

遠隔作業支援シナリオにおける注視位置可視化の効果

樋口 啓太* 米谷 竜* 佐藤 洋一*

概要. 実世界における作業員に対して遠隔地にいる他者（指示者）の指示を伝達する遠隔作業支援シナリオにおいて、指示者の視線情報を利用した新たな支援の枠組みを提案する。本枠組みにおいて、指示者は作業空間の映像を逐次監視でき、さらに映像に対する注視行動やハンドジェスチャを、プロジェクタやヘッドマウントディスプレイ（HMD）を通して作業員にフィードバックすることができる。本研究では、1) 机上での組み立て作業を想定した固定視点カメラ・プロジェクタベースの支援システム、2) 移動を伴う作業を想定したウェアラブルカメラ・HMD ベースの支援システムを構築・評価し、指示者による作業空間への指示がより高速、あるいは少ない失敗回数で行えることを示した。

1 はじめに

本研究では、実世界で作業をする人間を遠隔にいる別の人間が支援をする**遠隔支援作業システム**の構築に取り組む。本システムにより、たとえば素人が遠隔地にいる専門家の支援を受けることで、その技能を一時的に獲得することや、意思決定能力を向上させることができると期待される。

葛岡ら [1] は、実際に作業するユーザを「作業員（被指示者）」、遠隔から支援するユーザを「指示者」と呼び、指示者によるハンドジェスチャを作業員のヘッドマウントディスプレイ（HMD）に提示することで両者の協調的作業を促すシステムを構築した。ハンドジェスチャは対象の形状や動きを表現できる一方、作業空間中の特定対象を指示（ポインティング）するためには必ずしも有効な手段とは言えない。たとえばモバイルカメラ [2] やヘッドマウントカメラ [3]、ショルダーマウントカメラ [4] を通して指示者が作業空間を把握するシステムでは、作業員が移動し作業空間が絶えず変化する状況において、随時的確なポインティングを行うことは難しい。

そこで我々では、従来用いられてきたハンドジェスチャに加えて、作業空間をモニタを介して閲覧する指示者の視線情報を計測し、作業員への指示に利用する新たな枠組みを提案する。人間は動的に変化する映像中でも容易に特定の対象物を注視できるため、視線情報を作業員に提示することで正確なポインティングが実現できると考えられる。また、視線を通して指示者の非明示的な意図を作業員が汲み取ることができれば、遠隔作業支援をより円滑にすることができる。本研究では、提案する枠組みを実現するシナリオとして 1) 作業空間を固定カメラにより取得し、固定プロジェクタによって注視・ジェスチャをフィードバックするシナリオ（固

定視点シナリオ）および 2) 作業員視点の主観カメラにより作業空間を取得し、HMD を通してフィードバックを行うシナリオ（一人称視点シナリオ）を想定し、それぞれを実現するシステムを設計する。また評価実験を通して、提案する枠組みが遠隔共同作業を円滑にすることを示す。

2 遠隔作業支援シナリオ

本章では、固定視点シナリオと一人称視点シナリオそれぞれの定義やユースケース、課題とシステム設計の指針を述べる。図 1 に、それぞれのシナリオのコンセプトと、本研究で実現したシステムを示す。

2.1 固定視点シナリオ

固定視点シナリオは工作機などの固定された作業空間において、組み立てや電子工作などの手作業を支援するシナリオである。作業員は机の前に座り、指示者の支援を受けながら作業を遂行する。作業員は指示者に対して適時質問をすることができ、指示者は理解度に応じた教示が可能である。また、音声対話を通してどのような物を作るかを議論しながら作業を進めることもできる。

固定視点シナリオでは指示投影用のプロジェクタを設置することにより、作業空間全体に指示を提示することができる。指示の内容としては、作業空間中における特定対象へのポインティングと、組み付ける・回すといった動作が考えられる。円滑な作業支援のためにはこれらの指示を作業員に的確かつ速く伝達することが重要であるが、従来のシステムのようにハンドジェスチャによってポインティングと動作指示の両方を伝達することは指示者にとって負担となることがある。

