

JackIn: 一人称視点と体外離脱視点を融合した人間-人間 オーグメンテーションの枠組み

笠原 俊一^{1,a)} 暦本 純一^{3,2,b)}

概要: 本論文では, JackIn と呼ぶ, 人間が他の人間の環境に没入し, その状況や体験を共有したり共同作業を行ったりする際のインタラクションの枠組みを提案する. このようなインタフェースは, 現場の作業を遠隔地の専門家が支援したり, 現地での体験を他の利用者に伝送する, 遠隔から作業者を誘導する, などの応用に適用できる. Body と呼ぶ, 実環境にいる利用者の一人称映像を遠隔地の利用者 (Ghost) が観測し状況を共有する. さらに, Ghost が自由な視点位置で環境を観測することを可能にするために, 一人称映像のシーケンスから SLAM に基づいて空間をモデリングし, 擬似的に Ghost の視点外から状況を観測することを可能にする「体外離脱視点」を提供する. これにより, Ghost は Body に追従しながらも独立な視点で環境を観測することができ, 一人称映像特有のモーションシックネス問題も解決することができる. また, Ghost が Body を効果的に支援できるように, Body の視野外にある事物をポインティングするための視覚化技法を提案する. 本論文ではこの発想に基づくシステムを試作し, JackIn インタフェースの有効性を利用者実験によって確認した.

JackIn: Integrating the First Person View with Out-of- Body Vision Generation for Human-Human Augmentation

SHUNICHI KASAHARA^{1,a)} JUN REKIMOTO^{3,2,b)}

Abstract: *JackIn* is a new human-human communication method that connects two or more people. With 1st-person view streaming from a person (called *Body*) wearing a see-through HMD with a camera, other person (called *Ghost*) intervenes into the shared 1st-person view. Using JackIn people's activities will be shared and assisted or guided by other people's expertise. Applications include several situations in daily life, such as teaching cooking remotely, shopping navigation, education in fabrication, and sharing experience with sports and live events. For better viewing experience for 1st-person view, we introduce the *Out-of-Body* view where 1st-person images are integrated to construct the scene around the Body, and the Ghost can virtually control the viewpoint to look around the space surrounding the Body and we also provide tele-pointing gesture interface. We conducted an experiment to evaluate how effective this architecture is, and found that Ghosts understand the spatial situation of the Body.

1. はじめに

遠隔コミュニケーション技術は地理的に離れた人どうしが会話したりビデオ映像を介して対面したりする機会を提

供してきた. さらに, テレプレゼンスあるいはテレイグジスタンスと呼ばれる技術によって, 遠隔地のロボットの視覚などの感覚を通じて, その場所にいるような感覚を伝送し, さらに遠隔地の物体を操作するようなインタフェースの研究が行われている [1].

一方, われわれは日常の生活で自分の体験を他人と共有したいという欲求や必要性を感じる場面は多い. たとえば旅行に行っていて感じている周りの情景をそのまま誰かに伝えたいという状況などである. 写真やビデオによってもある程度このような状況を伝達することは可能だが, 「その場にいるような」という感覚とはまだ隔たりがある.

¹ ソニー株式会社
Sony Corporation. 2-10-1 Osaki, Shinagawa-ku, Tokyo 141-8610, Japan

² ソニーコンピュータサイエンス研究所
Sony Computer Science Laboratories, Inc. 3-14-13 Higashi-otanda, Shinagawa-ku, Tokyo 141-0022, Japan

³ 東京大学大学院情報学環
III, The University of Tokyo

a) Shunichi.Kasahara@jp.sony.com

b) rekimoto@acm.org

GoPro [20] のような装着可能な広角ビデオカメラの登場によって、このような体験の記録や伝送がある程度可能になってきた。個人が体験した情景をそのまま他人に伝送する、一人称 (1st person) の情報伝送やストリーミングは今後の画像伝送の一つの方向性であると考えられる。

この発想を発展させると、一人称情報伝送を双方向に行うことが考えられる。たとえば、被災地で活動している人間が、専門家の能力を必要としたとき、自分の周辺状況を遠隔地の専門家に一人称情報として伝え、さらに専門家のアドバイスなどを得る、といった可能性である。いわば、専門家が自分の体の中に入り込んで状況を共有しながらインタラクションを行うことができるようになる。

このような、一人の人間の周辺状況を他者に伝達し、さらに状況を共有して他人が遠隔地から作業支援を含むようなインタラクションを行うことの総称を、本論文では **JackIn** と呼ぶことにする (図 1)。“jack in” は、もともとは William Gibson の小説 *Neuromancer* で、電脳空間 (cyberspace) に没入する行為の表現として用いられていた [19]。本論文ではこの考えを発展させて、人間が他の人間に没入する状況を Jack In と呼ぶことにしている。

以下の議論では、現場にいて状況を提供する側の人間を **Body**、ネットワーク経由でその状況を受け取り、作業指示を与える側の人間を **Ghost** と呼ぶことにする。たとえば災害地現場で Body が現場を歩き回り、その感覚や状況を (専門家である) Ghost がネットワーク経由で受け取り、指示を与える。Ghost は一人である場合、複数人である場合、多人数である (一人称映像のストリーミングを多くの視聴者が鑑賞する) 場合が考えられる。

JackIn は拡張現実感 (augmented reality, AR) の一種と考えることができる。Body は透過型 HMD (head-mounted display)などを介して Ghost の誘導や指示を見る。これは AR で一般的な重畳表示の一種である。ただし、通常の AR では付与される状況はコンピュータやデータベースによるが、JackIn では人間が他の人間を拡張 (augment) している点が異なる。

