Vol.2017-HCI-173 No.21
Vol.2017-EC-44 No.21
2017/6/2

対面協調作業に適した相互モーションキャプチャシステムの開発

中村 淳之[*1]　清川 清[*2]　Photchara Ratsamee[*1, *3]　間下 以大[*1, *3]
浦西 友樹[*1, *3]　竹村 治雄[*1, *3]

A Mutual Motion Capture System for Face-to-face Collaboration

Atsuyuki Nakamura[*1], Kiyoshi Kiyokawa[*2], Photchara Ratsamee[1, *3], Tomohiro Mashita[*1, *3], Yuki Uranishi[*1, *3] and Haruo Takemura[*1, *3]

Abstract - In recent years, motion capture (MoCap) technology to measure the movement of the body has been used in many fields. Moreover, MoCap targeting multiple people is becoming necessary in multi-user VR. It is desirable that MoCap requires no wearable devices to capture natural motion easily. Some systems require no wearable devices using an RGB-D camera fixed in the environment, but the working range of the user is limited. Therefore, in this research, proposed is a MoCap technique for a multi-user VR environment using Head Mounted Displays (HMDs), that does not limit the working range of the user nor require any wearable devices. In the proposed technique, an RGB-D camera is attached to each HMD and MoCap is carried out mutually. The MoCap accuracy is improved by modifying the depth image. A prototype system has been implemented to evaluate the effectiveness of the proposed method and MoCap accuracy has been compared with two conditions, with and without depth information correction while rotating the RGB-D camera. As a result, it was confirmed that the proposed method could decrease the number of frames with erroneous MoCap by 49% to 100% in comparison with the case without depth image conversion.

Keywords : Virtual Reality, Motion Capture, Collaboration, Head Mounted Display and RGB-D Camera

1. はじめに

近年，Microsoft 社の HoloLens[1] や Google 社の Google Glass[2] などのスタンドアローンのヘッドマウントディスプレイ(Head Mounted Display, HMD)が登場し普及してきている．スタンドアローンのシステムであるため，HMD の利用シーンが時と場所に限定されず広がっており，HMD を用いた協調作業もますます利用されていくと考えられる．従って，バーチャルリアリティ(Virtual Reality, VR)や拡張現実(Augmented Reality, AR)を用いた協調作業についても時と場所に限定されずに利用可能であることが望ましい．

同一地点における協調作業の場合，お互いの身体を見ながら作業を行うほうが，作業効率を上げることができる．Billinghurst らの研究[1]によると，壁面スクリーンを用いた協調作業支援システムより，HMD を用いた協調型作業支援システムを用いたほうが，ユーザが現実における対面協調作業の振る舞いに近くなることを示している．Szalavári らが開発した Studierstube[2]では，AR によって表示された情報を，複数のユーザがお互いの様子を見ながら共有することができる．しかし，従来システムの多くは HMD 以外にも装着が必要なデバイスがある．VR や AR を用いた協調作業において，準備に手間がかからず，現実と同じような体の動きにより直感的な操作を行うことができるモーションキャプチャシステムが望まれる．

現在ではモーションキャプチャに関する様々な研究が行われている．また，モーションキャプチャを行うデバイスも数多く開発され，Microsoft 社の Kinect[3] に代表されるように，一般家庭向け製品も普及してきている．エンドユーザが手軽に利用でき，かつ自然な動きを再現するには，行動範囲に制限がなく，体に何も装着せずにモーションキャプチャを行えることが望ましい．モーションキャプチャは主に，装着したマーカやセンサを外部のデバイスで認識する手法，装着したセンサのみを用いて認識する手法，マーカを用いず外部に設置したカメラやセンサで認識する手法の 3 種類がある．

マーカやセンサを用いる手法として光学式のマーカを用いる OptiTrack 社のモーションキャプチャシステム[3]や，磁気式のセンサを用いる Polhemus 社の LIBERTY[4]が挙げられる．複数のカメラやセンサで認識するので，高い精度で計測を行うことができる．しかし，マーカの着脱が煩雑であり，複数のカメラやセンサが同時に認識できる範囲内でしかモーションキャプチャが行えない．装着したセンサを用いる手法として Noitom 社の Perception