そこで本研究では円滑な共同作業を実現するために、作業空間への指示者の視線位置投影によって、作業員への効率的なポインティング位置伝達を実現するシステムを構築する。本システムでは、上方に

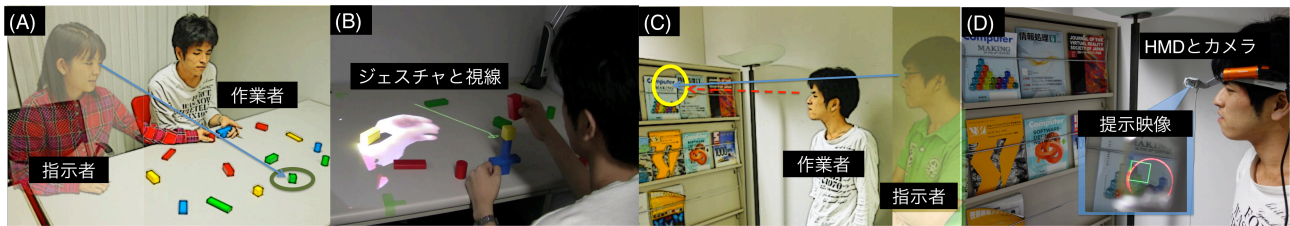


図 1. 本研究のコンセプトと実装：(A) 固定視点シナリオのコンセプト (B) 固定視点システムでの作業者への指示提示 (C) 一人称視点シナリオのコンセプト (D) 一人称視点システムにおける視線ポインティング

設置したカメラにより、机と作業者の行動を撮影し、指示者側に映像・音声を伝送・提示する。指示者はその映像をもとに作業者へ指示をする。作業者側には音声に加え、手の動作と視線位置をプロジェクタからの映像投影によって伝えることができる。

2.2 一人称視点シナリオ

一人称視点シナリオは、ウェアラブルカメラを装着した作業者が、遠隔の指示者と共同で行動するシナリオである。本シナリオは作業者にカメラが装着されているため、作業空間の広さに制約がなく、組み立てや修理といったタスクから、買い出しや物件探しなどのための相談など、固定視点シナリオに比べて多くのユースケースが想定される。しかしながら、固定された投影装置を使うことができないため、作業者への指示提示方法が課題となる。

本研究では、指示提示のために光学シースルー型のHMDを採用する。他の候補としてビデオシースルー型HMDやスマートフォンなどが考えられるが、ビデオシースルー型HMDは顔を覆ってしまうため日常空間で活動するには問題がある。また、スマートフォンには作業中に機器を保持しなければならない、機器画面と作業空間の間の頻繁な視線移動が必要であるという問題がある。そのため、実世界中における作業支援においては、視野に直接映像を重畳できる光学シースルー型HMDによる映像提示が適切である。

一人称視点シナリオでは、指示者が作業空間中のどのオブジェクトに注目しているのかを正しく伝え、ポインティングによる共同注意（ジョイントアテンション）を形成し意思疎通を図ることが重要である。しかし、ウェアラブルカメラにより撮影された映像は作業者の動きにより、シーンの移動が頻繁に起こるため、特定のオブジェクトにポインティングを続けることは難しい。本研究では、作業者の視線によるポインティングを導入することで、この問題の解決を図る。これにより指示者は、作業者の装着したカメラ映像中に映る対象物を注視することにより、ポインティング位置を伝達できる。また、固定視点シナリオと同様に、音声およびジェスチャによる指示も加えて導入することで、対象への作業内容も伝

達することができる。

3 プロトタイプシステムの構成

固定視点シナリオと一人称視点シナリオにおける視線位置提示の効果を明らかにするために、実験のためのプロトタイプシステムを開発した。それぞれのシナリオを実現するシステムを「固定視点システム」「一人称視点システム」と呼称する。二つのシステムともに、指示者側と作業者側のアプリケーションが相互に通信することで、作業環境を指示者に伝送し、作業者への指示伝達を実現している。

