

全周囲映像を用いたヒューマンテレプレゼンスのための 体験共有システム

永井 翔平^{2,a)} 笠原 俊一^{2,1,b)} 暦本 純一^{3,1,c)}

概要: ウェアラブルカメラの登場により、装着者の体験を一人称視点で記録し、共有することが可能になった。しかし、このような映像には、映像の揺れによるモーションシicknessや、視野が装着者の姿勢によって制限されてしまう等の問題がある。LiveSphereは、人間の体験を他者に共有しコミュニケーションを行うためのシステムである。装着者の体験は複数のカメラによって全周囲映像として記録され、伝送される。記録された映像は、画像処理によりリアルタイムに回転運動成分を取り除くことで、装着者の頭部運動が分離される。これにより、モーションシicknessが改善されるとともに、装着者と独立した環境の観察が可能になる。本稿では、システムの実装及び評価実験、パイロットスタディをおこない、ヒューマンテレプレゼンスにおけるインタラクションデザインに関する洞察を与える。

Immersive Experience Transmission with Omnidirectional Image for Human-Telepresence

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

Abstract: The development of wearable camera has enabled people to capture their first-person video as their experiences and share it. However, the view is limited to the head direction, and the captured video can be so shaky as to cause motion sickness. LiveSphere is the system where people share their experiences and communicate with others. It has multiple camera and shoots and share omnidirectional video as person's experiences. The wearer's ego-motion is decouple from the video using omnidirectional video stabilization. That alleviates motion sickness and realizes independent viewing. In this work, we implemented and evaluate the system. we also discussed our evaluation and pilot study to provide the insight for "human telepresence" interaction design.

1. はじめに

体験を完全に記録・伝送し共有することは、メディアテクノロジーにおける究極のゴールの一つである。日常の出来事や旅行、スポーツや趣味など自分の体験を記録し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。こ

れまで、写真やビデオ等の記録メディアは、そのような体験を記録し他者に伝えるという役割を担ってきた。また、GoPro^{*1} や ActionCam^{*2} などのウェアラブルカメラの登場によって、装着者の体験を一人称視点として記録し、伝送することが可能になった。SF映画 Brainstorm[7]にも描かれているように、一人称の全人的な体験の記録と伝送は、自身では経験できない体験を仮想的に体験・伝送することを可能にする。このような一人称の情報伝送やストーリーミングは、体験を伝送するメディアの有効な形の一つである。一方、自分以外の存在からの体験を伝送する方法として、テレプレゼンス技術がある。これは、遠隔地のロボットの

¹ ソニーコンピュータサイエンス研究所
Sony Computer Science Laboratories, Inc.
² 東京大学学際情報学府
Interdisciplinary Information Studies
³ 東京大学情報学環
Interfaculty Initiative in Information Studies
a) shohei.nagai14@gmail.com
b) kasahara@csl.sony.co.jp
c) rekimoto@acm.org

^{*1} GoPro <http://gopro.com/>
^{*2} ActionCam <http://www.actioncambysony.com/>

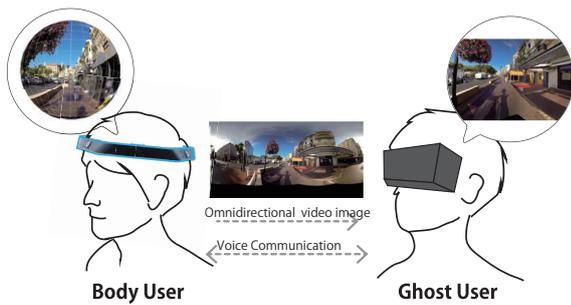


図 1 LiveSphere 概念図
Fig. 1 LiveSphere overview

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したり、環境においてインタラクションをするための技術で、さまざまな研究が行われている [28], [29].

本稿では、これらを統合し、「ヒューマンテレプレゼンス」というコンセプトを提案する。これは、一人称の体験を記録・伝送し、他者にリアルタイムに共有することで、コミュニケーションを行うという可能性である。これにより、一方的な体験伝送だけでなく、遠隔地にいる人と体験を共有しながら、それに基づいてコミュニケーションを行うことが出来る。例えば、機械の操作を行う装着者がその状況を遠隔地にいるエキスパートに共有し、指示を受けるといったことが可能になる。

しかし、このようなヒューマンテレプレゼンスには、考慮すべき課題も存在する。その一つが装着者の身体運動に伴う映像の揺れや意図しない向きの変化である。このようなウェアラブルカメラの映像の揺れや動きは、モーションシックネスを引き起こす原因になり、大きなディスプレイやHMD (Head Mount Display) で見るときその影響はさらに大きくなる [13]. このようなウェアラブルカメラによって得られる映像は、装着者の身体運動が畳み込まれた映像であると捉えることが出来る。つまり、装着者の視覚周辺環境と、その人の身体的な動きによって一人称視点点が構成されているのである。そして、モーションシックネスを引き起こす大きな要因のひとつは、この映像に畳み込まれた装着者の身体運動と観察者の身体運動の不整合である [16]. 他者に体験として映像を共有する上で、共有者のモーションシックネスを防ぐためにも、ウェアラブルカメラ映像から、既に畳み込まれている装着者の身体運動を分離することが必要である。

ヒューマンテレプレゼンスにおいて、考えなければならないもう一つの問題は、装着者の身体の状態に制限されたカメラの視界である。ウェアラブルカメラによって記録される一人称視点映像は、装着者の頭部が向いている方向しか映し出すことが出来ない。また、画像履歴から擬似的に広角映像を構成する手法 [6], [14] も存在するが、過去の映像を用いるので、同時刻の異なる方向や、装着者が向いた

ことのない方向を見ることは出来ない。観察者の見ることが出来る視界は装着者の身体の状態によって制限されてしまう。これは、観察者が自分の意志で環境を観察するテレプレゼンスを実現する上で大きな問題である。この問題は、人間がマスターで機械がスレーブとなり、スレーブの機械が人間の動きを忠実に再現する機械へのテレプレゼンスに対し、人間に対するテレプレゼンスでは、テレプレゼンスされる側の人間がする側の人間に従って動くとは限らないということに起因する。ヒューマンテレプレゼンスを実現する上で、この制限を取払い、視覚体験において、両者の独立性を許すインタラクションデザインが必要である。

