Vol.2019-GN-106 No.52
Vol.2019-CDS-24 No.52
Vol.2019-DCC-21 No.52
2019/1/25

身体的特徴を活かした 複数視点モーションキャプチャ・システムの検討

山﨑賢人[1]　阿倍 博信[2]

概要：廉価なモーションキャプチャの登場により，人や物体の動きを手軽にデータ化できるようになった．しかしこれら多くのデバイスは単一視点から三次元的に座標値を推定するにとどまるため，対象物が遮蔽されるなどキャプチャできる領域が限定的であった．このような課題を解決するため，複数台のモーションキャプチャを用いた複数視点モーションキャプチャ・システムを提案する．本稿では，複数視点から得られた各座標値を統合するにあたり，人の身体的特徴を活かした信頼度によって統合する手法について報告する．

キーワード：モーションキャプチャ，点群，多視点

A Study of Multi-View Motion Capture System Based on Physical Features

KENTO YAMAZAKI[1]　HIRONOBU ABE[2]

Abstract: This paper describes multi-view motion capture system. Recently, it is easy to convert digital data from the human or object motion, since a motion capture at a low edition was released. However almost this motion capture has a single-view so that its range area is limited. For example it cannot capture the motion when human is covered the object. In this study, we proposed the multi-view motion capture system which define of the credibility based on physical features, and integrate of each coordinate value based on its credibility.

Keywords: Motion Capture, Point Cloud, Multi-view

1. はじめに

現実空間の人や物体の動きを記録するモーションキャプチャ技術は様々な分野で活用されている．例えば，人の行動解析や，CG のアニメーション制作，インタラクティブなシステムにおけるジェスチャ入力にいたるまで様々なコンテンツに用いられている．
モーションキャプチャには様々な方式があり，代表的なものに光学式モーションキャプチャがある [1]．光学式は任意の領域を囲むように複数台のカメラを設置し，このカメラで計測する人や物体に貼付したマーカをトラッキングすることでモーションを計測する方式である（図 1 参照）．しかしこれらの多くは高価かつ大掛かりな機材のため，使用にはまだまだハードルが高かった．

しかし 2010 年に大きなブレイクスルーが発生する．Microsoft 社から家庭用ゲーム機のジェスチャ入力インタフェースとしてリリースされた Kinect は廉価なモーションキャプチャとして多くの研究に影響を与えた．Kinect は RGB カメラとデプスセンサを搭載し，スケルトンモデルを基に人のモーションを推定す [2]．Kinect 登場以降，様々なデプスセンサを用いたモーションキャプチャがリリース

[1] 三菱電機株式会社 情報技術総合研究所
Information Technology R&D Center, Mitsubishi Electric Corporation
[2] 東京電機大学 システムデザイン工学部
School of System Design and Technology, Tokyo Denki University

Vol.2019-GN-106 No.52
Vol.2019-CDS-24 No.52
Vol.2019-DCC-21 No.52
2019/1/25

されている．

一方で，RGB カメラによるモーションの推定も多く提案されている．代表的なものに Cao らが提案した OpenPOSE がある [3]．しかしこれらは単眼カメラであるため得られる推定値はカメラ画像上の二次元の値である．この値を三次元の値として取得するには，マルチビューステレオなど様々な手法が必要であるため，2 台以上のカメラであってもデプス情報は単一視点と同意となる場合がある．

このようにモーションキャプチャが手軽なデバイスで実現できるようになってきたが，どうしても視点の数は限定的である．そのため，遮蔽物によって隠れている領域など，モーションを推定ができない領域が発生する．これらを解決するため，複数視点を用いたモーションキャプチャ・システムを提案する．このとき，各モーションキャプチャから得られた推定値を統合するために人の身体的特徴を活用する．

2. 関連研究

複数台のモーションキャプチャを連携させることで，モーションの推定精度を向上させるシステムは様々提案されている．吉本らは複数台のモーションキャプチャの内，人がモーションキャプチャに対して最も正面を向いているデバイスを選択することで，最も高精度に推定された値に切り替えるシステムを提案した [4]．宮武らは複数台のモーションキャプチャで得られた座標値を統合するシステムを提案した [5]．このとき，統合には各推定値におけるフレーム間の移動距離に応じて評価した信頼度を定義し，この値をもって統合している．