3.1 固定視点シナリオを実現するシステム

固定視点システムは指示者の視線位置と手による指示を作業環境上に投影するシステムである。図2にシステム概要図を示す。作業空間の周囲にはカメラ（固定視点カメラ）とプロジェクタが設置されており、カメラによる環境の取得と、プロジェクタによる情報の投影される。指示者には作業環境を写した映像がディスプレイにより提示される。指示者の視線位置を取得するためディスプレイにはアイトラッカー（Tobii EyeX）が取り付けられている。また、指示者の右背後には手による指示を撮影するためのカメラ（ジェスチャカメラ）が取り付けられている。アイトラッカー利用時、頭部固定のためのチンアームは使用していないため、指示者は自由な頭部姿勢で指示ができる（図3(A)）。

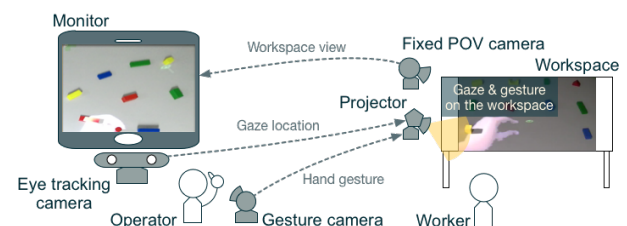


図 2. 固定視点システムの構成

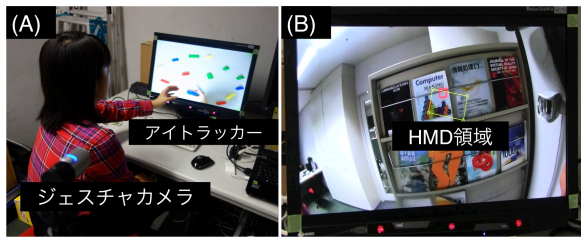


図 3. 指示者側のインターフェース:(A) 固定視点システム (B) 一人称視点システム

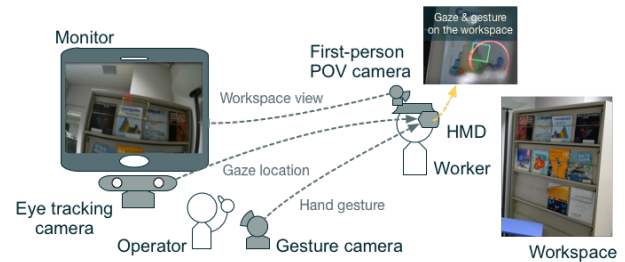


図 4. 一人称視点システムの構成

3.1.1 プロジェクタ-カメラ間のキャリブレーション

指示者への作業環境提示と、指示投影の位置合わせのためにプロジェクタ-環境カメラ間でキャリブレーションをする。作業環境を撮影するカメラによりされた映像中の4点と、それに対応するプロジェクタで投影されている4点を手動で設定することにより、キャリブレーションのための透視投映行列を計算する。得られたキャリブレーション行列から指示者への提示映像を生成する。位置合わせがされているため、指示者の視線位置や手の位置は作業空間上の同じ位置に投映される。

3.1.2 視線と手による指示の可視化

固定視点システムはアイトラッカーにより取得されたディスプレイ上の視線位置を可視化する。視線の動きには注視やサッケードなどがあり、視線位置は滞留と追従、高速な移動を繰り返す。サッケードの最大速度は500[deg/sec]にも達するとされ、視線の位置と移動を適切に提示しなければ、作業者は視線を見失ってしまう可能性がある。視線の可視化は円による視線位置と、線による移動軌跡によって表現される。

指示のための手の動きは、ジェスチャカメラにより撮影される。ジェスチャカメラには偏光フィルタが取り付けられており、ディスプレイに映された映像を遮断する[5]。撮影された画像中のディスプレイの角4点を手動で設定し、さらにアイトラッカーの発する光を遮断するための閾値による2値化により、手のみを切り出すマスクイメージを作成する。画像にマスクを適用し、ディスプレイの角4点と対応するプロジェクタの4点から作成した行列による透視投影変換を適用することにより、プロジェクタ投映用の映像を作成する。ジェスチャカメラはディスプレイの前に出された指示者の両手を捉えることができる。