本稿では、上記の問題を解決する体験伝送システムとして、LiveSphere を提案する。LiveSphere は、複数のカメラを備えたヘッドギアによって装着者のいる環境の全周囲映像を撮影する。撮影する全周囲映像は装着者の身体運動が畳み込まれているが、スタビライズにより装着者の身体運動と分離され、身体運動が畳み込まれていない状態で他者に伝送される。伝送された側は、HMD を用いて、頭部の動きに連動して映像を表示することにより、あたかも自分が実際に現地にいてヘッドギア装着者の身体運動とは独立して自由にその環境を見回す様な体験をすることが出来る。これにより、上記のモーションシックネスを解決するとともに、両者の独立性を持ったインタラクションを実現することが可能になる。

以下の章では、まず始めに LiveSphere の概要と映像から身体運動を分離するための画像処理、LiveSphere の実装について述べ、関連研究にも言及する。その後、LiveSphere システムがどのように問題を解決したのかを評価した実験について、その結果と考察を示すとともに、そこから得られた洞察をもとにしたリアルタイムのインタラクションのためのインターフェースデザインと、それに基づいて実装したシステムのパイロットスタディについて報告する。

2. LiveSphere アーキテクチャ

2.1 システム概要

ヘッドギアは6つのカメラを持ち、それぞれのカメラが水平 120° 、垂直 90° の画角を持つ。これらのカメラによって記録された映像は、全天球映像映像へスティッチングされ、画像処理により頭部運動成分の分離が行われる。この映像と頭部運動データは、Body からネットワークを通じて Ghost 側に伝送される。Ghost 側では、ヘッドモーションとラッキングが行える HMD (例えば、Oculus Rift^{*3}) によって映像を提示される。Ghost は Body の頭部の向きに制限されることなく、Body のいる視覚環境を自由に見回すことが出来る。また、Body-Ghost 間の音声によるコミュニケーションも行うことが出来る。

*3 Oculus rift DK-2 : <http://www.oculus.com/>



図 2 LiveSphere ヘッドギア：6台のカメラにより全周囲映像を記録する。

Fig. 2 LiveSphere headgear : Provide omnidirectional images of wearer's surrounding environment.

LiveSphere は、二つのパートから構成される (図 1)。ここで、体験を映像として記録・共有する側の人間を **Body**、映像を共有される側の人間を **Ghost** と呼ぶことにする。現場にいて実際に活動を行う Body はヘッドギア (図 2) によって全周囲映像を記録し、Ghost は HMD を使って伝送されてきた映像を見る。

2.2 映像の回転成分の分離

ウェアラブルカメラ映像のブレや、既に畳み込まれている装着者の身体運動と観察者の身体運動の不整合を発生させる映像の動きについて3種類に分類することができる。

1. 振動：Body が歩行などの動作を行う際にカメラが揺れることによって発生する映像の振動。
2. 無意識の動き：Body が周囲を見回したり、うなずいたりなど無意識な動作をすることによって引き起こされる映像の動き。文脈がなく、事前に予測する事は困難。
3. 意図的な動き：Body の意図的な動きによって発生する映像の動き。Body の意図を反映しており、文脈から予測することが可能。しかし、Ghost が Body と異なる意図 (例えば、Body と違う対象を見ようとしている等) を持っている場合、Ghost の観察の障害になる。

一般に、映像のモーションスタビライズは 1. を想定している。しかし、ウェアラブルカメラ特有の問題として 2. を考慮する必要がある。さらに、Live Sphere でのヒューマンテプレゼンスを目的とする場合、装着者の意思と分離するためには 3. を解決する必要がある。

2.3 Gyro-vision

LiveSphere では、一般的なモーションスタビライゼーションを拡張し、映像から装着者の身体運動成分を分離する処理を行う。ヘッドギアによって取得された全周囲映像に対して画像処理を行うことで、映像から Body の頭部回転の推定を行う。推定された回転と逆に全周囲映像を回転させることによって、回転成分が分離された映像を生成することが出来る (図 3)。ここで、この一人称視点映像

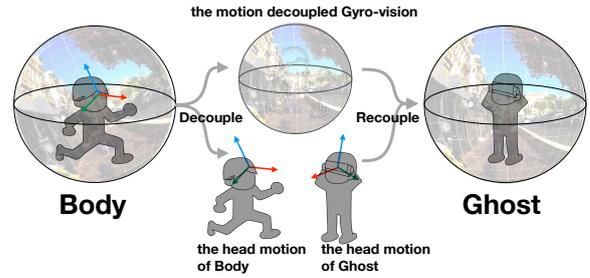


図 3 Gyro-vision : 頭部運動が分離された全周囲映像

Fig. 3 "Gyro-vision" : First-person omnidirectional video without any head directional motion.

から Body の動きを分離した全周囲映像を、**Gyro-vision** と呼ぶことにする。Gyro-vision は、Body の身体運動が畳み込まれていない状態の映像であり、この Gyro-vision に Ghost の運動を畳み込んで表示することにより、Ghost の一人称視点映像として見る事が可能になる。

2.4 Gyro-vision 生成アルゴリズム

ここでは、Body の身体運動を推定し、Gyro-vision を生成するためのアルゴリズムの詳細を述べる (図 4)。この処理のいくつかの部分はモーションセンサーによって代替可能だが、センサーからの信号と画像取得の同期の問題により、映像の回転を正確に推定できず、映像の細かな揺れを引き起こしてしまう。また、激しい運動には対応できないという問題が存在する。そのため、本手法では画像のみを用いて、全周囲映像の回転推定を行う。ヘッドギアのカメラから得られた映像は、正距円筒図法を用いて全周囲映像に合成される。正距円筒図法は全周囲映像を表現するために用いられる標準的な手法のひとつであり、世界地図にも用いられる。まず、毎フレーム $l(t)$ において、shi らの手法 [27] により画像特徴点群を算出する。ここで、画像特徴点 $p_n(t) (n = 1000)$ は、互いの距離が画像全体に対して可能な限りスパースになるように抽出される。なお、正距円筒画像における、高緯度部分 (画像の上下端) は歪みが大きいいため、ROI から除く。次に、pyramidal KLT 法 [4] により、それぞれの映像フレーム上の各々の画像特徴点 $p_n(t)$ のオプティカルフロー $f_n(t)$ を求める。これにより、フレーム $l(t+1)$ における画像特徴点群 $p_n(t+1) = p_n(t) + f_n(t)$ が求まる。次に、算出された 2 次元空間の全天球画像上の点群 $p_n(t+1), p_n(t)$ を 3 次元空間の球面の点群 $P_n(t+1), P_n(t)$ に変換する。ここで、点が配置される球の半径は、この後の処理には影響しないため任意の値とする。ここで、3 次元空間に配置された点群の移動を記述するアフィン変換 $P_n(t+1) = P_n(t)M(t+1)$ を表す、アフィン変換行列 $M(t+1)$ を RANSAC により算出する。この変換行列 $M(t+1)$ から、映像フレーム $l(t)$ から $l(t+1)$ の回転を表すクォータニオン $dQ(t+1)$ を算出する。ここで推定誤差を下記のように定義する。推定誤差がしきい値を超えた場

