる.

3．3 次元点群のノイズ除去とデータサイズの圧縮

Point Cloud Library [4] を用いて，3 次元点群のノイズ除去やデータサイズの圧縮といった後処理を行なう．復元された作業空間の 3 次元点群は，図 3.2 のように多数のノイズ点を含んでいるため，近傍に他の点群が存在しないような疎な点を除去する処理を行なう. また，本研究において密な形状復元で得られた点群はデータサイズが非常に大きいため，octree(図 3.3) 構造を適用して，見た目の質を大きく損なわずにデータサイズの圧縮を行なう.octree では点群全体を内包する立方体を根として，点が密にあるところに対して，再帰的に上位階層の立方体を 8 分割した小さな立方体を当てはめていくことによって微細な構造を表現する.octree 構造を適用することによって，点群と他の物体との衝突を立方体との衝突として定義することも可能となる. データサイズの圧縮は，最も小さな立方体に含まれる点群を，1 つ上位階層の立方体に含まれる 1 点として表すことによって行なう. この際，点群が持つ色情報は平均化される.

最終的に得られた 3 次元点群はスケール不定であり，また点群の座標軸は実際の 3 次元空間と一致していない. 後者は，前段の形状復元において，十分な対応点を持つ良い初期画像ペアを形成しているカメラを世界座標原点と座標軸として用いたためである．本研究では，3 次元点群の重心を原点として，点群に PCA を適用して最小固有値を持つ軸を空間の z 方向ベクトルとして近似することで，実空間に近い座標軸を得た．また，この際に点群のバウンディングボックスも得た.

3.3.2　ウェアラブルカメラの位置姿勢の復元

本節では，協調作業を撮影した各ウェアラブルカメラの位置姿勢を復元する方法を述べる．カメラの位置と姿勢の復元は，事前に背景空間の 3 次元モデルを構築した際に用いたリファレンス画像とその画像特徴点の 3 次元位置の情報 (特徴点の 3 次元地図) を参照して行われる. このよう

[4] http://pointclouds.org/

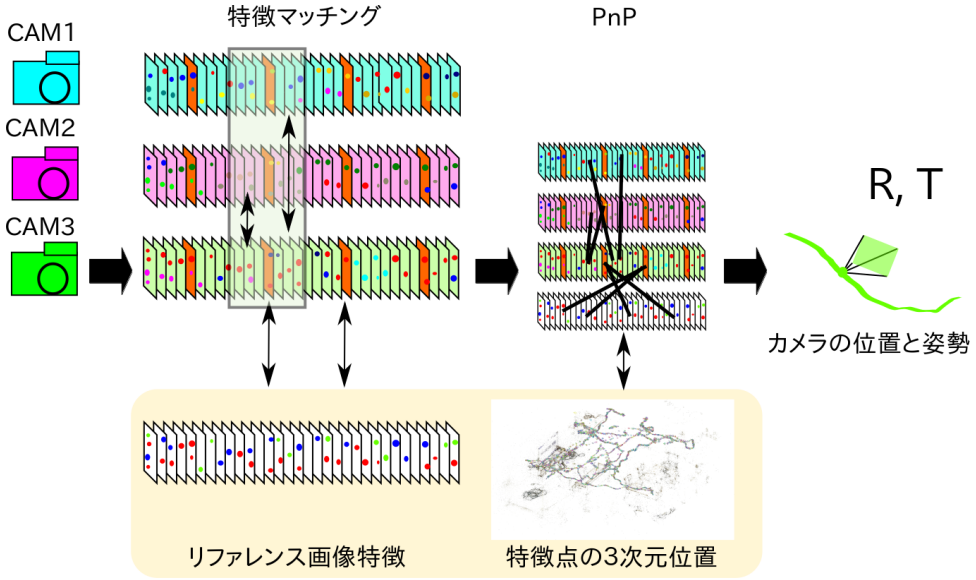

図 3.4: ウェアラブルカメラの位置姿勢の復元の流れ．1 つのカメラについての位置姿勢推定の流れを示す (実際には全てのカメラが同時に最適化される)．各カメラ位置姿勢を推定するにあたって, 淡黄透過色で示すように同一カメラや他のカメラの前後の連続したフレームや, 3 次元モデルの構築に用いたリファレンス画像との特徴マッチングを行なう．この際, 特徴マッチングのコスト削減のためリファレンス画像との特徴マッチングを行なうのは橙色で示した一部のフレームのみとしている．

な SfM ベースで事前に作成した 3 次元地図を用いて単眼一人称カメラの自己位置推定を行なう試みは [42] でなされており，本研究でもこの手法に則る．カメラ位置姿勢の復元も VisualSFM のフレームワークを用いて実装する．ある撮影者のあるフレームにおけるカメラの位置姿勢を復元するにあたって，まず，リファレンス画像や，同一撮影者の映像の前後のフレーム，そして同時点での他の撮影者の映像のフレームとの SIFT 特徴のマッチングを行なう．協調作業の映像では，前景として他の撮影者や操作対象の物体など，作業空間の背景モデルの構築時には存在しなかったものが多数映っている．そのため，少ない背景から作業空間との位置合わせを行なうにあたって，なるべく多くの対応点が得られるように同時点での他の撮影者のフレームや同一撮影者の前後のフレームも用いている．次に，これらの特徴点の対応情報と，リファレンス画像の各画像特徴点に対応付けられている 3 次元位置の情報を用いて，Perspective-n-Point 問題を解いてカメラ位置と姿勢の初期解を得る．このようにして得られた各カメラの位置姿勢は，3.3.1 節で述べたように Bundle Adjustment を用いて最適化することができる．本研究では，特徴マッチングのコストを削減するため，リファレンス画像との特徴マッチングは 20 フレームに 1 回とする．また，同一カメラや他のカメラのフレームとのマッチングは，前後の連続した 41 フレームを対象に 4 フレーム毎に行なった (図 3.4)．これらの数値は厳密なチューニングは行なっていないものの，復元に成功したフレーム数が著しく減らないように，また復元の処理時間が大きく増加しないように試行錯誤した結果，画像マッチングの総数が 100 万程度に収まるように設定した．

上述のカメラ位置姿勢推定では，十分な対応点が得られないことが原因となって多くの復元失敗フレームを含んでいる．そのため，復元に失敗したフレームに対しては，前後の復元に成功したフレームのカメラ位置姿勢情報を用いて線形補間を行なっている．また，点群のバウンディングボックス外に位置推定されたカメラは除外した．

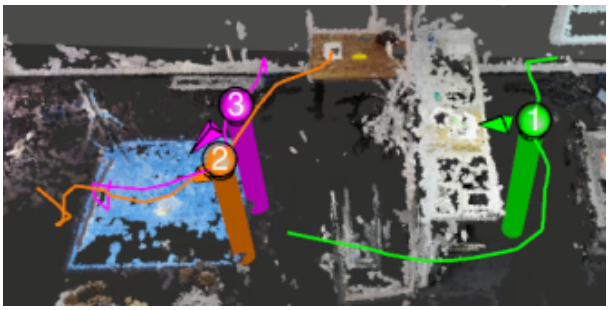

図 3.5: カメラの位置・姿勢・移動軌跡の可視化の様子．図 1.3 の拡大図である．

3.3.3　閲覧システムの可視化

閲覧システムの可視化は図 1.3 に示したように行なう．複数の一人称視点映像を並べて配置し，それに加えて作業空間の 3 次元点群と各カメラの位置・姿勢・移動軌跡を表示する．図 1.3 のうちカメラの位置，姿勢と移動軌跡の可視化の様子を，図 3.5 に拡大して示す．3 次元復元された作業空間の点群は，真上から約 45 度の俯瞰位置から見下ろしたものを常に表示時の視点とする．さらに，最も多くのカメラの方向に近い視点から作業空間を見るようにバウンディングボックスを z 軸周りに 90 度，または 180 度回転させて初期表示視点とする．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，作業空間の床面（バウンディングボックスの底面）からの高さを円柱でそれぞれ表示し，(C) カメラの向きを立体矢印で，そして (D) カメラの移動軌跡を線で表示する．カメラの移動軌跡は現在位置の前後約 5 秒分の長さを初期表示として提示する．これらの各撮影者の可視化要素と各撮影者の一人称視点映像との対応は色と撮影者 ID のラベルによって行なう．例えば，図 1.3 ではワークスペースビューに緑色で表示したカメラの位置と向き，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おり，また窓枠とカメラ位置の両方に撮影者 ID が付与されている．

3.3.4　閲覧インターフェースの実装

本研究では，3.3.3 節で述べたような可視化機能を有する閲覧インターフェースを開発した．本節ではインターフェースの機能や操作方法について詳細に述べる．閲覧インターフェースの外観は図 3.6 のようになっている．
インターフェースは統合 GUI フレームワークである Qt を用いて C++で開発しており，マルチプラットフォームに対応している．また，ワークスペースビューの描画は OpenGL シェーダを用いて高速に行なっているため，閲覧インターフェースはラップトップ [5] 上でも快適に動作する．インターフェースは基本的な動画再生機能を有しており，映像の再生と一時停止，リセット機能のほか早送りと巻き戻し，そしてシークバーによるスクロール機能が実装されている (A)．これらの操作によって，複数の一人称視点映像の再生位置とワークスペースビュー上のカメラの位置や向き，移動軌跡の表示は同時に変化する．それぞれの一人称視点映像に関しては，マウスのドラッグ操作によって映像窓の表示位置と表示サイズを自由に変化させることができる (B)．一方，ワークスペースビューに関しては，マウス操作による基本的な 3 次元操作機能を提供しており，ドラッグ操作による z 軸周りの回転移動と並進移動のほか，ホイール操作によって拡大縮小を行なうこと

[5] ThinkPad X230, CPU: Intel Core i5-3320M 2.60GHz, GPU: Intel HD Graphics 4000, Memory: 8GB, OS: Ubuntu 16.04

ができる (C). 回転移動では，マウスの 2 次元的な移動ベクトルに直行するベクトルを回転軸とした回転を行ない，画面の一端から一端までドラッグした場合に 180 度回転するようになっている. また，ビューの拡大縮小では，点群の点と点との間隔が拡がってモデルの視認性や見た目の質が低下しないように，拡大率に比例して点の表示サイズを拡大している. カメラの移動軌跡は，サイド画面のスクロールバーを操作することによって，前方と後方の軌跡を表示する長さをそれぞれ任意に 0 秒から 30 秒まで変化させることができる (D).

以上が本インターフェースの基本的な機能である. 本インターフェースには，本研究では有用性を検証していない機能を他にも実装されており，たとえば一人称視点映像と同時に撮影者の視線のデータを入力した場合に，視線方向や 3 次元モデルとの交点を表示したり，[30] のように複数人の視野の共通領域を 3 次元で可視化する機能を有する. これらの機能が協調作業の解析にどのように役立つかについての評価や議論は本研究では取り扱わず，今後の検討課題とする.

ロガー機能

本インターフェースでは，閲覧者がどのような閲覧行動を取ったかについての分析を行なうために，閲覧者が行なった全てのマウス操作を記録する機能を有している. また，閲覧者の両眼を計測して画面上のどの位置を見ていたかを 2 次元の座標で与える据え置き型のアイトラッカー [6] と連携しており，閲覧者が画面上のどのウィジェットのどの位置を見ていたかを分析することができる. 閲覧者の視線情報と操作情報のログを用いて閲覧者の操作を再現し，その際の閲覧者の視線を画面上に重畳表示する機能も実装されている.