また、JackIn はテレプレゼンスの一種とも考えることができる。通常のテレプレゼンスはロボットのような機械に “jack In” して機械の視点から世界を眺めることができるインタフェースであるが、本研究が対象としているのは、人間が他の人間に JackIn する状況である。機械へのテレプレゼンスの場合、人間がマスターで機械がスレーブとなり、スレーブである機械は人間の動きを忠実に再現することを前提としている。一方、人間が人間に JackIn する場合、Body は Ghost に従って動くとは限らないので、両者の独立性を許すインタフェースをどう設計するかが従来型のテレプレゼンスとは異なる研究課題となる。

2. JackIn のユーザインタフェース

JackIn を実際に行う場合のユーザインタフェースについて、以下の二つの要素から検討する。

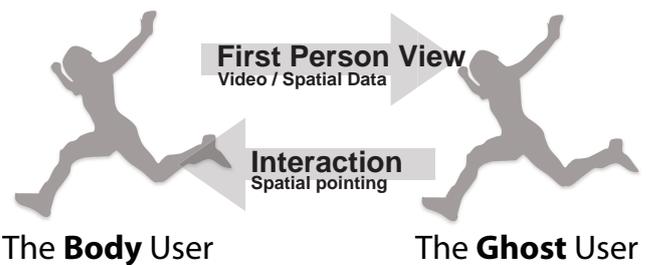


図 1 JackIn 概念図: “Body” とよぶ利用者の一人称映像を “Ghost” とよぶ遠隔地の利用者が共有する。Ghost は誘導や作業指示などを Body に対して行うことができる。

2.1 一人称映像のストリーミングと体外離脱視点の提供

JakIn の第一の要素は作業員 (Body) の状況の伝達である。本研究ではその第一段階として視覚情報の伝達に特に着目する。ビデオ映像を介して遠隔作業支援を行う研究は過去に多く行われている [4], [9]。たとえば遠隔地にいる専門家が作業員の状況を把握して作業指示を行うような場面が想定されていた。作業員の状況をより正確に把握するために、作業員からの視点 (一人称映像) を頭部搭載カメラによって取得し、伝送する事例もある [5]。一人称映像により、作業員が現場を移動する場合でも、作業員の見ている状況を伝送することが可能になる。

しかし、作業員 (我々の用語での Body) は自分の意思で周辺を見渡したり頭の方角を変えるので、その映像をそのまま観測すると、遠隔支持者 (Ghost) は揺れの激しい映像を見ることになり、モーションシクネス [15] を感じてしまう。また、Body が着目しているのではない別の箇所を Ghost が観測したい場合がある。たとえば作業に必要な道具を指示する場合、Body の現在の視野の外にある物を Ghost が指示したい状況などである。

この二つの課題を解決するために、本研究では一人称視点と体外離脱視点を統合したストリーミングを導入する。体外離脱視点とは、Body の視点以外の視点から現場を観測することを意味する。体外離脱視点により、Ghost は Body の視点や頭部方向に束縛されず現場を観測できることになる。体外離脱視点は、現場に頭部搭載以外のカメラを準備することによっても実現可能であるが、Body と共に移動する別のカメラを準備することは装備が大掛かりになり現実的ではない。

体外離脱視点生成の方法として、Body の周囲環境の 3 次元再構成を行なう手法が考えられる。例として、距離画像センサを用いて 3 次元再構成を行なう手法がある。 [12]。しかし、これらの距離画像センサを用いたアプローチは、屋外での使用には適さないという問題点がある。

また、過去の研究事例ではカメラ映像を二次元的につなぎ合わせた映像を提示し、遠隔支援者へ Body の周辺環境提示するものなどが提案されていた [7], [13]。しかし、遠隔支援や体験の伝送を考慮に入れた場合、前述のように Ghost による能動的な視点制御が必要である。

そこで本研究では、Body の一人称映像から擬似的に体外離脱視点映像を作り出す方式と体外離脱の視点移動に

より Ghost が Body の動きとは独立に周囲環境を観測できる方式を提案する。具体的には、SLAM(simultaneous localization and Mapping) 手法 [21] により、1 人称映像の連続画像から周辺の空間を擬似的に構築し、構築された空間を見る仮想カメラを Ghost が制御することで体外離脱視点を観測する。

2.2 作業員への遠隔指示

JackIn の第二の要素は遠隔者 (Ghost) と作業員 (Body) のインタラクションである。とくに、Ghost が何らかの指示を Body に与える場合のインタフェースに着目する。一般に遠隔コミュニケーションにおいては、作業員の周辺の事物を指し示す必要がある場合が多い。たとえば「そのボタンを押して」「その果物ではなく、こちらの果物を買って来て (スーパーマーケットで買い物の遠隔指示を与える場合)」など、「その」「これ」といった指示語を使う場合が多い。この場合、「その」が何を指し示しているのかを両者で共有する必要がある ([6])。

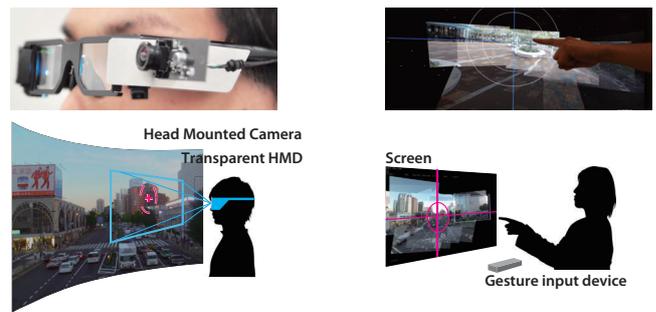
例として、Ghost が観測しているビデオのピクセル座標上に指示を行なっても、Body のカメラ映像は常に変化するので意味をなさない。指示した点が Body 側の空間上の 3 次元的座標となる必要がある。過去の研究事例では Ghost のハンドジェスチャーや指示を Body 周辺を捉えたカメラ映像に重畳するもの [2], [3], [8] が提案されていた。しかし、この方法では Body は実世界そのものに対しての指示を観測できない。また、Body の現在の視界の外ものを指示することが困難であった。