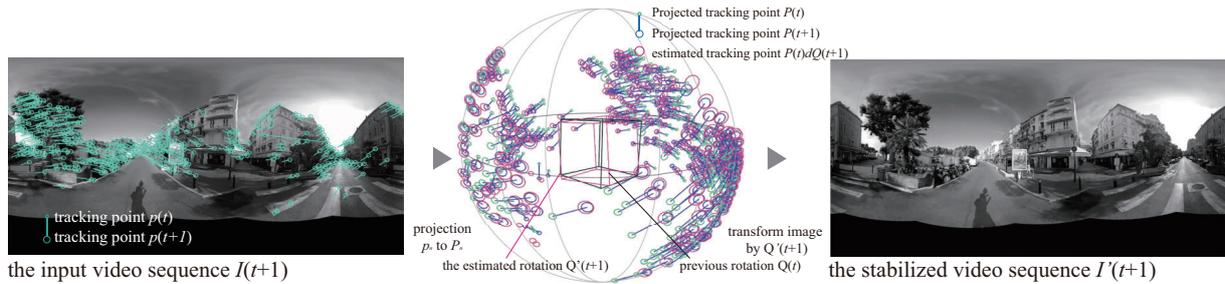


図 4 頭部回転運動推定のための画像処理と Gyro-vision の生成

Fig. 4 Image processing procedure for estimation of head rotation of Body and the generation of Gyro-vision.

合、回転の推定が失敗していると判断し、カルマンフィルタによる推定値を用いる。

$$Err(t+1) = P_n(t+1) - dQ(t+1)P_n(t)$$

以上により、得られた毎フレーム間の回転量 $dQ(i)$ ($i = s, \dots, t$) をスタビライズ開始時間 s から掛け合わせることで、スタビライズ開始姿勢を基準とした、時刻 t での回転量が下記のように求まる。

$$Q(t) = \prod_{i=s}^t dQ(i)$$

そして、時刻 t における回転量 $Q(t)$ の逆回転 $Q(t)^{-1}$ を用いて、正距円筒画像 $l(t)$ を回転処理することにより、開始時間 s を基準として回転成分を除いた正距円筒画像 $l'(t)$ を生成することが出来る。

$$I'(t) = I(t)Q(t)^{-1}$$

ここで、フレーム $l'(t)$ が Gyro-vision のシーケンス (図 5) を表し、 $Q(t)$ のシーケンスが映像から分離された Body の頭部運動を表す。

3. システム実装

LiveSphere システムは、リアルタイムと記録・再生機構の二つの実装を持ち、ヘッドギアも USB カメラによって構成されたものから、HD カメラによるものまで複数のバージョンを製作した (図 7)。

リアルタイムアーキテクチャ (図 6) では、6つの HD カメラ (GoPro) それぞれから解像度 1920*1080, 30Hz の映像ストリームが HDMI キャプチャー (BlackMagic *4) を通じて、thunderbolt によってコンピュータ (MacPro) に伝送される。コンピュータ側では、伝送された6つの映像ストリームから GLSL シェーダによって GPU 上で正距円筒映像がリアルタイムにステッチングされる。ここで、ステッチングの際の HD カメラのキャリブレーションは、OCamCalib[24] を用いて行った。

*4 UltraStudio Mini Recorder

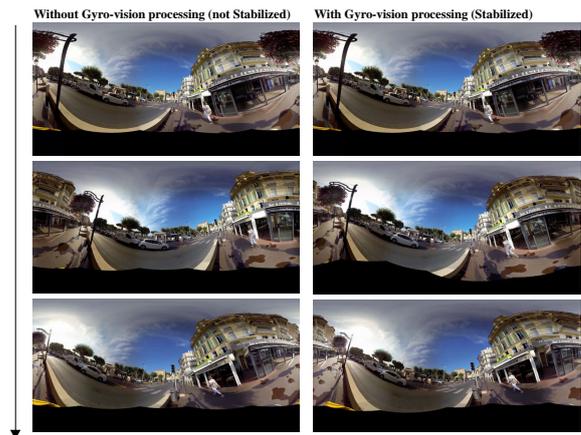


図 5 Gyro-vision シーケンスの例

Fig. 5 Example Gyro-vision sequence.

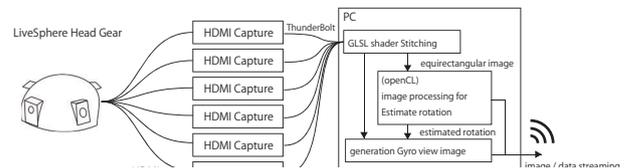


図 7 LiveSphere ヘッドギアのバリエーション。(a) と (b) は、USB カメラを用いた軽量バージョン。(c) と (d) は、HD カメラを用いたバージョン。

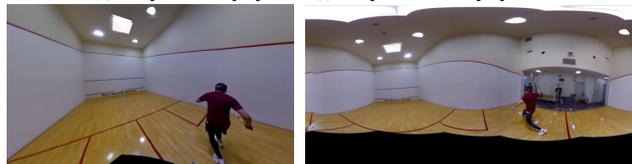
Fig. 7 Variation of LiveSphere headgear: (a) and (b) are lighter version with usb cameras. (c) and (d) are compatible design for realtime achitecture with six HD cameras.

ここで得られた全周囲映像ストリームは、上記の画像処理により Gyro-vision に変換される。なお、この処理は OpenCL による GPU アクセレーションによりリアルタイム (30Hz) で実行される。その後、得られた Gyro-vision

For scenario: (a) walking in the town



For scenario: (b) ball sports united purpose and (c) ball sports individual purpose



View from usual perspective

Equirectangular image

図 8 各テストシナリオにおける全周囲映像のサンプルフレーム：(a) 町中の散歩、(b) 球技、ボールの追跡、(c) 球技、相手プレイヤーの追跡

Fig. 8 Omnidirectional video sample for test scenario: (a) walking in the town, (b) ball sports, trace the ball, and (c) ball sports, trace another player.