3.2 一人称視点シナリオを実現するシステム

一人称視点システムは指示者の視線位置を、作業者が装着したHMDに提示することにより、効率的にポインティングを伝達するシステムである。また、

細かな作業指示を音声と手の動きにより伝えることができる。図4に一人称視点システムの概要図を示す。作業者は作業空間撮影用のウェアラブルカメラ(Panasonic A1H)と指示提示用の光学シースルー型HMD(Brother AirScouter WD-200S)を装着する。指示者側の構成は固定視点シナリオと同様である。ウェアラブルカメラの映像と、視線と手の動きによる指示はどちらもリアルタイムで伝送される。

3.2.1 HMD-カメラ間のキャリブレーション

指示者の視線による指示位置と、作業者がHMDを通して見る指示位置を合わせるためにHMD-ウェアラブルカメラ間のキャリブレーションをする。キャリブレーションは加藤らの二次元マーカーを用いたキャリブレーションを採用した[6]。作業者はHMDとカメラを装着するとき、HMDに表示された四角い枠と印刷されたマーカーの位置を合わせることでキャリブレーションをする。それにより、作業者から見たHMDの映像提示領域とアイトラッカーにより取得された視線の位置合わせができる。また、指示者に提示するウェアラブルカメラで撮影した映像中に、HMDの映像提示領域を描画することができる(図3(B))。

3.2.2 光学シースルー型HMDへの指示提示

光学シースルー型HMDの限られた視野内で、指示者の視線位置を提示しなければならないという問題に対して、方向インジケータと位置インジケータの併用による位置の提示を提案する。図5(A)のように、方向インジケータは12本の曲線から構成される円形で表現される。赤い辺は指示者の視線位置の方向を示しており、赤い辺が増えるほどHMD提示領域に近い部分を見ていることを示している(図5(B))。図5(C)中の緑の四角が位置インジケータであり、指示者が実際に見ている位置を示している。指示インジケータは視線位置とHMD領域が重なった時にのみ表れ、作業者と指示者が互いに同じ場所をみていることを意味する。

一人称視点システムでは、図5(D)のようにジェスチャカメラの映像から切り出された手をHMDに提示する。切り出された手の映像が直接提示される

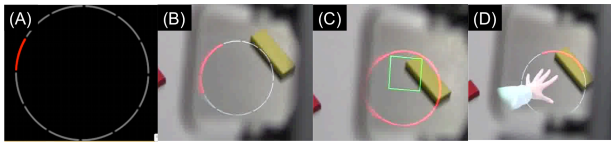


図 5. 光学シースルー型 HMD による視線位置とジェスチャの提示

ため、手の形や動きによる作業指示をすることができる。

4 実験 1: 固定視点シナリオでの視線提示による組み立て作業指示に関する実験

固定視点シナリオにおける視線位置提示の効果を測定するために、組み立て作業における指示がジェスチャのみを提示した場合と、ジェスチャと視線の両方提示した場合でどのように変化するかを評価する。本実験の仮説は「視線提示により作業指示が効率化するため、作業の完了時間が短くなる」である。その検証として、積み木を使った簡易な組み立てタスクを実施し「ジェスチャのみ提示」と「ジェスチャ+視線を提示」を比較する。また、実験の観察や、被験者からのフィードバックにより、固定視点シナリオにおける視線提示の効果を明らかにする。実験は二人一組からなり、被験者は男性の大学院生 8 名 (4 組) である。

4.1 実験環境

本実験のタスクは積み木の組み立てであり、指示者が指定された形状の組み立てを作業者に指示する。組み立てる対象の積み木は事前に決められており、図 6 に示される 4 種類である。全ての積み木が 10 個のブロックから構成されている。指示者には組み立て完成済みの対象を別々の角度から撮影した 3 枚の写真が渡される。指示者は計測開始と同時に写真をみることができる。固定視点には 25 個の積み木が配置されている。その中から必要なものを作業者に指示し、組み付け方を教示する。実験は開始から最後のブロックを組み付け終わるまでを 1 つのセッションとして、タスク完了時間を計測する。

実験開始前に固定視点システムの概要を説明し、被験者の二名はどちらが先に指示者になるかを決定する。指示者となった被験者は「ジェスチャのみ提示」と「ジェスチャ+視線を提示」により、作業者となった被験者に指示をする。その後、作業者と指示者の役割を交代し、同様の実験をする。このとき全ての実験セッションの前に 5 分程度の練習をする。また、提示方法使用の順番はランダム化される。全てのセッション終了後に、被験者からフィードバックを得るための聞き取りを実施した。

