Vol.2018-CVIM-210 No.45
2018/1/19

時系列情報を考慮した人体骨格追跡と評価

中島 雅貴[1,a)]　小篠 裕子[1,b)]　斎藤 英雄[1,c)]

概要：人物画像から検出した骨格情報を用いた研究は数多く提案されているが，人体に遮蔽がある環境下では，必ずしも完全な骨格情報が取得できるとは限らない．本研究では，人体に遮蔽がある人物動画像データセットを構築し，人体骨格検出において強力な手法のひとつである OpenPose を用いて，人体遮蔽が人体骨格検出に及ぼす影響について考察する．行動認識など，検出した人体骨格の時系列情報を用いる研究には，人体の関節追跡が必要となる．しかし，人体に遮蔽がある環境下では，常に全ての人体関節が検出可能であるとは限らないため，関節毎に関節を追跡する必要がある．そこで本稿では，構築したデータセットに対し，時系列情報を考慮して関節を追跡し，その結果を考察する．

キーワード：骨格検出，人体遮蔽，関節追跡

Human-Joint Tracking and Evaluation Considering Time Series Information

Masaki Nakashima[1,a)]　Yuko Ozasa[1,b)]　Hideo Saito[1,c)]

1. はじめに

人体骨格情報は，姿勢推定や行動認識，行動予測などの研究分野に応用され，高齢者や患者の見守りや，不審者の行動監視などのシステムとして，私たちの生活に適用されている．ここで，人体骨格情報とは，人体を構成する関節の位置を指す．また，本稿では便宜上，首や肩などの関節のみならず，目や鼻などの顔のパーツも合わせて，人体を構成する関節と表現する．

人体骨格を検出する強力な手法として，OpenPose[1] がある．多くの従来研究 [2], [3], [4] では，Microsoft 社 Kinect[5] などの 3 次元人体骨格の画像センサから取得される深度画像を用いて人体骨格を検出しているものがほとんどであったが，OpenPose は RGB 画像から人体骨格を精度よく検出することが知られている．OpenPose は，Convolutional Pose Machines (CPM) [6] を拡張した手法であり，RGB 画像中に写る複数人の骨格を同時に検出可能である．

OpenPose から得られた人体骨格情報を用いた研究は数多く提案されている [7], [8] が，全ての人体関節情報が必ずしも検出されるとは限らない．例えば，対象となる人体に遮蔽がある場合，関節情報が正しく推定されるとは限らない．人体骨格検出精度は，人体骨格を用いた全ての手法の精度に大きく影響することが想定されるが，人体に遮蔽がある環境下での人体骨格精度について十分な議論がなされているとは言い難い．

人体に遮蔽が起こる環境下で，OpenPose による人体骨格検出について検証するためには，人体に遮蔽が起こり，かつ，複数人が同時に写りこむ状況を撮影した RGB 画像データセットが必要である．しかし，従来の人物画像データセット [9], [10] には，このような撮影条件を十分に満たすものは存在しない．本稿では，複数人の人体に遮蔽が起こる状況を撮影した RGB 動画像データセットを構築し，構築したデータセットに対して OpenPose を用いて人体骨格を検出した結果を考察する．

行動認識など，検出した人体骨格の時系列情報を用いる研究には，人体の関節追跡が必要となる．しかし，人体に

[1] 慶應義塾大学 Keio University
Yokohama-shi, Kohoku-ku, Kanagawa 223–8522, Japan
[a)] nakashima@hvrl.ics.keio.ac.jp
[b)] yuko.ozasa@keio.jp
[c)] hs@keio.jp

図 1　データセット撮影シーン.

図 2　OpenPose による骨格検出結果例.