の正距円筒映像は Motion JPEG によって圧縮され、映像から分離された Body の回転データとともに Zmq プロトコルを用いて、Wifi でリアルタイムに無線伝送される。

記録・再生機構では、それぞれの HD カメラからの映像を SD カードに記録し、あらかじめステッチングと画像処理を行い Gyro-vision を生成する。この機構では時間的制約がないので、映像のフレームレートに制限はなく、HD カメラの性能に応じて、100~120fps の映像を扱うことも可能である。

4. 評価実験

LiveSphere において Gyro-vision は Body の身体運動を映像から分離し、その後の HMD によってトラッキングされた Ghost の身体運動との再合成により、Ghost へ身体運動に連動した映像体験を提供する。これは、モーションシックネスを防ぐという観点からも独立した映像体験によるインタラクションを提供するという観点からも極めて重要な要素である。本稿では、この Gyro-vision によって、どの程度モーションシックネスが解消されるか、独立した映像体験がどのように行われるかを実験し、評価を行った。

4.1 実験手法

LiveSphere では記録された体験の共有だけでなく、人間から人間へのテレプレゼンス、すなわちヒューマンテレプレゼンスの実現を目指している。本稿では、このヒューマンテレプレゼンスにおけるアクティビティや目的の違いが映像体験に与える影響に関して評価するために、以下の3つの異なるシナリオを想定し実験を行った。(a) 一緒に旅行する。(b) 一緒にスポーツを体験する。(c) スポーツの補助。本実験では、実験環境を統一するために LiveSphere の実装として記録再生機構を用いた。

4.1.1 シナリオ (a)：町中の散歩

このシナリオでは、観光地での観光客視点の全周囲映像を共有することで、旅行の体験を共有することを想定する。Body が見知らぬ町を散歩しており、昼食を食べる場所を探している(図 8-(a))。この映像を Ghost に共有することにより仮想的な旅行体験が実現される。提示される全周囲映像は、カンヌの町をおよそ 90m/min で歩きながら LiveSphere ヘッドギアで記録したものである。推定された映像の回転速度はオイラー角で平均 13.5 (ロール)、12.0 (ピッチ)、19.4 (ヨー) deg/sec である。また、最高速度は、137.5 (ロール)、185.5 (ピッチ)、353.0 (ヨー) deg/sec だった。被験者には、Ghost として映像から昼食を食べるためのレストランを出来る限り多く探すように指示を出した。

4.1.2 シナリオ (b)：球技、ボールの追跡

このシナリオでは、選手視点の全周囲映像をリアルタイムに見て、スポーツの体験を共有することを想定する。Body はスカッシュをプレイする選手で、その選手視点の映像が Ghost に提供される(図 8-(b))。Ghost は選手である Body が行っているのと同様にボールの動きを目で追う。提示される全周囲映像は、Body が 6.4m × 9.75m のコートでもう一人のプレイヤーとスカッシュをプレイしているのを LiveSphere ヘッドギアで記録したものである。推定された映像の回転速度はオイラー角で平均 20.3 (ロール)、15.7 (ピッチ)、50.7 (ヨー) deg/sec である。また、最高速度は、63.5 (ロール)、531.2 (ピッチ)、793.4 (ヨー) deg/sec だった。被験者には、Ghost として出来る限りボールを目で追い続けるよう指示を出した。

4.1.3 シナリオ (c)：球技、相手プレイヤーの追跡

このシナリオでは、選手視点の全周囲映像をリアルタイムに見て、選手とは異なる周辺状況の把握を行い、アシスタンスや状況判断を行うことを想定する。Body はシナリオ (b) と同様にスカッシュをプレイする選手で、Ghost はその選手の視点映像から相手プレイヤーを見て戦略を立て Body に指示を出す。このシナリオでは、シナリオの違いによる映像体験の違いを比較するためにシナリオ (b) と同じ全周囲映像を用いた。被験者には、Ghost として出来る限り相手の選手を目で追い続けるように指示を出した。

実験では、上記の3つのシナリオについて、Gyro-vision (スタビライズを行った) 全周囲映像とスタビライズを行っていない全周囲映像でタスクを行った。被験者はそれぞれのシナリオで Ghost の役割をし、HMD として Oculus Rift DK2 を用いた。HMD のフレッシュレート、映像のフレームレートともに 60Hz である。これらは、1 シナリオにスタビライズあり・なしでタスクを行い、その後、他の2つのシナリオについて同様にタスクを行う。各々のタスクは2分間で、間に30分間の休憩を取りながら合計6回行った。また、実験シナリオの順序や、1 シナリオにお

るスタビライズあり・なしの順序は、学習による順序効果が相殺されるように被験者ごとに変えて実験を行った。それぞれのタスクの前後において、SSQ(Simulator Sickness Questionnaire)[17]による主観評価を行った。SSQは映像酔いの主観評価のために広く使われている手法である。ここでSSQの値が大きいほど、それぞれの症状(吐き気、眼球運動、ふらつき感)の感覚が強いことを示している。被験者は、書面によるインフォームドコンセントを得た年齢が20~38までの健常成人12名である。また、タスク中に被験者の体調が悪くなった場合はいつでもタスクを終了して良いことを伝え、その場合はその時点でSSQへの回答を結果とした。さらに、記録された映像から算出されたBodyの頭部運動のデータと、HMDから取得されたGhostである被験者の頭部運動のデータの時系列比較も行った。また、全てのタスクが終了した後は、口頭によるインタビューも行った。

4.2 実験結果

以下に、実験の結果を示す。ここからは、どのシナリオかをA, B, C, スタビライズあり・なしをS, Nを使って表すこととする(例えば、シナリオ(a)のスタビライズありをA-S)。