本研究では複数台のモーションキャプチャから得られた座標値に人の身体的特徴を基に定義した信頼度を用いて座標値を統合する手法を提案する．

3. 提案手法

3.1 方針

多くのモーションキャプチャは人の関節で構成した骨格の座標値を推定する．しかし推定した座標値は単一視点からのセンサ情報を基に，スケルトンモデルに当てはめているため，ロバスト性に欠ける場合がある．そのため推定した座標値はロバスト性を評価した数値（信頼度）が付加されていることが多いが，この数値は一つの指標でしかなく正しいとは限らない．こうした背景からロバスト性を向上させるため，視点を増やすこととした．

視点を増やすために，複数台のモーションキャプチャを用いるが，推定した関節の座標値は各モーションキャプチャ座標系の数値であるため，任意の座標系（世界座標系）を定義し，この世界座標系に各座標値を変換する必要がある．

変換した各座標値を統合するためには，文献 [5] と同様に，各座標値に対応した信頼度を用いる．本研究ではこの信頼度を人の身体的特徴である次の前提を基に定義する．

【関節間の長さ】
計測時間における人の体格変化は無視できる範囲であると仮定した場合，推定された関節間の長さは不変と見なすことができる．

【各関節の可動域】
人の関節の可動域は骨の構造により限定されている．例えば図 2 に示すとおり，膝関節は正面に向かって鋭角になることはない．

3.2 キャリブレーション

各モーションキャプチャで推定した座標値を世界座標系に変換するためには，各モーションキャプチャ座標系と世界座標系との回転・並進成分が既知である必要がある．本研究で使用するモーションキャプチャは固定で設置したと仮定した場合，事前にキャリブレーションすることで回転・並進成分を推定する．

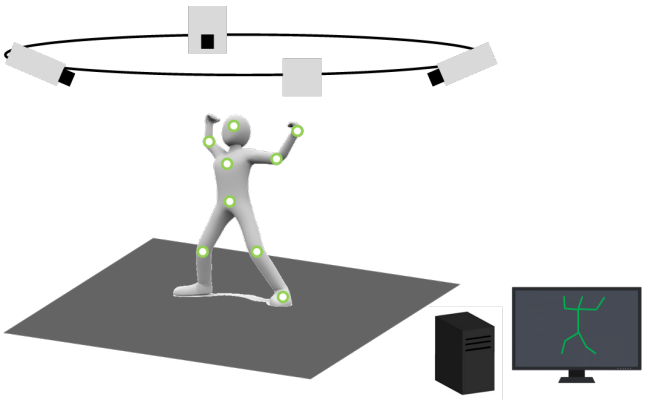


図 1: 光学式モーションキャプチャのイメージ
Fig. 1 Image of optical motion capture

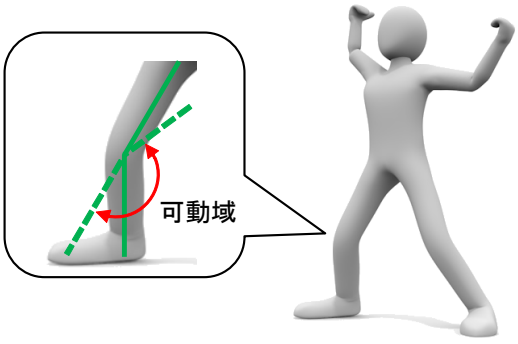


図 2: 各関節の可動域
Fig. 2 Range of joint motion

Vol.2019-GN-106 No.52
Vol.2019-CDS-24 No.52
Vol.2019-DCC-21 No.52
2019/1/25

本稿における回転・並進成分の推定手法はモーションキャプチャがデプスセンサを搭載し，点群を取得できることを前提として述べる．

回転成分を R, 並進成分 t とおいたとき, 各モーションキャプチャ座標系 (x_m, y_m, z_m)，と世界座標系 (X_w, Y_w, Z_w) には式 1 が成り立つ（図 3 参照）．したがって各モーションキャプチャ座標系の 3 点と，それに対応する世界座標系の 3 点から R と t を算出することは可能である．

$$\begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \begin{pmatrix} R_{3\times3} & | & t_{3\times1} \end{pmatrix} \begin{pmatrix} x_m \\ y_m \\ z_m \\ 1 \end{pmatrix} \quad (1)$$

3.3 信頼度の定義

モーションキャプチャが推定した関節の座標値に信頼度を定義し，この数値を基に座標値を統合する．モーションキャプチャ k（k=1〜K）は関節 n（n=1〜N）の推定座標値 $J_{kn}(x_{kn}, y_{kn}, z_{kn})$，信頼度 C_{kn} としたとき，統合された座標値 $J_{n\_int}(x_{n\_int}, y_{n\_int}, z_{n\_int})$ は式 2 で表すことができる．

$$J_{n\_int}(x, y, z) = \frac{\sum_{k=1}^{K}(J_{kn} \times C_{kn})}{\sum_{k=1}^{K} C_{kn}} \quad (2)$$