*1: 大阪大学 大学院情報科学研究科
*2: 奈良先端科学技術大学院大学 情報科学研究科
*3: 大阪大学 サイバーメディアセンター
*1: Graduate School of Information Science and Technology, Osaka University
*2: Graduate School of Information Science, Nara Institute of Science and Technology.
*3: Cybermedia Center, Osaka University

1: https://www.microsoft.com/ja-jp/hololens

2: https://www.google.com/glass/start/

3: http://www.xbox.com/en-US/xbox-360/accessories/kinect

Vol.2017-HCI-173 No.21
Vol.2017-EC-44 No.21
2017/6/2

Neuron[5]や Slyper らが開発したシステム[6]が挙げられる．センサを体の各部位に装着するだけであるので，行動範囲が制限されないが，装着が煩雑であり，装着感がある．外部に設置したカメラやセンサを用いる手法として田中らの研究[7]，Microsoft 社の Kinect[8]が挙げられる．体に装着するデバイスは不要であるが，カメラやセンサを環境に設置した場所でしかモーションキャプチャを行えず，行動範囲が制限されてしまう．

そこで本研究では，複数のユーザがそれぞれ HMD を装着するような VR あるいは AR システムを対象として，これに適したモーションキャプチャシステムを開発する．このモーションキャプチャシステムでは，それぞれのユーザが装着する HMD に，他のユーザのモーションキャプチャを行うデバイスを取り付け，相互にモーションキャプチャを行う．これにより，体に取り付ける機器を最少化して着脱の煩雑さを軽減し，行動範囲を限定しないモーションキャプチャシステムを実現する．

2. 提案システム

2.1 システム概要

提案するシステムでは，HMD を装着した VR や AR における対面協調作業を想定している．そこで，HMD にモーションキャプチャデバイスを取り付けてモーションキャプチャを行う．それぞれのユーザが装着する HMD に他のユーザのモーションキャプチャを行うモーションキャプチャデバイスが取り付けられているため，相互的なモーションキャプチャが可能となる．図 1 のように他のユーザのモーションキャプチャデータをユーザ間で共有することで，全てのユーザのモーションキャプチャを行う．

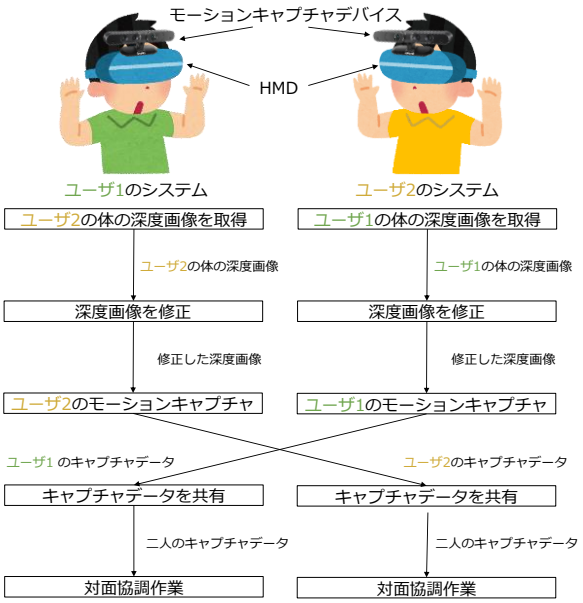


図 1　提案システム

Fig.1　Proposed system

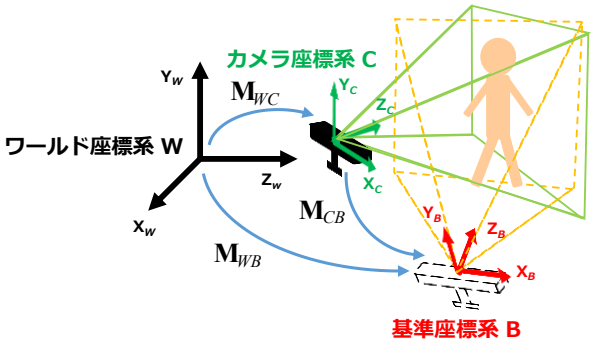


図 2　提案手法における座標系

Fig.2　Coordinate systems in the proposed method

2.2 提案手法

提案システムではモーションキャプチャデバイスとして RGB-D カメラを HMD に取り付けて用いる．しかし，RGB-D カメラを用いるモーションキャプチャ手法ではデバイスを固定して使用することを想定しているため，HMD が移動すると正確なモーションキャプチャができないと予想される．そこで本システムでは，取得した深度画像に修正を加えることで，モーションキャプチャ精度の低下を軽減する手法を提案する．