4.2.1 SSQ分析

表1に各シナリオのスタビライズのあり・なしにおけるタスク前後でのSSQ値の変化の平均値を示す。ここから、それぞれのシナリオにおいてGyro-visionにより、モーション sicknessの症状が低減されていることが分かる。また、図9に各シナリオのスタビライズあり・なしにおけるSSQの吐き気(Nausea)、眼球運動(culomotor)、ふらつき感(Disorientation)のスコアと、各項目に重み付けした合計値(Total score)を示す。ここで、シナリオ(a)では、スタビライズなしの全周囲映像(A-N)と、スタビライズありの全周囲映像(A-S)には、有意差が見られた($p < 0.05$)。また、シナリオ(b)でも、スタビライズなし(C-N)とスタビライズあり(C-S)の間に有意差が見られた($p < 0.05$)。C-Nにおいては、SSQの平均値は最も大きくなり、12人の被験者のうち6人が2分間のタスク時間の前にタスクを中断した。SSQの個別の項目に関しては、シナリオ(a)のDisorientationと、シナリオ(c)の全ての項目に置いてスタビライズのあり・なしによる大きな差が見られた。逆に、シナリオ(b)においては、SSQ値に有意差が見られず自体も、他の2つのシナリオと比較しても小さな値となった。

また、指定したタスクが行いやすかったかという質問に対して、シナリオ(a)とシナリオ(c)では有意差が見られた($p < 0.05$)が、シナリオ(b)では見られなかった。

4.2.2 時系列分析

次にそれぞれの実験において、被験者がどのような動きをしたのかを検証するために、ヘッドトラッキングに

Average(SD)	A	B	C
No GyroVision	27.1(26.8)	12.2 (23.2)	30.5 (25.2)
GyroVision	11.8(15.2)	5.6(13.2)	0.6 (11.26)

表1 SSQ値の変化値の平均と標準偏差。(a)町の散歩。(b)球技、ボールの追跡。(c)球技、相手プレイヤーの追跡

Table 1 Average and standard deviation of SSQ score change values:(a) walking in the town, (b) ball sports, trace the ball, (c) ball sports, trace another player.

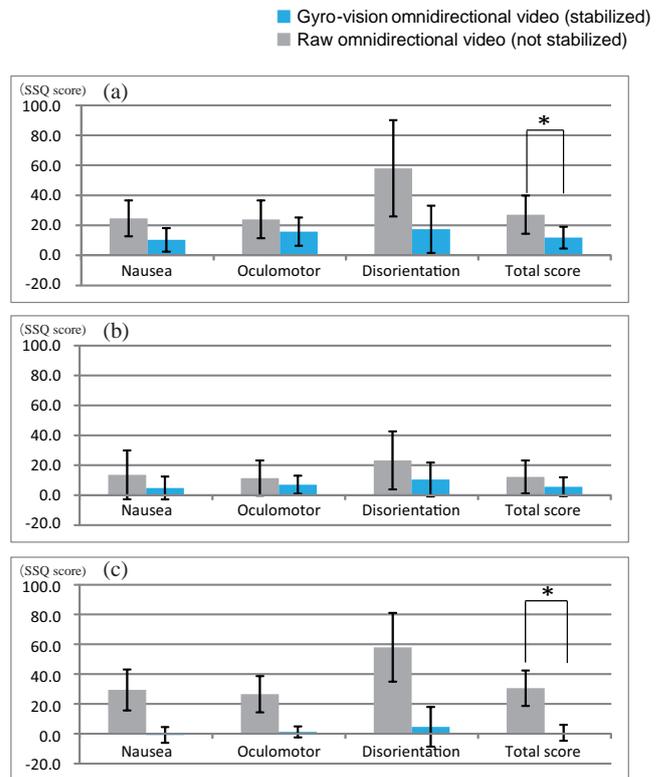


図9 SSQ値のの変化の詳細：吐き気、眼球運動、ふらつき感、合計値

Fig. 9 Detailed results for SSQ : Nausea, Oculomotor discomfort, Disorientation and Total score.

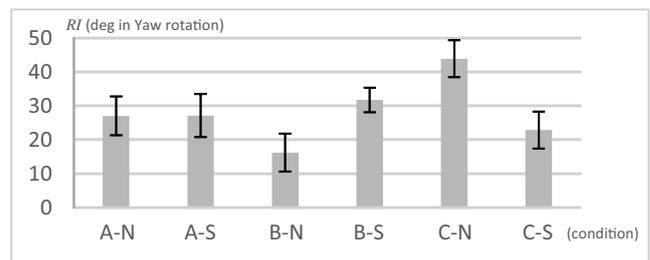


図10 結果概要：それぞれのシナリオにおけるGhostのヨー回転の激しさRI。

Fig. 10 Result overview: Yaw axis rotational intensity RI of Ghost in each scenarios.

よって得られた被験者の頭部方向の時系列データから被験者の頭部の運動の激しさRIを分析した。ここで、ある時刻tにおける激しさRI(t)は、Ghostの頭部の向きR(t)と、その時系列平均R_{ave}(t)の差の絶対値と定義する。時

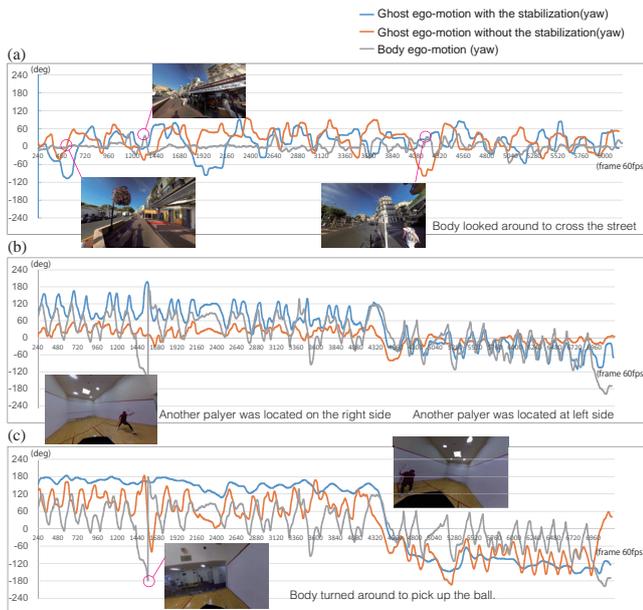


図 11 ヨー回転の時系列データ

Fig. 11 Time series data of Yaw axis rotation.