多くのモーションキャプチャは推定のロバスト性の指標として信頼度を付加しているものが多い．本提案では，この信頼度に乗ずることも可能である．

【関節間の長さ】

モーションキャプチャ k で推定できる関節 n と隣接する関節 $n-1$ との間の長さ l_n としたとき，事前に計測した関節間の長さを l_{n_pre}，モーションキャプチャ k で推定した任意のフレームの関節間の長さを $l_{n\_f\_k}$ とおくことができる．図 4 に示すとおり 2 つの長さを比較して信頼度を決定する（式 3 参照）．

$$C_{kn} = \mathrm{a}|l_{n\_pre} - l_{n\_f\_k}| \text{（a：正の定数）} \quad (3)$$

また人は複数の関節から成り立っているため，中心から離れれば離れるほど，中心に近い関節の信頼度の値に影響されるため，信頼度 C_{kn} は式 4 で表すことができる．

$$C_{k(n+1)} = C_{kn}\mathrm{a}|l_{n\_pre} - l_{n\_f\_i}| \text{（a：正の定数）} \quad (4)$$

【各関節の可動域】

各関節の可動域も同様に事前に計測した任意の関節の可動域の最大値を $R_{MAX\_pre\_n}$，最小値を $R_{min\_pre\_n}$，モーションキャプチャで計測した対応する関節間の可動域を $R_{f\_n\_i}$ としたとき，式 5 が成り立つ．

$$C_{kn} = \begin{cases} \mathrm{a}_{MAX}|R_{MAX\_pre\_n} - R_{f\_n\_k}| \text{（a：正の定数）} \\ \mathrm{a}_{min}|R_{min\_pre\_n} - R_{f\_n\_k}| \text{（a：正の定数）} \end{cases} \quad (5)$$

4. 評価と考察

4.1 評価機器

本評価では，Kinect v2 をモーションキャプチャとして 2 台（Kinect1 と Kinect2）使用した．Kinect v2 は図 5 に示すとおり 25 の関節の三次元座標値を取得可能である．また Kinect v2 で取得した座標値は次の 3 つの状態情報が付加されている．

(1) NoTracked
(2) Inferred
(3) Tracked

本評価では NoTracked の信頼度 0，Inferred を 0.3，Tracked を 0.7 と定義した．次に Kinect v2 は PC1 台につき 1 台のみ接続可能でため，図 6 に示すとおり，クライアント・サーバモデルを用いたローカルネットワークを構

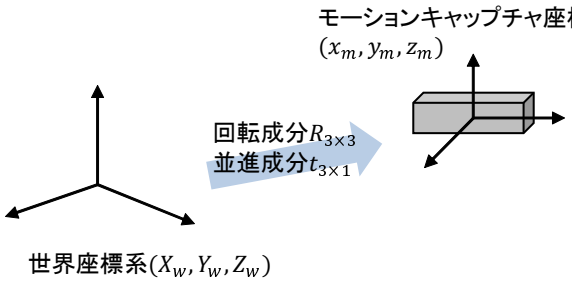


図 3: 座標系
Fig. 3 Coordinates

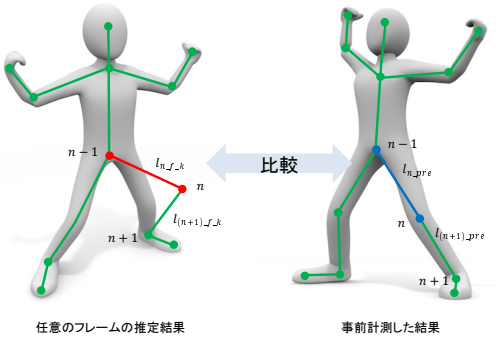


図 4: 関節間の長さ
Fig. 4 Length between joins

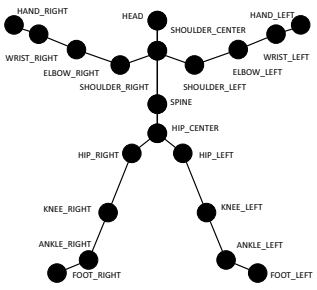


図 5: 認識可能な関節
Fig. 5 Recognition of the human skeleton

築した．

図 7 に示すのは各 Kinect のキャリブレーション時の光景である．本評価ではモーションキャプチャに Kinect v2 を使用することから，点群が取得可能である．Kinect1 の座標系を世界座標系と仮定し，取得した各 Kinect の点群から手動で対応する 3 点を選び，回転成分と並進成分を推定する．しかし，手動で対応点を選んだことから微妙な誤差が生じる．そこで Kinect2 の点群を世界座標系に変換後，各 Kinect で取得した点群に ICP (Iterative Closest Point) アルゴリズムを用いて回転成分と並進成分を高精度に求めた [6]．