この二つの問題を解決する手段として、本研究では (Body に対してではなく) 空間に対して固定できるポインティングの手法と、Body の視野外に指示を与える手法を提案する。

3. JackIn アーキテクチャ

以下では、前説までに説明した JackIn の基本要素の具体的な実現手段について説明する (図 2)。

JackIn アーキテクチャにおける体外離脱視点は 3 つの主な特徴を持つ。1. 体外離脱視点は Body の一人称視点映像のビデオフレームの空間的なステッチングにより構成され、一人称視点映像よりも広い視野を提供する (図 3)。2. Ghost は体外離脱視点における視点のコントロールを、Body とは独立してジェスチャにより制御できる。3. Ghost は体外離脱視点を通して、Body の空間に対してポインティングでき、そのポインティングは Body が見る実空間に重畳されて観測。これらの機構は、一人称視点映像の空間認識 (SLAM) により求められる。実空間と Body のカメラ推定位置および実空間における画像特徴点の空間位置に基づいて実現される (図 2)。JackIn アーキテクチャのハードウェア構成は、Body 側における光学透過型の Head mounted display および頭部搭載カメラと、Ghost 側におけるジェスチャ検出デバイスとスクリーンから構成される。



The Body User The Ghost User

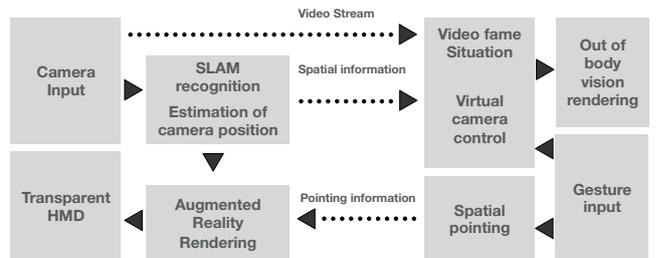


図 2 JackIn のアーキテクチャ構成。Body は光学透過型の Head mounted display および頭部搭載カメラを装着。Ghost はジェスチャ検出デバイスとスクリーンを用いる

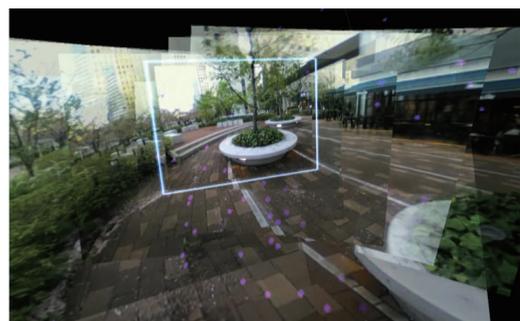


図 3 体外離脱視点は空間的に配置された過去ビデオフレームと、リアルタイムのビデオフレームから構成される。ここでは、図中の中央にあるフレームがリアルタイムのビデオフレームを表している。

3.1 体外離脱視点生成

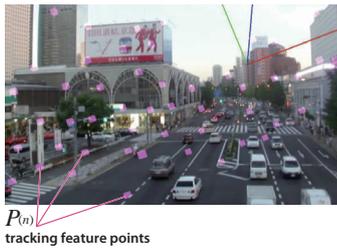
JackIn システムは、頭部搭載カメラの映像に対してリアルタイムに SLAM による空間認識を行い、現在のビデオフレームと過去のビデオフレームを空間的につなぎ合わせることで、空間に定位した仮想的に広い視野映像を生成する (図 3)。以下では 3 次元における位置・姿勢を表現する量として、16 次元のマトリクス表現を用いて説明をする。

空間が認識された場合、基準となる空間座標に対する Body の推定カメラの位置・姿勢 (M_{bc}) と認識された空間における画像特徴点の 3 次元点データ群 P_0, P_1, \dots, P_N が得られる (図 4 (a))。

ここで時刻 t における、ビデオフレームの空間的な配置位置 $M_f(t)$ は、 $M_f(t) = M_{bc}(t)T(d)$ として、定義される (図 4 (b))。 $T(d)$ はカメラの視線方向への平行移動量を示しており、頭部搭載カメラの画角と、画像特徴点の 3 次元点群データ P_0, P_1, \dots, P_N のうちカメラ視線中心に近い点 P' と Body のカメラ位置との距離 d から決定され、過去フ

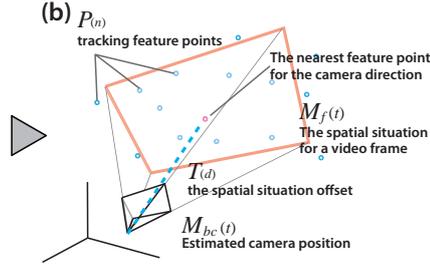
SLAM recognition

(a)



Video frame situation

(b)



Realtime spatial video stitching

(c)

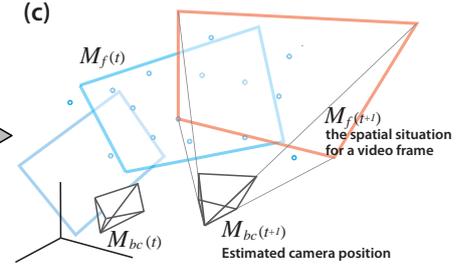


図 4 頭部搭載カメラの映像に対してリアルタイムに SLAM による空間認識を行い、現在のビデオフレームと過去のビデオフレームを空間的につなぎ合わせて空間に定位した仮想的に広い視野映像を生成する