系列平均 $R_{ave}(t)$ は時刻 $t - 1sec$ から時刻 $t + 1sec$ の平均値である (60fps において, $n = 120$ サンプル).

$$RI_t = \left| R_t - \frac{\sum_{i=t-1sec}^{t+1sec} R_i}{n} \right|$$

ここでは, Ghost, Body 両方の結果から, オイラー角においてヨー回転が支配的であることが分かった. ここで, 図 10 に, Ghost である被験者の頭部運動におけるヨー回転の激しさ RI の平均値を示す. さらに, Ghost である被験者が Body の身体運動に対して独立して観測を行えているかを調べるために, Ghost と Body のヨー軸周りの頭部回転の時系列比較を行った. 得られた結果を, 図 11 に示す.

シナリオ (a) では, 図 10 から A-N と A-S には有意な差は見られず, 動きの活発さに大きな差がないことが分かる. 一方で, 図 11-(a) の結果は, Ghost が Body の身体運動から独立して, その環境を見回し観測していることを示している.

シナリオ (b) では, 運動の激しさ RI は B-S に比べて, B-N の方が小さい. これは, B-N において, Ghost のボールを目で追うというタスクが Body の意図と合致しており, Ghost が能動的に行わないでも, Body の動きによって映像が自然にボールを追いかけるためであると考えられる. 逆に, B-S は Body の身体運動が分離された映像なので, Ghost が能動的にボールを追いかける必要がある. このことは, 図 11-(b) で, Body の運動と B-S の Ghost の運動が同期したシーケンスであることから明らかである. また, B-N においても振幅こそ小さいが Body の運動と同期した Ghost の運動が見られたことも述べておく.

シナリオ (c) では, シナリオ (b) とは逆に, 回転の活発度 RI は C-N に比べて, C-S の方が小さかった. このこと

は, 図 11-(c) から読み取ることが出来る. これは, C-S においては Gyro-vision によって, 相手のプレイヤーが常に一定の方向に存在するため, Body の運動に関わらず, 安定して相手プレイヤーを観察することが出来るためであると考えられる. 一方, C-N では, 映像中の Body の運動によって映像中のプレイヤーのいる方向は常に変化するため, Ghost はそれを追いつけなければならない, このような結果になったと考えられる. ここで, B-S や B-N において観測された動きと C-N を比較すると, Ghost の動きが Body の動きと同期しておらず, 逆の位相での動きが発生している場合もある. これが, C-N において, 大きな SSQ 値を引き起こした原因ではないかと考えられる.

5. 考察

ここでは, 前章の実験の結果を受けて複数の観点から考察を行うとともに, そこで得られた洞察からヒューマンテプレゼンスにおけるインタラクションデザインに関して提言を行う.

5.1 同期と非同期

B-N と C-N を比較すると, 両者 Gyro-vision がないにも関わらず, B-N では C-N に比べてあまりモーションシックネスが発生していない. また, B-N と B-S では, 大きな差は見られなかったが, 7人の被験者が B-N において「体験にリアリティを感じた」とコメントしている. また, 頭部運動 (図 11) の分析から Body と Ghost の動きに同期が見られる場合に, モーションシックネスの発生が抑えられるとともに, より臨場感が増すこと, 反対に, 同期が少ない場合にモーションシックネスが引き起こされるとともに, 体験の臨場感が低下することが示唆される. このことから, ヒューマンテプレゼンスにおいては, 動きの同期を促すことにより, モーションシックネスが軽減するとともに, より臨場感の高い体験を提供できる可能性がある. また, Body と完全に同じ動きをしなくても, 動きの方向性が一致しているだけでもこのような事象がおこっていることから, 身体感覚の拡張により, 他人の体験の自己化が可能なることを示唆している. これを応用すると, 通常再現が不可能な身体運動に関しても, 映像に同期した身体運動を Ghost に対して促すことで, 高い臨場感で同時体験を可能にすることが出来る.

5.1.1 独立性インタラクションのための Gyro-vision

A-N において, Ghost が予想しない動き, 例えば, Body が車を確認して急に横を見る, 道を渡るときに左右を確認するなどの時に Body と逆方向の動きを Ghost が行った場合に, 激しい酔いが発生した. この事象は C-N においても強く感じたと報告されている. これらのことより, Ghost と Body の動きの不整合が強いモーションシックネスを引き起こすことが示唆された. 一方で, 今回のタスクのよう

に、Bodyの周辺環境を独立して観測し、コミュニケーションすることがヒューマンテレプレゼンスにおいては重要であり、Gyro-visionによる身体運動の分離が有効であると考えられる。

5.2 Bodyの頭部方向の可視化

インタビューにおいて、複数の被験者がC-Nに比べて、A-Nの方が強い酔いを感じたと報告している。この原因として、A-Nにおいては、「Bodyがどのような動きをするのか予想が出来ず、そのため瞬間的に発生する動きの不整合が大きくなり、強い酔いを感じた。」というコメントが得られた。このことから、周辺の映像だけでなく、Body自身の状態も映像の中に表示することにより、Bodyの意図を把握しやすくするとともに、次にどのような身体挙動を取るのか予測できるようにすることが重要であると考えられる。そこで、LiveSphereシステムでGyro-visionとBodyの頭部姿勢を同時伝送することで、Ghost側で全天球映像に重畳してBodyの頭部が向いている方向を示すグラフィックスを表示する(図12)。



図12 全周囲映像上のBodyの頭部方向の重畳表示

Fig. 12 Indication of Body's head direction overlaid in omnidirectional video.

5.3 モーションシックネスの改善

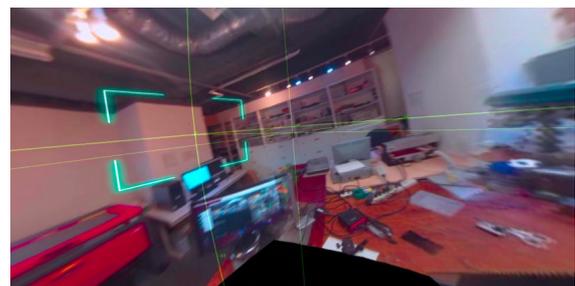
数人の被験者がGyro-visionによって回転を分離しても、モーションシックネスに陥ったとコメントした。インタビューの結果、その原因は大きく2つに分類された。1つは、平行移動における急激な速度の変化である。この変化は、シナリオ(b)とシナリオ(c)において、ボールを拾うため走り出す際や、シナリオ(a)において、あたりを見回すために停止する際に観察された。もう一つは、垂直方向の振動である。A-Sにおいて複数の被験者が、垂直方向の振動に対して、不快感を持ったとコメントした。これらは、Bodyの平行運動に起因しており、現状のGyro-visionアルゴリズムでは取り除くことが出来ない。このような平行移動とその加速度に起因するモーションシックネスに関しても、今後の研究の課題とする。