次に各 Kinect で推定した関節の座標値はクライアント PC から統合サーバへ UDP によるストリーミングで行う．送信内容がデバイス ID と 25 点の座標値であるため，通信遅延は無視できるものとした．

図 8(a) に示すのは統合サーバにおいて，Kinect1 と Kinect2 の計測結果を表示したものである．

4.2 評価方法

モーションキャプチャ 2 台から取得した関節の座標値を 2 種類の手法で統合し，これらの値を用いて事前に計測した長さと比較する．事前計測の方法は，任意の Kinect の前に立ち，関節間の長さを推定した図 9 参照)．

評価に用いる手法は 2 種類である．この 2 種類のうち，前者は付加された状態情報のみを用いて座標値を統合する手法である（既存手法）．後者は前者の手法に 3 章で述べた「関節間の長さ」の提案手法を追加した（提案手法，図 8(b) 参照)．

Kinect v2 で推定した 25 の関節のうち，次の 2 点の関節間の長さを求め，事前計測結果と比較した．比較に使用したフレーム数は 50 である．

(1) HIP_RIGHT と KNEE_RIGHT との距離
(2) KNEE_RIGHT と ANKLE_RIGHT との距離

4.3 評価結果

評価結果は図 10 に示すとおり，本研究で提案した手法のほうが事前計測した結果に近い距離を示している．したがって各関節の座標値を統合するときに，事前計測結果に近いものに重きをおき統合していることがわかる．

したがって複数視点から推定された座標値は提案手法を用いることで向上したといえる．

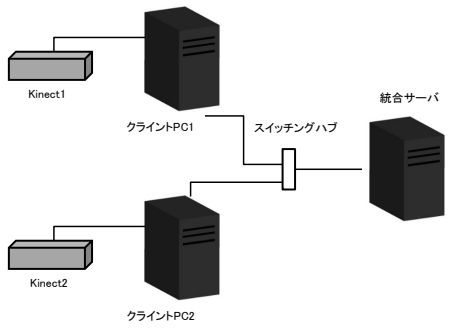


図 6: 機器とネットワーク構成
Fig. 6 Equipment and network configuration

4.4 考察

複数視点からモーションをキャプチャすることによって座標値の推定精度が向上することがわかった．本実験では事前計測に Kinect を用いて簡易的にのみ行った．案手法は，この事前計測結果に依存するため，この事前計測で誤差が生じた場合，誤った推定結果となる．したがって，より高精度に事前計測する必要があることがわかった．

5. おわりに

本稿では複数視点によるモーションキャプチャ・システムについて述べた．複数台のモーションキャプチャを使用ため，推定した座標値を統合する必要がある．統合するにあたって人の身体的特徴を活かした信頼度を定義することによって高精度に統合する手法について検討した．

本稿においては「関節間の長さ」のみを実装・評価した．今後は「各関節の可動域」も実装・評価し，他の手法との比較評価を行う予定である．また光学式モーションキャプチャなどとの比較評価も行う．

参考文献

[1] 中澤：モーションキャプチャ，映像情報メディア学会誌, Vol. 63, No. 9, pp. 1224 - 1227, 2009.
[2] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, R. Moore: Real-time human pose recognition in parts from single depth images, Comm. of the ACM, Vol. 56, No. 4, pp. 116 - 124, 2013.
[3] Z. Cao, T. Simon, S. Wei, and Y. Sheikh: Realtime multi-person 2d pose estimation using part affinity, Proc. IEEE Conf. on Computer Vison and Pattern Recognition, pp. 1302 – 1310, 2017.
[4] 吉本，高野，岡田：複数 Kinect ボーン情報を組み合わせた高精度モーションキャプチャシステムの構築，情報処理学会第 75 回全国大会講演論文集，6ZB-2, pp. 219 - 220, 2013.
[5] 宮武，大坪，吉田：複数台の Kinect を用いたモーションキャプチャの実現，HAI シンポジウム，G-22, 2016.
[6] P. J. Besl and N. D. Mckya: Amethod for registration of 3-D shapes, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 239 – 256, 1992.

