16INT004
2016/3/2

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

小宮山 凌平[1,a)]　味八木 崇[1,b)]　暦本 純一[1,2,c)]

概要：従来のテレプレゼンスシステムを用いた遠隔共同作業には，利用者がある特定の視点からの一人称映像しか利用できず，遠隔作業現場の状況を把握しづらいという問題があった．その解決手段として，本研究では一人称映像と三人称映像の間を自由に移行できる枠組み *JackIn Space* を提案する．遠隔作業者 (サロゲート) の頭部に取り付けた広角一人称カメラと遠隔環境に取り付けた複数のデプスセンサを利用することで，一人称映像からなめらかに体外離脱視点である三人称映像へと遷移することを可能にし，サロゲートとつながった利用者は仮想的に現場周囲の環境を見回すことができる．また，利用者は遠隔環境にいる別のサロゲートの一人称映像にも入り込むことが可能で，環境を別の視点からも観測することができる．本提案の有効性を検証するため，本論文では作成したプロトタイプシステムを用いて評価実験をおこなった．この結果から，提案システムがより自然な視点選択を提供し，より効果的な遠隔共同作業を支援することが示された．

JackIn Space: Supporting the Seamless Transition Between the First and the Third Person's View for Effective Telepresence Collaborations

Ryohei Komiyama[1,a)]　Takashi Miyaki[1,b)]　Jun Rekimoto[1,2,c)]

Abstract: Traditional telepresence systems only supported first person's view and users had difficulty in recognizing the surrounding situation of the remote workspace. The *JackIn Space* concept addresses this problem by seamlessly integrating the first person's view with the third person's view. With a head-mounted first person camera and multiple depth sensors installed in the environment, the surrogate user's fist person vision smoothly changes to the out-of-body third person's view, and the user who connects to the surrogate user can virtually look around the environment. The user can also dive into the other remote user's first person view to look at the environment from the different perspective. Our evaluation supports that out prototype system provides more natural view position selection, and thus supports better remote collaborations.

1. はじめに

テレプレゼンスとは，遠隔地にいる作業者 (人やロボット，以下「サロゲート」) に没入的に接続する遠隔コミュニケーションの方法である．典型的なものには，サロゲートロボットに利用者が接続するものがある [1]．利用者が，まるで自分がロボットになって遠隔地にいるかのように感じられるように，ロボットから得られる視聴覚情報は利用者に伝達され，利用者の動作はロボットで再現される．この概念は，人を物理的な距離による制約から解放し，遠隔医療や危険地帯の検査・機械操作といった様々な応用を可能にする．

近年，人間と人間をつなぐテレプレゼンス方式の提案によって，この概念は拡張されている [2]．典型的には，ヘッドマウントカメラなどのセンサをつけた人間から，そのカメラから得られる一人称映像が遠隔地にいる共同作業者に伝送される．例えば，実際の現場にいる作業者は，周囲の

[1] 東京大学大学院情報学環
The University of Tokyo, Tokyo, Japan
[2] 株式会社ソニーコンピュータサイエンス研究所
Sony Computer Science Laboratories, Inc, Tokyo, Japan
[a)] komikomi.mikomiko@gmail.com
[b)] miyaki@acm.org
[c)] rekimoto@acm.org

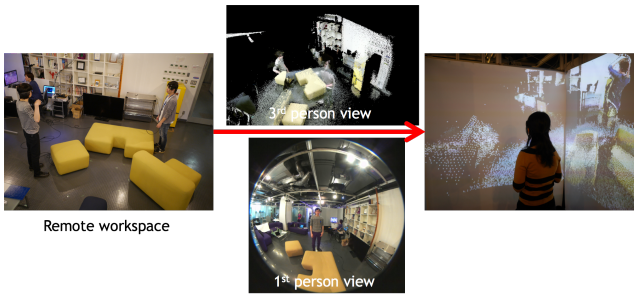

図 1 *JackIn Space*: 一人称・三人称映像間の連続的な遷移が可能なテレプレゼンス環境

状況を伝送することで遠隔地にいる専門家の技能支援を受けることができる．この場合，ロボットではなく人間が，別の人間のためのサロゲートとなっている．

どちらの場合も，既存のテレプレゼンスシステムでは，主に一人称映像と呼ばれるサロゲート視点の映像のみを利用していた．しかし，一人称映像が唯一の視覚情報である場合，いくつかの欠点が残る．

第一に，サロゲートの動きによる映像酔いが引き起こされる．これは人間-人間間のテレプレゼンスで特に顕著である．作業現場のサロゲートが自ら環境を見回すとき，一人称映像は頭部動作により激しく揺れてしまい，遠隔地の利用者は快適に映像を見ることができない．この問題は，ヘッドマウントカメラに全周囲カメラを利用し，映像に安定化処理を施すことで解決出来る [3], [4].

次に，空間の把握に関する問題が挙げられる．実際に同じ現場にいる場合の共同作業では，利用者は周囲の人間，もの，さらにはそれを取り囲む環境との位置関係を簡単に把握することができる．例えば，機械整備の共同作業をしている場合，我々はその空間内を自由に歩き回ることで現場の状況を理解できる．一方で，テレプレゼンスを用いた共同作業の場合，遠隔地にいる人間は歩き回る自由が制限されてしまうため，サロゲートの視点からしか作業をすることができず，周囲の状況を把握することは難しい．もしその場に二人以上の人間がいたり別のテレプレゼンス手段 (他のロボット，ドローンや環境に取り付けられたカメラ) があった場合，それらの視点を「渡り歩く」ことができれば環境の把握に非常に役に立つ．しかし，ある視点から別の視点に突然切り替わるだけでは，視点間の空間的な位置把握が困難になると予想される．