6. リアルタイムLiveSphereのパイロットスタディ

実験によって得られた結果をもとに、LiveSphereのリアルタイムシステムの実装し、パイロットスタディを行った。これは、実際ユーザにリアルタイムLiveSphereを使ってもらいフィードバックを集めるとともに、リアルタイム構成におけるGyro-visionの有効性を検証するためである。パイロットスタディでは、一人の被験者がBodyとなり、ヘッドギアをつけ研究所の工作室において、もう一人の被験者がGhostとなり別の部屋にいる(図13)。両方の被験者は散らかった工作室の中から特定のいくつかの物を探すように指示を出す。パイロットスタディには6人の被験者が参加し、それぞれBodyとGhost両方の役割を体験した。その結果、リアルタイムLiveSphereにおいてもGyro-visionによるモーションシックネスの改善や独立的な探索活動を確認することが出来た。また、評価実験の考察から実装を行ったBodyの頭部の方向の可視化は、会話の結果Bodyがどのように動いているのか、Ghostが直感的に分かるので、コミュニケーションをする上で非常に有効であるというコメントが得られた。しかし、数人の被験者においては、会話中にGhostがどこを見ているのか分からないことに対して不安感を示し、「Ghostに自分がしている事のすべてを見られている感じがする。」というコメントもあった。今後の研究で、これらのインタラクションデザインの課題を解決し、LiveSphereによるヒューマンテレプレゼンスを改善していこう考えている。



1. An omnidirectional image of the pilot study setup



2. View image for Ghost with the visualization of the Body head motion

図13 LiveSphereリアルタイム構成のパイロットスタディ
Fig. 13 Pilot study for LiveSphere real-time architecture.

7. 関連研究

7.1 テレプレゼンス

これまで、遠隔地にあるセンサやアクチュエータまた、システムを操作するためのインタフェースに関して多くの研究がなされてきた [26]. 特に、テレプレゼンスの領域では、ネットワークやセンシング、ディスプレイの技術の向上により、人間に、実際に自分が遠隔地にいるかのような感覚を与えることを可能にした [29]. また、遠隔作業のためのマスタースレーブ型のテレプレゼンスシステムでは、人間が遠隔地の環境とインタラクションを行うことが可能になった [28]. これらの技術は、人間の動きによりロボットの動きと同期させ直接操作することを可能にした. 例えば、TORSO[30] は、ロボットと人間の顔の動きを連動させることによって、視覚的テレプレゼンスを実現した. また、Flying Head[12] は、UAV(Unmanned Aerial Vehicle)の動きを人間の動きに連動させる. ここで、UAV に搭載されたカメラの映像を HMD によりリアルタイムに見ることにより、あたかも自分の身体運動が拡張されたような感覚を得ることが出来る. これらの研究において、人間とロボット間のマスタースレーブ型のロボットテレプレゼンスが形成されている.

7.2 ヒューマンテレプレゼンスの初期研究

初めて人間へのテレプレゼンスを行った研究としては、視聴覚交換マシンがある [11]. これは、2 人の人間が HMD とカメラを装着し、互いのカメラの映像を HMD によって見せることによって、互いが相手の中に入り込む体験を提供する. また、The Machine to be Another[3] では、入り込まれる側の人間が入り込む側の人間と同期して動くことにより、入り込む側の人間に本当に自分であるかのような体験を提供する. これらの研究は人間の自己同一性の問題を提起するとともに、ヒューマンテレプレゼンスにおけるインタラクションデザインの可能性を示している.

7.3 一人称視点映像を用いたインタラクション

リモートコラボレーションの領域において、一人称視点映像を用いた様々なシステムの研究が行われている. 頭部搭載型カメラによる一人称視点映像の共有は、装着者が見た者や装着者の意図を共有することを可能にするが、一方で、Fussell ら [8] は、一人称視点映像を使うことによるいくつかの問題を指摘している. 例えば、映像の振動や同じ視野に制限されてしまうこと、また観測者が周辺環境を把握するためには現場にいる人に向きを変えてもらわなければならないこと等である. これまで、一人称視点映像を使った様々な応用例があり、知識の拡張 [21] やスキル転送 [15] などの研究が行われている. また、一人称視点映像

から仮想的により広い視野を提供するという研究も行われており、Cheng ら [6] の研究では、一連の一人称視点映像をスティッチングしていくことによってより広い視野を提供している. また、笠原ら [14] は、画像を 3 次元的にスティッチングしていくことにより仮想的に空間の再構築を行った、しかし、ここで用いられる画像処理は、大きな動きや屋外での活動において十分ロバストとは言えない.

7.4 モーションシックネス

視覚情報や一人称視点映像によるテレプレゼンスでは、観測者がウェアラブルカメラや没入型の視覚環境の映像を見る際に酔いを引き起こすことがある. これは、乗り物酔い等のモーションシックネス [23] とほとんど同じ症状で、近くされた動きと前庭系の動きの感覚の不一致によって発生する [22]. バーチャルリアリティにおいては、シミュレーションシックネス [18] と呼ばれ、フライトシミュレータなどでよく見られる. モーションシックネスを引き起こす主な原因の一つは、知覚の矛盾である. 視覚によって認識される自己の運動と、前庭系によって知覚される慣性力に不整合が生じるときに、めまいや方向感覚の喪失、吐き気等のモーションシックネスの症状が発生する [16]. さらに、HMD 等の没入型の視覚環境においては、モーションシックネスはさらに強いものとなる [13]. 一方で、Sharples らの研究 [25] では、ヘッドトラッキングによる頭部動きに連動した映像表示により症状が改善されることが示されており、低遅延のヘッドトラッキングや、ディスプレイのフレームレートの向上によりこれらの問題は改善されている. しかし、ウェアラブルカメラにおいては、このモーションシックネスは十分に解消されていない状況である.

7.5 全周囲画像処理と映像スタビライズ

画像スタビライズは、一人称視点映像ストリーミングにおける基本的な技術である. Kopf ら [20] は一人称視点映像から滑らかなタイムラプス映像が作れる手法を提案した. この手法は、ビデオのフレームを減らすことなく映像の揺れの問題を解決した. また、全周囲映像は、遠隔支援システム [5] などの情報共有のためのみならず、カメラの自己位置推定や運動推定にも用いられてきた. カメラの回転運動を推定するための手法は、画像のみから行う手法 [10] から、モーションセンサを用いる手法 [1], まで数多く存在する. Bazin ら [2] は、回転だけでなく平行移動も推定できる手法を提案した. これらの技術は、ロボットや乗り物の操作などに応用されている [9]. また、近藤ら [19] はウェアラブルな全周囲カメラを提案したが、モーションシックネスや装着性などの問題は解決されていない.