Vol.2019-GN-106 No.52
Vol.2019-CDS-24 No.52
Vol.2019-DCC-21 No.52
2019/1/25

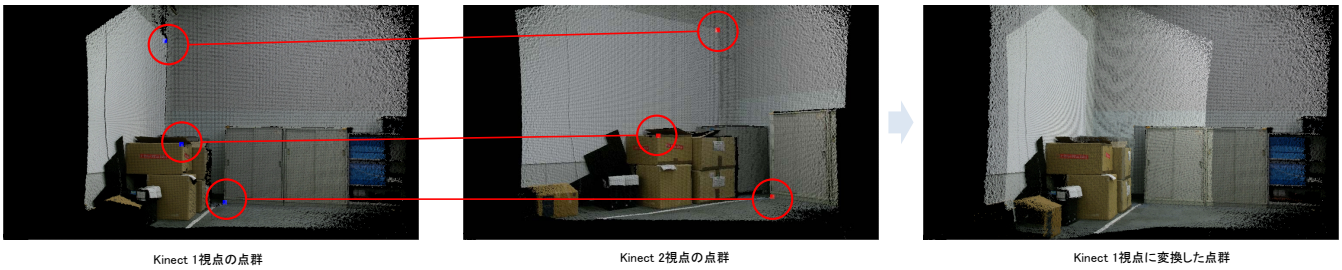
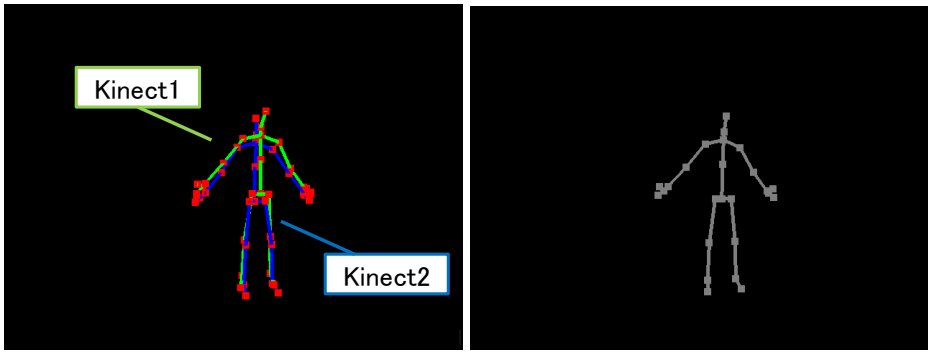


図 7: キャリブレーション風景
Fig. 7 Calibration



(a) 統合前　(b) 統合後

図 8: モーションキャプチャ・システム
Fig. 8 Motion capture system

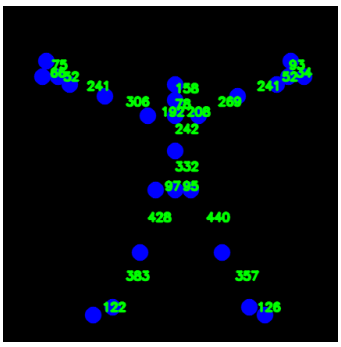
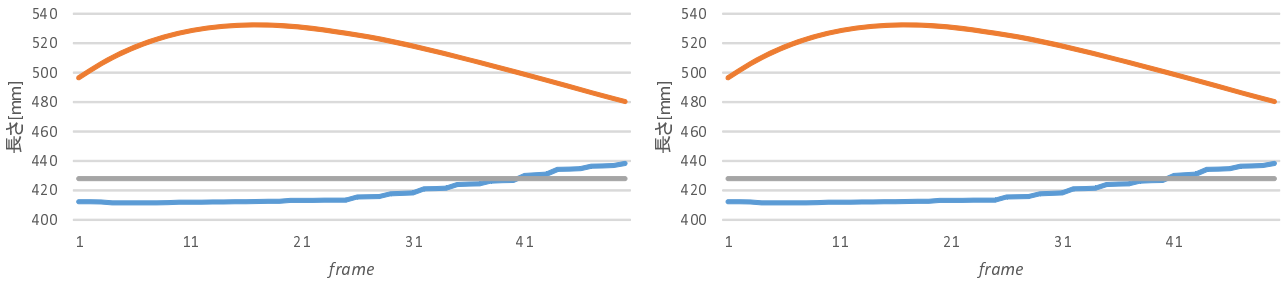


図 9: 事前計測結果
Fig. 9 Pre-measurement



(a) HIP_RIGHT と KNEE_RIGHT との距離　(b) KNEE_RIGHT と ANKLE_RIGHT との距離

図 10: 評価結果
Fig. 10 Evaluation result