フレームと現在のリアルタイムのフレームが接合するようにレンダリングされる。

毎時、既に配置されているビデオフレームと新しい配置するフレームを比較し、既存のフレームよりも一定時間以上新しく、他のフレームとの空間位置が離れている場合に、新規に空間へ配置する (図 4 (c)). このようにして空間に位置づけられた画像フレームは、後述する Ghost が設定する仮想視点から表示することができ、Body の体外に離脱した視点からの映像となる。つまり、体外離脱視点は Body が過去に見回した映像の蓄積であり、またその中で Body の現在の視点からの画像は常に最新の状態に保たれている。

3.2 体外離脱視点映像におけるインタラクション

体外離脱視点映像は過去ビデオフレームとリアルタイムのビデオフレームが配置されている仮想的な空間を、Ghost がコントロールする仮想的なカメラ視点 (M_v) からの映像としてレンダリングされる。空間に配置されたフレームのレンダリングでは、ステッチングのエラーによる視覚的な不整合を軽減するために、それぞれのフレームは仮想カメラの姿勢と配置されているフレームとの姿勢差分を元に透視性が決定しレンダリングされる。

空間認識が失敗した場合には、 M_v は一定時間維持されつつ、他の過去フレームは徐々に消失し、リアルタイムフレームのみが描画更新される。一定時間内に認識が復帰した場合には、過去フレームの描画を復帰し、一定時間以上認識が復帰しない場合には、空間認識を初期化し再び体外離脱視点の構築を開始する。

ここで、体外離脱視点におけるモードは大別して 3 つのモード；Follow モード、視点移動モードと Pointing モードから構成され、それぞれのモードは Ghost ユーザのジェスチャ操作によってコントロールされる (図 5)。

Follow モード： Follow モードは、Ghost は操作せずに体外離脱視点映像を見るモードであり、仮想カメラの位置姿勢は図 5 (a) のように決定される。SLAM により推定された Body ユーザのカメラ位置・姿勢 $M_{bc}(t)$ から、平行移動成分 $T_{bc}(t)$ 、回転成分 $R_{bc}(t)$ それぞれ k_T 、 k_R の係数として平滑化された、 $T'_{bc}(t)$ および $R'_{bc}(t)$ から $M'_{bc}(t)$ を得る。

$$T'_{bc}(t+1) = k_T T_{bc}(t) + (1 - k_T) T'_{bc}(t)$$

$$R'_{bc}(t+1) = k_R R_{bc}(t) + (1 - k_R) R'_{bc}(t)$$

仮想カメラ位置・姿勢 $M_v(t)$ は時系列的に平滑化した姿勢 $M'_{bc}(t)$ から、より広い範囲を見れるように一定のオフセット T_{offset} 分、仮想的に後方に移動した位置に毎時近づくように制御される。係数 k_v により、どの程度 Body ユーザのカメラに追従するか制御される。ここでは、仮想カメラ位置・姿勢 $M_v(t)$ から見たときには、リアルタイムフレームが取まっているときには、 k_v は小さい値 (つまり、仮想カメラは動かずに空間に定位した映像を見る) となり、リアルタイムフレームが $M_v(t)$ が捉える範囲外に移動した場合、 k_v は大きい値となり、リアルタイムのフレームを追うように制御される。

$$M_v(t+1) = k_v M'_{bc}(t) T_{offset} + (1 - k_v) M_v(t)$$

これにより、Ghost は Body のカメラの動きに対して平滑された動きで、空間的に定位しステッチされた状態でのリアルタイム映像と過去のビデオフレームを同時に観測できる。また、Ghost は Body よりも空間を広く見回せると同時に、Body の現在の視野がどこにあるかを知ることが出来るようになる。これらの処理はユーザの介入なく自動的に更新されるので、Body、Ghost とともに視野の設定制御をする必要がない。

視点移動モード： Ghost がスクリーンに対して、一定以上離れた距離でポインティングを行なった場合に、視点移動モードとなり、仮想カメラの視点移動を行なう (図 5 (b)). ユーザはジェスチャ検出デバイスから得たユーザの指位置を利用して、スクリーンへのポインティングを行なう。

スクリーンにおけるポインティング位置と体外離脱視点のフレームとのヒット判定から、Ghost ユーザのポインティングが指し示しているフォーカスフレーム (f') が決定される。そのフレームの位置 (M'_f) から一定のオフセット $T(d)_{offset}$ 分、仮想的に後方に移動した位置が仮想カメラのターゲット位置として決定され、仮想カメラはターゲット位置へアニメーションを伴い移動する。

このモードでは、Ghost が Body のカメラ方向とは独立して仮想カメラの方向をコントロールできる。なお、ここでのオフセット $T(d)_{offset}$ は、Follow モードに比べてより

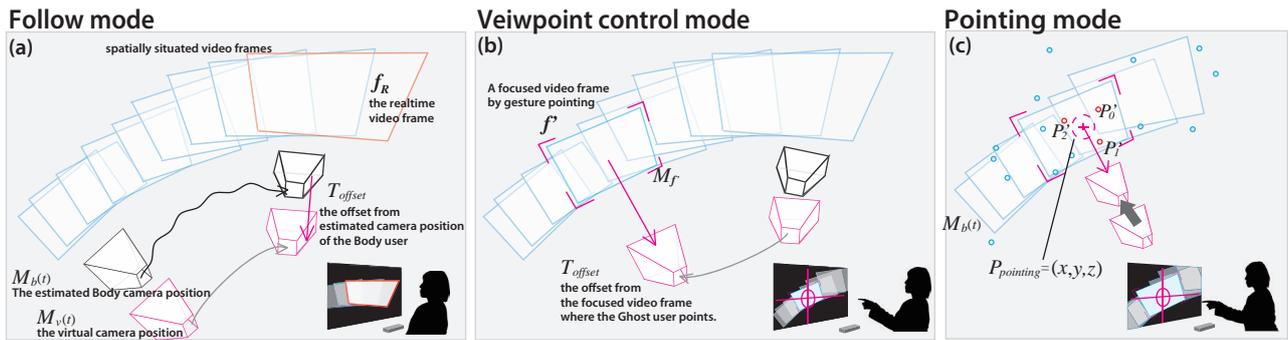


図 5 Follow モードにおける Ghost 視点の移動. Ghost はジェスチャ入力により, 体外離脱視点映像の視点をコントロールする.

