Vol.2019-CG-176 No.2
Vol.2019-DCC-23 No.2
Vol.2019-CVIM-219 No.2
2019/11/7

Random Tree Walkを用いた高速・高精度な3次元人物姿勢トラッキング

楊 森[1,a)]　渡辺 義浩[1]

概要：人物の高速な姿勢推定は，様々なインタラクティブな応用で遅延を抑えることができるため，重要な役割を果たすと考えられる．特に既存手法の中で，Random Tree Walk (RTW) を用いたものは1000fpsを達成できることが示されている．一方，同手法は骨格構造に基づいた逐次推定に基づいている．このため，もし同構造的に初期の関節位置の推定を失敗すると，後続する関節位置に誤差が蓄積する問題があった．また，このような逐次推定は，関節毎に処理を並列化することが困難であった．同問題を解決するために，本稿では時系列情報を利用する．提案するトラッキング手法は，1時刻前に推定された各関節位置から始めても現在位置の推定が収束するようにRTWを学習するものである．これによって，関節毎の推定処理の並列化や関節間の蓄積誤差の回避が可能である．

1. はじめに

人物のマーカレス姿勢推定は，マンマシンインタフェース，動作解析，監視，アニメーション，バーチャルリアリティ・拡張現実型のゲームなど，様々な応用で広く利用されている．このような応用分野では，低遅延でのインタラクションが重要視されている．同背景の下，より高速な3次元での姿勢推定のニーズが高まっている．

近年の人物のマーカレス3次元姿勢推定の多くは，RGB画像やデプス画像を入力としている．このうち，RGB画像を用いたアプローチは，3次元の情報を直接的に取得することが難しいとともに，自己遮蔽時における曖昧性を解決することが難しい．2次元での推定結果から3次元の姿勢を復元する手法もあるが，処理時間が増大する問題がある [1], [2]．そこで，本稿ではデプス画像を用いたアプローチに着目する．

一般に，デプス画像を用いたアプローチには，トラッキングと検出の2種類がある．トラッキング型の手法は時系列情報を利用するものである．同手法は，事前に定義された身体モデルを観測データに一致させる最小化問題を解くことによって姿勢を推定する．特に，Iterative Closest Point (ICP) 法 [3], [4] や Gaussian mixture model (GMM) [5] の2つに基づくものが主流である．しかし，従来のトラッキング型の手法は，身体モデルや運動制約に強く依存しており，柔軟な推定が難しかった．これに対して，本稿もトラッキング型のアプローチを採用しているが，各関節を独立に推定するものとなっている．

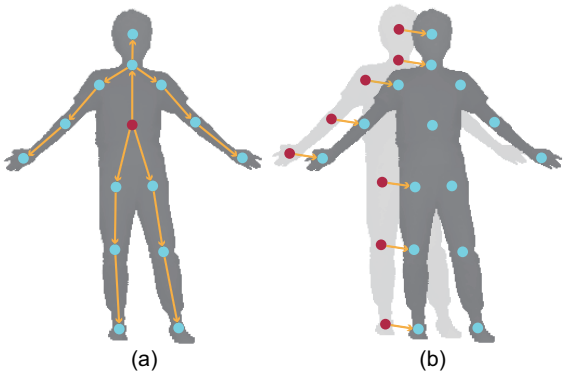

図 1　RTW を用いた姿勢推定の概念図．(a) 従来手法のアプローチ．(b) 本稿で提案するアプローチ．

次に，検出型のアプローチについて述べる．同手法では，1枚のデプス画像から関節位置が推定される．Shottonらは，各ピクセルを身体パーツに分類する識別器をランダムフォレストによって学習し，ミーンシフトを用いて関節位置を推定する手法を提案した [6]．さらに，近年ではConvolutional Neural Network (CNN) を用いた手法も提案されている [7], [8], [9]．一方，検出型のアプローチの中でも特に計算効率が高いものとして，Random Tree Walk (RTW) に基づく手法が提案されている [10]．同手法は注目する関節位置へ向かう方向を回帰木によって推定するものであり，大幅な高速化を達成している．しかし，同手法

[1] 東京工業大学
Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama, Kanagawa 226–8503, Japan
[a)] yang.s.af@m.titech.ac.jp

Vol.2019-CG-176 No.2
Vol.2019-DCC-23 No.2
Vol.2019-CVIM-219 No.2
2019/11/7

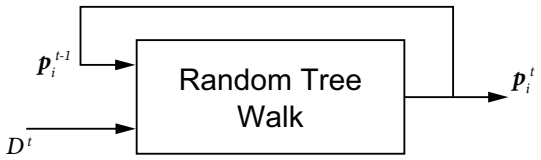

図 2 提案するトラッキングアプローチの概念図 (例として i 番目の関節についてのみ示す)