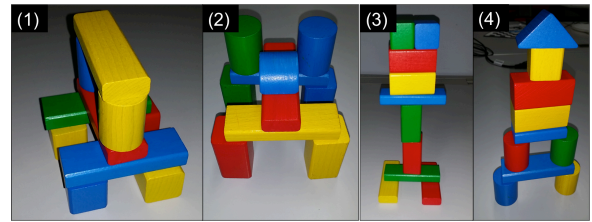


図 6. 実験 1 で組み立てる 4 種類のブロック (1-4)

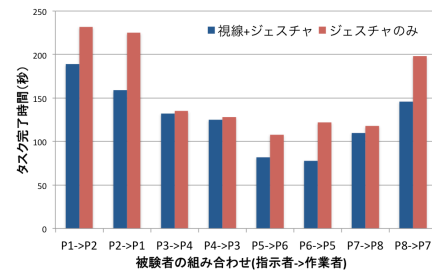


図 7. 実験 1 の結果

4.2 結果

図 7 にそれぞれの被験者の組におけるタスク完了時間 (秒) を示す。「ジェスチャのみ提示」と「ジェスチャ+視線を提示」の結果に対し、ノンパラメトリックな検定手法であるマン・ホイットニーの U 検定を適用した。その結果、 $p = 0.01$ となり 2 つの提示方法の結果に有意差があり、仮説が指示された。

4.3 被験者の観察とフィードバック

ジェスチャ+視線を提示において、6 名の被験者は必要なブロックを指示するときに視線と発話（「これを取ってください」など）を用いて指示を行った。P3 と P4 の組みでは、あまり発話を行わず視線とジェスチャを組み合わせた指示をしていたため、ジェスチャのみ提示とあまり指示の速度が変わらなかった。視線と発話を使った 6 名の被験者からは「視線により指示位置が伝わっていると感じていたので、取ってほしい物をみた状態で発話することにより指示を出した」「視線を使うことにより二つ以上の物を同時に指示することができた」といったフィードバックがあった。視線とジェスチャを組み合わせた指示を行っていた 2 名の被験者からは「視線が伝わっていることは分かったが、指示の確定のためにジェスチャを使った」「最初は自分の視線が伝わっているかわからなかったが、後から視線だけで指示できそうだとわかった」といった感想があった。

作業者として指示を受けていた被験者は、指示対象のブロックをほとんど間違えずに選択していた。視線とジェスチャが両方提示されていたときの感想を聞いたところ「相手がどこを見ているかわかったので、相手の発話による指示に合わせて見ているブ

ロックをとればよかった」「視線が提示されている方が相手が指示をしているのか、取るブロックを探しているのかわかって安心して作業できた」「視線によって次にどのブロックがくるのか予想できた」などの感想があった。

ジェスチャ+視線を提示では、ジェスチャはブロックの順番や方向を教示するときに使われることが多かった。視線のみでは向きや順番などを伝えることができないため、作業者が対象のブロックを保持した後や、口頭での指示を誤った後などに、指示者はジェスチャによる教示を行った。しかし、立体形状を組み立てているため、二次元的なジェスチャの投影のみでは正確な指示に時間がかかるという課題もあった。

4.4 考察

実験の結果から、視線提示により指示対象物へのポインティングが速くなり、組み立て作業が効率化することが示された。指示者は指示対象物を見て発話をするだけで、作業者に意思を伝達することができる。また、視線は高速で移動するため、2つ以上の対象物へのポインティングも効率的にできる。一方で、ジェスチャのみ提示している場合は、手腕を動かして指差し必要であるため視線と比べてポインティングに時間がかかってしまう。視線提示をしている場合、ジェスチャは組み付け方の教示に使われている。そのため、指示者は視線とジェスチャという2つのモーダルによる入力を効率的に使い分けができていたといえる。また、作業者は視線が提示されていることにより意図を事前に推測できるため、指示者からの指示を的確に理解することができることが示唆された。