[6] http://developer.tobii.com/tag/eyex-controller/

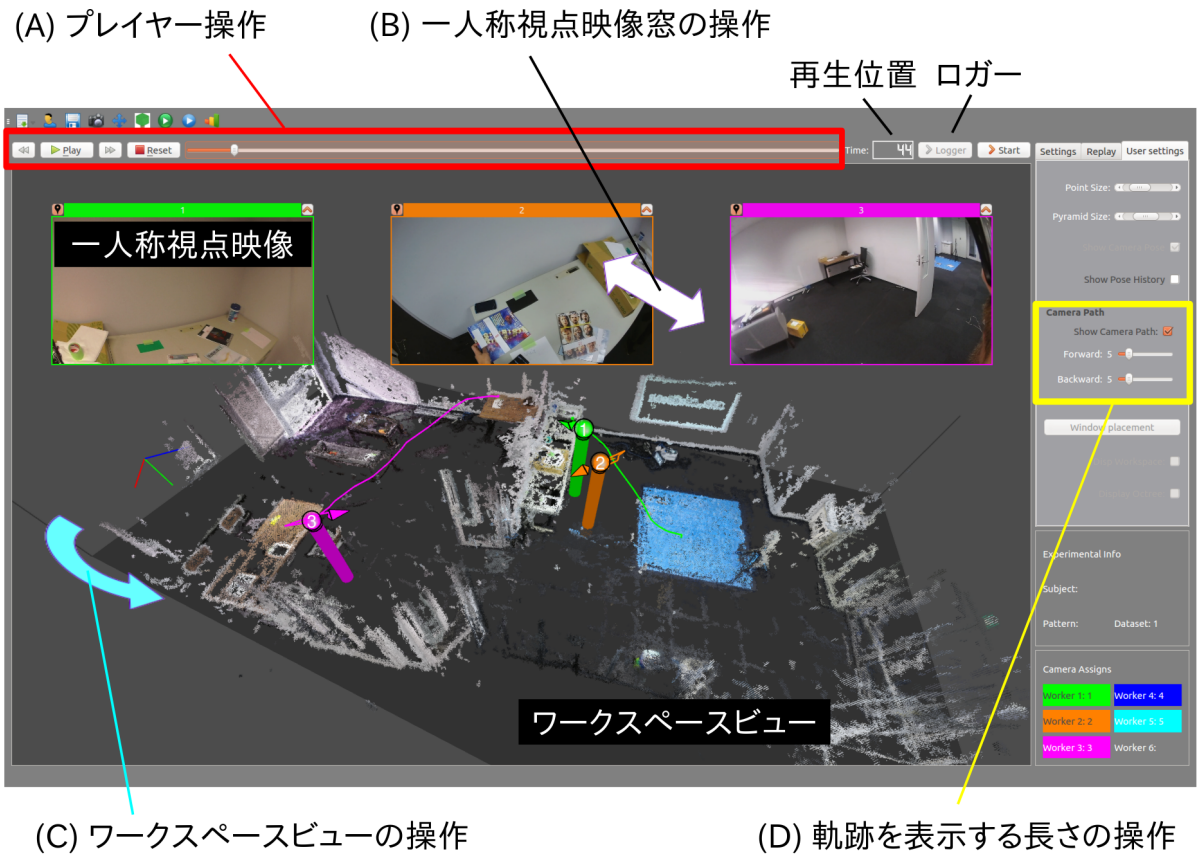

図 3.6: 閲覧インターフェースの外観．インターフェースは (A) 基本的な動画再生機能, (B) 各一人称視点映像の表示サイズや表示位置を変更する機能, (C) ワークスペースビューの回転や拡大縮小, 並進移動を行なう機能, そして (D) カメラの移動軌跡の長さの変更する機能を有する．

第4章

ユーザ評価実験

本章では，本研究で実施したユーザ評価実験について，設定した仮説や被験者に課したタスク，構築した協調作業データセット，実験条件や評価指標，そして被験者へのアンケートやインタビューでの質問項目について詳細を述べる.

このユーザ評価実験の目的は，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，提案するワークスペースビューが撮影者位置情報の効果的な把握を支援するかどうかを確認し，さらにワークスペースビューがもたらした効果を詳細に分析することである. そのため，ユーザインターフェース操作による影響や，被験者個人の閲覧スタイルの差による影響を排除するために，インターフェースで協調作業の映像やワークスペースビューを再生している様子を画面キャプチャして，それを被験者に提示するという型式で行なうこととする.

ユーザ評価実験では，上述の目的のため 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，それらを検証する. 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，一人称視点映像を閲覧している時間割合が高くなって，ワークスペースビューは補助的な閲覧にとどま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)，一人称視点映像同士を見比べる回数が変化するような役割担当者推定タスクに取り組んだりしてもらい (タスク 5)，ユーザに課す負荷を様々に変化させる. そのような場合においても，ワークスペースビューの参照によって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検証する. また，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，ユーザの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.

4.1 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正確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，およびグループの把握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. 各タスクで用いる映像の 1 カットと解答記入の例を図 4.3 に示す.

タスク 1: 撮影者の絶対位置変化の把握

複数の撮影者が作業空間に存在する状況下での特定の 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する (図 4.3(1)). 実験用映像は 12 秒程度の長さであり，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撮影者の移動方向を示唆するカメラの向き，そして前後約 5 秒のカメラの移動軌跡を活用して，撮影者の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する（図 4.3(2)，1 名分が既に記入済）. 実験用映像は 30 秒程度の長さであり 3 名分の一人称視点映像が提示される. 実験用映像には，相対的位置関係が映像の再生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，撮影者の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，その中の指定した 1 つのグループについて全メンバーを白地図に記入する (図 4.3(3)). 実験用映像は 30 秒程度の長さであり 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，それとも映像の再生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用いることでグループの候補となる近接する撮影者の一団を容易に発見することができ，カメラの向きと移動軌跡を手掛かりとして用いることで，実際に協調行動を取っているかどうかを判断することができると考えられる（たとえば位置の近い 2 名の撮影者が一緒に移動している場合，ワークスペースビューからその 2 名を割り出して，2 名分だけの映像に着目して実際に協調しているかどうかを容易に判断できる，と考えられる）. このように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.

4.2　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

撮影者の移動が少なく手元でどのような作業を行なっているかを把握する場合や，3 次元モデルに表示されないような小さな物体を探す場合などにおいては，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ワークスペースビューの参照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かどうかを確認する. 提示する一人称視点映像の数を変化させたり，一人称視点映像同士を見比べる回数が変化するような様々な役割担当者推定タスクに取り組んでもらい (タスク 4, タスク 5)，多面的に検証する.

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際の物体の移動経路を白地図に記入する (図 4.3(4)). 提示する一人称視点映像の数は 1・3・5 と変化させる（FPV1, FPV3, FPV5）.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 3 名または 5 名の中の 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を利用することで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. また，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時間とともに変化する場合には，固定位置に提示されてい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がある. そのような場合においても，ワークスペースビュー上のカメラ位置表示と映像との視線移動は色対応を発見するのみであり，また，1 名の撮影者の位置把握に必要な視線移動回数はワークスペースビューと映像との 1 往復のみである一方で，ベースラインで同様に 1 名の撮影者位置を把握する際の映像の見比べ回数は組み合わせ増加が予想される. このことから，ワークスペースビュー上のカメラ位置表示を参照して撮影者位置を容易に把握できることによる見比べ回数の減少効果は，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，特定の役割を担当している撮影者を把握して白地図に記入する (図 4.3(5)). 実験用映像はそれぞれ 50 秒程度の長さである. 本人の一人称映像のみから役割担当者が決定できる場合 (ROLE1: 雑誌整理作業におけるタグ配布者)，本人以外の一人称映像から役割担当者を決定する必要がある場合 (ROLE2: 清掃作業においてゴミ袋を縛っている人物/机を囲んだ乾杯でのコップの回収者や配布者)，そして，同時に複数の撮影者の役割を決定する必要がある場合 (ROLE3: ブロックの組み立て作業における 2 名の「何もしない」撮影者) の 3 つのバリエーションを設ける.ROLE2 では同時に操作対象物体の位置も記入する. 提示する一人称視点映像の数はそれぞれ 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. 役割把握の最中であっても，映像に映る他の人物が誰であり，どこにいるかについてワークスペースビューのカメラ位置とカメラの向きを参照することによって常に容易に把握することができると考えられ，映像中の人物の ID と位置を把握するための一人称視点映像間の見比べ行動が軽減されると考えられる. また映像から役割担当者を見つけた際にも，その人物が誰でありどこにいるかを容易に把握することができると考えられ，このことによってタスク正答率と主観難易度に悪影響を及ぼさないことが期待される.

	P1	P2	P3	P4	P5
Reference images	5670	3584	4735	1821	1864
Reconstruction time[min.]	244	164	236	225	289

表 4.1: 各作業場所の 3 次元モデルの構築に用いた画像数と復元に要した時間.

4.3　データセットの構築

本研究では 3 名または 5 名で行われる実用的な協調作業の様子を頭に装着したウェアラブルカメラ [1] で記録して，8 種類のデータセット (A-H) を構築した. 各データセットで行われる協調作業は，(A) 作業空間内の複数箇所に配置した箱の梱包と運搬，(B) 作業空間内の複数箇所に設置されたポスターの見回り，(C) 机を囲んだ乾杯，(D)2 人 1 組での 1 枚の紙への描画，(E)2 人 1 組での箱の運搬，(F) 部屋の清掃，(G) ブロックの組み立て，そして，(H) 空間内の各所に散置された複数の雑誌や箱を一箇所に集荷して整理し，決められた場所へと配達する作業である. 各協調作業では複数人による協調行動が頻繁に発生し，また，C・D・G 以外では各撮影者の大きな移動が伴っている. ユーザ評価実験における各タスクではこれらの 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，各データセットに関して，提案インターフェースを用いて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. 各データセットについて，映像の長さ，撮影者の人数や作業場所，カメラの復元に要した時間，および各タスクがどのデータセットから切り出しを行なったかについて図 4.4 に示す. 各データセットはそれぞれ 30–630 秒程度の長さがあり，図 4.1 に示した異なる 5 つの作業場所のうち少なくとも 1 箇所以上で収録された. 作業空間と撮影者のカメラ位置および姿勢の 3 次元復元には，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. 作業場所 P1-P3 とデータセット A-G は脚注 [2]，作業場所 P4-P5 とデータセット H は脚注 [3] の環境で復元した. 作業空間の 3 次元復元に要した時間や，復元に用いた画像数については表 4.1 に示す. 作業空間の 3 次元復元には 164-289 分の時間を要し，協調作業データセットのカメラ位置姿勢の復元には 74–2756 分を要した. また，各協調作業の開始時と終了時には，全撮影者がタイマーが表示されているスクリーンを注視しており，複数の撮影者の一人称視点映像はその部分を用いて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.

4.4　実験の詳細

4.4.1　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）の両方を交互に閲覧してもらう対照比較実験を行なった. 実験開始時には一人称視点映像やワークスペースビューに関する導入を行ない，また本番タスクとは異なる映像を使用した練習問題を出題して，被験者が一人称視点映像やワークスペースビューの閲覧に慣れる時間を設けた. 被験者はタスク群 1(タスク 1-3) またはタスク群 2(タス

[1] Panasonic HXA-100, 60fps, 1280x720
[2] CPU: Intel Core i7-4790K 4.00GHz, GPU: NVIDIA GeForce GTX750, OS: Ubuntu 14.04
[3] CPU: Intel Core i7-2600 3.40GHz, GPU: NVIDIA GeForce GT530, OS: Ubuntu 14.04

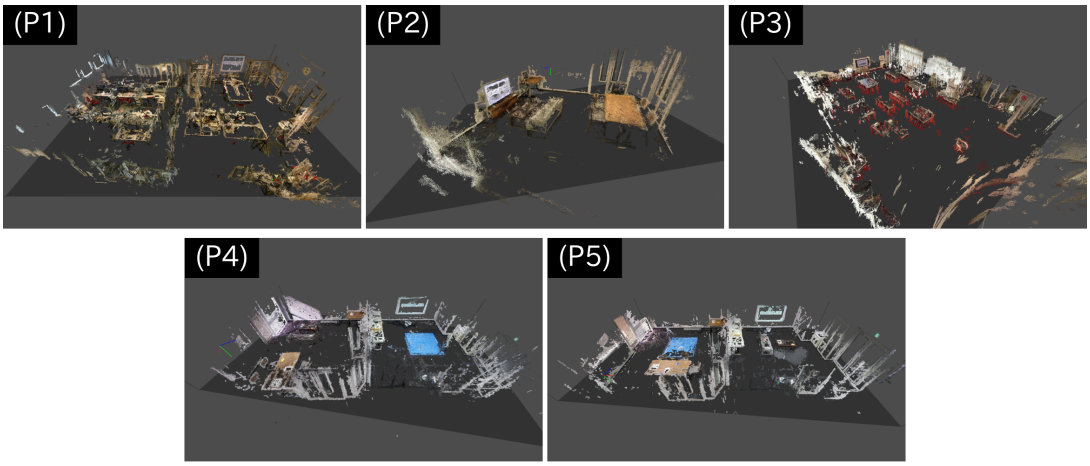

図 4.1: 協調作業を記録した 5 つの作業場所について, それぞれを復元した 3 次元モデルでその外観を示す. (P1)2 部屋で構成されるオフィス, (P2) ダイニングルーム, (P3) 多数の机が配置されたホワイエ（広い空間）, (P4)2 部屋で構成される実験室, (P5)2 部屋で構成される実験室 (P4 のレイアウトを変更したもの) の 5 つの作業場所で協調作業を記録した.