2.2.1 深度情報の座標変換

安定したモーションキャプチャが可能な深度画像に修正するため，深度情報の座標変換を行う．図 2 のように RGB-D カメラの座標系，深度画像の基準となる座標系がある．実環境の座標系をワールド座標系，RGB-D カメラの座標系をカメラ座標系，深度画像の基準の座標系を基準座標系と呼び，それぞれを *W*，*C*，*B* と表す．基準座標系 *B* は，実際の RGB-D カメラの近傍で，かつモーションキャプチャに適すると考えられる座標系である．ワールド座標系からカメラ座標系への変換行列 $\mathbf{M}_{WC}$ は，ワールド座標における RGB-D カメラの位置や移動量から求められる．本研究では，移動量の算出に HMD に内蔵されている慣性センサを用いる．取得した移動量の回転成分 **R**，平行移動成分 **t** を (1)式，(2)式とする．

$$\mathbf{R} = \mathbf{R}_x(\varphi)\mathbf{R}_y(\theta)\mathbf{R}_z(\psi) \quad (1)$$

$$\mathbf{t} = (tx, ty, tz) \quad (2)$$

ここで $\mathbf{R}_x$，$\mathbf{R}_y$，$\mathbf{R}_z$ はそれぞれ X 軸周り，Y 軸周り，Z 軸周りの回転行列を表す．すると，変換行列 $\mathbf{M}_{WC}$ は(3)式のように表される．

$$\mathbf{M}_{WC} = \begin{pmatrix} & \mathbf{R} & & \mathbf{t} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3)$$

安定したモーションキャプチャが可能な深度画像に修正するため，深度情報の座標変換を行う．座標変換はカメラ座標系から基準座標系への座標変換となる．ここで，RGB-D カメラによるモーションキャプチャシステムでは，デバイスがある高さで水平に設置され，対面ユーザを適切な距離からほぼ正面方向で捉えている場合には正

しく動作するものと仮定する．すると，基準座標系はワールド座標系からの変換を，床面上の位置 X，Y および方位角 θ の 3 自由度で指定できる座標系と定めることができる．よってワールド座標系から基準座標系への座標変換行列 $\mathbf{M}_{WB}$ は(4)式のように表される．

$$\mathbf{M}_{WB} = \begin{pmatrix} & \mathbf{R}' & & \mathbf{t}' \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (4)$$

ここで $\mathbf{R}'$，$\mathbf{t}'$は，(5)式，(6)式と表される．

$$\mathbf{R}' = \mathbf{R}_y(\theta') \quad (5)$$

$$\mathbf{t}' = (tx', ty, tz') \quad (6)$$

θ'，tx'，tz'は基準座標系にカメラを配置した際に，対面ユーザのモーションキャプチャが正確に行われるような値を設定する．よって，カメラ座標系から基準座標系への座標変換行列 $\mathbf{M}_{CB}$ により，深度情報を修正する．$\mathbf{M}_{CB}$ は $\mathbf{M}_{WC}$ と $\mathbf{M}_{WB}$ により(7)式のように求められる．

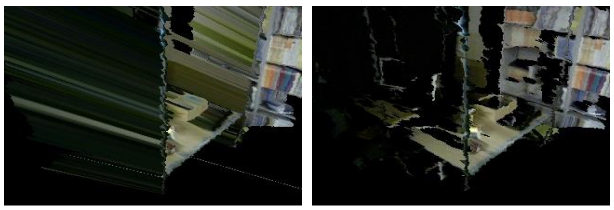
$$\mathbf{M}_{CB} = \mathbf{M}_{WC}^{-1}\mathbf{M}_{WB} \quad (7)$$

