4R-03

Viterbiアルゴリズムを用いた3次元骨格位置推定の高精度化

齋藤一誠† 中村 友昭† 八田 俊之†† 藤田 渉†† 渡邉 信太郎†† 三輪 祥太郎††
†電気通信大学 ††三菱電機株式会社 先端技術総合研究所

1 はじめに

生産現場における作業を効率化するためには，人の動作の計測・解析が必要となる [1]．人の動作の計測には光学式モーションキャプチャ等を使用することで高精度に計測することができるが，マーカーを身体に貼り付ける必要があり，計測準備が煩雑になるといった欠点がある．一方，近年では，安価なRGB-Dセンサと OpenPose[2] や Mediapipe[3] といった二次元の骨格検出技術によって，容易に人の動作を計測することが可能となった [4]．しかし，1台の RGB-D センサだけでは一方向の映像しか撮影できずオクルージョンが発生するといった問題や，二次元の骨格検出による誤検出により計測精度が低下するといった問題がある．そこで，本稿では3台の RGB-D センサと OpenPose を使用し，複数の骨格候補点を計算し，前後のフレームの整合性から誤差が最小となる候補を選択する手法を提案する．実験では，セル生産作業を3台の Intel RealSense Depth Camera D455 で撮影した RGB-D データを用いて，より高精度に骨格の計測が可能であることを示す．

2 提案手法

図1が提案手法の概要である．

2.1 OpenPose による骨格候補点の検出

3台の各カメラの RGB 画像から OpenPose を用いて各骨格を検出する．次に，各骨格の二次元座標を Depth 画像を用いて3次元座標へと変換する．この際，骨格の二次元座標に対応する Depth が得られていない場合や，骨格検出の誤差の影響で正確な位置が検出できない場合がある．そこで図 2(a) の黄色の点のように各骨格位置（図中の赤点）の近傍の点 $q_{t,n}^{(c)}$ を骨格位置の候補とする．ただし，$c = 1, 2, 3$ はカメラのインデックス，$t = 0, 1, \cdots T$ はフレームのインデックス，$n = 1, 2, \cdots, N$ は候補のインデックスある．また，二次元座標に対応する Depth が存在せず，候補点全ての位置 $q_{t,1}^{(c)}, \cdots, q_{t,N}^{(c)}$ が計算出来ない場合は，$t-1$ 番目の候補により補完し $q_{t,n}^{(c)} = q_{t-1,n}^{(c)}$ とする．

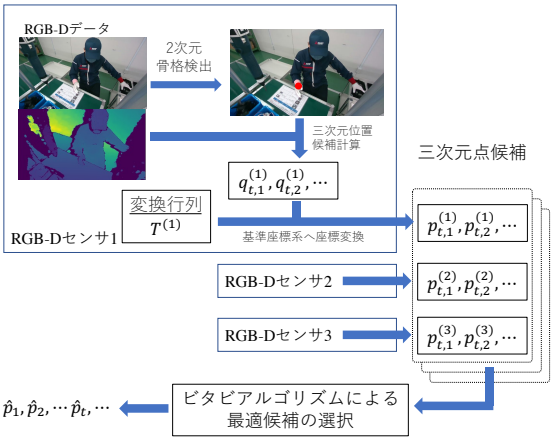
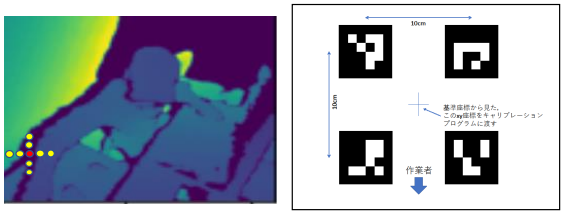


図 1: 提案手法の概要



(a) (b)

図 2: (a) 骨格近傍の候補点の計算例，(b) キャリブレーションシート

2.2 基準座標系への変換

3台のカメラは異なる方向から撮影しているため，1つの基準座標系への変換が必要となる．そこで，事前に以下の手順で各カメラの三次元座標を基準座標へ変換するための変換行列 $T^{(c)}$ を計算する．

1. 図 2(b) の AR マーカーのカメラ座標系での位置 $q_{AR,i}^{(c)}$ を計測する．
2. シート中央を原点とした座標系を基準座標系とする．キャリブレーションシート上の AR マーカーの位置は既知であるため，基準座標系での AR マーカー位置 $p_{AR,i}^{(c)}$ は以下となる．

$$p_{AR,1}^{c} = (0.05, 0.05, 0) \quad (1)$$
$$p_{AR,2}^{c} = (0.05, -0.05, 0) \quad (2)$$
$$p_{AR,3}^{c} = (-0.05, 0.05, 0) \quad (3)$$
$$p_{AR,4}^{c} = (-0.05, -0.05, 0) \quad (4)$$