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の手法について連続して 2 種類（タスク 1-3）または 3 種類（タスク 4-5）の映像を閲覧してそれぞれの映像に対して出題される問題に解答し，手法を入れ替えてもう一方を同様に閲覧した. 被験者は各映像の再生終了後にキャンバス上の白地図に問題の解答を記入し，解答提出後にその問題の難易度に関してアンケートに回答した. また，各タスクの解答終了時には，ワークスペースビューにおけるカメラの各可視化要素 (位置・向き・移動軌跡) がそのタスクにおいて役に立ったかどうか（提案手法の場合），または映像中の何に注目して解答を導いたか（ベースライン）に関するアンケートにも回答した. 全タスクの終了時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行なった.

4.4.2　実験環境と実験条件

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 4.2 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（解像度 1920x1200）上に提示され，再生はモニタ上のボタンによって行なわれる. 各映像の再生は 1 回のみとし，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 秒間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. 各タスクに関する問題内容の提示や解答入力も同一のモニタ上にて行なわれる. 解答入力はモニタ上の 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. 被験者は再生ボタンを押す前に問題の内容を熟読して理解し，またキャンバスに表示されている白地図を閲覧して，作業空間のレイアウトに関して事前に十分に理解するように指示されている. また本実

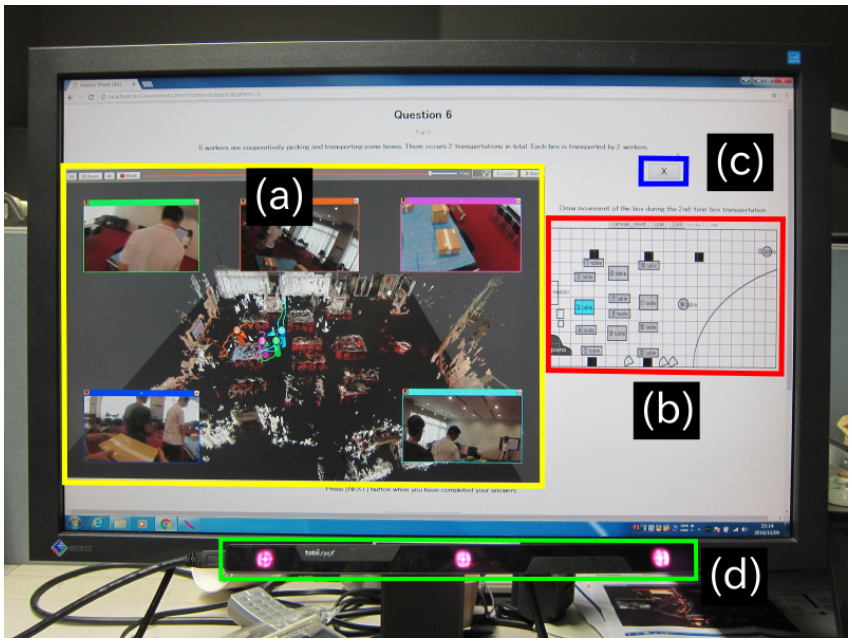

図 4.2: 実験用インターフェース – 24 インチモニタ上に (a) 実験用映像と (b) 解答入力用キャンバスが表示されている．被験者は (c) 再生ボタンで映像を一度だけ再生できる．閲覧中の被験者の視線は (d) アイトラッカで記録される．

験では，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[4] を用いて被験者の視線を約 60Hz で記録した．また実験後に実施したインタビューでは，その様子を録音した．

4.4.3　視線情報の解析

ユーザ評価実験では，被験者の視線情報をアイトラッカで記録し，映像閲覧中の被験者の視線移動について解析する．本実験で用いるアイトラッカでは，被験者の視線情報がモニタ上の視線座標として与えられる．本研究では，この視線座標データを用いて被験者が実験用映像の再生中にどのウィジェットを見ていたかについてのウィジェットの遷移と，画面上での被験者の視線移動量を分析する．しかしながら，本実験で用いた Tobii EyeX Controller で記録した視線データは正確度で $< 0.6°$，精度で $< 0.25°$ の誤差を含んでいる [43]．計測誤差の影響を軽減するため，また閲覧者がウィジェット境界付近を閲覧していた場合に，境界を跨いだ視線のブレが過剰なウィジェット遷移として加算されないようにするため，さらには，実際には「見ずに」視線を通過させただけのウィジェットを「見た」ウィジェットとしてカウントしないようにするために，本研究では注視点ベースで解析を行なうこととする．

人間の視線は，fixation(注視: 固視, 停留とも呼ばれる) と saccade（サッカード）の 2 つの状態に大別される (図 4.5(a))．ある注視点から別の注視点への大きな移動がサッカードとして定義される．注視区間の検出には，[44, 4] らと同様にウェーブレット解析ベースの手法を用いて注視と注視との境界となるサッカードイベントを検出し，100ms を下限とする注視区間を抽出して，各注視区間での視線座標を平均して注視点位置として用いた．2 次元点列として与えられる視線座標データを X, Y 方向に分割して，それぞれについて前後 9 点によるメディアンフィルタ処理を加えた視

[4] http://developer.tobii.com/tag/eyex-controller/

(1) 実験中アンケートでの質問項目	
EQ1	このタスクを遂行する難易度はどうだったか?
EQ2-1	(提案手法使用時　　) カメラの可視化 (位置・向き・移動軌跡) は役に立ったか?
EQ2-2	(ベースライン使用時) このタスクの問題に解答するために何を参照したか?
(2) 実験終了時アンケートでの質問項目	
EQ3	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?
EQ4	ワークスペースビューにおいて 3 次元モデルを閲覧する視点は適切だったか?
(3) インタビューでの質問項目	
IQ1	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?
IQ2	役割推定タスクではどのようにワークスペースビューを利用したか？
IQ3	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／またはワークスペースビューがあることによって却って混乱したか?
IQ4	提示する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？
IQ5	提案手法への改善要望点

表 4.2: アンケートとインタビューでの被験者への質問項目. (1) 実験中アンケートは, 各問題や各タスクにおける各手法 (ベースライン/提案手法) の解答が終了する度に実施される (EQ1–2). (2) 実験終了時アンケートは全てのタスクが終了した際に実施される (EQ3–4). (3) 実験終了後にはインタビューを実施して, 5 項目について質問する (IQ1–5).

線座標データ $s(t)$ に対して,

$$c_b^a(s) = \int s(t)\overline{\psi\left(\frac{t-b}{a}\right)}dt \tag{4.1}$$

で表されるウェーブレット変換を行なった (図 4.5(b)). 本研究ではスケールパラメータを $a = 5$ とした. また，$\psi(\cdot)$ は mother wavelet で，本研究では

$$\psi(x) = \begin{cases} 1 & (0 < x < \frac{1}{2}) \\ -1 & (\frac{1}{2} < x < 1) \end{cases} \tag{4.2}$$

と表される haar mother wavelet を用いた. 変換後には X 方向，Y 方向それぞれに閾値処理 (60 とした) をウェーブレット係数 c_b に対して適用することによって各方向のサッカードイベントを検出した (図 4.5)(c). この注視点データを用いて注視点移動量とウィジェットの遷移回数を算出する. ここで遷移回数をカウントする対象となるウィジェットには，実験用映像中の全ての一人称視点映像とワークスペースビューのほか，解答入力用のキャンバスが含まれる. 注視点がモニタ上の実験用映像からも解答入力用キャンバスからも外れた場合は，その間の注視点は注視点移動総量には含めず，また注視点が外れたことに伴って発生するウィジェット遷移はカウントしない.

4.5　評価方法

ユーザ評価実験では，各タスクでの被験者の正答率や主観難易度を用いた量的評価と，インタビューやアンケートの分析による質的評価の両方を用いる. 本節ではそれぞれの評価指標について，その詳細を述べる.

4.5.1　量的評価

量的評価では，ベースラインと提案手法の両方で得られた対応のある量的指標を用いて 2 手法を比較する.

客観評価

以下に示した採点基準に沿って手動で採点し，各タスクにおける正答率を算出する.

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: 白地図上に既に記入されている 1 名から見た他の 2 名の方向が 8 方位区分で全て正しい場合にのみ 1 点を与える（この項目では ID の正誤は問わない）. また，既に記入されている 1 名から時計回りに見て他の 2 名の ID の並び順序が正しい場合に 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・椅子など）とする.

タスク 4: タスク 1 と同様の基準を用いて採点する.

タスク 5: ROLE1 では役割を担当する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，撮影者位置から見た操作物体の方向が 8 方位区分で正しい場合に 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

各タスクにおける各問題の終了時に，被験者に主観的な難易度を「難しい」を 1,「易しい」を 7 とした 7 段階で回答してもらい，それらを主観評価指標として用いる.

4.5.2　質的評価

質的評価では，実験中や実験終了以降に被験者に実施したアンケートやインタビューを元に，提案手法が各タスクを遂行する上でどの要素がどのように役に立ったか，あるいはどの要素が役に立たず困惑する原因になったか，さらに提案手法の課題は何であるかやユーザの閲覧行動がどのように変化したか，などを多面的に分析する. アンケートおよびインタビューで被験者に質問した各項目については表 4.2 に示す. アンケート項目のうち EQ1 の主観難易度は，ベースラインと提案手法とで対応のある量的指標として，上述した量的評価における主観評価指標として使用する.EQ2-1 では提案手法を使用している場合に，カメラの各可視化要素がどの程度役に立ったかについて,「役に立たなかった」を 1，「役に立った」を 7 とした 7 段階で回答してもらった.EQ2-2 ではベースラインを使用している場合に，映像中の何を参照して解答を導いたかについて，(1) 撮影者の手の外見，(2) 顔，(3) 衣服，(4) 手や頭の動き，(5) 手元で操作するような小さな物体，(6) 机や椅子などの大きな物体，(7) 撮影者の場所移動（足元の位置の移動），および (8) 壁や天井な

どの部屋の構造物の 8 項目の中から該当するもの全てにチェックをつけてもらった.8 項目以外のものを参照した場合は自由記述で回答欄に記入してもらった.EQ3 と EQ4 では，実験終了時にワークスペースビューによる可視化やワークスペースビューを見る視点が適切だったかどうかについて,「悪い」を 1,「良い」を 7 とした 7 段階でそれぞれ回答してもらった. インタビューでは IQ1–5 の 5 項目について各被験者に質問した.

図 4.3: ユーザ評価実験で被験者に課したタスク一覧

	A	B	C	D
				
	箱の梱包と 2人1組での運搬	ポスターの見回り	机を囲んだ乾杯	2人1組での描画
撮影者数(人)	5	5	5	5
映像の長さ(秒)	A1: 134	B1: 80	C1: 44 C2: 57	D1: 41 D2: 41
作業場所	A1: (P3)	B1: (P3)	C1: (P1) C2: (P2)	D1: (P1) D2: (P2)
復元時間(分)	A1: 280	B1: 180	C1: 160 C2: 88	D1: 143 D2: 74
使用タスク	TASK1, TASK4	TASK1	TASK2	TASK3

	E	F	G	H
				
	2人1組での 箱の運搬	ブロックの組み立て	部屋の清掃	箱と雑誌の集配と 整理
撮影者数(人)	5	5	5	3
映像の長さ(秒)	E1: 32	F1: 65 F2: 64	G1: 177	H1: 632 H2: 567
作業場所	E1: (P1)	F1: (P1) F2: (P3)	G1: (P1)	H1: (P4) H2: (P5)
復元時間(分)	E1: 93	F1: 288 F2: 229	G1: 330	H1: 2576 H2: 1871
使用タスク	TASK3	TASK5	TASK3, TASK5	TASK2, TASK4, TASK5

図 4.4: 本研究で用いた協調作業データセットの詳細 – 本研究で構築した A-H の 8 種類の協調作業データセットにおける, 撮影者人数, 映像の長さ, 作業場所, カメラの復元に要した時間, および各データセットを使用したタスク.　各データセットは図 4.1 に示した (P1-5) の作業場所のうち一箇所以上で収録された.