5 実験 2: 一人称視点シナリオにおけるポインティングの効率化に関する実験

一人称視点シナリオにおいて、視線位置提示によりカメラが動いている状態でも、ポインティング位置を効率的に伝達できるかの評価を行う。そのため、提案手法として「視線提示のみ」と、対象手法として「ジェスチャ提示のみ」を比較する。本実験の仮説は「視線提示によりポインティングが効率化され、正確かつ迅速に共同注意を形成できる」である。被験者は指示者となり、固定の作業者に指示をする。被験者は20-30歳代の8名(男性6名、女性2名)である。

5.1 実験環境

本実験では図8(A)のように作業者の前に複数のブロックが配置されている環境下において、特定の対象物へのポインティングを、間違いなく迅速に伝達できるかを評価する。ポインティングの対象となるブロックは、赤いブロックを中心として楕円状に

18個並べられている。各ブロック間の距離は、15cm程度である。実験開始時、作業者は中心の赤いブロックを注視しており、指示者は指定されたブロックに共同注意を向けるよう指示をする。ブロックの指定は図8(B)のような一つのブロックを四角で囲んでいる写真を見せることにより、対象を指示者に伝える。対象となるブロックは、ランダムに選ばれる。ディスプレイは指示者の前に設置されており、視線を使う場合にはディスプレイと正対する状態で、ジェスチャを使う場合には体によるジェスチャカメラの遮蔽が起こらないようディスプレイから少し横に移動した状態で指示をする。以下に実験の手順を示す。

手順1: 指示者に対象となるブロックを見せる。

手順2: 指示者に実験の開始を伝える。指示者は視線またはジェスチャ(順番はランダム)による指示と同時に「これです」という発話をする。

手順3: 作業者は指示をもとに対象のブロックを探し「これですか?」と指示者に尋ねる

手順4: 指示者は正しければ「はい」と答え、ポインティングによる共同注意が形成されたとして実験終了となる。間違っていれば「いいえ」と答え、[手順3]に戻る。

手順4が終了するまでの間違えた回数とその完了時間を計測する。本実験では9回の指示を1セッションとし、その合計値を一人当たりの結果とする。被験者はそれぞれの指示方法による実験の前に、最初に3回の練習をする。また、作業者は固定であるため、実験を重ねることによる学習効果をなくすことを目的とし、実験開始前に1時間程度の練習セッションを経験している。

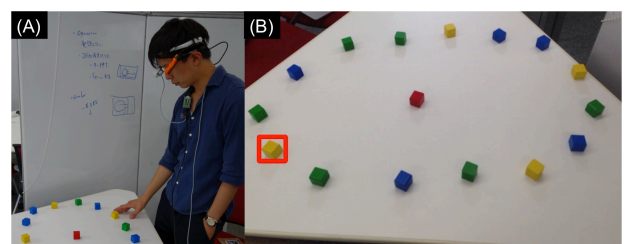


図 8. 実験 2 の環境：(A) 実験風景 (B) 被験者（指示者）へ最初に提示される写真

5.2 結果

図9に各被験者における、間違えた回数と完了時間の合計値を示す。全ての被験者において視線による提示の方が、間違えた回数が少なかった。間違えた回数の合計値に、マン・ホイットニーのU検定を適用したところ、 $p = 0.01$ となり2つの提示方法の結果に有意差があり、仮説が支持された。また、完了時間に関しても、マン・ホイットニーのU検定を