これらの課題に対処するために，本論文では一人称映像と三人称映像，さらには作業環境内の様々な視点における映像間の連続的な遷移を可能にする *JackIn Space* を提案する (図 1)．この概念の実現には遠隔地の人間や作業環境の三次元情報を取得する必要があるが，その手段として，複数台のデプスセンサ (Kinect 2) を利用し情報を再構成することで，一人称映像から三人称映像へとなめらかに遷移することが可能なシステムを実現した．三人称映像への遷移が完了すると，現場の状況を把握するために，視点を自由に動かすことが可能になる．また，この視点移動により，再び元の一人称映像に入り込んだり，さらに別の人間やロボットの一人称映像へと移行することも可能である．

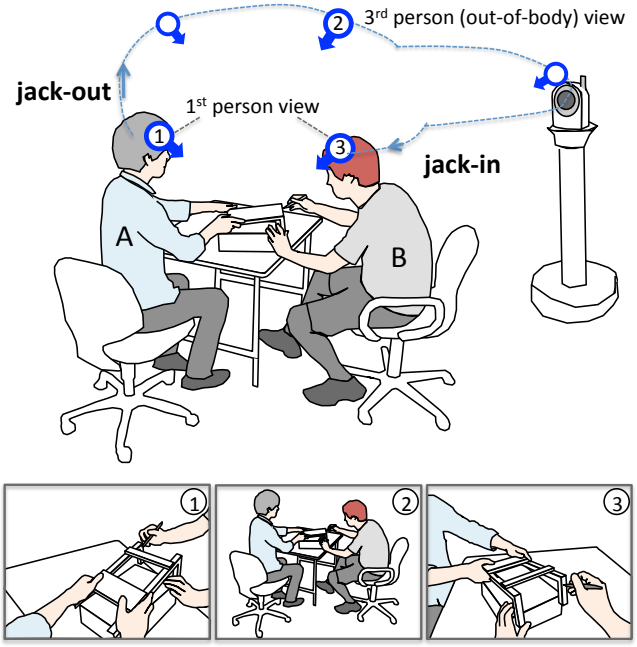

図 2 *JackIn Space* によるテレプレゼンスの概念: (i) 遠隔地の *ghost* は，*body*A の視点における一人称映像を見ている (1). (ii) *jack out* 操作を行うと，*ghost* の視点が *body*A の視点から離れ，三人称映像を見ることができる (2). (iii) *ghost* は自由に視点を変えることができ，他の *body* やテレプレゼンスロボットに *jack in* し直すこともできる．(iv) *ghost* は次の *body*B に狙いを定め，*jack in* 操作を実行する．(v) *ghost* の視点が *body*B の視点へと移動する (3).

2. JACKIN SPACE

JackIn Space は一人称・三人称映像間の連続的な遷移を支援するフレームワークである (図 2)．このフレームワークの中で，*jack in* は遠隔地のサロゲート (人間やロボット) とつながることを意味し，*jack out* は *jack in* 状態から抜け出すことを意味する．これらの造語は，1984 年の William Gibson による SF 小説で，Cyberpunk [5] の代名詞的作品である “Neuromancer” に由来する．ただし，本来の *jack in* が主に電脳空間への接続を意味するのに対し，我々はこの用語を拡張し，テレプレゼンス環境への接続の意味を含めて用いる．

加えて，実際の現場で作業をするサロゲート (人間やロボット) を *body*，遠隔地からネットワークを通じて *body* につながる (*jack in* する) 人間のことを *ghost* とそれぞれ呼ぶことにする．

従来のテレプレゼンスシステムでは，基本的には遠隔地

のサロゲートをひとつしか想定していないので，*jack out* 操作について明確には議論されていない．一方で我々は，*jack in* と *jack out* の操作を可能にすることが，従来のテレプレゼンスの概念に様々な利点を付加することになると考えている．

第一に，遠隔地の人間 (*ghost*) が現場の状況をより容易に把握できるようになる．一人称映像を利用することで，利用者はまるで本当に遠隔地に存在しているかのように感じられるが，周囲の状況を理解するためにはさらに別の手段が必要となる．この場合，一人称視点から抜け出すこと (*jack out*) が役に立つ．また，2 つ以上のサロゲート（*body* ユーザー，テレプレゼンスロボット，環境に設置されたカメラなど）が利用可能であった場合，ある視点から別の視点への連続的な遷移も，状況の把握に役立つ．例えば，被災地の調査をする場合，遠隔地にいる専門家は複数人の現場作業員それぞれに *jack in* することができ，さらに三人称視点を得るために *jack out* することもできる．

このフレームワークを実現するため，プロトタイプシステムでは，三人称映像取得のために複数の Kinect 2 [6] を環境に設置し，一人称映像取得のためにそれぞれの *body* がヘッドマウントカメラを装着している．*body* の頭部位置がトラッキングされているため，仮想的に一人称・三人称映像間の連続的な遷移が可能となる．

3. 関連研究

3.1 実世界のキャプチャと再構成

複数のセンサを用いて実世界をキャプチャし再構成する方法は，*virtualized reality* の概念が提唱されて以来 [7]，長い間研究されてきた．遠隔地のオフィス間を接続し，1 つの三次元空間として統合した “The Office of the Future” は，仮想現実を用いた遠隔コミュニケーションの一つの例である [8]．これらの考えはまた，多くの SF 小説で繰り返されてきた．例えば，A．C．Clarke は “都市と星” の中で，ダイアスパーと呼ばれる三次元複製された遠い未来の都市を表現している [9]．この複製された都市空間では，どの時代のどんな場所をも見回すことができる．これに似たアイディアは “Déjà-Vu” という映画にも登場し，この世界では，現実の全ての都市がリアルタイムでキャプチャ・アーカイブされ，システムの利用者は都市に対して 4 次元的な視覚を持つことができる [10]．また，Gelernter は現実世界の複製である “Mirror Worlds” の概念を提唱していて，その世界が情報空間と関わる主要な方法になると著している [11]．現代のセンシング技術と計算能力とを統合することで，これらの “実世界の複製” のアイディアは実現可能となり，テレプレゼンスはサロゲートの視点に縛られることがなくなるはずである．