座標変換後の深度情報を用いたモーションキャプチャでの体の部位の位置座標は，基準座標系における座標である．ゆえにモーションキャプチャデータをワールド座標系に座標変換を行う必要がある．対面ユーザ上の点 $\boldsymbol{P}$ を考える．モーションキャプチャにより基準座標系上での位置 $\boldsymbol{P}_B$ を取得できる．ワールド座標系での点の座標 $\boldsymbol{P}_W$ は，変換行列 $\mathbf{M}_{WB}$ を用いて(8)式のように求められる．

$$\mathbf{P}_W = \mathbf{M}_{WB}^{-1}\mathbf{P}_B \quad (8)$$

2.2.2　深度画像の補完

取得した深度情報を座標変換する際に，深度画像の画素とその深度値をもとに点群を 3 次元空間に配置する．一般に基準座標系から見た点群の深度画像には穴ができ，深度情報が一部欠損してしまう．これを防ぐため，点群の各近傍点から三角形を構成しメッシュを生成する．これにより隣接点間を補完できるので，深度情報の欠損を抑え，モーションキャプチャへの影響を低減できる．しかし，すべての点に対してメッシュを生成すると，図 3(a)に示すように隣り合う点が離れている箇所では実際には存在しない面が生成されてしまう．そこで，隣り合う点の距離に対する閾値 th を設定することで，図 3(b)が示すように不要な面の生成を抑制する．



(a) 全ての点に対しメッシュを生成したデータ　(b) 閾値を設定したメッシュを生成したデータ

図 3　閾値を設定したメッシュデータ
Fig.3　Mesh data with a threshold

3.　試作システム

3.1　試作システム

提案手法の有効性を評価するための試作システムを作成し，実験を行う．試作システムのモジュール図を図 4 に示す．試作システムに使用した計算機の構成は CPU が Core i7-6700k，メモリが 8GB，GPU は NVIDIA GeForce GTX1080 8GB である．RGB-D カメラは ASUS 社の Xtion PRO LIVE(解像度:320×240，フレームレート:60fps，重量:210g)を使用し，HMD は Oculus VR 社の Oculus Rift DK2(片目解像度:1080×1200，重量:440g)を使用した．モーションキャプチャに用いたライブラリは OpenNI 1.5.4.0 for Windows，NiTE 1.5.2.21 for Windows，PrimeSense Sencor 5.1.2.1 for Windows である．また，OpenNI 1.5.4.0 をインストールしたときに同梱されている OpenGL 1.1, glut 3.7 を用いて，深度情報を点群として 3 次元座標に配置し，座標変換を行う．

実験を行った試作システムでは，基準座標系はワールド座標系に固定し，$\theta = 0$，$tx' = tx$，$tz' = tz$ として $\mathbf{M}_{CB}$ を求め座標変換を行う．また，メッシュ生成に関する閾値 th を 10cm とする．

3.2　予備実験

初めに，試作システムにおいて深度情報の座標変換やメッシュ生成が適切に行われているかを検証する予備実験を行った．同時に，提案システムでは VR や AR を用いた対面協調作業を対象としているため，試作システム

図 4　試作システムのモジュール図
Fig.4　Modules of a prototype system

Vol.2017-HCI-173 No.21
Vol.2017-EC-44 No.21
2017/6/2

の処理時間を検証する予備実験を行った．深度情報の座標変換の実験では図 5 に示すような，RGB-D カメラを HMD に装着したヘッドセットを回転させる．また回転は図 5 のピッチ軸，ヨー軸，ロール軸の周りにそれぞれ回転させ，座標変換を行った．

例としてロール軸の周りに回転させた様子を図 6(c)に示す．図 6(d)のように座標変換を行うことにより，深度情報を取得した範囲において，回転前(図 6(b))と回転後(図 6(d))で同じような深度画像を生成できることを確認した．一方で，回転により視界から外れた範囲は深度情報が欠損した．

また，深度情報の補完の検証では，図 6(a)と同じ環境において，図 7 の(a)，(b)に示すように点群の座標変換において発生していた穴が補完により減少していることを確認した．

処理時間の検証では，図 8 に示す静止したユーザのボディトラッキングにおいてシステムの 1 秒間のフレーム数を計測し比較した．比較する条件は深度情報の座標変換なし条件，深度情報の座標変換あり・補完なし条件，深度情報の座標変換あり・補完あり条件の 3 条件とする．1 フレーム当たりの処理時間をそれぞれ図 9 のグラフに示す．3 次元データ生成は，3 次元の点群およびメッシュデータ生成を表し，ワールド座標への変換はモーションキャプチャデータの基準座標からワールド座標への座標変換を表す．1 秒当たりのフレーム数に注目すると変換なし条件では約 60 フレーム/秒なのに対し，変換あり・補完なし条件では約 52 フレーム/秒，変換あり・補完あり条件では約 42 秒/フレームとなった．結果から提案手法により処理時間は増加したが，実時間で処理を行えることを確認した．