遮蔽がある環境下では，常に全ての人体関節が検出可能であるとは限らないため，関節毎に関節を追跡する必要がある．本稿では，構築したデータセットに対して，ある 1 つの関節を追跡した結果を紹介する．OpenPose は 1 枚の RGB 画像に写る複数人の骨格を検出するワンショットの手法であるため，動画像においても，フレーム毎に骨格を検出するのみであり，同一人物の骨格に同じ ID を割り当て，人物を追跡するわけではない．そこで，関節追跡の際は，対象となるフレームの前フレームの関節追跡結果を用いることで，時系列情報を考慮して関節を追跡する．

2. 人体遮蔽データセット

本稿では，人体の遮蔽が骨格検出及び関節追跡に及ぼす影響を検証するため，RGB 動画像データセットを構築する．複数人が写りこみ，かつ人体の遮蔽が起こるシーンを RGB カメラにより撮影し，データセットに用いる．具体的には，ロボットカフェにて複数人の来店客が注文カウンターにて飲み物の注文する様子を，ロボットカフェの部屋上部に設置した RGB カメラで動画撮影した．来店客は複数人から成るグループで来店し，注文中の来店客グループ以外にも，注文待機している来店客グループや，ロボットカフェの店員も写りこむ状況であり，常に複数人の人体に遮蔽が起こっていた．動画は 2 回に分けて撮影され，動画の解像度は $1,920px \times 1,080px$，フレームレートは $30fps$ であった．各動画をそれぞれデータセット A，データセット B とする．データセット A は合計 175,398 フレーム，データセット B は合計 188,476 フレームであった．実験の様子を撮影したフレームの一例として，データセット A の 65,000 フレーム目を図 1 に示す．注文カウンターは図 1 の右部にあり，この付近に人が密集する傾向があった．

3. 評価実験

本稿で構築した人体遮蔽データセットの評価として，OpenPose を用いて人体骨格を検出し，その結果を考察する．また，OpenPose で検出された 1 関節の時系列情報を用いて，その関節を追跡した結果を考察する．

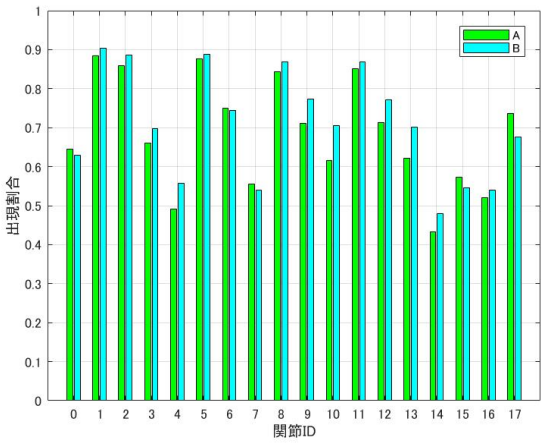

図 3　データセット A，B における各関節の出現割合.

3.1 データセット評価

データセット A と B に対して，OpenPose による骨格検出を行った結果，データセット A からは合計 805,552 個，データセット B からは合計 927,090 個の関節が検出された．骨格検出の際の学習データには，Microsoft COCO Dataset[11] を用いた．図 2 に動画 A の 65,000 フレーム目の骨格検出結果を示す．OpenPose で検出可能な関節数は 18 個であるが，各人の骨格検出結果を見ると，少ない人で 3 関節，多い人で 18 関節が検出されていることが分かる．

図 3 に，データセット A，B における各関節の出現割合を示す．黄緑色棒グラフはデータセット A，水色棒グラフはデータセット B の結果を示している．図の横軸は，関節 ID($=0, 1, \cdots, 17$) を示している．関節 ID は，ID= 0 から順番に，鼻，首，右肩，右肘，右手首，左肩，左肘，左手首，右腰，右膝，右足首，左腰，左膝，左足首，右目，左目，右耳，左耳を指している．図の縦軸は，各関節の出現割合を示している．出現割合は，(各関節が検出された人数)/(各データセットで検出された人数) により求めた．両データセットにおいて最も出現割合が高かった関節は首であった．