近年，RGB-D(depth) センサの低価格化により，複数台の Kinect を用いた現実世界の仮想化が注目されている [12]．Kinect 2 は複数台使用しても互いに干渉しづらい time-of-flight 方式のセンシング方法を採用しているため，広範囲の安定した距離画像の取得が可能であり，このようなアプリケーションに適している．

3.2 一人称と三人称のテレプレゼンス

一人称視点の人間-人間間テレプレゼンスは，従来の人間-ロボット間テレプレゼンスに代わる新しい方法として注目を集めている [2], [3], [4], [13]．この場合，現場の人間（我々の用語における *body*）はヘッドマウントカメラを身につけ，自らの視点における映像をキャプチャし伝送する．また，頭部の回転による映像の揺れを補正するために，映像の安定化も行われる．こうすることで，遠隔地で映像を見る人間 (*ghost*) の視野を *body* の頭の動きから分離することができる．*JackIn Space* では，さらに *ghost* は視点位置さえも *body* と独立して選択できるため，この分離の自由度を拡張していると捉えることができる．

Time Follower’s Vision [14] は， ロボットの仮想的な三人称映像を提供することで，遠隔地のロボット操作者を支援する．この映像は複合現実感技術を用いて作成されている．ロボットが撮影した現在の画像を，過去の画像に重ね合わせることで，環境に三人称視点カメラを取り付けることなく，操作者が仮想的に合成された三人称映像を見ながらロボットを操作できる．

スポーツトレーニングの分野では，三人称映像を記録することの大切さが広く認識されている．いくつかのシステムは体外離脱視点からの映像を利用するため，カメラを取り付けたドローンや棒を使ってトレーニングを受ける人間の体をキャプチャしている [15], [16]．こうして得られた体外離脱視点映像は，トレーニングを受けている人間や，遠隔地のコーチに伝送され利用される．客観的な三人称映像はトレーニングを受ける人間の動きを理解するのに効果的な方法である．我々は *JackIn Space* によって，コーチがトレーニングを受ける人間に *jack in* したり，場合によっては *jack out* して三人称映像も利用しながら，スポーツトレーニングの指導が効果的にできるようになることを期待している．

3.3 映像の遷移

コンピュータゲームの世界においては，カメラのなめらかな遷移手法や，一人称・三人称カメラ間の遷移手法が確立されている [17]．一人称視点を用いたコンピュータゲームの多くは，周囲の環境の中にゲームキャラクターを映し出す三人称視点モードを同時にサポートしている．*JackIn Space* は，そのような知見をテレプレゼンスによる遠隔共同作業の枠組みに取り込んでいる．

Rhizomatiks research は South by Southwest(SXSW’15) における音楽と映像のパフォーマンスで，通常のビデオ

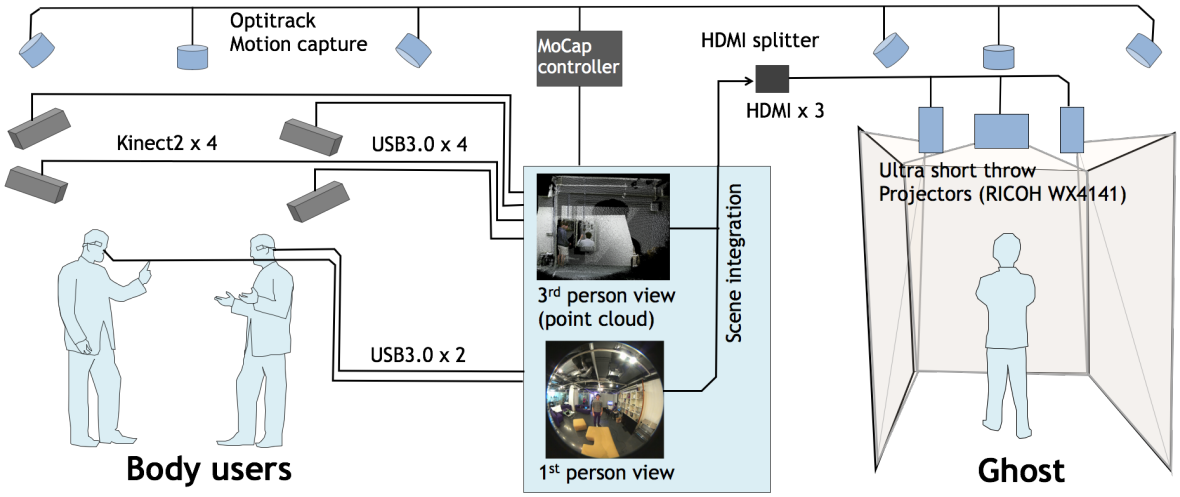

図 3 *JackIn Space* のシステム構成

映像と三次元空間とを融合させた映像表現を発表している [18]．このとき，パフォーマンスを行った劇場の三次元モデルは事前に計測され，本番で用いたビデオカメラの位置と向きもトラッキングされていた．これによって実際の映像と三次元モデルを連続的に合成することが可能となり，現実世界と仮想世界が統合されたことで強烈な印象を与えている．*JackIn Space* は，事前に環境の計測をすることなく，実際のカメラ映像（一人称映像）とポイントクラウド（三人称映像）の間のなめらかな遷移を提供する．