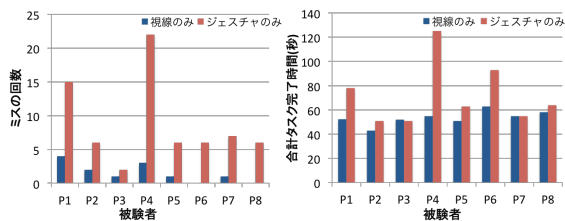


図 9. 実験 2 の結果

適用したところ、 $p = 0.03$ となり、こちらも 2 つの提示方法の結果に有意差があった。

5.3 考察

実験の結果から、視線提示を用いた方がジェスチャ提示よりも効率的にポインティングを伝達できたといえる。これは、作業者の頭部運動に伴いウェアラブルカメラの映像中の指示対象が移動している場合でも、視線ならば指示対象を継続的に追跡できるためである。そのため、HMD に提示される視線による指示位置も対象を捉え続けるため、作業者は容易に指示対象ブロックを推測できる。視線による指示の精度はアイトラッカーの精度依存しており、2 度未満である場合が多かった。一方、ジェスチャ提示の場合では、移動する対象を指示し続けることが難しいため、指示者は最初に提示されたジェスチャのみを用いて対象を探さなければならなかった。指示対象の候補が複雑に配置されていた場合、この傾向はより顕著になると考えられる。今後の実験では、視線による効率的なポインティングが、実践的なタスクにおいてどのように活用されるかを観察したい。

一人称視点システムでは作業者の視線追跡をしていないため、共同注意を形成できるのは作業者が視野中の HMD 提示領域に対象物を捉えたときのみである。そのため一人称視点システムは、指示者からの共同注意形成を効率化しているといえる。一方で、もし作業者の視線も追跡することができれば、作業者からの共同注意形成が可能となると考えられる。

6 関連研究

アイトラッカが高性能・安価になったことから、視線情報を用いたインタラクション技術の研究が盛んになりつつある。山中らは、注視していないことを利用したカーソル移動の高速化手法を提案し、作業が効率化されることを明らかにした [7]。また、Pfeuffer らにより視線情報を用いたタッチ入力の高速度の研究が行われている [8]。画面上のコンテンツに対しての共同編集を促進するために、Carletta らは画面に編集者たちの視線位置を提示するシステムを開発した [9]。本研究では実世界での作業を支援する二つの遠隔作業支援シナリオに着目し、指示者の視線位置を作業者に提示できる二つのシステムを

構築した。また、視線提示によって、実世界上へのポインティングが効率化することを明らかにした。

7 結論

遠隔作業支援シナリオとは実世界のタスクにあたる「作業者」を「指示者」が遠隔から支援するものである。本研究では、遠隔作業支援シナリオにおける視線位置提示の効果を明らかにするために、二つの実験システムを構築し、その検証をした。対象とした作業支援のシナリオは、固定視点での組み立てや電子工作などを支援する「固定視点シナリオ」と、ウェアラブルカメラを装着した作業者を支援する「一人称視点シナリオ」である。実験の結果から固定視点シナリオでは、視線提示は高速なポインティングを実現できること、作業者が指示者の意図を推測するのを助けることが示された。一人称視点シナリオでは、視線提示により特定の対象物へのポインティングによる共同注意の形成が効率化できることが示された。本研究の知見は遠隔作業支援システムの構築において、指示者側に固定型アイトラッカーを取り付けるだけで取り入れることができる。

参考文献

- [1] 葛岡 英明, 庄司 裕子. 空間型共同作業の評価方法の提案とその利用. 電子情報通信学会論文誌 A 基礎・境界 vol.77, no.6, 1994.
- [2] Kim, Seungwon, et al. Improving co-presence with augmented visual communication cues for sharing experience through video conference. ISMAR'14, 2014.
- [3] Kasahara, Shunichi, and Jun Rekimoto. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. AH'14, 2014.
- [4] Kurata, Takeshi, et al. Remote collaboration using a shoulder-worn active camera/laser. ISWC'04, 2004.
- [5] Fukuchi, Kentaro, et al. Pac-pac: pinching gesture recognition for tabletop entertainment system. AVI'10, 2010.
- [6] Kato, Hirokazu, and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. IWAR'99, 1999.
- [7] 山中 祥太, 栗原 一貴, 宮下 芳明. 注視していないことを利用したポインティング高速化手法とその評価, コンピュータソフトウェア (日本ソフトウェア科学会論文誌), Vol.30, No.3, pp.53-63, 2013.
- [8] Pfeuffer, Ken, et al. Gaze-touch: combining gaze with multi-touch for interaction on the same surface. UIST'14, 2014.
- [9] Carletta, Jean, et al. "Eyetracking for two-person tasks with manipulation of a virtual world." Behavior Research Methods 42.1, 2010.