図 4，図 5 中の青色棒グラフは，データセット A，B の 1 人当りに検出された関節数の出現割合を示す．横軸は，検

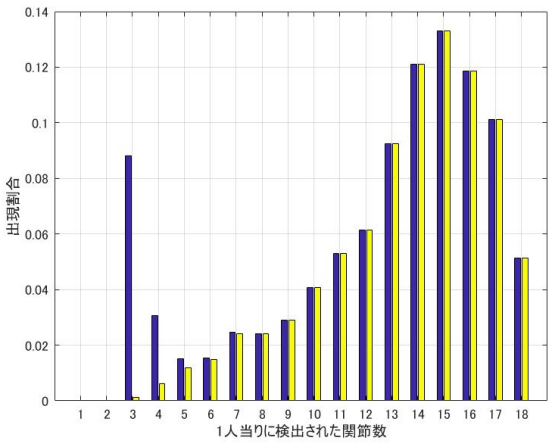

図 4 データセット A の 1 人当りの関節数の出現割合.

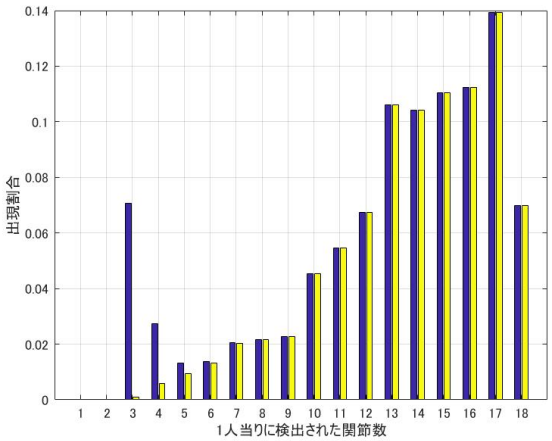

図 5 データセット B の 1 人当りの関節数の出現割合.

出された関節の個数であり，縦軸は，出現割合である．出現割合は，(x 個の関節のみが検出された人数)/(各データセットで検出された人数) により算出した．ここで x はグラフ横軸の値である．図中の黄色棒グラフは，青棒グラフのうち，首関節が検出された割合を示している．図 4，図 5 より，8 個以上の関節が検出された骨格には，必ず首関節が含まれることがわかる．

3.2 1 関節での骨格追跡

本実験では，構築したデータセットに対して，1 関節のみを用いて，その関節を追跡し，結果について考察する．3.1 節の実験結果より，データセット A，B において最も安定して検出された関節は首関節であった．本研究では関節追跡評価の第一歩として，首関節を追跡する．

図 6 に首関節追跡の概念図を示す．図中の $Neck^0_{t-1}$ は，$t-1$ フレーム目の人物 ID=0 番目の人物の首関節座標を示している．$t-1$ フレーム目の画像から OpenPose により検出された複数人の全首関節と，t フレーム目の画像から検出された全首関節間のユークリッド距離を算出し，$t-1$ フレーム目で検出された首関節のうち，最も距離の近い首関節を t フレーム目の首関節と同一人物の首関節であるとみなすことで，首関節を追跡した．

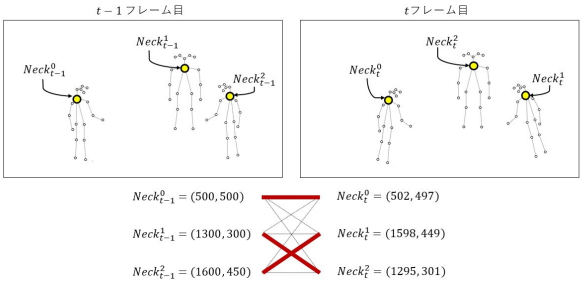

図 6 首関節追跡の概念図.