後方に移動するように設定される. これにより, 移動中により広い視野での周囲の観察が可能になる.

Pointing モード: ジェスチャ検出デバイスから得られたユーザの指の位置が, 一定以上スクリーンに近づいた場合に, Pointing モードが開始される.

Pointing モードでは, 仮想カメラの制御は視点移動モードと同じアルゴリズムで決定されるが, フォーカスフレームにズームして見える位置に仮想カメラの位置が制御される. これにより, Ghost はよりフォーカスフレームを細かく観察できる.

さらに, このモードでは Body 側の空間に対して固定されるポインティングを行なう (図 5 (c)). スクリーンにおける 2 次元のポインティング座標と, 画像特徴点群のスクリーン上での 2 次元位置を比較し, ポインティング座標と 2 次元上で最近傍となる 3 点 P'_0, P'_1, P'_2 を画像特徴点群データ P_0, P_1, \dots, P_N から選択する.

ポインティングの 3 次元的な座標はこの 3 点が張る平面上の点と仮定して, ポインティング座標 $P'(x, y, z)$ を下記のように求める.

$$P'(x, y, z) = a(P'_2 - P'_0) + b(P'_1 - P'_0)$$

ここで, a と b を求めることで, 3 点が張る平面上の点が得られる. これにより, Ghost はビデオフレームのピクセルにポインティングをするのではなく, ビデオフレームか

ら認識された空間座標における 3 次元の点として, ポインティングできる.

3.3 Body / Ghost 間のインタフェース

Body は前述の Ghost が指定した空間的なポインティングを, 透過型 HMD を通じて実空間に重なった状態で見ることになる. 重畳映像と肉眼の関係は利用に先立ってキャリブレーションされている. 図 6) は Body がシースルーグラス越しに見る状態の例を示している.

現状で入手できる透過型 HMD の視野角には限界があり, 肉眼の視野に対する透過型ディスプレイの表示領域が限られており, かつ Ghost は体外離脱視点映像により, より広い視野を見ている. したがって, Ghost が空間に対して指示を出したり空間の事物を指定する場合には, 単に透過型ディスプレイに情報を重畳できるだけでは不十分で, 透過型 HMD の「視野外」情報を指示するインタフェースが必要である.

そこで JackIn アーキテクチャでは, 携帯電話などの小型ディスプレイ用の視覚化技法である Halo [14] を応用した AR ナビゲーションインタフェースを実現している. Halo は, 地図ブラウジングなどを対象としている技法で, スクリーン外にある地点を示す場合に, その地点を中心点とする円弧をスクリーンに描く. 地点が画面外にある場合でも, 表示される円弧の方向と曲率から, スクリーン外のおおよその位置を直感的に知ることができる.

空間ポインティングが HMD 表示領域外にあり, HMD 表示領域中心からの距離がある一定以上の距離にある場合, その方向を示す矢印が表示される (図 7- a).

Body がその方向に頭を動かし, 指示した点が視野の外ではあるが一定の範囲内に納まった場合は, 空間ポインティングの位置と HMD 表示領域中心との距離が半径となる円が描画される (図 7- b).

Body は, 円弧の中心が視野内に入るように頭を動かして行き, 指示された点を確認することができる. 指示点と視界のずれに応じて円弧の半径が変化していく (視界に近いほど半径が小さくなり, 円弧の曲率が増す) ので, Body は直感的に視野外にある指示点の位置を認識することができる (図 7- c).

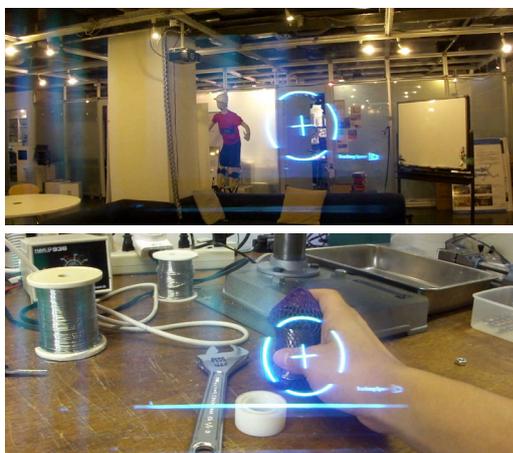


図 6 Ghost による遠隔ポインティング: Body は透過型 HMD を介してこの表示を観測する.

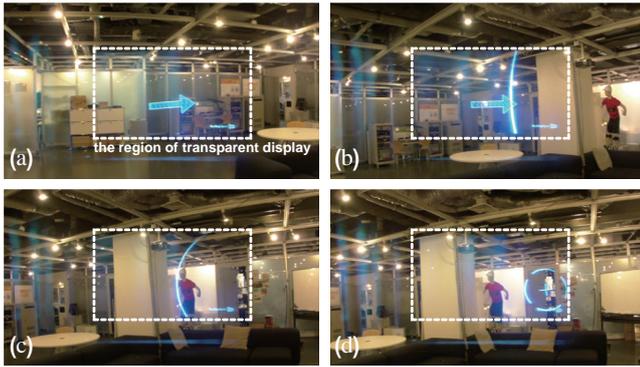


図 7 視野外の事物のポインティング

・最終的に指示点が視野内に納まったときには別のインディケータに変化し、目標点の位置を示す(図 7-d).