4. JACKIN SPACE のシステム構成

これまでに示した *JackIn Space* の概念を評価するため，二人の *body* と一人の *ghost* を扱うシステムを構築した．システムの構成を図 3 に示す．

4.1 Body の機器と環境

Body は図 5 に示すヘッドセットを身に付ける．このヘッドセットは，魚眼カメラ (e-consystems See3Cam USB3.0 カメラ，レンズ視野角 185°) と，位置・向きのトラッキングをするための赤外反射マーカーからなる．この魚眼カメラで *body* の一人称映像を撮影する．

作業領域の大きさは $3m \times 3m$ であり，天井には 4 台の RGB-D センサ (Kinect 2) と NaturalPoint 社製のモーションキャプチャシステム Optitrack が取り付けられている．これら 4 台の Kinect は床を見下ろすように設置してあり，三人称映像のためのポイントクラウドを取得するために利用される．モーションキャプチャセンサと Kinect は事前にキャリブレーション済みで，全てのセンサの座標系は統一されている．

Kinect 1 に対して Kinect 2 は互いに干渉しないため，複数台でのセンシングが可能となった．これは，Kinect 2 がセンシング方法として time-of-flight 方式を採用しており，

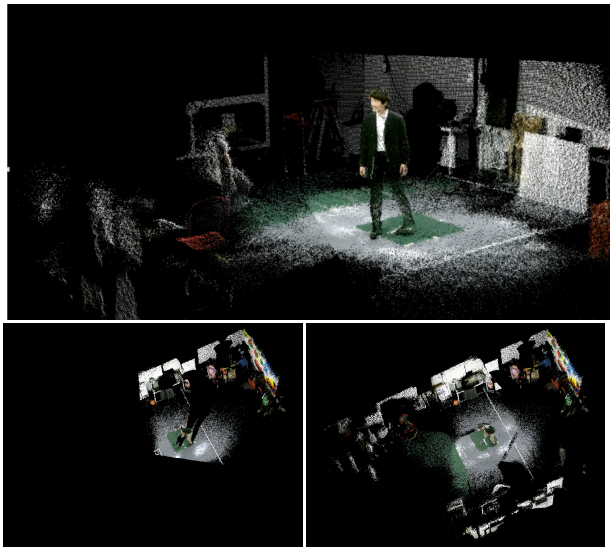

図 4 複数台の Kinect から生成された作業環境のポイントクラウド映像: (左下) 1 台の Kinect から生成されたポイントクラウド; (右下) 4 台の Kinect から生成されたポイントクラウド．

赤外照射時間が非常に短いためだと考察している．4 台の Kinect からの RGB-D 情報は，全ての点が三次元位置情報を持った 1 つのポイントクラウドに統合される (図 4)．1 台の Kinect だけではセンシングの “影” が生じてしまうが，複数台の Kinect を用いたセンシングでは互いに補うことができる．このポイントクラウドが *body* 周りの作業環境を表現する．

頭部に取り付けられた魚眼カメラから得られた画像は，ポイントクラウド空間内の球面テクスチャにマップされる．一人称視点モードを利用しているとき，*ghost* は球面にマップされたテクスチャをパノラマ映像として見ることができる．三人称視点モードに切り替えると，この球面テクスチャは徐々に透明に変化し，*ghost* はポイントクラウドで表現された環境を見ることができる．

4.2 一人称映像の安定化

Body の頭部動作により激しく揺れてしまう一人称映像は，*ghost* の映像酔いを引き起こす．この問題を解決するために，画像処理による映像の安定化 [3] が有効である．このアルゴリズムではまず，一人称映像におけるオプティカルフローを求め，それぞれに対応するカメラの回転量（クォータニオン）を推定する．それらの推定値から外れ値を取り除き，平均値を求め，それを全体としてのカメラ回転量とする．この回転に対する逆変換を一人称映像に適用することで，映像の揺れを抑制することができる．適用する変換の回転角は徐々に 0 に近づくため，*ghost* の視線方向を *body* のものに追従させながらも，激しい揺れだけを取り除くことができる．

4.3 Ghost の機器と環境

Ghost の視覚環境には，CAVE [19] 方式の 3 面プロジェクション環境を用いる．この環境は，3 面のスクリーン，3 台のプロジェクタ (RICOH WX4141)，HDMI スプリッタ，モーションキャプチャシステム，そして 3D グラスからなる．

3 台のプロジェクタは，時分割方式のステレオ投影に対応している．全てのプロジェクタが 1 つの HDMI スプリッタを通して繋がっているため，画像更新のタイミングが同期され，3D グラスをつけた *ghost* は，複数プロジェクタで表示される立体映像を見ることができる．

他の視覚環境として，ヘッドマウントディスプレイ (HMD) などの選択肢も考えられたが，*ghost* が *body* から *jack out* する際に目の前に現れる *body* の存在感を最重要に考えたため，CAVE 方式の構成を選択した．また，*ghost* を (*ghost* 自身の) 環境から完全に切り離してしまう HMD は共同作業には向かないとも考えている．共同作業中，*ghost* は *body* を支援するために，*ghost* 自身の環境にある情報を必要とすることがある．それゆえ，通常の環境へのアクセス性もまた重要である．その他にも，ヘッドトラッキングを用いた一画面構成の Fish-Tank VR [20] も考えられるが，CAVE 方式の構成を選択することで，より大きな視野角を確保できる．

一人称視点モードの使用中，*ghost* は *body* の視点における半球映像を見ているため，自由に周囲を見回すことができる．また，*body* の魚眼カメラの視野角が，通常のカメラの視野角に比べて大きいため，従来のテレプレゼンスシステムと違い，*ghost* の視線方向が *body* の視線方向と常に一致している必要はない．