Improving the accuracy of 3D skeletal position estimation using Viterbi algorithm
†,Issei SAITO†,Tomoaki NAKAMURA†,Toshiyuki HATTA††,Shintaro WATANABE††and Shotaro MIWA††
†The University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan
†Advanced Technology R&D Center, Mitsubishi Electric Corporation
{, nakamura}@radish.ee.uec.ac.jp

3. カメラ座標系 $q_{AR,i}^{(c)}$ での AR マーカー位置を，基準座標系での位置 $p_{AR,i}^{(c)}$ へ変換する同時変換行列 T を推定する．

$$T^{(c)} = \arg\min_{T} \sum_{i}^{4} (T q_{AR,i}^{(c)} - p_{AR,i}^{(c)})^2 \quad (5)$$

このようにして求めた同時変換行列を用いて，各骨格位置を基準座標系へと変換する．

$$p_{t,n}^{(c)} = T^{(c)} q_{t,n}^{(c)} \quad (6)$$

これにより，各骨格の基準座標系での複数の候補点 $p_{t,n}^{(c)}$ を得ることができる．

2.3 Viterbi アルゴリズムを用いた候補の選択

次に，計算された候補の中から最適な 1 つを選択する．時系列の作業データはフレームレートが十分高く，フレーム間で各骨格位置の移動量が小さいと考えられる．そこで，骨格の移動量と位置の整合性（右手首と右肘または左手首と左肘の距離）を基準に候補を選択する．まず移動量を基準として，距離の総和が最小となる候補をビタビアルゴリズムにより選択する．

$$\hat{p_1}, \hat{p_2}, \cdots, \hat{p_t}, \cdots = \arg\min_{p1,p2,\cdots} \sum_{t=1}^{T-1} (p_t - p_{t+1})^2 \quad (7)$$

$$p_t \in \{p_{t,n}^{(c)} | 1 \leq c \leq 3, 1 \leq n \leq N\} \quad (8)$$

適切な候補が含まれていれば，距離の総和の最小化によって適切な候補を選択することができる．しかし，適切な骨格座標の候補が存在しないフレームでは，適切でない候補の中から 1 点が選択されることになる．距離の総和を最小化しているため，その前後のフレームでも不適切な候補が連続的に選ばれてしまう場合がある．そのため骨格の整合性（右手首と右肘または左手首と左肘の距離）を基準に，適切な候補が含まれない時刻を検出前後のフレームからその時刻の候補を補完し，再度ビタビアルゴリズムを実行し，適切な候補の選択を行う．

3 実験

セル生産作業を 3 台の Intel RealSense Depth Camera D455 で撮影した 5000 フレームの RGB-D データを用い，提案手法の有効性を検証した．骨格検出には OpenPose を用いて，各カメラから得られる 3 次元座標の候補数を $N = 25$ とした．比較手法として，3 台のカメラから得られる骨格位置の平均を最終的な骨格位置とする手法を用いた．

各手法において 100 枚ごとに可視化し，右手首の位置の検出精度を求めた．表 1 がその結果である．提案手法により骨格推定の精度が向上したことがわかる．図 3 が，比較手法では正しく検出することができなかったフレームにおける提案手法の検出結果の一例である．赤色の球，緑色の球がそれぞれ右肘，右手首に対応しており，それぞれ正しく検出できていることがわかる．

表 1: 右手首の推定精度

	比較手法	提案手法
精度（%）	62	98

図 3: 可視化例

4 おわりに

本稿では骨格の三次元位置推定の高精度化を目的とし，複数のカメラから得られる RGB-D データから複数の骨格候補点を計算し，前後のフレームの整合性から誤差が最小となる候補を選択する手法を提案した．実験では，3 台の RGB-D カメラから撮影したデータに本手法を適用し，単純な方法では適切な位置推定が難しいフレームであっても，提案手法によって正しい位置を推定できることを確認した．また，本手法によって推定された骨格位置を，階層的な時系列構造の解析が可能な GP-HSMM-DAA[5] により解析することで，作業に含まれる階層的な構造を教師なしで推定できるか検証する予定である．

参考文献

[1] 八田俊之，友田翼．玉置哲也，三輪祥太郎: GP-HSMM による効率的な作業行動の分節化，信学技報, Vol. 122, No. 260，pp. 79-84, 2022

[2] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei and Yaser Sheikh: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019

[3] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg and Matthias Grundmann: MediaPipe: A Framework for Building Perception Pipelines, arXiv:1906.08172 , 2019

[4] 春名 弘一, 昆 恵介, 稲垣 潤, 佐藤 洋一郎: マーカーレスモーションキャプチャによる三次元動作解析の応用例, 日本義肢装具学会誌, Vol. 35, No. 1, pp. 17–23, 2019

[5] 長野匡隼，中村友昭: GP-HSMM に基づく二重分節化モデルによる連 続音声の教師なし構造学習，日本ロボット学会学術講演会，4F3-08，Sep. 2022