上記のような、グラフィックフィードバックのナビゲーションを通して、Body ユーザはアノテーションが HMD 表示領域外に存在していても、空間的に位置を知覚し指示の場所を知ることができる。

4. システム実装

使用した HMD は LUMUS DK-32 [11] で画素数が 1280×720 、対角視野角は 40° だった。頭部には水平視野角 80° のカメラを搭載して一人称視点画像を取得している。この HMD とカメラは Body が携帯するラップトップコンピュータに接続され、そのコンピュータ上で SLAM 処理による空間認識を行っている。Body 側のコンピュータから Ghost 側へは画像を motion jpeg で圧縮したものを無線伝送し、VGA(画素数 640×480) の画像を 20fps のレートで伝送している。無線 LAN を介した伝送遅延は 400msec 程度である。SLAM 処理には SmartAR ライブラリ [10] を利用している。

Ghost サイドでは、ユーザーはジェスチャ入力デバイスとして Leap motion [18] を用いる。また、体外離脱視点映像は 40 インチの TV スクリーンに描画される。Ghost サイドの JackIn アプリケーションは、Body ユーザ側から提供されたビデオフレームと空間認識の結果より体外離脱視点映像生成と、仮想視点のコントロール、およびポインティング情報の送信を行なう。それぞれの情報は Wifi 経由で UDP プロトコルを通して交換される。

5. 利用者評価実験

JackIn システムにおける体外離脱視点映像の有効性を検証するため、ユーザーテストを行なった。さらに、体外離脱視点を含む JackIn システムが、どのように Ghost と Body 間のコミュニケーションを変化させるかを観察するため、Body が装着するカメラの方向変化と Ghost の視点移動操作の記録、および実験中のビデオ録画から観察を行なった。

タスクデザイン: 本実験のため、図 8 のような使用するべき道具が散在している実験室において、Ghost が Body に



図 8 利用者評価実験環境

使用するべきアイテムを指示する場面を想定した遠隔指示を実験のタスクとした。対象アイテムは、Ghost も Body の言語表現による特定が困難な、半組み立て状態のレゴブロックとした。ここで、エキスパートとしての Ghost は指示するべき道具がどこにあるのかに関して未知である。

それぞれの被験者は Ghost 側として、別室にいる Body の周辺にある複数のアイテムから指示をしたアイテムを取得するように指示をする。Ghost 側では、1) 体外離脱視点を含んだ JackIn システム、2) 体外離脱視点を含まない(一人称映像のみ) JackIn システムの二つの条件に対して比較を行なう。なお 2つの条件において、体外離脱視点以外の機能は同じとした。Ghost と Body ユーザの組み合わせにおける前提知識による分散を少なくするために、実験を通して一人の被験者を Body とした。ここで、Body ユーザは行なわれている実験の条件は知らされない。

実験条件と手順: 実験では Body 側の実験室のワークスペースに、固有 ID を持つ 12 個の半組み立てレゴがランダムに配置されている(図 8)。また、全く同じアイテムが Ghost 側にも用意される。1 回のタスクでは、Ghost が実験オペレータから指定された一つのレゴアイテムを手元で参照しつつ、JackIn システムを使用して Body に同じアイテムを手取る様に指示をする。Body が Ghost から指定されたターゲットを取得した時点でタスク終了とする。

それぞれの実験条件につき 3 回のタスクを行う。実験における条件の順序はランダムに決定し順序による影響を考慮した。本実験においては、20 歳~40 歳、日常的にコンピューターを使用している 10 人の被験者を採用し、実験条件に対して被験者内配置とした。被験者の内 5 人は Microsoft Kinect や Leap Motion などの空間ジェスチャデバイスを使用した経験がある被験者であった。実験の後にシステムにおけるアンケートと、アンケート回答内容におけるインタビューを行なった。

実験結果: 表 1 はアンケート項目それぞれに対する結果を示して。各アンケートの質問項目は、Q1:”対象のアイテムを用意に見つけれられたか”、Q2:”遠隔地の状況を把握できたか”、Q3:”システムを簡単に使用できたか”と Q4:”作業しやすかったか”と設定した。各項目に対するの回答は、同意した場合は 7 を、同意できない場合には 1 とした。回答における、中央値、標準偏差および同意側(5 以上)への回答割合(表中では%と示す)の結果を表 1 に示す。特に差が観察された結果として、Q2”遠隔地の状況を把握できたか”において体外離脱視点なしでは 10 人中 4 人のみ同

表 1 各アンケート項目に対する結果

質問項目	体外離脱視点を含まない			体外離脱視点を含む		
	Mean	SD	%	Mean	SD	%
Q1	5.0	1.48	6/10	6.0	1.07	8/10
Q2	4.0	1.35	4/10	6.0	0.63	10/10
Q3	6.0	1.08	9/10	5.0	1.15	7/10
Q4	5.5	1.16	8/10	5.5	0.70	10/10

意側に回答しているのに対して、体外離脱視点ありでは全員が同意する側に回答した。

このアンケート結果における Q2 の顕著な差より、体外離脱視点により Ghost はより Body の周囲空間を把握できることを確認できた。また、このアンケート項目に対してのインタビューで、8 人の被験者は” 周囲環境が把握できるため、より安心して映像を見ることができた” とコメントした。また、” 複数のアイテムを比較してより目標アイテムに似ている方を指示したい時に、Body とは独立して視点を動かして映像中のアイテムを確認し指示できたため、より自信を持って会話できた” といったコメントも得られた。