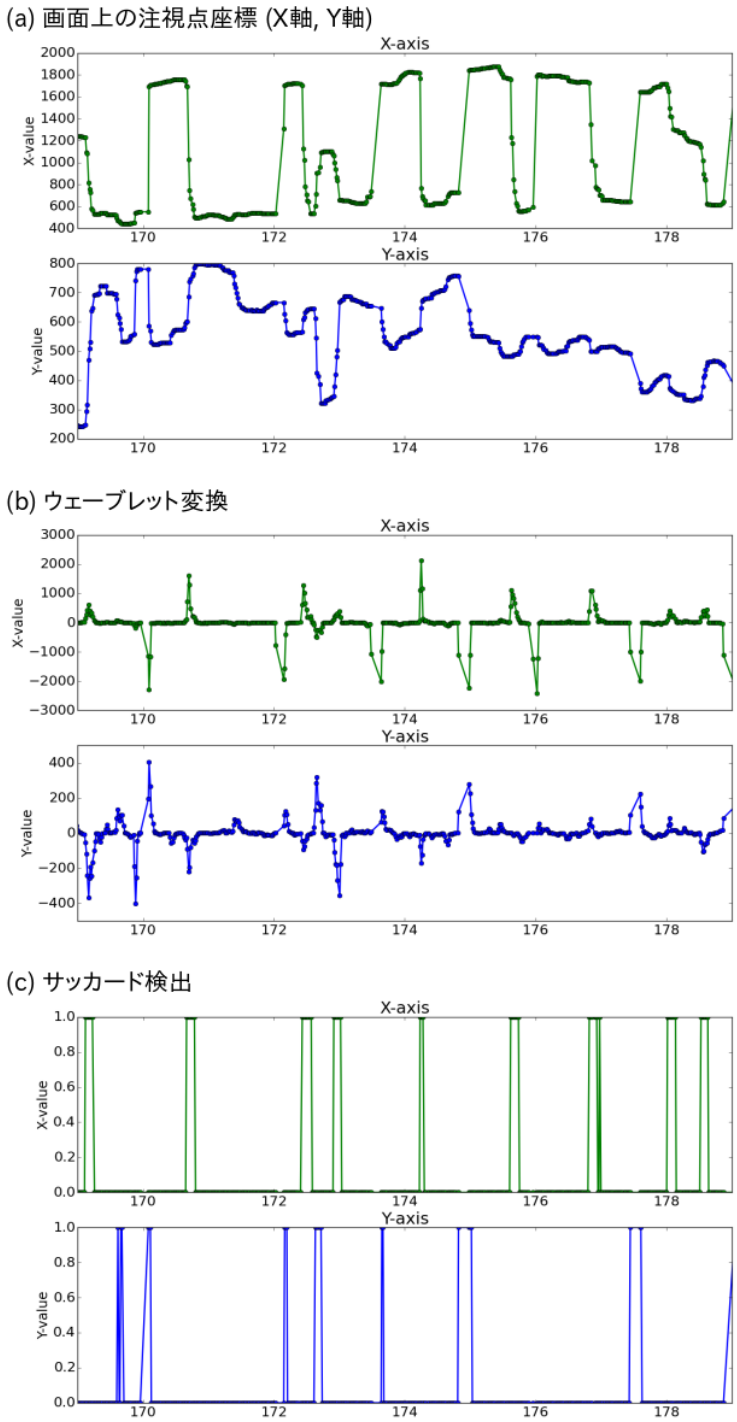

図 4.5: ウェーブレット解析によるサッカードの検出. ある被験者の映像閲覧時の視線データのうち約 10 秒分を取り出して図示する. (a)X 方向, Y 方向それぞれの視線座標. 不連続にジャンプしている部分がサッカードに該当する. (b) ウェーブレット変換後の視線データ. (c)(b) に閾値処理を加えて各方向のサッカードイベントを検出した結果.

第5章

ユーザ評価実験の結果と考察

5.1 ユーザ評価実験の結果

本節では，ユーザ評価実験の結果について詳細に述べる. 以降では，まず各タスクについての量的評価結果とアンケートによる質的評価結果を述べ，次に実験終了時アンケートと視線解析の結果，そしてインタビューの結果を順に述べる.

5.1.1 タスク 1-3 の結果

量的評価結果

各タスクの正答率 (0-1) と主観難易度 (1-7) をそれぞれ図 5.1(a) 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を確認した（それぞれ $p < .05$, $p < .001$）. 主観難易度は，タスク 1・2・3 のいずれにおいても提案手法では有意に易化することを確認した (それぞれ $p < .01$, $p < .001$, $p < .01$). また，各タスクにおける採点項目ごとの正答率もそれぞれ算出した (図 5.2(1–3)). タスク 1 では，撮影者の移動経路のうち移動開始点の正答率が提案手法では有意に上昇した ($p < .05$). タスク 2 では，白地図に事前に記入された 1 名の撮影者から見た，他の 2 名の撮影者 ID の並び順序を特定する項目と，事前記入済の 1 名から見た他の 2 名の相対位置を特定する項目の両方で有意に正答率が上昇することを確認した (それぞれ $p < .01$, $p < .05$).

ベースラインにおけるアンケート結果

タスク 1-3 のベースラインにおいて，被験者が問題に解答するために参考にしたものについてのアンケート (EQ2-2) の結果を，表 5.1 上段に示す. ここでは 4 名以上の撮影者が用いたものを列挙する. タスク 1 では被験者 8 名中 8 名が (6) 作業空間に静置されている机や椅子などの大きな物体を利用し，7 名が (8) 作業空間の壁や窓等の背景情報を利用して問題に解答した. タスク 2 では 8 名中 6 名が (2) 撮影者の顔または (6) 作業空間に静置されている大きな物体を利用した. タスク 3 では 8 名のうち 8 名が (6) 作業空間に静置されている大きな物体を利用した. また，5 名が (5) 撮影者が手元で操作している物体を利用し，各 4 名が (2) 撮影者の顔や (3) 衣服，(4) 手や頭の動きを利用した.

	(1)Hands	(2)Faces	(3)Clothes	(4)Head or Hand movements	(5)Small objects	(6)Scene objects	(7)Location movements	(8)Scene structures
TASK1	0	0	0	2	0	**8**	1	**7**
TASK2	1	**6**	3	3	0	**6**	2	3
TASK3	2	4	4	4	**5**	**8**	0	2
TASK4								
FPV1	0	0	0	3	1	**6**	2	**6**
FPV3	1	0	0	3	3	**6**	3	**5**
FPV5	1	0	1	2	3	**5**	**4**	**4**
TASK5								
ROLE1	1	0	0	3	3	**5**	0	**4**
ROLE2	2	2	1	3	**4**	**5**	1	3
ROLE3	2	2	0	**4**	2	**6**	0	**4**

表 5.1: ベースラインにおいて被験者が用いた手掛かり – (1) 手の外見, (2) 顔, (3) 衣服, (4) 手や頭の動き, (5) 手元で操作するような小さな物体, (6) 机や椅子などの大きな物体, (7) 撮影者の場所移動 (足元の位置移動), および (8) 壁や天井などの部屋の構造物の 8 項目について, 問題を解答するために参考にした, とアンケート (EQ2-2) でチェックをつけた被験者の人数をそれぞれ示す. 各タスクについて人数が多かった上位 2 項目を太字で示す.

提案手法におけるアンケート結果

タスク 1-3 の提案手法において，ワークスペースビュー上のカメラの各可視化要素（位置・向き・移動軌跡）が各問題に解答するにあたって役に立ったかどうかについて，アンケート (EQ2-1) の結果を図 5.3(a) に示す.

5.1.2　タスク 4 の結果

量的評価結果

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 5.1(b) に示す. タスク全体の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，および FPV3 での提案手法において全被験者が満点となった. 一方，項目ごとの正答率（図 5.2(4)）では，運搬物体の移動終点の正答率が FPV3 の場合に提案手法において有意に上昇した ($p < .05$). また，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 ($p < .05$). さらに，ベースラインと提案手法のそれぞれにおいて，一人称視点映像数を変化させたことが正答率や難易度に影響したかどうかを検証するために Friedman 検定を実施した結果，ベースラインでの正答率と難易度，および提案手法での正答率と難易度の全ての場合において有意となった (それぞれ $p < .05$, $p < .01$, $p < .05$, $p < .01$). 全てのペアに対して Wilcoxon の符号順位検定と Bonferroni 調整による多重比較を行なった結果，ベースラインの難易度において FPV1 と FPV5, FPV3 と FPV5 との間に有意差が

確認された (それぞれ $p < .1$, $p < .1$). また，提案手法の難易度においても FPV1 と FPV5 との間に有意差が確認された ($p < .1$).

ベースラインにおけるアンケート結果

FPV1,FPV3,FPV5 のそれぞれについて，ベースラインにおいて被験者が問題に解答するために参考にしたものについて尋ねたアンケート (EQ2-2) の結果を表 5.1 に示す．4 名以上の被験者が利用したと回答した項目を列挙する．FPV1 では被験者 8 名中 6 名が (6) 作業空間に静置されている机や椅子などの大きな物体，または (8) 作業空間の壁や窓，ドアなどの背景情報を利用した．FPV3 では 8 名中 6 名が (6) 作業空間に静置されている机や椅子などの大きな物体，5 名が (8) 作業空間の壁や窓，ドアなどの背景情報を利用した．FPV5 では 8 名中 5 名が (6) 作業空間に静置されている机や椅子などの大きな物体，4 名が (7) 撮影者の場所移動または (8) 作業空間の壁や窓，ドアなどの背景情報を利用した．

提案手法におけるアンケート結果

FPV1,3,5 のそれぞれについて，提案手法においてワークスペースビュー上のカメラの各可視化要素（位置・向き・移動軌跡）が問題の解答に役立ったかどうかについて質問したアンケート (EQ2-1) の結果を図 5.3(b) に示す．

5.1.3　タスク 5

量的評価結果

ROLE1-3 のそれぞれにおける正答率および主観難易度を図 5.1(c) に示す．また，役割担当者を特定する項目の正答率を図 5.2(5–左) に，役割担当者や物体の位置を把握する項目の正答率を図 5.2(5–右) にそれぞれ示す．役割担当者特定の正答率では ROLE1-3 の全ての場合に有意差は確認されなかった．一方，位置把握の正答率では提案手法において ROLE3 の場合に有意に上昇し ($p < .1$)，主観難易度では ROLE3 の場合に提案手法において有意に易化した ($p < .05$).

ベースラインにおけるアンケート結果

ROLE1-3 のそれぞれについて，ベースラインにおいて被験者が問題に解答するために参考にしたものについて質問したアンケート (EQ2-2) の結果を表 5.1 に示す．4 名以上の被験者が利用したと回答した項目を列挙する．ROLE1 は被験者 8 名中 5 名が (6) 作業空間に静置されている机や椅子などの大きな物体，4 名が (8) 作業空間の壁や窓，ドアなどの背景情報を利用した．ROLE2 では 8 名中 5 名が (6) 作業空間に静置されている机や椅子などの大きな物体，4 名が (5) 撮影者が手元で操作している物体を利用した．ROLE3 では 8 名中 6 名が (6) 作業空間に静置されている机や椅子などの大きな物体，4 名が (4) 撮影者の頭や体の動き，または (8) 作業空間の壁や窓，ドアなどの背景情報を利用した．

提案手法におけるアンケート結果

ROLE1-3 のそれぞれについて，提案手法においてワークスペースビュー上のカメラの各可視化要素（位置・向き・移動軌跡）が問題の解答に役立ったかどうかについて質問したアンケート (EQ2-1) の結果を図 5.3(c) に示す．

5.1.4　実験終了時アンケートの結果

実験終了時アンケートでは，(EQ3) 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，(EQ4) 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どうかを最も不適切を 1 とした 1-7 の 7 段階で尋ねた.EQ3 では 6.3 ± 0.8，EQ4 では 6.4 ± 0.7 という結果を得た.

5.1.5　被験者の視線解析の結果

アイトラッカで記録した被験者の視線解析の結果について，実験用映像の再生中に注視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を表 5.3 に示す. また，全タスクについて注視点をもとに算出した単位時間あたりの注視点移動量とウィジェットの遷移回数を図 5.2 にそれぞれ示す．ウィジェットの遷移回数は一人称視点映像間の遷移 (2-1), ワークスペースビューへの遷移 (2-2), キャンバスへの遷移 (2-3) の各項目についてもそれぞれ算出している．アイトラッカで視線を記録した被験者はタスク 1-3 が 4 名，タスク 4,5 が 7 名である.

5.1.6　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 名から次の 5 つの質問 (質問 2 と 4 は 8 名) に関する回答を得た. 以下では得られた回答を集約する.

IQ1 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？

撮影者位置:　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).
撮影者の向き:　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.
撮影者の移動軌跡:　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).

IQ2 役割担当者決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.

IQ3 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.
却って混乱を招いた場面: 特になし (6 名)，カメラ位置推定の誤り (3 名)，カメラ向き推定の誤り (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).

IQ4 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大して作業量が増大 (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作業量が増大 (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).

IQ5 提案手法への改善要望点

カメラ位置・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).