8. 結論

本稿では、LiveSphere と呼ぶ、ウェアラブルな全方位カ

メラを用いた体験共有のためのシステムを提案した。そして、システムを用いたヒューマンテレプレゼンスを提案し、その中でモーションシックネスの改善と両者が独立性を保ったインタラクションを実現した。評価実験の結果は、LiveSphere システムが提供する Gyro-vision が、記録した全周囲映像から装着者の身体運動を分離し、観測者の身体運動を畳み込んで表示することで、これらの問題を解決するとともに、ヒューマンテレプレゼンスというコンセプトが実現可能であることを示した。また、実験で得られたフィードバックやパイロットスタディからは、ヒューマンテレプレゼンスにおける多くの洞察を得た。

参考文献

- [1] Albrecht, T., Tan, T., West, G. A., Ly, T. (2010, December). Omnidirectional video stabilisation on a virtual camera using sensor fusion. In *Control Automation Robotics Vision (ICARCV), 2010 11th International Conference on* (pp. 2067-2072). IEEE.
- [2] Bazin, J. C., Démonceaux, C., Vasseur, P., Kweon, I. (2012). Rotation estimation and vanishing point extraction by omnidirectional vision in urban environment. *The International Journal of Robotics Research*, 31(1), 63-81.
- [3] Bertrand, P., Gonzales, D., Pointeau, A., Cherece C., The Machine to be Another - Embodied Telepresence using human performers WIP Demo, In *TEI 2014 Proceedings in the ACM* (2014)
- [4] Bouguet, J. Y. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. Intel Corporation, 5. (2001).
- [5] Boulton, T. E. (1998, July). Remote reality via omnidirectional imaging. In *ACM SIGGRAPH 98 Conference abstracts and applications* (p. 253). ACM.
- [6] Cheng, L. T., Robinson, J. Dealing with speed and robustness issues for video-based registration on a wearable computing platform. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on* (pp. 84-91). IEEE. (1998).
- [7] Douglas Trumbull (director), "Brainstorm," Advanced Robotics, Metro-Goldwyn- Mayer, 1983.
- [8] Fussell, S. R., Kraut, R. E., Siegel, J. Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (pp. 21-30). ACM. (2000).
- [9] Gandhi, T., Trivedi, M. (2005). Parametric ego-motion estimation for vehicle surround analysis using an omnidirectional camera. *Machine Vision and Applications*, 16(2), 85-95.
- [10] Gluckman, J., Nayar, S. K. (1998, January). Ego-motion and omnidirectional cameras. In *Computer Vision, 1998. Sixth International Conference on* (pp. 999-1005). IEEE.
- [11] Hachiya, K. Inter Dis-communication Machine. *Prix Ars Electronica*, 96, 138-139. (1993).
- [12] Higuchi, K., Rekimoto, J. Flying Head: head-synchronized unmanned aerial vehicle control for flying telepresence. In *SIGGRAPH Asia 2012 Emerging Technologies* (p. 12). ACM. (2012).
- [13] Howarth, P. A., Costello, P. J. The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. *Displays*, 18(2), 107-116. (1997).
- [14] Kasahara, S., Rekimoto, J. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference* (p. 46). ACM. (2014).
- [15] Kawasaki, H., Iizuka, H., Okamoto, S., Ando, H., Maeda, T. Collaboration and skill transmission by first-person perspective view sharing system. *RO-MAN, 2010 IEEE*, 125-131. (2010).
- [16] Kennedy, R. S., Drexler, J., Kennedy, R. C. Research in visually induced motion sickness. *Applied ergonomics*, 41(4), 494-503. (2010).
- [17] Kennedy, R. S., Lane, N. E., Berbaum, K. S., Lilienthal, M. G. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3), 203-220. (1993).
- [18] Kolasinski, E. M. Simulator Sickness in Virtual Environments. (1995).
- [19] Kondo, K., Mukaigawa, Y., Yagi, Y. (2009, November). Wearable imaging system for capturing omnidirectional movies from a first-person perspective. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (pp. 11-18). ACM.
- [20] Kopf, J., Cohen, M. F., Szeliski, R. First-person hyperlapse videos. *ACM Transactions on Graphics (TOG)*, 33(4), 78. (2014).
- [21] Kortuem, G., Bauer, M., Segall, Z. (1999). NETMAN: the design of a collaborative wearable computer system. *Mobile Networks and Applications*, 4(1), 49-58.
- [22] Oman, C. M. Motion sickness: a synthesis and evaluation of the sensory conflict theory. *Canadian Journal of Physiology and Pharmacology*, 68(2), 294-303. (1990).
- [23] Reason, J. T., Brand, J. J. Motion sickness. Academic press. (1975).
- [24] Scaramuzza, D., Martinelli, A., Siegwart, R. A toolbox for easily calibrating omnidirectional cameras. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (pp. 5695-5701). IEEE. (2006).
- [25] Sharples, S., Cobb, S., Moody, A., Wilson, J. R. Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays*, 29(2), 58-69. (2008).
- [26] Sheridan, T. B. Teleoperation, telerobotics and telepresence: A progress report. *Control Engineering Practice*, 3(2), 205-214. (1995).
- [27] Shi, J., Tomasi, C. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on* (pp. 593-600). IEEE. (1994)
- [28] Tachi, S., Arai, H., Maeda, T. Tele-existence master-slave system for remote manipulation. II. In *Decision and Control, 1990., Proceedings of the 29th IEEE Conference on* (pp. 85-90). IEEE. (1990).
- [29] Tachi, S., Tanie, K., Komoriya, K., Kaneko, M. Tele-existence (I): Design and evaluation of a visual display with sensation of presence. *Theory and Practice of Robots and Manipulators*, 245-254. (1985).
- [30] Watanabe, K., Kawabuchi, I., Kawakami, N., Maeda, T., Tachi, S. TORSO: Development of a teleexistence visual system using a 6-dof robot head. *Advanced Robotics*, 22(10), 1053-1073. (2008).