インタビューの中から約半数の被験者より” ジェスチャー入力操作のためには、もうすこし練習が必要であった” とコメントがあった。これはジェスチャー入力の習熟度、ポインティングモードにおける 2 次元の座標からの 3 次元座標への変換でのエラーが影響していると考えられる。また、タスク完了時間は、タスク条件に関わらず 5 秒から 120 秒までの大きいばらつきを持った。これらは、ビデオと会話の記録より、ターゲットアイテムの配置場所とアイテムの配置されている姿勢に大きく依存していると観察されたため、定量的な評価に用いることができないと判断した。

インタラクションシーケンスの観察と考察: 我々は、JackIn システムがどのようなインタラクションを発生させるかを観察するために、Ghost と Body の様子を記録した動画、Ghost の制御する仮想カメラ方向と Body の頭部搭載カメラの方向の記録を照らし合わせた。

仮想カメラと Body のカメラの動きのアクティブ性を解析するために、我々は毎フレームごとにカメラ方向の変化量をプロットした。それらは、図 9 の (a) のグラフに対応する。この量が大きいほど、その時刻におけるカメラの動きが大きいと観察できる。また、毎フレームにおける、仮想カメラと Body の頭部搭載カメラの方向の差の絶対値をプロットした。それらは、図 9 における (b) のグラフに対応する。この量が大きいほど、Ghost と Body が異なる方向を向いていると分かる。

図 9 は実験を通して、多く観測されたインタラクションの例を示している。タスク開始の 2 秒後から Ghost はそれらしいアイテムを探し、即座に仮想カメラをコントロールし、ラベル A の時刻においてポインティングジェスチャーを行い、Body にアイテムを取るよう指示をした。

ポインティングの指示の後に Body は即座に頭の向きを変更し、ラベル B において、Ghost の仮想カメラ位置と Body の頭の方向が一致し、Ghost は指定したアイテムが正しいかを Body のリアルタイム映像から確認した。しか

し、Ghost は指定したアイテムが異なる物であると発見し、Ghost は Body にアイテムの特徴を伝え、Ghost と Body は共に独立的に見回して探索を開始した。ラベル B からラベル C まで、グラフ (a) における連続して発生する小さなピークと、グラフ (b) における比較的大きな値から観察できるように、Ghost と Body は互いに独立して周囲を探索した。

その中で、時刻 17-20sec, 30-34sec においては、Ghost は仮想カメラを余り動かさず、ほとんど同じ方向を見続けて、体外離脱視点における過去フレーム中に映っているアイテムに集中し、手元にあるアイテムと映っているアイテムを比較している場面も見受けられた。その後、ラベル C の時刻において、Ghost が別のもっともらしいターゲットを発見し、ポインティングを行なうと、グラフィックガイダンスによって Body もその方向に頭部方向を合わせ、アイテムを確認して、答えであると確認をした (ラベル D)。このようなインタラクションのシーケンスは実験において多く見受けられた。このように、体外離脱視点により Ghost が独立して視野空間を探索する事象が観察された。

先行研究における議論からも示されるように、一人称視点映像と体外離脱視点映像の両方の条件において、遠隔ポインティングを伴って、Ghost による” これ”、” あっち” や” こっちこっち” などの指示語が観察された。これにより、Ghost と Body は空間的な共通認識を確立している様子が観察された。例として、両者がポインティングして確認したアイテムが間違った物であった場合、” そこから左” のように、既に共通認識として確立している空間的な参照点からの表現で他のアイテムへの指示を行っていた。

また、多くの被験者はポインティングと言葉による指定を組み合わせて、ポインティングの位置が多少ずれていたとしても、共通認識した場所からの相対位置などを指定することで目的を達成している様子が観察された。このように、実空間へのポインティングはある一点を指し示すだけではなく、ある空間的な共通認識を確立するためにも有効であると観察された。

指示語について、二つの実験条件で表現の差が存在する点が観察された。体外離脱視点映像でポインティングを行なう場合、Ghost は” これ” のような位置を示す指示語だけではなく、” こちら” のように、方向を示す指示語を用いている様子が観察された。この 2 つの使い分けにおいて興味深い点は、そのポインティングの点が、リアルタイムの映像フレーム内に存在するか否かで使い分けられているように観察された点である。Ghost としての被験者はリアルタイムの映像フレームが実際に Body が見ていてかつ拡張表示されている領域であると捉えて、上記のような指示語の使い分けを行なったのではないかと考えられる。

このように実験結果および観察より、Body の身体運動で動いてしまう一人称視点映像が空間に定位されることで空間把握が用意になり、モーションシックネスの改善につながることで、Ghost が Body の頭部の動きと独立して、遠

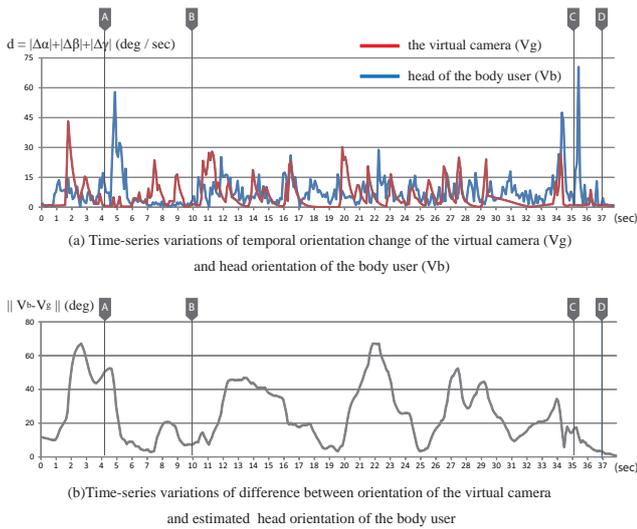


図9 Bodyが装着するカメラの方向とGhostの仮想カメラ方向の時系列変化のプロット(a), および仮想カメラとBodyの頭部搭載カメラの方向の差の絶対値の時系列プロット(b)