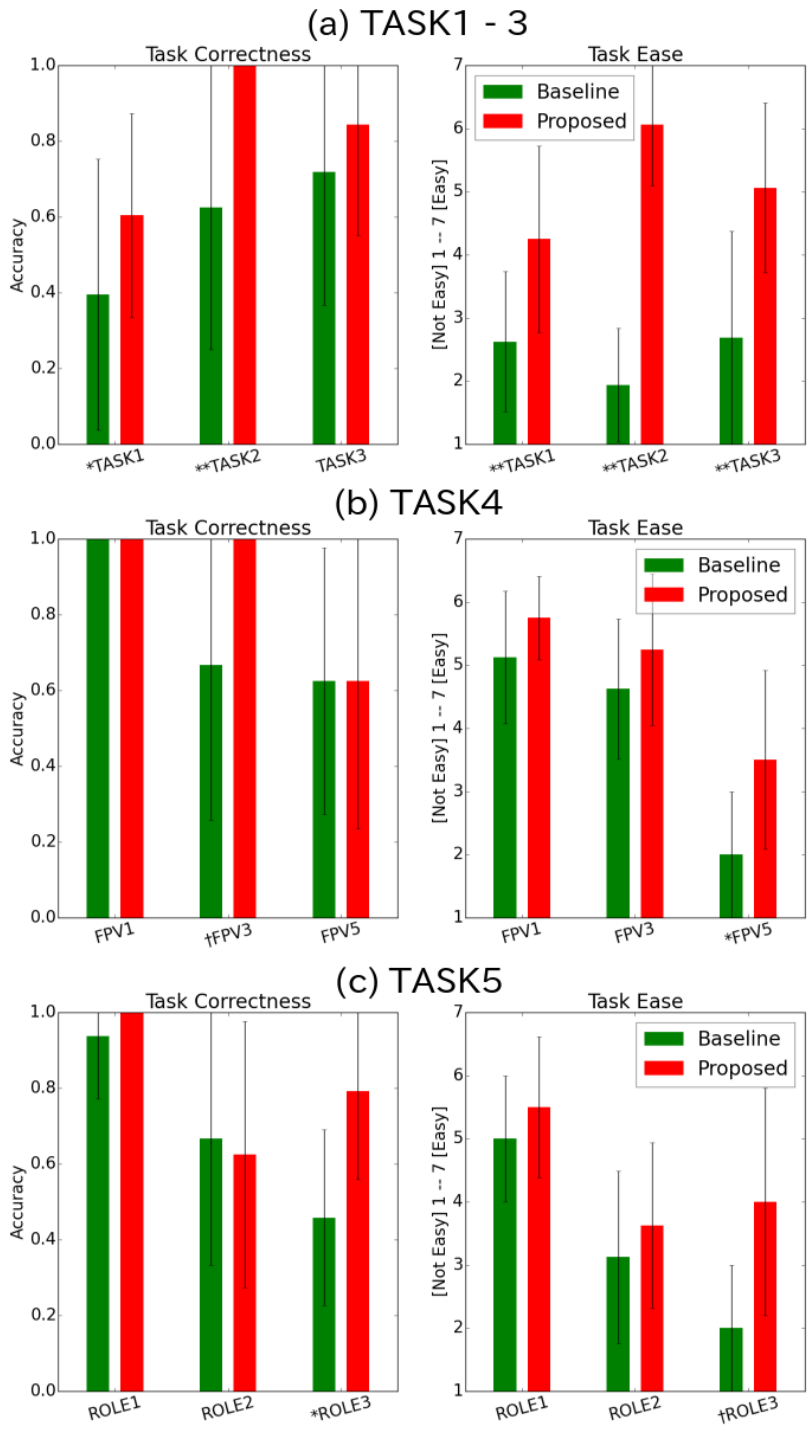

図 5.1: 評価実験における各タスクの正答率と難易度. (a) タスク 1-3, (b) タスク 4, (c) タスク 5 の結果をそれぞれ示す. (†, ∗, ∗∗) はそれぞれ有意水準 (.1, .05, .01) を示す.

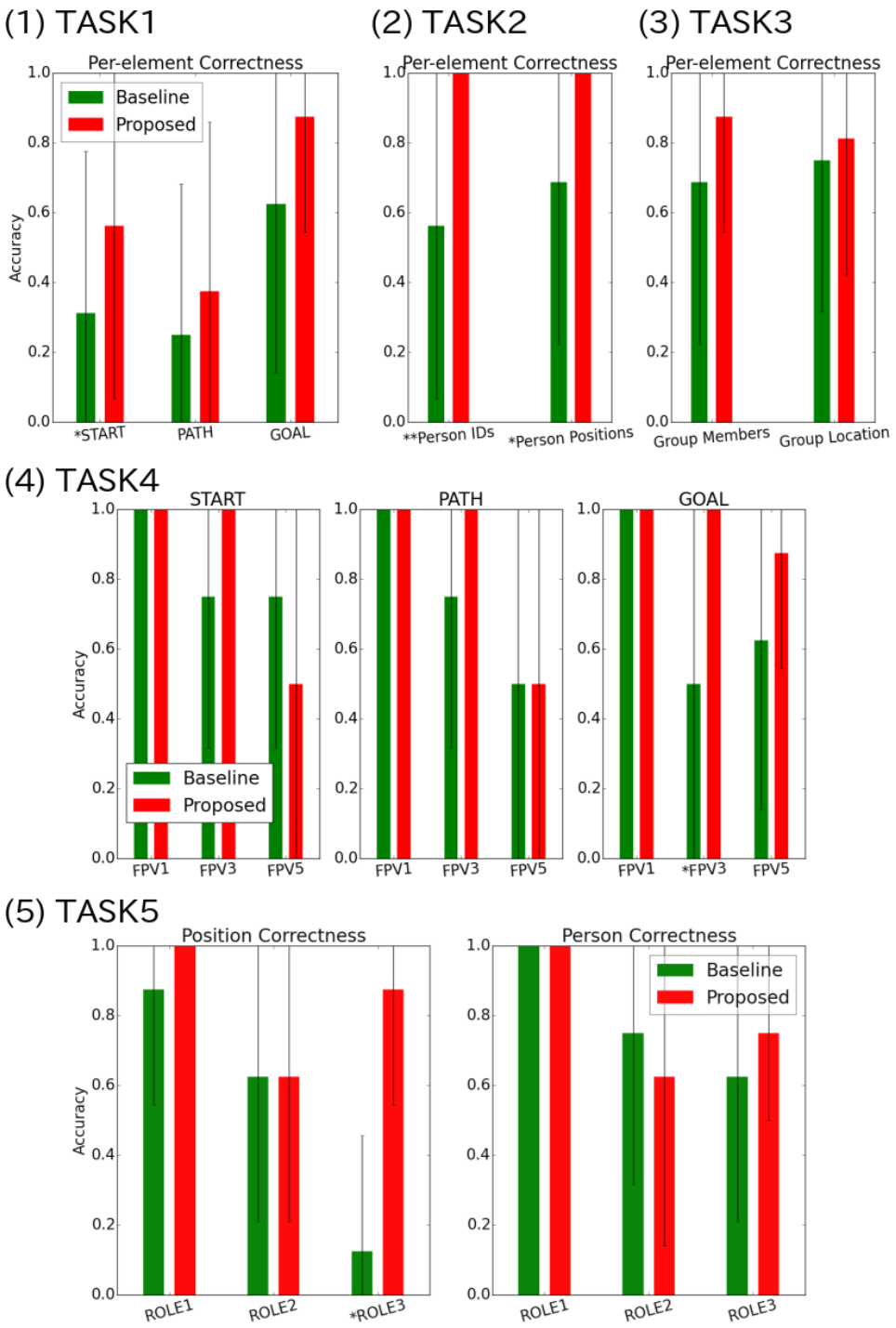

図 5.2: 各タスクにおける採点項目ごとの正答率. (1) タスク 1 における始点・途中経路・終点の正答率, (2) タスク 2 における撮影者 ID の並び順序特定と相対位置特定の正答率, (3) タスク 3 におけるグループメンバーの ID 特定と協調作業場所特定の正答率, (4) タスク 4 における始点・途中経路・終点の正答率, (5) タスク 5 における役割担当者の ID 特定と位置特定の正答率をそれぞれ示す. (†, ∗, ∗∗) はそれぞれ有意水準 (.1, .05, .01) を示す.

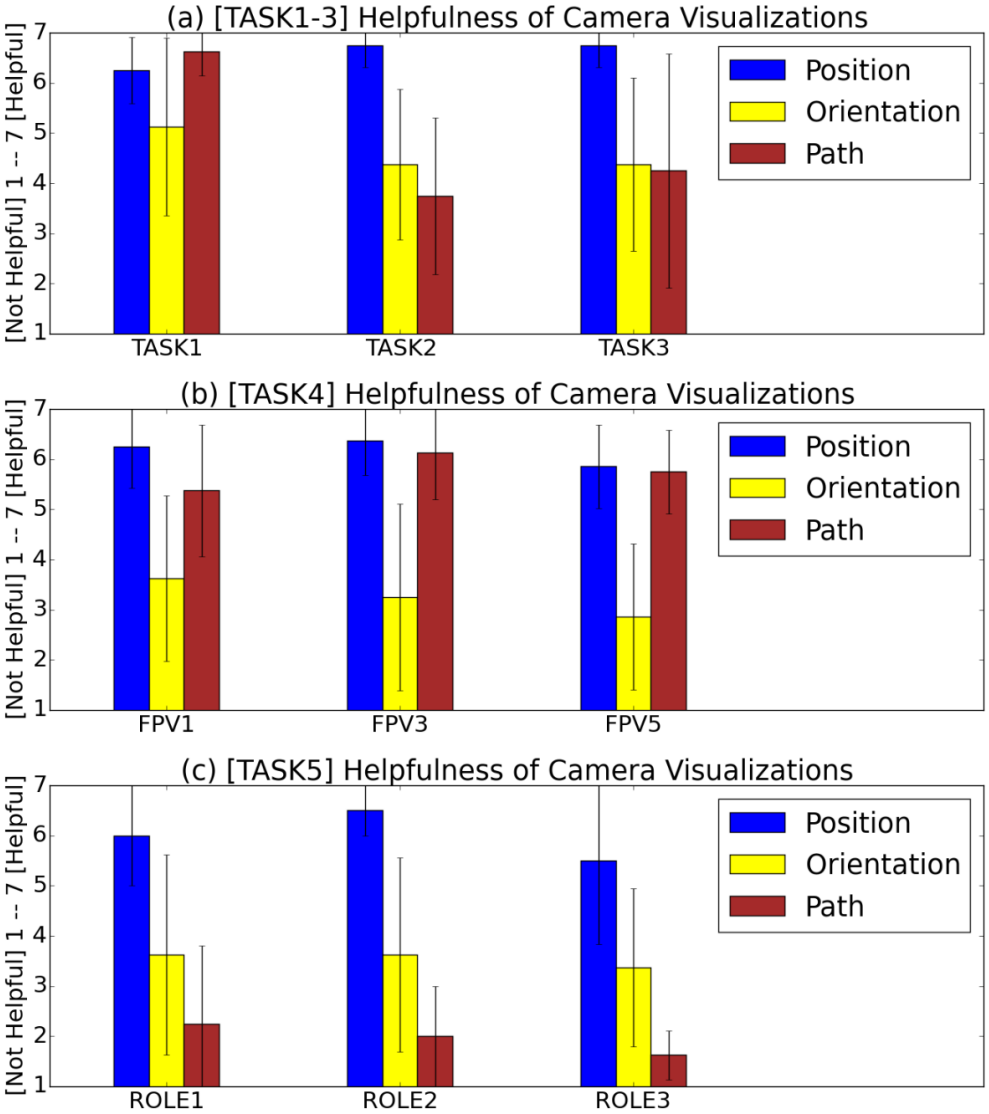

図 5.3: 提案手法におけるカメラの各可視化要素の有用性についてのアンケート (EQ2-1) の結果. ワークスペースビューにおけるカメラの位置・方向・移動軌跡の各可視化要素が, 各タスクを実行するにあたって役に立ったかどうかを被験者に 7 段階で評価してもらった. (a) タスク 1-3, (b) タスク 4, (c) タスク 5 のそれぞれについて結果を示す.