4. 実験

4.1 実験条件と結果

次に提案手法によって RGB-D カメラが動いた時のモーションキャプチャ精度の低下を軽減できるかを検証する本実験を行った．

処理時間の検証と同じ静止した対面ユーザを対象としてモーションキャプチャを行っている状態で，予備実験と同様にヘッドセットをピッチ軸，ヨー軸，ロール軸の周りに回転させた場合のモーションキャプチャ精度を検証する．各軸における回転の方法は表 1 に示すとおりであり，手動で回転させる．比較として，予備実験における処理実験の検証で比較した 3 条件でモーションキャプチャの精度の比較を行う．なお，RGB-D カメラで取得した動画像のデータはファイルに保存しておき，これを再利用することにより条件間で公平な評価を実施する．モーションキャプチャ精度の評価には，図 8 に示すようにユーザの頭，両手，両足首に取り付けたマーカと深度画像における体の各部位の 2 次元座標のユークリッド距離を誤差として取り扱う．各条件における，モーションキャプチャの誤差の実験結果を図 10 に示し，例としてロール軸回転における 5.5 秒後のモーションキャプチャの様子を図 11 の(a)から(c)に示す．

図 5　実験における回転軸

Fig.5　Rotation axes

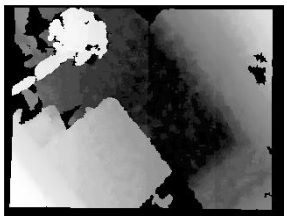
表 1　回転の動作

Table.1　Motion of rotation

動作	時間
静止	1 秒
左回りに 10 度回転	1 秒
静止	1 秒
右回りに 20 度回転	1 秒
静止	1 秒
左回りに 10 度回転	1 秒
静止	3 秒

(a) 予備実験の実験環境

(b) 実験環境の深度画像



(c) 座標変換前の深度画

(d) 座標変換後の深度画

図 6　座標変換の予備実験

Fig.6　Experiment of transformation of depth data

(a)補完前の深度情報

(b)補完後の深度情報

図 7　補完の予備実験

Fig.7　Expriment of hole filling

図 8　実験における対象ユーザ

Fig.8　A subject of main experiment

Vol.2017-HCI-173 No.21
Vol.2017-EC-44 No.21
2017/6/2

4.2 考察

本システムを用いた場合，変換なし条件に比べて静止ユーザに対してのモーションキャプチャの精度が安定し，誤差が少なくなっていることを確認した．図 10 のロール軸回転におけるグラフを見ると，変換なし条件において誤差が非常に大きいことが確認できる．これは図 11 の(a)を見ると，背景が動くことにより背景もユーザとして認識していることが原因であることが分かる．一方で，変換あり・補完なし条件および変換あり・補完あり条件では背景が動かないので背景をユーザとして認識することがないため，誤差を抑制できていることが分かる．

変換なし条件では回転軸によって，誤差の変化に大きく差が出ている．ピッチ軸回転，ロール軸回転では回転が始まると大きく増加するが，ヨー軸回転では他の手法に比べて大きな差がないことがわかる．このような結果になった要因は，試作システムに実装しているモーションキャプチャシステムはユーザが左右に動くことは想定しているが，著しく上下に移動，または RGB-D カメラのロール軸を回転軸として回転することを想定していないためである[8]．このようなことから，モーションキャプチャシステムはユーザのシステムが想定していない程大きな上下運動には弱いといえる．

誤推定頻度として腕や足の幅以上の誤差となっているフレームを数える．各条件における誤推定頻度を図 12 に示す．図 12 に示すように変換あり・補完なし条件の誤推定頻度は変換なし条件と比べて，頭，左手，右手，左足，右足の各部位においてそれぞれ，17%，57%，25%，4%，0%軽減した．また，変換あり・補完あり条件の誤推定頻度は変換なし条件と比べて，頭，左手，右手，左足，右足の各部位においてそれぞれ 100%，79%，100%，49%，49%軽減している．このことから RGB-D カメラを HMD に装着した場合，頭の動きを相殺する座標変換がモーションキャプチャ精度低下の軽減に有効であることが分かる．変換あり・補完なし条件では，変換あり・補完なし条件に比べて誤推定頻度を軽減していることが分かる．よって，深度情報の座標変換を行った場合，補完を行うことがさらに有用であることが確認できる．一方で，提案手法を用いた条件では，用いない条件に比べて，ユーザではない背景を人物と誤検出しやすいことが判明した．ヘッドセットを回転させる前に誤検出していることから，点群またはメッシュデータから深度画像を生成する際に発生する計算誤差が影響することが原因であると考えられる．