全てのプロジェクタが 3 次元プロジェクションに対応しているため，三人称視点モードを利用している *ghost* は，*body* の三次元環境に入り込むことができる．一人称視点モードは三次元プロジェクションに対応していないが，より鮮明な映像を提供できるため，細かい部分の確認に便利

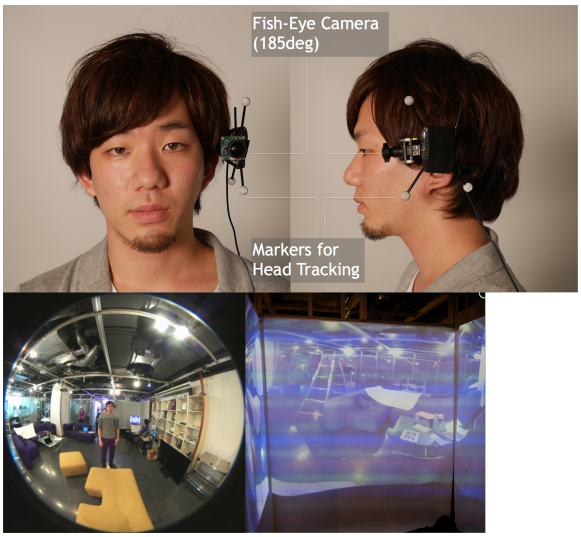

図 5 *body* のヘッドセットの構成

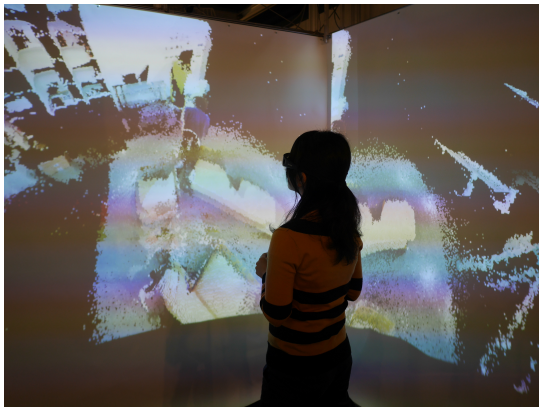

図 6 *body* の三人称映像を見る *ghost*

である．これらの一人称映像と三人称映像を組み合わせることで，ある *body* に *jack in* する，その *body* から *jack out* した後に環境を見回す，また別の *body* に *jack in* して詳細部分を確認する，といった一連の体験が可能になる．

4.4 処理システムの共有

現在の *JackIn Space* システムは，Ubuntu Linux が動作する単体のデスクトップコンピュータ (Intel Core i7-4790K, GeForce GTX TITAN X graphics card，32 GB memory) で，*body* のヘッドマウントカメラと環境センサ (Kinect とモーションキャプチャシステム) からの入力情報を処理し，*ghost* のための視覚情報の生成と提示を行っている．この処理の実現のため，コンピュータ上では 2 つの主要なプロセスが動作している．1 つは *body* 環境のためのプロセス，もう 1 つは *ghost* 環境のためのプロセスである．

概念を証明するためのプロトタイプであるため，本論文でのシステムは，*body* と *ghost* がお互い離れた場所にいる状況には対応しておらず，全てのセンサとディスプレイは同じコンピュータに接続されている．本論文では一人称・三人称映像間の連続的な遷移を重視しており，遠隔分散環

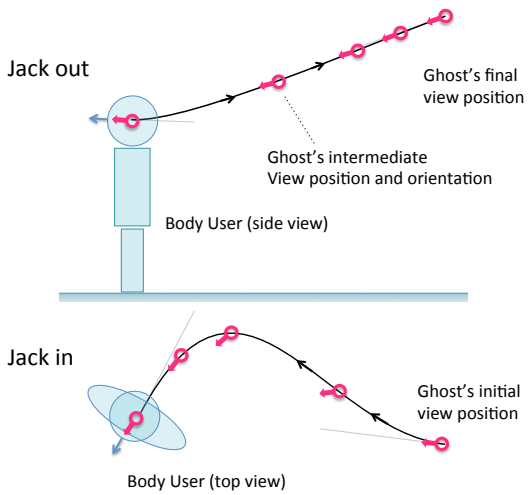

図 7 *jack in*, *jack out* 時の視点・視線の動き: (上) 三人称視点モードに切り替えるとき (*jack out*), 視点の移動経路はベジェ曲線に基づいて生成される．また，遷移の連続性を保つために，移動中の視線方向は初期値と最終値から補間される．これによって *ghost* は三人称視点へとなめらかに到達することができる．(下) 一人称視点モードに切り替えるとき (*jack in*) も同様である．

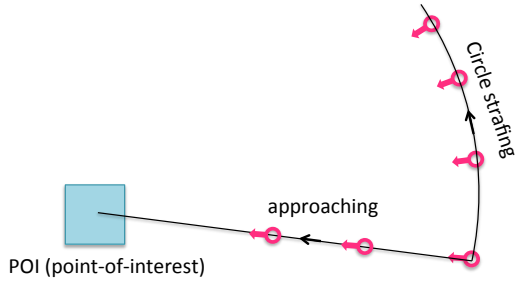

図 8 三人称視点モード利用時の POI に基づいた視点移動

境に適応するための設計上の問題については，議論の節で補足する．

5. JACK IN と JACK OUT のインタラクション

これまでに示したように，*ghost* は一人称視点モードと三人称視点モードをなめらかに切り替えることが可能である．

はじめに，*ghost* が，ある *body* の一人称映像を見ている状況を考える．*ghost* が *jack out* 操作を実行したとき，視点は自動的に，*body* の頭の位置 (一人称視点) から後方かつ少し上昇した地点へと移動する．これは，*ghost* が *jack out* した後に，*body* の頭や体を見ることができるようにするためである．*jack out* 終了後，三人称視点モードにおける視線方向の初期値は，*body* の頭を向いている．*jack out* 時の映像の遷移をなめらかにするために，線形補間を用いている (図 7，図 9)．