は，**図 1**(a) に示すように，身体の中心から順に関節位置を推定するものとなっている．例えば，推定された肘の位置をもとに，手の位置が推定される．このように骨格構造に強く依存しているため，関節位置の推定に一度失敗すると，それに続く関節位置には累積誤差が発生する問題を抱えている．

本稿では RTW を用いたトラッキング型の人物姿勢推定手法を提案する．具体的には，図 1(b) に示すように，従来の手法における推定開始点を前フレームの各関節位置に変更するものである．特に，このようなアプローチを支える技術として，近年では 400 ∼ 1,000 fps レベルの高速なデプスセンシングが新たに実現されている [11], [12]．このような技術の導入を想定して，連続する 2 フレーム間の姿勢の動きが小さいという前提の下で手法を設計する．評価実験で従来手法と比較した結果，提案手法はより高速に，より高精度の姿勢推定を達成できることを確認した．

2. RTW を用いた人物姿勢トラッキング

本稿は，検出型のアプローチで用いられていた RTW に基づく手法を，トラッキング型のアプローチへ拡張するものである．本稿では，骨格は頭，首，脊椎，両肩，両肘，両手，両臀部，両膝，両足首の計 15 個の関節で構成されるとする．

2.1 Random Tree Walk に基づく従来手法

RTW は反復型の推定手法である [10]．同手法は，各関節に対して，現在の推定位置からより正しい推定位置へ向かう単位ベクトル $\hat{\boldsymbol{u}}$ を回帰木によって推定する．同ベクトル $\hat{\boldsymbol{u}}$ は，葉ノードに登録された候補のクラスタからランダムに選択される．初期位置 $\boldsymbol{q}_i^0$ から開始して，このランダムウォークのプロセスを繰り返すことで推定を達成する．なお，$\boldsymbol{q}$ は同プロセスで得られる位置を表す．m 回目の反復時には下式に基づいて，得られた単位ベクトル $\hat{\boldsymbol{u}}$ と移動量 $dist_s$ によって $\boldsymbol{q}_i^m$ を $\boldsymbol{q}_i^{m+1}$ へと移動させる．

$$\boldsymbol{q}_i^{m+1} = \boldsymbol{q}_i^m + \hat{\boldsymbol{u}} \cdot dist_s \tag{1}$$

この反復は各フレームで N 回実施される．最終的な推定位置 $\boldsymbol{p}_i$ は，同反復における位置 $\boldsymbol{q}_i^1$,...,$\boldsymbol{q}_i^N$ を平均化することで得られる．

従来の RTW では，i 番目の関節の推定のための初期位置 $\boldsymbol{q}_i^0$ は，骨格構造的に隣接する関節の推定結果が用いられていた．例えば手の場合，肘の推定位置が初期位置として用いられる．このアプローチは，手や足首のように，より先端でより速く動く関節に対して誤差を発生させる要因となると考えられる．

2.2 RTW を用いたトラッキング

提案手法では，RTW を関節位置のトラッキングに利用する．本手法は身体モデルや運動制約などを必要としないため，各関節を独立に推定できるものである．**図 2** に本手法の概念図を示す．式で表すと下記となる．

$$\boldsymbol{p}_i^t = RTW(\boldsymbol{p}_i^{t-1}, D^t) \tag{2}$$

本手法では，i 番目の関節位置を求めるために，RTW は時刻 $t-1$ に推定された同関節位置によって初期化される．このように，時刻 t のデプス画像 D^t と時刻 $t-1$ で推定された i 番目の関節の位置 $\boldsymbol{p}_i^{t-1}$ から，時刻 t の同関節の位置 $\boldsymbol{p}_i^t$ を取得する．これによって，前節で述べた骨格構造順で推定することによる累積誤差の問題を回避することができる．

また，高速なデプスセンシングの導入を想定した，2 フレームの関節位置 $\boldsymbol{p}_i^{t-1}$ と $\boldsymbol{p}_i^t$ が近い前提を用いることで，前節の手法に比べて反復回数 N を大幅に減らすことができる．結果として，処理速度の向上が期待できる．また，同様の前提によって学習時においても効果が望める．従来手法では学習時の教師データが隣接する関節位置付近までを覆うように分布させる必要があった．これに対して本手法では 1 時刻前の位置までを覆えばよい．この点によって，学習対象を単純し，冗長性を軽減することができると考えられる．