5. 終わりに

本研究では HMD を装着した複数のユーザが AR や VR を用いて行う対面協調作業を対象とし，HMD に RGB-D カメラを装着してユーザが相互にモーションキャプチャを行うシステムを提案した．このシステムでは，安定したモーションキャプチャを実現するため，その移動と回

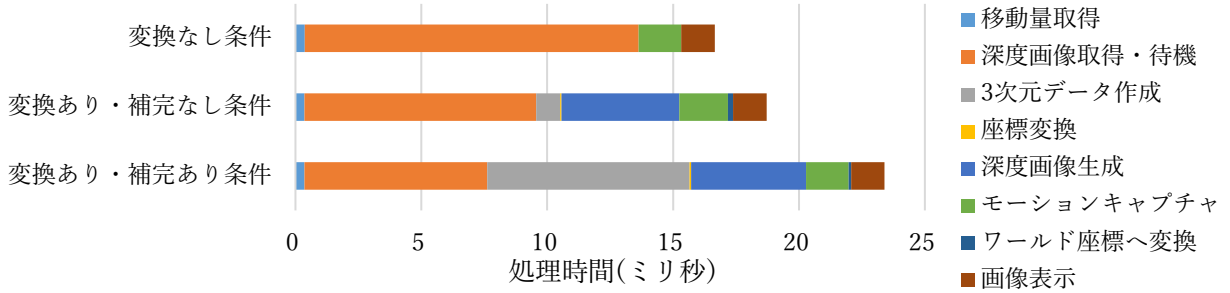


図 9　各条件での処理時間

Fig.9　Process time

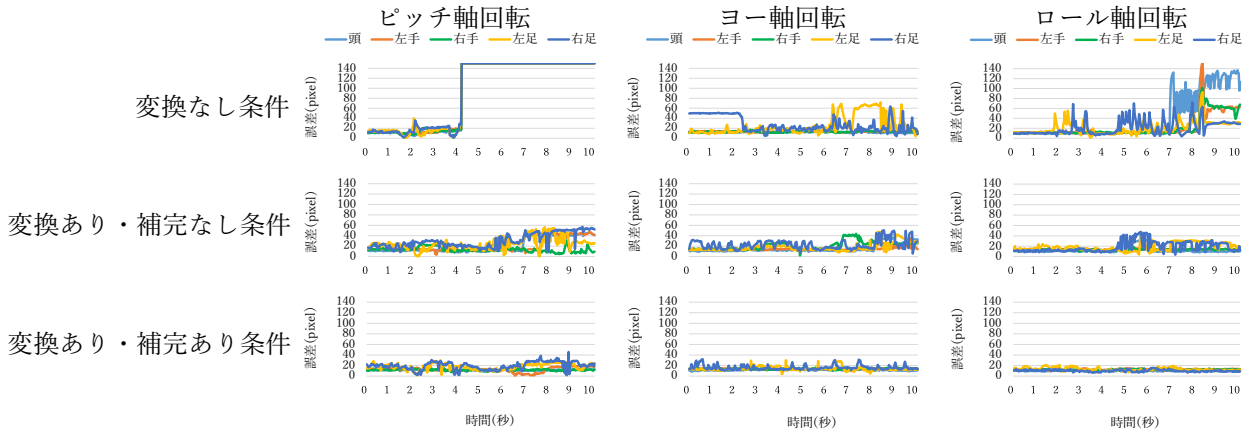
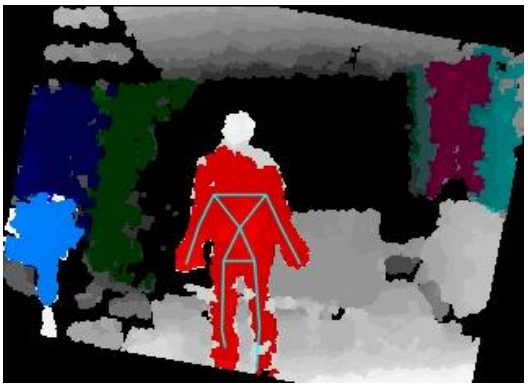