	TASK1–3								
	TASK1			TASK2			TASK3		
	Baseline		Proposed	Baseline		Proposed	Baseline		Proposed
(1)Fixation movements	$353_{\pm 99}$	†	$597_{\pm 273}$	$602_{\pm 157}$	†	$475_{\pm 166}$	$540_{\pm 115}$		$548_{\pm 118}$
(2)Widget transitions	$0.53_{\pm 0.30}$		$0.63_{\pm 0.45}$	$1.69_{\pm 0.43}$	*	$0.76_{\pm 0.31}$	$1.35_{\pm 0.32}$	*	$1.01_{\pm 0.28}$
(2-1 FPV-FPV)	N/A		N/A	$1.03_{\pm 0.37}$	*	$0.17_{\pm 0.12}$	$1.21_{\pm 0.36}$	*	$0.44_{\pm 0.27}$
(2-2 Workspace-view)	N/A		$0.36_{\pm 0.24}$	N/A		$0.42_{\pm 0.21}$	N/A		$0.48_{\pm 0.24}$
(2-3 Canvas)	$0.53_{\pm 0.30}$	*	$0.26_{\pm 0.23}$	$0.44_{\pm 0.30}$	*	$0.13_{\pm 0.08}$	$0.14_{\pm 0.16}$		$0.07_{\pm 0.06}$
	TASK4								
	FPV1			FPV3			FPV5		
	Baseline		Proposed	Baseline		Proposed	Baseline		Proposed
(1)Fixation movements	$619_{\pm 185}$	*	$336_{\pm 89}$	$737_{\pm 140}$	†	$570_{\pm 65}$	$776_{\pm 176}$		$685_{\pm 85}$
(2)Widget transitions	$0.80_{\pm 0.26}$		$0.92_{\pm 0.18}$	$1.80_{\pm 0.29}$	*	$1.28_{\pm 0.15}$	$1.80_{\pm 0.42}$	*	$1.25_{\pm 0.16}$
(2-1 FPV-FPV)	N/A		N/A	$1.25_{\pm 0.35}$	*	$0.49_{\pm 0.11}$	$1.29_{\pm 0.43}$	*	$0.53_{\pm 0.30}$
(2-2 Workspace-view)	N/A		$0.79_{\pm 0.15}$	N/A		$0.64_{\pm 0.12}$	N/A		$0.56_{\pm 0.35}$
(2-3 Canvas)	$0.78_{\pm 0.26}$	*	$0.10_{\pm 0.06}$	$0.51_{\pm 0.22}$	*	$0.13_{\pm 0.11}$	$0.50_{\pm 0.33}$	*	$0.15_{\pm 0.15}$
	TASK5								
	ROLE1			ROLE2			ROLE3		
	Baseline		Proposed	Baseline		Proposed	Baseline		Proposed
(1)Fixation movements	$523_{\pm 177}$		$435_{\pm 155}$	$506_{\pm 177}$		$451_{\pm 121}$	$497_{\pm 113}$		$463_{\pm 88}$
(2)Widget transitions	$0.99_{\pm 0.14}$		$0.97_{\pm 0.16}$	$1.35_{\pm 0.32}$	*	$0.97_{\pm 0.23}$	$1.25_{\pm 0.37}$		$1.19_{\pm 0.28}$
(2-1 FPV-FPV)	$0.56_{\pm 0.18}$	*	$0.26_{\pm 0.11}$	$1.01_{\pm 0.26}$	*	$0.49_{\pm 0.21}$	$0.95_{\pm 0.29}$	*	$0.53_{\pm 0.29}$
(2-2 Workspace-view)	N/A		$0.54_{\pm 0.18}$	N/A		$0.34_{\pm 0.11}$	N/A		$0.52_{\pm 0.28}$
(2-3 Canvas)	$0.43_{\pm 0.22}$	*	$0.13_{\pm 0.12}$	$0.31_{\pm 0.22}$	†	$0.14_{\pm 0.07}$	$0.29_{\pm 0.17}$	†	$0.12_{\pm 0.11}$

表 5.2: 各タスクにおける注視点解析の結果. (1) 実験用映像の再生中に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. ウィジェットには実験用映像中の各一人称視点映像, ワークスペースビュー, サイドバー, そして映像外のキャンバスを含む. (2) ではさらに, 一人称視点映像間の遷移回数 (2-1), ワークスペースビューへの遷移回数 (2-2), キャンバスへの遷移回数 (2-3) についても示す. また, (†, ∗, ∗∗) はそれぞれ有意水準 (.1, .05, .01) を示す.

TASK1–3			TASK4			TASK5		
TASK1	TASK2	TASK3	FPV1	FPV3	FPV5	ROLE1	ROLE2	ROLE3
$69.4_{\pm 15.7}$	$44.8_{\pm 24.2}$	$38.4_{\pm 17.8}$	$32.1_{\pm 9.2}$	$31.9_{\pm 9.4}$	$35.0_{\pm 21.8}$	$27.9_{\pm 16.5}$	$16.3_{\pm 5.7}$	$25.2_{\pm 14.0}$

表 5.3: 実験用映像の再生中に注視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 (%).

5.2　評価実験結果の考察

本節では，評価実験における量的評価結果を用いて設定した仮説を検証し，さらに被験者へのアンケートやインタビューで得られた質的評価結果を用いて，提案した可視化手法の効果や可視化手法を改善するための有意義な知見について議論する.

5.2.1　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正確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(タスク 1–3) の全てのタスクにおいて，図 5.1(a) および図 5.2(1–3) より，項目別の正答率でも全体の正答率でも上昇傾向が確認された. またいずれのタスクにおいても主観難易度が有意に易化した. 以上の点から仮説 1 は支持された. 以下では各タスクに関する考察を述べる.

タスク 1

タスク 1 では，被験者は図 4.1(P3) の多数の机が並んだ広い空間における 1 名の撮影者の移動経路を把握した. アンケートの結果 (表 5.1) から，ベースラインでは被験者は机や作業空間の背景情報 (柱や窓，エレベータホールなど) を主な手掛かりとして用いて撮影者の位置と移動経路を推定していたと考えられる. しかしながら，作業空間 (P3) は実験で用いた (P1–5) の中で最も複雑なレイアウトを呈する環境であり，また外見の類似している机や窓が多数並んでいるため，これらを手掛かりにして撮影者の絶対位置，とりわけ映像からでは手掛かりの得ることの難しい後方の移動経路と移動開始地点を推定することが困難であったと考えられる. また，注視点解析の結果（表 5.2）からは，被験者は映像の再生時間中でも頻繁にキャンバスに注視点を移していたことが観測された. 被験者は事前の白地図の閲覧によって空間レイアウトの概要を把握していたと考えられるが，一人称視点で撮影者の位置移動を推定するためには，映像中で得た断片的な背景情報を白地図と常に照らしあわせる作業が必要であったことが示唆される. また，被験者の描いた移動経路の形状が正解と大きく異なっている事例や，解答提出時に「全くわからないので適当に描いた」と述べた被験者の例も確認されており，移動開始点や終点の絶対位置だけでなく，移動開始点から見た移動経路の形状把握（たとえば前方に直進した後に右方へ方向転換する，など）においても，一人称視点映像のみでは困難が伴っていたことが示唆される.

一方，提案手法では項目別に見ると移動開始点の正答率が有意に上昇している. このことは，ワークスペースビューに表示されたカメラ位置と移動軌跡の参照によって，一人称視点映像のみからでは確認の難しい撮影者の後方の絶対位置の把握を特に効果的に支援できたことを示唆している. また，途中経路の正答率は提案手法では改善傾向にあるものの，始点・終点の正答率と比べて低い傾向となった. このことは作業空間の複雑なレイアウトが影響していると考えられ，提案手法では机一つ分隣の経路を描くといった誤答例が多く見られた一方で，正解と極端に異なる形状や「全くわからない」と述べた被験者の例は確認されなかった. さらに，本タスクではカメラの向きの利用度が高く，撮影者の移動方向の把握を効果的に支援できたと考えられる. 以上のことから，ワークスペースビューに表示されたカメラ位置とカメラの移動軌跡，そしてカメラの向きを参照することによって，このような類似した背景テクスチャと複雑なレイアウトを持つ空間内での移動経路を把握する難しさを効果的に軽減できたと考えられる.

タスク 2

タスク 2 では，被験者はブルーシートを囲んだ 3 名の撮影者の相対的位置関係や，5 名が机を囲んでいる中での 3 名の相対的位置関係を把握した．アンケートの結果 (表 5.1) から，ベースラインでは被験者は主にシーン中の大きな物体（ブルーシートや机）や撮影者の顔を主な手掛かりとして相対的位置関係を推定していたことが示唆される．本タスクで用いたデータセットでは撮影者が頻繁に俯いて手元での作業を行なっているため，映像に他の撮影者の顔が常時映っているとは限らず，顔の対応を取って各撮影者が誰であるかを把握することは非常に困難であったと考えられる．

一方で，提案手法では全被験者が満点となるなどタスク正答率が大幅に改善した．項目別でも撮影者 ID の並び順序と撮影者位置の両方で正答率が有意に改善しており，特に撮影者 ID の並び順序において大幅な改善が見られた．また，注視点解析の結果 (表 5.2) からは一人称視点映像同士の見比べ回数が大幅に減少したことが確認された．ワークスペースビューの参照によって映像同士を見比べて顔の対応を取る作業や背景情報から撮影者の位置を確認する作業を大幅に軽減できたことが示唆される．以上のことから，本タスクではワークスペースビューを閲覧して主にカメラ位置の可視化を参照することによって，撮影者の相対的位置関係の把握に関して効果的な支援ができたと考えられる．

タスク 3

タスク 3 では，被験者は 5 名が机を囲んで 2 人 1 組に分かれて絵を描いている場面や，5 名が 2 人 1 組になって箱を運搬している場面でのグループを把握した．ベースラインで多数の被験者が用いた手掛かりは，撮影者の顔，衣服，手や頭部の動き，そしてシーン中の大きな物体（机など）と多岐にわたった．被験者は，シーン中の大きな物体を用いて各グループや各撮影者の絶対位置を推定し，撮影者の顔や，紙・箱・ペンといった撮影者の手元の物体，手の見た目や衣服，そして手や頭部の動きを手掛かりとして用いて誰と誰が協調作業を行なっているかを推定する，という閲覧スタイルを取ったと考えられる．注視点解析の結果 (表 5.2) から，本タスクではベースラインにおいてもキャンバスの参照回数は少ない傾向であった一方，一人称視点映像同士の見比べ回数は多かった．映像の背景情報を白地図と照らし合わせて撮影者位置を確認する作業は少なかった一方で，本タスクで用いた実験用映像では各グループが同一内容の作業を行なっていて各撮影者の手元の映像が類似しているため，また撮影者の頭部の動きによって手元が十分に映らない場合もあるため，複数の映像を見比べて，撮影者の手の動きや手元の物体の対応付けによってグループ分けを行なうことは難しい状況であったと考えられる．

提案手法では有意ではないものの各項目別の正答率においても改善傾向が観測された．また，注視点解析の結果からは一人称視点映像同士の見比べ回数が大幅に減少したことが確認された．以上のことから，グループ分けに際して複数の一人称視点映像を見比べて手や手元と物体等の対応を取るといった閲覧の負担は，ワークスペースビューを参照して主にカメラ位置を用いることによって，効果的に軽減することができたと考えられる．

5.2.2　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 4 の FPV5 における移動開始地点の把握とタスク 5 の ROLE2 における役割担当者決定の項目以外では，全項目において正答率の悪化傾向は見られなかった (図 5.2(4,

5)). また，正答率の悪化傾向が見られたこれらの項目においても有意差は確認されなかった．さらに，各タスクのいずれの場合においても被験者の主観難易度には易化傾向が見られ，FPV5 と ROLE3 では有意だった (図 5.1(b, c)．以上の点から仮説 2 は支持された. 以下では各タスクについて考察を述べる.

タスク 4

タスク 4 では，被験者に提示する一人称視点映像の数を変化させて (FPV1, FPV3, FPV5)，被験者は物体の運搬経路を把握するという同一内容のタスクに繰り返し取り組んだ．本タスクではタスク 1 よりも実験用映像の再生時間が長く，また，被験者は箱や雑誌が運搬されている部分を発見するために一人称視点映像の閲覧に比較的多くの時間を割く必要がある. ワークスペースビューの閲覧割合はタスク群 1 に比べて低い傾向となった (表 5.3)．ベースラインでは，FPV1,3,5 のいずれの場合においてもシーン中の大きな物体と背景構造物を経路把握の手掛かりに用いた被験者が多数であった.FPV5 では撮影者の足元の位置移動を箱の移動経路の把握に用いた例も多数見られた.FPV1 と FPV3 で用いた作業場所 (P4, P5) は，タスク 1 で用いた (P3) と比較すると狭い空間かつ単純なレイアウトであり，テクスチャの変化にも富んでいる. そのため，被験者はタスク 1 よりも容易にタスクを遂行できたと考えられ，FPV1 では全被験者が満点となった. 一方 FPV3 では，FPV1 と比べて被験者の正答率は低下傾向にあり，一人称視点映像の見比べ回数は FPV5 と同程度にまで増大したにもかかわらず，被験者の主観難易度ではそれらと一致した難化傾向は見られなかった．多重比較結果においても主観難易度の有意差が確認されたのは FPV1 と FPV5, および FPV3 と FPV5 との間であった．2 名の被験者が適切な情報量だったと回答したことなどから被験者への心理的負担は少なかったことが示唆されており，FPV3 では実際のタスクの正確度との間にギャップがあった考えられる．また，FPV5 の場合には，作業場所 (P3) の複雑なレイアウトに加えて，インタビューの結果でも確認されたように映像が増加することによって体感的な作業量が増大し，主観的な難易度も上昇したと考えられる.