3. 評価実験

3.1 実験条件

本実験では，予備実験として高速デプスセンサの代わりに，Kinect v2 を用いた．センサのフレームレートは 30fps である．前節で述べたフレーム間の姿勢変化が小さい前提を満たすために，実験データを収集する際はゆっくり動いた．データセットは，解像度 512 × 424 の 1800 枚のデプス画像と各画像における正しい関節位置から成る．同データセットには 6 種類の異なる動作シーケンスが含まれる．各シーケンスは 300 枚のデプス画像から成る．本実験では，5 つのシーケンスを学習に用い，残る 1 つのシーケンスを評価に用いた．学習時には leave-one-out 交差検証を用いた．

設定したパラメータについて述べる．教師データの分布範囲は 15pixels とした．葉ノードのクラスタ数は 20 とした．また葉ノードのサンプル数は 400 とした．さらに反復

Vol.2019-CG-176 No.2
Vol.2019-DCC-23 No.2
Vol.2019-CVIM-219 No.2
2019/11/7

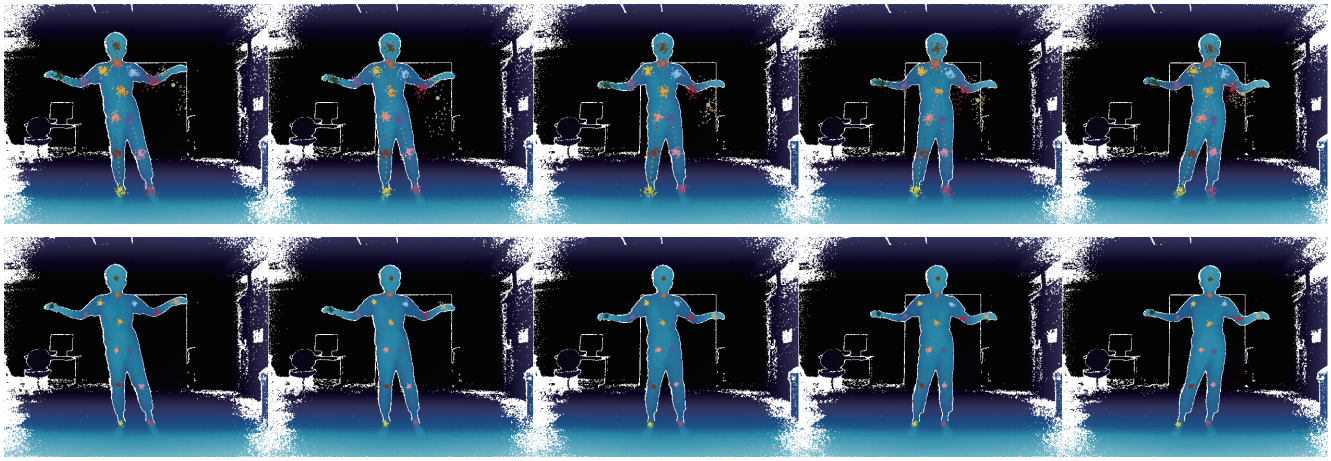

図 3 上段：従来手法による推定結果．下段：提案手法による推定結果．各ドットはランダムウォークの軌跡を表している．ここでは 10 フレーム毎の時系列画像を示す．

回数 N を 16，移動量 $dist_s$ を 5cm とした．一方，比較に用いた従来の RTW 手法に対しては，教師データの分布範囲を 65pixels，反復回数 N を 64 回，移動量 $dist_s$ を 10cm とした．これらのパラメータは，精度と計算効率の間のトレードオフを考慮して，適切に設定したものである．すべての実験は Xeon E5-1650v4 (6 cores, 12threads, 3.6GHz) を搭載したコンピュータ上で行った．

3.2 結果

RTW を用いた従来手法 [10] と本手法を比較した結果について述べる．**図 3** に定性的な比較結果を示す．同図では，反復時の推定位置 $\boldsymbol{q}_i^1,...,\boldsymbol{q}_i^N$ を，関節毎に異なる色のドットで示している．水平方向に配置された画像は，10 フレーム毎の時系列画像である．同結果より，左の手や肘で従来手法は失敗しているが，本手法は正確に推定されていることが分かる．

図 4 に定量評価結果を示す．ここでは，人物の姿勢推定評価で良く用いられる，10cm 平均精度 (mean average precision: mAP) を採用した．この指標では，もし推定された関節位置と正しい位置の距離が 10cm 以下であれば，正しく推定されたとみなす．mAP は正しく推定された関節の割合を示すものである．同結果より，特に手や肘において本手法はより高い精度を達成していることが分かる．前節で述べたように，この向上の主な理由は RTW の推定開始点を各関節独立に，前フレームの推定位置としたためであると考えられる．