隔地の状態を探索できることが観察された。

6. 結論と今後の展望

本論文ではJackInと呼ぶ、人間が他の人間の状況を共有する際のインタラクションの枠組みを提供した。一人称映像およびそれから生成される体外離脱視点を介して他人の状況を観測し、遠隔地から利用者を誘導するインタフェースについて報告した。利用者評価実験によって体外離脱視点による効果を明らかにした。今後はさらに現実的な状況において有効性を確認していく予定である。

JackInは人間(Body)の能力を他の人間(Ghost)が強化拡張するという可能性を持っている。今回のインタラクションは実環境での誘導やポインティングなどをまず対象としているが、能力拡張という観点からは他の操作も支援対象となりうる。たとえば道具を持つ方法や操作する方法などを遠隔から指示する際のインタラクションなどはまだ検討の余地がある。今回のシステムは、人間を主に視覚情報(や音声)によって誘導していたが、人間-人間の接続インタフェースの可能性は他の感覚にも展開できる可能性がある。たとえば振動フィードバック、仮想力覚[22]や筋電刺激[17]などによってBodyに対して身体的な誘導を与える可能性も検討したい。また、今回は主にGhostが一人の場合のインタラクションについて検討したが、Ghostが複数ないし多人数の場合の集合知的なサポートにも可能性があると考えている。今後の課題としたい[16]。

参考文献

[1] Tachi, S., Arai, H., Maeda, T., Oyama, E., Tsunemoto, N. and Inoue, Y., "Tele-existence in real world and virtual world," *Advanced Robotics*, 1991.193-198
 [2] Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Giuliano Macioccoi. 2013. BeThere: 3D mobile collaboration with spatial input. In *Proc. CHI 2013*. ACM Press (2013), pp.179-188.

[3] Tecchia, F., Alem, L. and Huang, W. 2012. 3D helping hands: a gesture based MR system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '12)*. pp.323-328.
 [4] Kuzuoka, H., Spatial workspace collaboration: a Shared-View video support system for remote collaboration capability, *CHI 1992*, pp.533-540, 1992.
 [5] Fussell, S.R., Setlock, L.D. and Kraut, R.E. 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks, *CHI 2013*, pp. 513-520. 2013.
 [6] Bauer, M., Kortuem, G. Segall, Z. "Where Are You Pointing At?" A Study of Remote Collaboration in a Wearable Videoconference System. *ISWC 1999*, pp.151-158, 1999.
 [7] Kurata, T., Sakata, N., Kouroggi, M., Kuzuoka, H. and Billingham, M. 2004. Remote Collaboration using a Shoulder-Worn Active Camera/Laser. In *Proc. ISWC 2004*. 62-69.
 [8] Gauglitz, S., Lee, C., Turk M. and Hllerer, T. Integrating the physical environment into mobile remote collaboration. In *Proc. MobileHCI '12*. ACM Press (2012) 241-250.
 [9] Ou, J., Fussell, S.R., Chen, X., Setlock, L.D. and Yang, J. Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In *Proc. ICMI '03*. ACM Press (2003), 242-249.
 [10] SmartAR (<http://www.sony.net/SonyInfo/News/Press/201105/11-058E/index.html>)
 [11] LUMUS (<http://www.lumus-optical.com/>)
 [12] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A. and Fitzgibbon, A. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proc. UIST '11*. ACM Press (2011) 559-568.
 [13] L. Cheng and J. Robinson, Dealing with speed and robustness issues for video-based registration on a wearable computing platform. In *Proc. ISWC98*, pages 84.91, 1998.
 [14] Baudisch, P. and Rosenholtz, R. Halo: a technique for visualizing off-screen objects. In *Proc. CHI '03*. ACM Press (2013) 481-488.
 [15] Benson, Alan J. (2002). "Motion Sickness". In Kent B. Pandoff and Robert E. Burr. *Medical Aspects of Harsh Environments 2*. Washington, D.C.: Borden Institute. pp. 1048-1083. Retrieved 4 Dec 2012.
 [16] Goldberg, K., Song, D., Khor, Y., Pescovitz, D., Levandowski, A., Himmelstein, J., Shih, J., Ho, A., Paulos, E. and Donath, J. Collaborative Online Teleoperation with Spatial Dynamic Voting and a Human "Tele-Actor" In *Proc. ICRA '02*. 1179-1184.
 [17] Tamaki, E., Miyaki, T. and Rekimoto, J. PossessedHand: techniques for controlling human hands using electrical muscles stimuli. In *Proc. CHI2011*, ACM Press (2011) 543-552.
 [18] Leap Motion (<https://www.leapmotion.com/>)
 [19] William Gibson, *Neuromancer*, Ace Science Fiction, 1984.
 [20] GoPro Official Website, <http://gopro.com/>
 [21] Leonard, J.J. and Durrant-whyte, H.F., Simultaneous map building and localization for an autonomous mobile robot. *Proceedings IROS'91*. IEEE/RSJ International Workshop, pp.1442-1447, 1991.
 [22] Rekimoto, J. "Traxion: A Tactile Interaction Device with Virtual Force Sensation", *ACM UIST 2013*, pp.427-432, 2013.