三人称視点モードになった *ghost* は，複数台の Kinect によって生成されたポイントクラウド空間内を，自由に見回すことができる．多くの仮想現実システムと同様に，*JackIn Space* システムにおいても様々な操作機器が考えられるが，現在は単純な小型リモコンを用いている．*ghost* の頭部位置と頭部方向がトラッキングされているため，*ghost* は環境内を自然に見回すことができる．

三人称視点モードの利用中は，*ghost* は Point-of-Interest (POI) (図 8) を起点にした操作をすることも可能である．POI には，あらかじめ定義された地点 (作業机の中心など) や，*body* の位置が含まれる．*ghost* はこれらの POI から 1 つを選択することで，その方向へのなめらかな移動と，POI を視点中心とした回転操作ができる．これらは仮想空間システム内の移動手法として利用されている方式を参考に実装した [21]．

他の *body* に *jack in* するために，*ghost* はまず，目的の *body* の方向を見るか (頭を目的の *body* に向ける)，POI の中から目的の *body* を選択する．すると，目的の *body* の頭の周りにロックオンマークが現れる．*ghost* が *jack in* 操作を実行したとき，*ghost* の視点はなめらかに *body* の視点へと移動する (図 7)．このとき，*ghost* が *body* に後ろから入り込んだと感じられるよう (連続的な変化で一人称映像に移行できるよう)，視点移動の経路はベジェ曲線を用いて生成し，最終的な視線方向を *body* の視線方向と一致させる．

6. 評価

JackIn Space の概念を評価するため，予備実験を行った．被験者は CAVE 環境の中で *ghost* となり，*body* 側の環境には 3 つのソファーと 2 人の *body* を用意しておく．被験者のタスクは，ソファーの配置を理解しスケッチすることである．各被験者に対してそれぞれ，*JackIn Space* システムを用いた場合と (2 人の *body* への *jack in* と *jack out* が可能)，1 人の *body* の一人称映像のみを利用できる場合の 2 通りの実験を行う．どちらの場合も，*ghost* は *body* に向いてほしい方向を音声で指示することができ，それぞれのタスクは 2 分間でおこなわれる．各タスクの前の簡単な実験説明も合わせると，1 人の被験者が要する実験時間は約 7 分間となる．タスクの終了後，被験者には，7 段階のリッカート尺度を持つ 4 つの質問に回答してもらう．図 10 に実際の評価実験の様子を示す．

実験には，コンピュータサイエンスのバックグラウンドを持つ 8 人の被験者が参加した．このうち 3 人はコンピュータサイエンス専攻の博士課程学生で，残りの 5 人はコンピュータエンジニアである．ソファーの配置は 2 種類用意し，ラテン方格を用いて，ソファーの配置と用いる方法の順序を決定した．

質問への回答を図 11 に示す．全体的に被験者からの反応は前向きなものであった．100%の被験者がテレプレゼンスにおける *jack in* と *jack out* の機能に好意的で (Q4, 平均 6)，87.5%が *jack out* により全体の状況を理解できた

図 9 映像の遷移の例: (上) 一人称映像から三人称映像 (*jack-out*), (下) 三人称映像から一人称映像 (*jack-in*)

図 10 実験構成

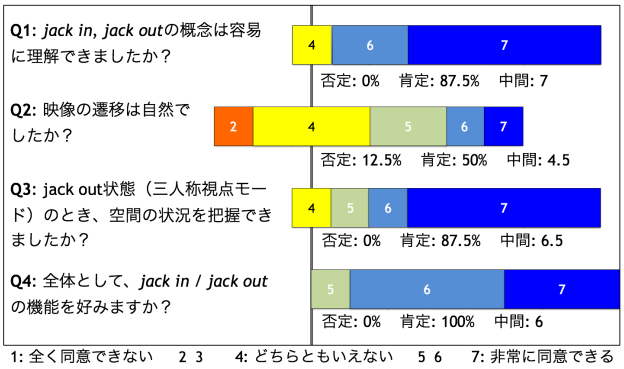

図 11 実験結果: 7 段階のリッカート尺度を持つ質問

と答えた (Q3，平均 6.5)．また，87.5%の被験者は *jack in* と *jack out* の概念を容易に理解できたと答えている (Q1，平均 7)．一方で，映像の遷移のなめらかさに関しては比較的低い評価を得た (Q2，12.5%が否定的で，50%が肯定的，平均 4.5)．これは，使いやすさを追求する上で，映像の遷移のアルゴリズムに改善の必要があることを示している．

スケッチの正確さでは，*JackIn Space* を利用した場合と通常の一人称映像のみを利用した場合で，明確な違いを見出せなかった (図 12)．これは，タスクが単純過ぎたことや (用いたのは 3 つのソファーのみ)，タスクを行うのに十分過ぎる時間を与えてしまったことが原因だと考えている．ほとんどの被験者が配置を理解できていた．一方で，一人称映像のみを用いた場合，「右を向いて」「下を向いて」といった，*ghost* から *body* への方向に関する指示語が多く観察された．これは，天井に近い位置から配置全体を見渡せる *JackIn Space* 環境ではあまり見受けられなかった結果である．

これらの実験結果をまとめると，一人称・三人称映像間の遷移表現を改善する必要がある一方で，*JackIn Space* の概念や機能は被験者に非常によく受け入れられた．