提案手法では FPV3 においても全被験者が満点となり，項目別では FPV3 での移動終点把握において有意に正答率が上昇した. また主観難易度では FPV5 で有意な改善が見られた．一方で FPV5 の移動開始点把握の項目では，有意でないものの正答率の悪化傾向が観測された．FPV5 で用いた実験用映像では，5 名の撮影者が複雑なレイアウトを持つ作業空間において，複数の箱の梱包と運搬を協力して行なっている．映像では箱を運搬していない撮影者もグループになって移動している部分が含まれているほか，運搬中の撮影者の映像においても頻繁に箱が映像の視野範囲から外れてしまう．一部の被験者では，このことによってワークスペースビューを参照しても箱の運搬のために移動している撮影者のグループを瞬時に判別できなかった，あるいはワークスペースビューを注視していたことによって一人称視点映像から運搬中の箱を発見するタイミングが遅れてしまったと考えられる. FPV5 では，映像が増加したことに加えてワークスペースビューがさらに追加されたことによって，どこを見るべきか分からないという困惑や，視線移動の困難の増大を感じた被験者も見られた．しかしながら，注視点解析の結果 (表 5.2) からは，一人称視点映像の見比べ回数やウィジェットの見比べ回数は減少していることが確認され，また注視点移動量も減少傾向にあることが確認されている．以上のことから，ワークスペースビューの参照によって，主にカメラ位置とカメラの移動軌跡を用いて多数の一人称視点映像を見比べて物体や撮影者の位置情報を整理する負担を効果的に軽減したと考えられる．また，タスク 4 はタスク 1 と類似したタスク内容であるが，タスク 1 と比較してカメラ向きの利用度が低い傾向にあった. 撮影者が運搬中の物体に視線を落とす，あるいは複数人で運搬する際に移動方向とカメラの向きが必

ずしも一致していなかったことが，物体の運搬経路の把握において役に立つ場面が少なかったと感じた原因の一つであると考えられる.

タスク 5

タスク 5 では役割担当者特定のために必要となる，複数映像の見比べ回数を変化させることによって被験の閲覧負荷を変化させた. ワークスペースビューの閲覧割合はタスク群 1 と比べて低い傾向であった (表 5.3). ベースラインでは，ROLE1–3 のいずれの場合においてもシーン中の大きな物体を手掛かりに用いた被験者が多数であった. 撮影者が手に持った物体を追う必要のある ROLE2 ではそれらの物体を，また複数人の手元での作業状態を詳細に追う必要のある ROLE3 では手元の動きを手掛かりに用いた被験者が多数見られた. さらに ROLE3 では，作業空間の背景構造物の情報を用いた被験者も多数見られ，それは共同作業場所の特定のためであったと推測される.ROLE3 では各撮影者の視野範囲が手元を中心とした共同作業領域に集中していることや，複数の撮影者の役割を把握するという閲覧負担が大きく，役割担当者の決定と同時に少ない背景部分から作業場所を把握することは困難であったと考えられる.

一方，提案手法では，主にカメラ位置を用いることによって役割担当者の把握を妨げることなく作業場所や該当人物の位置の把握を効果的に支援した. ROLE2 では有意でないものの役割担当者決定の項目において正答率の悪化傾向が見られたが，ROLE2 では役割担当者を本人以外の映像から推定する必要があり，ワークスペースビューの向き表示を使用しなかった，あるいは使用しても映像中の人物の特定に役に立たなかった場合に閲覧の手間が増大したと考えられる. 一人称視点映像のみから役割担当者を決定してからワークスペースビューを参照して該当人物の ID と場所を決定する，という方針を取った被験者では位置解決に充てる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 名の被験者はインタビューで上記のように回答している. また，注視点解析の結果 (表 5.2) からは，ROLE1–3 の全ての場合において一人称視点映像間の見比べ回数が有意に減少したことが確認されており，ワークスペースビューの参照によって，位置特定のために一人称視点映像を見比べるという負担を効果的に軽減できたと考えられる.

5.2.3　注視点解析の結果から得られた知見

注視点解析の結果 (表 5.2) については各タスクの考察で部分的に述べてきたが，ここでは今回の評価実験で観測された注視点移動に関して総観的に述べる. 本実験では解答入力用のキャンバスに表示された白地図を事前に閲覧して作業空間のレイアウトを事前に把握するように被験者に指示したが，ベースラインでは映像再生中にもキャンバスに頻繁に視線を移す閲覧行動が確認された. 一方で，提案手法ではキャンバスへの視線遷移の大部分がワークスペースビューへの遷移に取って代わったことが示唆される. また，提示映像が 1 つであるタスク 1 とタスク 4 の FPV1 以外の全ての場合において一人称視点映像間の遷移回数が有意に減少しており，総遷移回数においても提案手法ではベースラインに比べてウィジェット総数が増加したにもかかわらず，タスク 1 とタスク 4 の FPV1 以外では減少傾向が確認されている. 以上のことから，ワークスペースビューは撮影者位置や作業場所の特定において，一人称視点映像を見比べて映像の背景情報を集積して白地図と照合する，という負担を軽減していることが示唆される. さらに，提案手法ではタスク 1 とタスク 3 以外において注視点移動量の減少傾向も確認された. ベースラインで参照の多かった解答入力用のキャンバスが実験用映像外にあるという点を考慮しても，ウィジェットの総遷移回数が減少していることから，ワークスペースビューによって視線移動の負担を効果的に軽減できたと考えられる.

5.2.4　ワークスペースビューの使用方法についての知見

カメラ位置の可視化は撮影者位置決定のほか撮影者位置を起点とした作業空間の把握等での活用が見られた．カメラの向きの可視化は作業場所や複数撮影者間での協調の有無の把握のほか，撮影者の移動方向の把握や，3 次元モデルと一人称視点映像との視点の対応など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なかった被験者や，意識的には使用しなかった被験者が多数であった．カメラの移動軌跡の可視化は，撮影者の移動経路の把握だけでなく，ペアの形成や位置の先読み，動きを用いた撮影者との紐付け等の使用方法が見られた．

5.2.5　可視化手法の改善すべき点

被験者インタビューにおいて最も多かった改善要望点は，カメラ位置姿勢の推定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．本研究で用いたカメラ位置姿勢の推定手法では，カメラ位置姿勢の復元に必要な十分な対応点が得られなかったことによって復元できなかったフレームを数多く含んでおり，復元に成功した前後のフレームを用いて単純線形補間を行なっている．このため，復元できないフレームが多数連続する場合には，位置や姿勢が大きくズレてしまう．本研究で用いたような SfM ベースの手法では，[28] のようにエピポーラ幾何を利用したより厳密な補間手法や，フレーム間の連続的移動を考慮してモーションモデルによる補間手法などを用いて，SfM で復元できないフレームを効果的に補間していく必要があると考えられる．一人称視点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ると考えられ，ワークスペースビュー内に一人称視点映像を配置するほか，各カメラ位置表示から一人称視点映像へのリードや，人間の目で識別しやすい表色系を用いた色対応の改善によってスムーズな視線移動を促進するといった対策が考えられる．このような対策は一人称視点映像数が増加して視線移動の負担が増大した場合に効果的に働くと考えられる．また，ワークスペースビューの閲覧視点 (作業空間の 3 次元モデルを閲覧している視点) と一人称視点映像が撮影された視点との対応がスムーズでなかった点も改善点として挙げられる．このことはカメラ向き表示の意識的な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，カメラ向きの意識的な利用の拡大を促すために，カメラ向き表示の視認性を高めるように可視化を改良するほか，一人称視点映像の閲覧中においてもカメラの向きや他の撮影者との位置関係を把握できるように，一人称視点映像自体の可視化を改善することも効果的であると考えられる (例えばある撮影者の映像に他のカメラの位置を投影したり，他のカメラが視野範囲になくてもどの方向にあるかが分かるように矢印で方向を示す，など)．また，本研究では協調作業の映像の事後的な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評価実験を実施して，撮影者位置把握への効果を確認した．Visual SLAM[45] やセンサヒュージョン等を活用した高速なカメラ位置姿勢の復元手法を用いることで，複数一人称視点映像による監視システムのようなリアルタイムシステムへの応用にも十分に検討の余地があると考えられる．

第6章

結論

6.1 本研究のまとめ

本研究では，ウェアラブルカメラを頭部に装着して撮影した一人称視点映像について，それらを複数用いて協調作業を記録するシナリオに着目した．そのような複数一人称視点映像という形態での作業の記録は，複数の撮影者が着目した重要物体などといった解析の観点からは様々な示唆に富むが，一方で人間がそのような映像を一度に閲覧しようとする際には，各撮影者の絶対位置だけでなく，複数の撮影者の相対的位置関係やグループといった，協調作業を理解する上で重要となる各撮影者の位置に関する情報を把握することが難しくなる．そのため，本研究では並列提示した複数の一人称視点映像に加えて，作業が行われている空間を 3 次元復元して提示し，その上に各カメラの位置と向き，および移動軌跡を可視化するワークスペースビューを備えた閲覧システムを提案した．

提案システムにおけるワークスペースビューの可視化効果を検証するためにユーザ評価実験を実施し，ワークスペースビューを閲覧することによって撮影者の位置に関する情報をより正しくかつ容易に理解できるようになるかどうか，また，ワークスペースビューを補助的に閲覧する場合において，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って一人称視点映像の閲覧自体を妨げないかどうか，という 2 つの観点から仮説を構築して検証し，ユーザ評価実験の結果それぞれの仮説が支持されたことを確認した．さらに，被験者からアンケートやインタビューで得た回答や，被験者の実験中の注視点解析の結果を用いて多面的に閲覧行動を分析した結果，一人称視点映像数が増加した場合や複数の一人称視点映像を見比べる難易度が上昇した場合などの，閲覧負荷が増大する場合においてもタスクの難易度に改善傾向が見られたほか，注視点移動量の減少傾向や一人称視点映像同士の見比べ回数の有意な減少が観測されるなど，被験者の閲覧負担を効果的に軽減していることが分かった．また，被験者からインタビューで得た可視化手法へのフィードバックを元に，可視化手法の改善点やシステムの今後の課題についての知見も得た．

6.2 可視化デザインの改善点

被験者からのフィードバックや閲覧行動の分析の結果得られた知見を用いて，今後同様の複数一人称視点映像閲覧システムを構築するにあたって，本研究で提案した可視化デザインをどのように改善するべきかについて述べる．

6.2.1　カメラ位置姿勢の推定精度

本研究で用いたカメラ位置姿勢の推定手法では，カメラの断続的に動きによって，Perspective-n-Point 問題を解くために必要な十分な数の良い対応点が得られないという問題が頻発した．カメラ位置姿勢の推定できないフレームが長時間連続した場合，本研究で用いたような線形補間ではカメラ位置姿勢が不自然な値となって復元され，ワークスペースビューを基点にした閲覧者の空間理解に悪影響を及ぼしうる．カメラ位置から各撮影者を連想させるような可視化では，カメラの位置姿勢の推定精度を向上させることが，閲覧者の作業空間や撮影者情報に対する理解の向上と直結する．

6.2.2　閲覧者の効率的な視線移動の考慮

本研究で提案したような複数の一人称映像を同時に提示する閲覧システムでは，閲覧者は多数の映像を見比べて情報を収集するという閲覧スタイルを取る．そのため，各映像間の物理的な距離を短縮したり，ある映像から別の映像への視線移動の際に困惑が生じないようにする必要がある．たとえば，各カメラの位置と視点方向が断続的に変化する場合には，ある撮影者の映像から別の撮影者の映像への視線の遷移において，目的の撮影者とそれに対応する映像を探す負担が増大すると考えられる．また，ワークスペースビューを参照する場合においても，各映像に対応するカメラ表示を探す手間の増大が予想される．そのため，ワークスペースビュー上の表示物から映像へのリードによってスムーズな視線移動を促したり，各一人称視点映像の配置に関しても詳細に検討する必要がある．また，本研究ではワークスペースビューを中心にその周囲に各一人称視点映像を配置した．一人称視点映像同士の見比べ回数の減少が注視点移動の減少に結びついており，ワークスペースビューを画面の中央に配置して，ワークスペースビューを基点として各一人称視点映像との間を行き来させるようなデザインが，視線移動を軽減する上では効果的であると考えられる．