図 5 に従来の RTW に基づく手法に対する相対的な速度性能を示す．本手法は 3.5 倍の高速化を達成することができる．さらに，各関節を並列に推定する処理を適用した場合，速度向上は 26 倍にまで達する．

4. まとめ

本稿では，RTW を用いた人物の姿勢トラッキングを提案した．本手法は，速度を向上するとともに，従来手法で問題となっていた関節間の推定における累積誤差の問題を回避することができる．

一方，本手法は時系列情報に基づいている．このため，一度推定に失敗すると，続くフレームでの推定に誤差が生じる恐れがある．そこで今後の課題として，検出型の手法と連携させることで，より安定的な動作を達成する予定である．

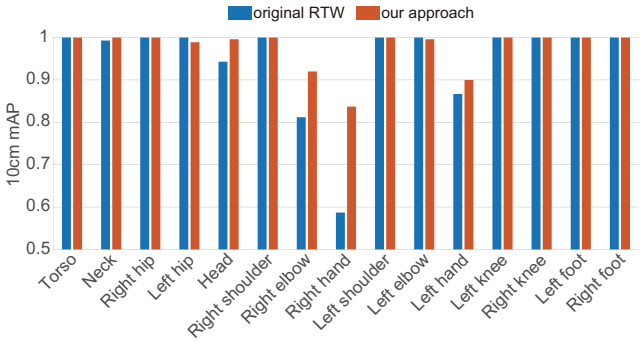

図 4 各関節の 10 cm 平均精度 (Mean Average Precision: mAP) の比較

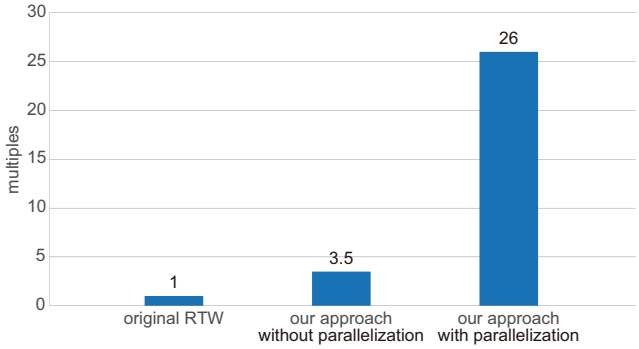

図 5 従来の RTW を用いた手法に対する速度性能 (従来手法の速度を 1 とした)

参考文献

[1] Chen C. H., Ramanan D.: *3d human pose estimation = 2d pose estimation + matching*, CVPR (2017)
[2] Martinez J., Hossain R., Romero J., Little J. J.: *A simple yet effective baseline for 3d human pose estimation*, ICCV (2017)
[3] Pellegrini S., Schindler K., Nardi D.: *A Generalisation of the ICP Algorithm for Articulated Bodies*, BMVC (2008)
[4] Ganapathi V., Plagemann C., Koller D., Thrun S.: *Real-Time Human Pose Tracking from Range Data*, ECCV (2012)
[5] Ye M., Yang R.: *Real-time simultaneous pose and shape estimation for articulated objects using a single depth camera*, CVPR (2014)
[6] Shotton J., Fitzgibbon A. W., Cook M., Sharp T., Finocchio M., Moore R., Kipman A., Blake A.: *Real-time human pose recognition in parts from single depth images*, CVPR (2011)
[7] Marín-Jiménez M. J., Romero-Ramirez F. J., Muñoz-Salinas R., Medina-Carnicer R.: *3D human pose estimation from depth maps using a deep combination of poses*, JVCIR (2018)
[8] Haque A., Peng B., Luo Z., Alahi A., Yeung S., Fei-Fei L.: *Towards viewpoint invariant 3d human pose estimation*, ECCV (2016)
[9] Moon G., Yong Chang J., Mu Lee K.: *V2v-posenet: Voxel-to-voxel prediction network for accurate 3d hand and human pose estimation from a single depth map*, CVPR (2018)
[10] Yub Jung H., Lee S., Seok Heo Y., Dong Yun I.: *Random tree walk toward instantaneous 3d human pose estimation*, CVPR (2015)
[11] Maruyama M., Tabata S., Watanabe Y., Ishikawa M.: *Multi-Pattern Embedded Phase Shifting using a High-speed Projector for Fast and Accurate Dynamic 3D Measurement*, WACV (2018)
[12] Tabata S., Maruyama M., Watanabe Y., Ishikawa M.: *Pixelwise Phase Unwrapping Based on Ordered Periods Phase Shift*, Sensors (2019)