7. 議論

7.1 前提とするセンシング環境

現在，提案システムは，*body* の身体を含めた三次元空間情報をキャプチャするために，複数台の RGB-D センサが環境に備え付けられていることを仮定している．この構成は，遠隔地の手術や研究などいくつかのアプリケーションには合理的であるが，可動性を必要とするアプリケーションを考えた場合その限りではない．よってこの研究の次の段階では，ウェアラブルなデプスセンサを用いて，*body* 周りの三次元空間を徐々に構築することを検討する．このための手法は近年盛んに研究されている（KinectFusion [22] など）．他のアプローチとしては，センシングのためのロボットを用いることなどが考えられる．これらのロボットは *body* の後を追い，周りの環境をセンシングする．また，*ghost* に別の視点からの一人称映像を提供するためにも役に立つ．

究極的には，一人称映像は必要なくなるかもしれない．一人称映像は三人称映像に含まれるためである．しかし，実際には一人称カメラから得られる映像の質と空間センサ (Kinect など) から得られる映像の質には大きな差異がある

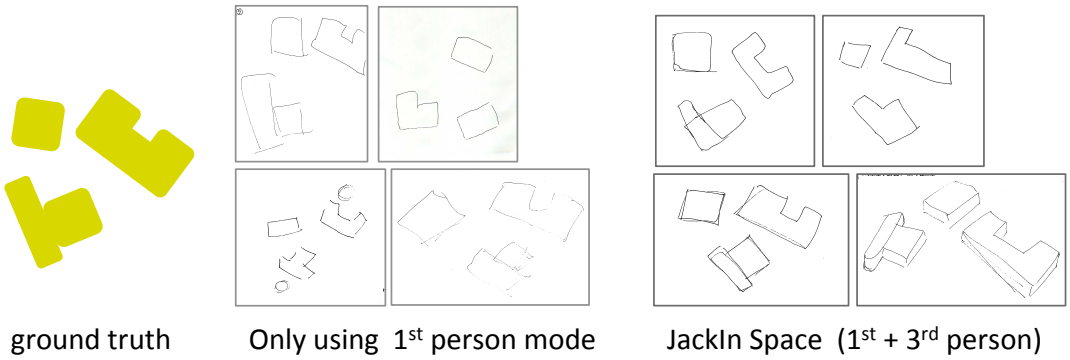

図 12 被験者によるスケッチ

ため，これら 2 つを組み合わせることが現時点での合理的な解決方法となっている．

7.2 通信量

システム構成の節で記述したように，現状のシステムは *body* と *ghost* のためのセンサが，USB3.0 や HDMI のような大容量の通信ケーブルで接続されている．*body* と *ghost* が分散したシステムを構築するためには，2 つの環境間で伝送されるデータ量がどの程度であるかを見積もり，削減することが重要である．

Kinect 2 は RGB-D 画像を生成する．非圧縮の場合，その 1 フレームのデータ量は 1.03MB である．環境をキャプチャするために 4 台の Kinect を利用すると仮定すると，30fps で伝送するデータ量の合計は 124MB(/sec) となる．伝送時にポイントクラウド形式を利用すると，各ポイントが三次元位置情報を持つため，データ量はさらに大きくなる．しかし，三次元空間情報の多くは静的なものであり，周囲の環境から人間を切り離すことができるため，この場合は更新すべきデータは大幅に小さくなる．実際の計測では，ポイントクラウド形式における人間 1 人あたりのデータ量は約 300KB であり，4 人分のデータを伝送する場合，合計のデータ量は 1.2MB となる．さらに，伝送時のデータ量を削減するために，他にも様々なデータ圧縮手法を適用できる可能性がある．

7.3 技能の共有や伝達

JackIn Space は，技能伝達の基盤としても利用できると考えている．スポーツや楽器演奏などの技術を教えるとき，先生は通常，実演しているところを生徒に見せ，それを真似させる．もしこのプロセスを伝送することができれば，遠隔学習の機会を増加させることができるかもしれない．加えて，一人称映像の共有が技能伝達には重要だと唱えている研究者もいる [23], [24]．*JackIn Space* を用いることで，学習者は先生の技術を三人称視点からも一人称視点からも観察することができる．CAVE 環境に先生の動作が等身大で表示された場合，学習者は先生の動作と自分の動作を比較することもできるだろう．また，これらを技術伝達における新しいインタラクティブな教材として記録し，パッケージ化することも考えられる．

7.4 自己と他者

JackIn Space が，人工的な体外離脱体験を研究するためのプラットフォームになることも期待している．例えば Lenggenhager らは，視覚と触覚を組み合わせることが体外離脱感覚（自分自身を外から見ていると感じる感覚）を生み出すと提唱している [25]．これらの研究は，自己と他者の認識を理解するのに有用である．本システムは一人称・三人称映像間の連続的な遷移を提供するため，体外離脱体験における認知作用の研究にも有用である．

8. 結論

本論文では，一人称・三人称映像間の連続的な遷移を可能にする遠隔共同作業システムを提案した．従来のテレプレゼンスシステムでは，1 つのサロゲートからの一人称映像しか利用できなかったのに対し，本研究では，一人称映像を用いた精密な作業の共有と三人称映像を用いた作業現場状況の把握の，両方を可能にするアプローチをとっている．この *JackIn Space* の概念は，一人称映像用のヘッドマウントカメラと，環境に取り付けられた複数台のデプスセンサによって実現されている．評価実験の結果から，この構成は映像の変化に改善の余地がある一方で，より効果的な遠隔共同作業支援を実現していることを示した．

参考文献

[1] Minsky, M.: Telepresence, *OMNI*, No. July, pp. 45–51 (1980).
[2] Kasahara, S. and Rekimoto, J.: JackIn: Integrating First-person View with Out-of-body Vision Generation for Human-human Augmentation, *Proceedings of the 5th Augmented Human International Conference*, AH '14, New York, NY, USA, ACM, pp. 46:1–46:8 (online), DOI: 10.1145/2582051.2582097 (2014).
[3] Nagai, S., Kasahara, S. and Rekimoto, J.: LiveSphere: Sharing the Surrounding Visual Environment for Immersive Experience in Remote Collaboration, *Proceed-*

ings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '15, New York, NY, USA, ACM, pp. 113–116 (online), DOI: 10.1145/2677199.2680549 (2015).