6.2.3　どの映像を見るべきかについての示唆

映像が複数あった場合には，どの撮影者の映像やどのグループの映像に着目すべきかについての示唆を与えることが必要である．すなわち，本研究のような，複数撮影者の詳細な行動の様子を全て提示する閲覧システムにおいても，全体的な作業の流れに関して要約的な理解を促進するような可視化デザインが必要であると考えられる．たとえば，空間の中で多数の撮影者が協調している領域をハイライトして協調行動を行なっている撮影者の集団を容易に把握できるようにするほか，撮影者の接近だけでなく協調行動や共同注視などのイベントの発生を先読みできるようにカメラの移動軌跡のうち重要部分をハイライトする，などが考えられる [46]．

6.2.4　各撮影者視点からの空間把握の補助

各一人称視点映像やワークスペースビューを行き来するにあたって，それぞれのカメラが作業空間の中でどの位置からどの方向をとらえているかを知覚することが，例えば映像に映っている人物が誰でどこにいるかを把握するためには重要である．本研究ではカメラの向きを可視化したが，多くの被験者では無意識的な補助的利用にとどまった．カメラの向きを閲覧者に意識的に知覚させるためには，本研究で用いたものよりもカメラの向きを強調表示して，視認性を高めるような可視化デザインを検討する必要がある．また，向きを意識しなくても映像に映っている他の撮

影者が誰であるかが分かるように，各一人称視点映像に各撮影者の位置や方向を投影するようなアプローチについても検討する必要がある．

6.3　今後の課題

本研究における今後の課題としては，可視化手法の改善とインターフェースとしての効果の検証を予定している．ユーザ評価実験で浮かび上がった可視化手法の改善点（カメラ位置姿勢推定の高精度化と高速化，映像やワークスペースビューの配置関係や連結性）に関して検討するほか，ユーザによる操作性を与えてインターフェースとしての機能を検証する．さらに，実際の協調作業の映像記録を長時間閲覧して内容を理解するような場面においても，作業空間やカメラの可視化がもたらす効果を検証するとともに，新たな課題を発見していきたいと考えている．

謝辞

本研究を行なうにあたって，指導教官の佐藤洋一先生には 3 年間大変お世話になりました. 研究の方向性で行き詰まった際には常に熱心に相談に乗っていただき，道標を示してくださいました. また，進路で困った際にも温かく相談に乗っていただきました. 本当にありがとうございました. 助教の米谷竜さんにはコンピュータビジョン分野の初学者として大学院に入学した頃から，コンピュータビジョンの師として勉強や研究面での技術的な相談のほか，論文の書き方でも熱心に指導していただき大変感謝しております. 本論文の執筆においても懇切丁寧に指導していただきました. また，助教の樋口啓太さんには，本研究を行なうにあたって特にお世話になりました. ヒューマン・コンピュータインタラクション分野の師としてユーザインターフェース構築にあたっての技術的な相談や，ユーザ評価実験のノウハウ，研究での議論やストーリーの立て方，また個人的なことでも熱心に相談に乗っていただきました. 大変感謝しております. 佐藤研究室秘書の鈴木咲恵さんと今川洋子さんには，学会出張の手続きや，実験データ収録場所の使用申請，また被験者への謝礼支出の手続きなど多方面において大変お世話になりました. そして，研究室で 2 年間を共に過ごし，共に勉強や研究について議論して，データセットを一緒に収録し，時には励ましあった松本大輝さん，村上晋太郎さん，神窪利絵さん， Nattawan Tantirujananont さん，丸山玄氣さん，計良宥志さん，中野雄介さんにも大変感謝しております. 長時間の被験者実験に協力してくださった研究室メンバーの皆様にも感謝しております. 最後に大学院生活を見守り, 支援してくれた両親や友人にも感謝の念を末筆ながらここに表して，本論文の結びとさせていただきます.

2017 年　2 月 3 日
杉田　祐樹

参考文献

[1] Hamed Pirsiavash and Deva Ramanan. Detecting activities of daily living in first-person camera views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2847–2854, 2012.

[2] Minghuang Ma, Haoqi Fan, and Kris M. Kitani. Going deeper into first-person activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[3] Minjie Cai, Kris M Kitani, and Yoichi Sato. A scalable approach for understanding the visual structures of hand grasps. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1360–1366, 2015.

[4] Keisuke Ogaki, Kris M Kitani, Yusuke Sugano, and Yoichi Sato. Coupling eye-motion and ego-motion features for first-person activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1–7, 2012.

[5] Hyun Soo Park, Jyh-Jing Hwang, Yedong Niu, and Jianbo Shi. Egocentric future localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4697–4705, 2016.

[6] Kris M. Kitani. Syed Zahir Bokhari. Long-term activity forecasting using first-person vision. In *Asian Conference on Computer Vision (ACCV)*, 2016.

[7] Michael S Ryoo and Larry Matthies. First-person activity recognition: What are they doing to me? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2730–2737, 2013.

[8] Yoshio Ishiguro, Adiyan Mujibiya, Takashi Miyaki, and Jun Rekimoto. Aided eyes: Eye activity sensing for daily life. In *Proceedings of the 1st Augmented Human International Conference (AH)*, p. 25, 2010.

[9] Tung-Sing Leung and Gerard Medioni. Visual navigation aid for the blind in dynamic environments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 153–158, 2014.

[10] Alircza Fathi, Jessica K Hodgins, and James M Rehg. Social interactions: A first-person perspective. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1226–1233, 2012.

[11] Ryo Yonetani, Kris M Kitani, and Yoichi Sato. Recognizing micro-actions and reactions from paired egocentric videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2629–2638, 2016.

[12] Kris Kitani Nicholas Rhinehart. Learning action maps of large environments via first-person vision. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[13] Hiroshi Kera, Ryo Yonetani, Keita Higuchi, and Yoichi Sato. Discovering objects of joint attention via first-person sensing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 7–15, 2016.

[14] Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI)*, pp. 513–520, 2003.

[15] Keita Higuch, Ryo Yonetani, and Yoichi Sato. Can eye help you?: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI)*, pp. 5180–5190, 2016.

[16] Shunichi Kasahara and Jun Rekimoto. Jackin: Integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference (AH)*, p. 46, 2014.

[17] Eleanor G Rieffel, Andreas Girgensohn, Don Kimber, Trista Chen, and Qiong Liu. Geometric tools for multicamera surveillance systems. In *First ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, pp. 132–139, 2007.

[18] Peter M Roth, Volker Settgast, Peter Widhalm, Marcel Lancelle, Josef Birchbauer, Norbert Brandle, Sven Havemann, and Horst Bischof. Next-generation 3d visualization for visual surveillance. In *Proceedings of the 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pp. 343–348, 2011.

[19] Philip DeCamp, George Shaw, Rony Kubat, and Deb Roy. An immersive system for browsing and visualizing surveillance video. In *Proceedings of the 18th ACM international conference on Multimedia*, pp. 371–380, 2010.

[20] Shuaicheng Liu, Lu Yuan, Ping Tan, and Jian Sun. Steadyflow: Spatially smooth optical flow for video stabilization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4209–4216, 2014.

[21] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, Vol. 81, pp. 674–679, 1981.

[22] Johannes Kopf, Michael F Cohen, and Richard Szeliski. First-person hyper-lapse videos. *ACM Transactions on Graphics (TOG)*, Vol. 33, No. 4, p. 78, 2014.

[23] Yair Poleg, Tavi Halperin, Chetan Arora, and Shmuel Peleg. Egosampling: Fast-forward and stereo for egocentric videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4768–4776, 2015.

[24] Tavi Halperin, Yair Poleg, Chetan Arora, and Shmuel Peleg. Egosampling: Wide view hyperlapse from single and multiple egocentric videos. *arXiv preprint arXiv:1604.07741*, 2016.

[25] Zheng Lu and Kristen Grauman. Story-driven summarization for egocentric video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2714–2721, 2013.

[26] Joydeep Ghosh, Yong Jae Lee, and Kristen Grauman. Discovering important people and objects for egocentric video summarization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1346–1353, 2012.

[27] Wen-Sheng Chu, Yale Song, and Alejandro Jaimes. Video co-summarization: Video summarization by visual co-occurrence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3584–3592, 2015.

[28] Ido Arev, Hyun Soo Park, Yaser Sheikh, Jessica Hodgins, and Ariel Shamir. Automatic editing of footage from multiple social cameras. *ACM Transactions on Graphics (TOG)*, Vol. 33, No. 4, p. 81, 2014.

[29] Harpreet S Sawhney, Aydin Arpa, Rakesh Kumar, Supun Samarasekera, Manoj Aggarwal, Steve Hsu, David Nister, and K Hanna. Video flashlights: real time rendering of multiple videos for immersive model visualization. In *Proceedings of the 13th Eurographics Workshop on Rendering*, pp. 157–168, 2002.

[30] Hyun S Park, Eakta Jain, and Yaser Sheikh. 3d social saliency from head-mounted cameras. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 431–439, 2012.

[31] Luca Ballan, Gabriel J Brostow, Jens Puwein, and Marc Pollefeys. Unstructured video-based rendering: Interactive exploration of casually captured videos. *ACM Transactions on Graphics (TOG)*, Vol. 29, No. 4, p. 87, 2010.

[32] Shunichi Kasahara, Mitsuhito Ando, Kiyoshi Suganuma, and Jun Rekimoto. Parallel eyes: Exploring human capability and behaviors with paralleled first person view sharing. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI)*, pp. 1561–1572, 2016.

[33] 小宮山凌平, 味八木崇, 暦本純一. Jackin space: 一人称・三人称映像間の連続的な遷移を可能にするテレプレゼンスシステム. インタラクション 2016 論文集, pp. 29–37, 2016.

[34] Richard Szeliski. *Computer vision: algorithms and applications*, pp. 303–332. Springer Science & Business Media, 2010.

[35] Richard Szeliski. *Computer vision: algorithms and applications*, pp. 489–499. Springer Science & Business Media, 2010.

[36] Carlo Tomasi and Takeo Kanade. *Detection and tracking of point features.* School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991.

[37] David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, Vol. 60, No. 2, pp. 91–110, 2004.

[38] Noah Snavely, Steven M Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision (IJCV)*, Vol. 80, No. 2, pp. 189–210, 2008.

[39] C. Wu, S. Agarwal, B. Curless, and S. M. Seitz. Multicore bundle adjustment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3057–3064, 2011.

[40] Changchang Wu. Towards linear-time incremental structure from motion. In *3DTV-Conference, 2013 International Conference on*, pp. 127–134, 2013.

[41] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 32, No. 8, pp. 1362–1376, 2010.

[42] 山崎俊太郎, 持丸正明, 金出武雄. 一人称ビジョンシステムのための自己位置推定法. 信学技報, Vol. 110, No. 27, pp. 73–78, 2010.

[43] Agostino Gibaldi, Mauricio Vanegas, Peter J Bex, and Guido Maiello. Evaluation of the tobii eyex eye tracking controller and matlab toolkit for research. *Behavior research methods*, pp. 1–24, 2016.

[44] Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard Trster. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 33, No. 4, pp. 741–753, 2011.

[45] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 225–234, 2007.

[46] Ralf P Botchen, Fabian Schick, and Thomas Ertl. Action-based multifield video visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, Vol. 14, No. 4, pp. 885–899, 2008.

発表文献

国内会議

[1] 杉田祐樹, 樋口啓太, 米谷竜, 佐藤洋一. 複数一人称視点映像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. 第 171 回情報処理学会ヒューマンコンピュータインタラクション研究会, 2017 年 1 月.

[2] 杉田祐樹, 樋口啓太, 米谷竜, 佐藤洋一. 複数一人称視点映像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. インタラクション 2017 (プレミアム発表), 2017 年 3 月発表予定.

本研究に含まれない文献

国内会議

[3] 杉田祐樹, 米谷竜, 佐藤洋一. 一人称視点映像における視線情報を活用した自己アクティビティ認識. 第 18 回画像の認識・理解のシンポジウム (MIRU2015), 2015 年 7 月.