対象となる人とカメラの間に他の人が横切るなど，数フレームの間，首関節が取得できないケースもある．最後に取得された首関節の座標と新たに検出された首関節の座標の距離が閾値 ($= 20,000$) 以下，かつ両首関節取得時のフレーム数の差（インターバル）が閾値 ($= 90$ フレーム) 以下の時は同一人物とみなし，新たに検出された首関節を持つ骨格に同じ人物 ID を割り当てた．

データセット A において OpenPose で関節検出した結果を図 7 に示す．各フレーム画像は，元のデータ画像に対し，右詰め $1,080px \times 1,080px$ 部分の結果を図に用いた．図 7 には，それぞれ 11,700，11,800，11,900，12,000 フレーム目の画像における関節検出結果が示されている．この時の関節検出結果を用いて，首関節を追跡した．データセット A において首関節追跡をした結果を図 8 に示す．データセット A の 11,700 フレーム目から 12,000 フレーム目間の追跡結果を 100 フレーム毎で区切り，区切りが開始する 4 フレーム分の画像に首関節追跡結果を重畳したものである．図中の 1 画像には，100 フレーム毎に追跡された全首関節の軌跡が描かれている．11,700 フレーム目では茶色で示された骨格が，11,800 フレーム目では黄土色で示された別骨格に遮られ，茶色で示された骨格が検出されていない．しかし，11,900 フレーム目では再び同じ色の骨格，つまり同じ人物 ID が割り当てられ，追跡できていることが分かる．その他，青色や水色，黄色などで示された骨格も，同一人物を正しく追跡できていることが分かる．

4. まとめ

本研究では，人体に遮蔽がある人物動画像データセットを構築し，人体骨格検出手法である OpenPose を用いて，人体遮蔽が人体骨格検出に及ぼす影響について考察した．全ての関節のうち，首関節の検出される頻度が最も高いことが分かった．首関節を追跡した結果，関節追跡の実現可能性を確かめることができた．今後は，全関節追跡を検証し，人体遮蔽がある状況下においてロバストな関節追跡手法を提案する予定である．

図 7　データセット A での OpenPose による関節検出結果例.

図 8　データセット A での首関節追跡結果例.

謝辞　本研究は，科学技術振興機構 (JST) の戦略的創造研究推進事業 (CREST) の支援によって実現した.

参考文献

[1] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields, *arXiv preprint arXiv:1611.08050* (2016).
[2] Zhang, Z.: Microsoft kinect sensor and its effect, *IEEE multimedia*, Vol. 19, No. 2, pp. 4–10 (2012).
[3] Rafi, U., Gall, J. and Leibe, B.: A semantic occlusion model for human pose estimation from a single depth image, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 67–74 (2015).
[4] Obdržálek, Š., Kurillo, G., Ofli, F., Bajcsy, R., Seto, E., Jimison, H. and Pavel, M.: Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population, *Engineering in medicine and biology society, 2012 annual international conference of the IEEE*, IEEE, pp. 1188–1193 (2012).
[5] Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M. and Moore, R.: Real-time human pose recognition in parts from single depth images, *Communications of the ACM*, Vol. 56, No. 1, pp. 116–124 (2013).
[6] Wei, S.-E., Ramakrishna, V., Kanade, T. and Sheikh, Y.: Convolutional pose machines, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4724–4732 (2016).
[7] Papandreou, G., Zhu, T., Kanazawa, N., Toshev, A., Tompson, J., Bregler, C. and Murphy, K.: Towards Accurate Multi-person Pose Estimation in the Wild, *arXiv preprint arXiv:1701.01779* (2017).
[8] Gkioxari, G., Girshick, R., Dollár, P. and He, K.: Detecting and Recognizing Human-Object Interactions, *arXiv preprint arXiv:1704.07333* (2017).
[9] Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition, 2005. IEEE Computer Society Conference on*, Vol. 1, IEEE, pp. 886–893 (2005).
[10] Hwang, S., Park, J., Kim, N., Choi, Y. and So Kweon, I.: Multispectral Pedestrian Detection: Benchmark Dataset and Baseline, *The IEEE Conference on Computer Vision and Pattern Recognition* (2015).
[11] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L.: Microsoft coco: Common objects in context, *European conference on computer vision*, Springer, pp. 740–755 (2014).