[4] Shiroma, N. and Oyama, E.: Asynchronous visual information sharing system with image stabilization, *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 2501–2506 (online), DOI: 10.1109/IROS.2010.5649928 (2010).

[5] Gibson, W.: *Neuromancer*, Ace Science Fiction, Canada (1984).

[6] Microsoft: Kinect for Xbox One (2013).

[7] Rander, P., Narayanan, P. J. and Kanade, T.: Virtualized Reality: Constructing Time-varying Virtual Worlds from Real World Events, *Proceedings of the 8th Conference on Visualization '97*, VIS '97, Los Alamitos, CA, USA, IEEE Computer Society Press, pp. 277–ff. (online), available from ⟨http://dl.acm.org/citation.cfm?id=266989.267081⟩ (1997).

[8] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L. and Fuchs, H.: The Office of the Future: A Unified Approach to Image-based Modeling and Spatially Immersive Displays, *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98, New York, NY, USA, ACM, pp. 179–188 (online), DOI: 10.1145/280814.280861 (1998).

[9] Clarke, A. C.: *The City and the Stars*, Frederick Muller Ltd (1956).

[10] Marsilii, B. and Rossio, T.: Dj Vu (2006 film), Buena Vista Pictures (2006).

[11] Gelernter, D.: *Mirror Worlds: or the Day Software Puts the Universe in a Shoebox...How It Will Happen and What It Will Mean*, Oxford University Press (1993).

[12] Dou, M. and Fuchs, H.: Temporally enhanced 3D capture of room-sized dynamic scenes with commodity depth cameras, *Virtual Reality (VR), 2014 iEEE*, pp. 39–44 (online), DOI: 10.1109/VR.2014.6802048 (2014).

[13] Shiroma, N. and Oyama, E.: Development of virtual viewing direction operation system with image stabilization for asynchronous visual information sharing, *RO-MAN, 2010 IEEE*, pp. 76–81 (online), DOI: 10.1109/ROMAN.2010.5598662 (2010).

[14] Sugimoto, M., Kagotani, G., Nii, H., Shiroma, N., Inami, M. and Matsuno, F.: Time Follower's Vision: A Teleoperation Interface with Past Images, *IEEE Comput. Graph. Appl.*, Vol. 25, No. 1, pp. 54–63 (online), DOI: 10.1109/MCG.2005.23 (2005).

[15] Higuchi, K., Shimada, T. and Rekimoto, J.: Flying Sports Assistant: External Visual Imagery Representation for Sports Training, *Proceedings of the 2Nd Augmented Human International Conference*, AH '11, New York, NY, USA, ACM, pp. 7:1–7:4 (online), DOI: 10.1145/1959826.1959833 (2011).

[16] Hasegawa, S., Ishijima, S., Kato, F., Mitake, H. and Sato, M.: Realtime Sonification of the Center of Gravity for Skiing, *Proceedings of the 3rd Augmented Human International Conference*, AH '12, New York, NY, USA, ACM, pp. 11:1–11:4 (online), DOI: 10.1145/2160125.2160136 (2012).

[17] Haigh-Hutchinson, M.: *Real-Time Cameras: A Guide for Game Designers and Developers*, CRC Press (2009).

[18] Vincent, J.: This futuristic performance by Japanese trio Perfume will make you lose your mind, *The Verge*, (online), available from ⟨http://www.theverge.com/2015/3/25/8282377/perfume-sxsw-performance-livestream⟩ (2015).

[19] Cruz-Neira, C., Sandin, D. J. and DeFanti, T. A.: Surround-screen Projection-based Virtual Reality: The Design and Implementation of the CAVE, *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, New York, NY, USA, ACM, pp. 135–142 (online), DOI: 10.1145/166117.166134 (1993).

[20] Arthur, K. W., Booth, K. S. and Ware, C.: Evaluating 3D Task Performance for Fish Tank Virtual Worlds, *ACM Trans. Inf. Syst.*, Vol. 11, No. 3, pp. 239–265 (online), DOI: 10.1145/159161.155359 (1993).

[21] Mackinlay, J. D., Card, S. K. and Robertson, G. G.: Rapid Controlled Movement Through a Virtual 3D Workspace, *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '90, New York, NY, USA, ACM, pp. 171–176 (online), DOI: 10.1145/97879.97898 (1990).

[22] 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, *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, New York, NY, USA, ACM, pp. 559–568 (online), DOI: 10.1145/2047196.2047270 (2011).

[23] Kawasaki, H., Iizuka, H., Okamoto, S., Ando, H. and Maeda, T.: Collaboration and skill transmission by first-person perspective view sharing system, *RO-MAN, 2010 IEEE*, pp. 125–131 (online), DOI: 10.1109/ROMAN.2010.5598668 (2010).

[24] Amores, J., Benavides, X. and Maes, P.: ShowMe: A Remote Collaboration System That Supports Immersive Gestural Communication, *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, ACM, pp. 1343–1348 (online), DOI: 10.1145/2702613.2732927 (2015).

[25] Lenggenhager, B., Tadi, T., Metzinger, T. and Blanke1, O.: Video Ergo Sum: Manipulating Bodily Self-Consciousness, *Science*, Vol. 317, No. 5841, pp. 1096–1099 (online), DOI: 10.1126/science.1143439 (2007).