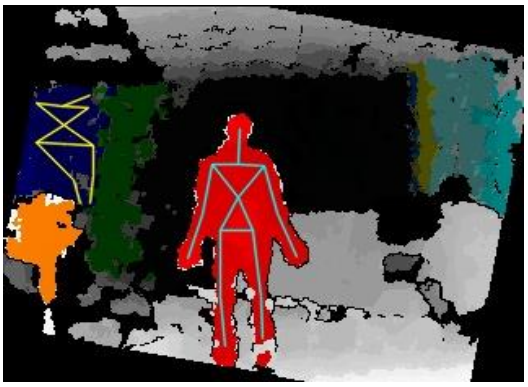
図 10　各軸回転における各部位のモーションキャプチャの誤差

Fig.10　Error of motion capture

(a) 変換なし条件



(b) 変換あり・補完なし条件



(c) 変換あり・補完あり条件

図 11　実験におけるモーションキャプチャの例
Fig.11　An example of motion capture in the main experiment

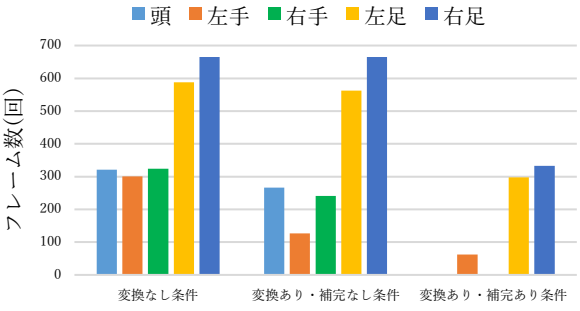


図 12　各条件における誤推定頻度
Fig.12　Number of frames with error

転を相殺し，座標変換で発生する深度情報の欠損を補完した深度画像を生成する．試作システムを通じて，提案手法によりデバイスが回転した場合でも，誤推定頻度を49%から最大 100 %軽減することができることを確認した．

今後の課題として，本研究ではヘッドセットの回転の移動に対応したシステムを開発し検証する必要がある．また 3.2.1 節で述べた，基準座標系において最適な θ'，tx'，tz'を見つける必要がある．本研究における実験では，試作システムの定量評価実験のみを行ったので，本システムを用いた対面協調作業でのユーザによる定性評価実験を行わなければならない．本システムを用いることにより対面協調作業における作業支援システムやエンターテイメント向けのアプリケーションを期待できる．また本システムを拡張し 3 人以上に対応することができれば，システムの適用範囲を拡げることができる．

謝辞

本研究の一部は JSPS 科研 JP16H02858 の助成を受けた．

参考文献

Billinghurst, M., Becher, D., et al.: Communication Behaviors in Co-located Collaborative AR Interfaces; International Journal of Human Computer Interaction, **Vol.**16, No.3, pp.395–423 (2003).

[1] Szalavári, Z., Schmalstieg, D., et al.: Studierstube An Environment for Collaboration in Augmented Reality; IEEE Virtual Reality, **Vol.**3, No.1, pp.37–48 (1998).

[2] OptiTrack; https://www.optitrack.co.jp/, Last Access Jan. 25, 2017

[3] LIBERTY; http://polhemus.com/motion-tracking/all-trackers/liberty, Last Access: Jan. 25, 2017.

[4] Perception Neuron: https://neuronmocap.com/ja/products /perception neuron, Last Access: Jan. 25, 2017.

[5] Slyper, R., Jessica, Hodgins, J.K.: Action Capture with Accelerometers; the ACM SIGGRAPH/Eurographics Symposium on Computer Animationm, pp.193–199 (2008).

[6] 田中, 中澤, 他: ボリュームデータの細線化とグラフマッチングを用いた事例ベース人体姿勢推定; 電子情報通信学会論文誌, **Vol.**6, No.6, pp.1580–1591 (2008).

[7] Shotton, J., Sharp, T., Kipman, A., et al.: Real-Time Human Pose Recognition in Parts from Single Depth Images; Communications of the ACM, **Vol.**56, No.1, pp.116–124 (2013).