

三次元点群データに基づく着座姿勢推定手法の提案

片山 ひかる¹ 水本 旭洋¹ Hamada Rizk¹ 山口 弘純¹

概要: デスクワークに従事する労働者が増加している現代において、工作中的に着座時間や姿勢を把握することは、健康維持のために重要である。しかし、既存の技術はカメラや椅子に取り付けるセンサに依存しており、プライバシーの侵害や、使用する環境に限られる等の問題がある。そこで本稿では、携帯可能な卓上設置型のセンサデバイスを用いて、カメラ画像や場所に依存せずに、着座姿勢を推定可能な手法を提案する。提案手法では、小型の LiDAR センサを用いて着座姿勢の 3 次元点群データを計測し、計測された 3 次元点群データから抽出された特徴量を入力とする深層学習モデルにより着座姿勢を推定する。卓上の 3 ヶ所にセンサデバイスを設置し収集した 3 次元点群データを用いて、着座推定モデルを構築した結果、F 値 0.87 以上の精度で 9 種類の異なる着座姿勢を推定できることを確認した。

1. はじめに

パーソナルコンピュータの普及により、デスクワーク中心に作業を行う労働者(デスクワーカー)が爆発的に増加した。近年では、天板の高さを変更可能で立ちながら作業が行える昇降机(スタンディングデスク)と呼ばれるデスクも販売されているが、日本国内においては普及途上であり、多くのオフィスでは従来通り着座による作業が行われている。しかしながら、長時間にわたる着座は、血行不良や代謝の低下により、高血圧や高脂血症、心臓病などを引き起こす可能性が高まることが分かっている [1,2]。特に、日本人は、世界的にみて着座時間が長い [3]、長時間の着座による健康リスクは大きな問題である。血行不良や代謝の低下は、姿勢の良し悪しに関わらず、同じ姿勢を長時間続けることで生じてしまうため、デスクワーカーの健康を維持するには適度な姿勢変更が必須である。しかしながら、デスクワーカーが作業中に、自身の姿勢やその変化を意識して、適度に姿勢を変更することは難しい。そのため、デスクワーカーの姿勢を継続的にモニタリングし、適度に姿勢の変更を促す技術の開発が求められている。

姿勢をモニタリングする手法は、センサベースの手法と画像ベースの手法の主に二種類に分けられる。センサベースの手法では、圧力センサや加速度センサなど複数のセンサを人や椅子に取り付け、計測したセンサデータに基づいて、ルールベース、もしくは、モデルベースで推定を行う。この手法では、センサの設置や維持といったコストが高い

ことに加え、人にセンサを取り付ける場合は装着負荷(手間や不快感)が生じる、椅子に取り付ける場合は人間工学を基に設計されたオフィスチェアの性能を損なう可能性があるなどの問題がある。加えて、このような手法は場所に依存することが多く、COVID-19 により爆発的に普及したハイブリッドな働き方には適していない。画像ベースの手法では、RGB カメラやデプスカメラなどで取得した画像を使用して推定を行う。RGB カメラを使用する場合は、一般的に高い推定精度で姿勢を推定可能である。しかしながら、デスクワーカーの顔や身体の RGB 画像を常に撮影する必要があるため、プライバシーに関する不快感を与える可能性が高い。また、デプスカメラを使用する場合は、深度画像(RGB-D 画像)を使用するため、プライバシーの侵襲は少ない。しかし、デバイスに RGB カメラのレンズが備わっている場合や、精度向上のために RGB 画像を併用する場合があります。カメラのレンズで顔を撮影され続けるという不快感は大きい。一方、近年、自動運転や農業など様々な分野において、レーザー光により計測した 3 次元点群データを用いて形状や距離を測定する LiDAR センサの活用が活発になっている。一般的な LiDAR デバイスは広範囲、高粒度な計測が求められる用途のために設計されているため、高価かつサイズが大きいなど携帯性に問題がある。LiDAR センサを利用した姿勢推定手法も提案されているが部屋全体の各人が立っているか、寝ているか、かがんでいるかなどの大まかな姿勢しか対象としていない。

そこで、本稿では、持ち運び可能な卓上設置型センサデバイスを用いて個々人の着座姿勢の推定を行う手法を提案する。提案手法では、我々の研究室で開発している小型

¹ 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University



図 1 提案手法のイメージ



図 2 センサデバイス (Hitonavi-μ) の外観

LiDAR センサを搭載したセンサデバイス Hitonavi-μ [4] を用いることで、図 1 のように作業時に卓上の好きな場所に設置するだけで継続的に着座姿勢を推定可能である。また、LiDAR センサにより、プライバシーの侵害リスクが低い 3 次元点群データを用いて、デスクワークの着座姿勢を推定できる。そして、3 次元点群データから効率的に特徴量を抽出し、特徴量を基に学習された深層学習モデルにより、着座姿勢を推定する。具体的には、入力された 3 次元点群データを、物体検出手法に用いられる Pillar Feature Net により、擬似画像と呼ばれる 2 次元画像に変換し、それを入力画像として深層学習や着座姿勢推定を行う。

提案手法の評価のため、卓上の 3 ヶ所にセンサデバイスを設置し、9 種類の着座姿勢ごとに 3 次元点群データを収集し、姿勢推定モデルを構築したところ、F 値 0.87 以上で姿勢を推定できることを確認した。

2. 関連研究

着座姿勢推定手法は、センサベースの手法、画像ベースの手法に分けられる。センサベースの手法では、椅子や人の身体に圧力センサや加速度センサなどのセンサを複数取り付けて姿勢を推定する。圧力センサを取り付ける手法は、座面や背もたれに、クッション [5] や、シート [6] を設置し、姿勢変化時の圧力の変化を基に着座姿勢の推定を行っているが、人間工学に基づいて設計された椅子本来の機能を損なう問題がある。また、加速度センサを取り付ける手法では、メッシュ素材の椅子に加速度センサを取り付ける手法 [7] や、人の体にセンサを取り付ける手法 [8] が提案されているが、前者は、メッシュ素材の椅子限定であり、後者は装着負荷や不快感を感じさせる。これらのセンサベースの手法は、複数のセンサを設置する必要があるため、総じて設置コストや維持コストが高いという問題や、測定場所が固定されるという問題がある。

画像ベースの手法では、センサを設置するのではなく、RGB カメラやデプスカメラなどで取得した画像を用いて姿勢を推定している。RGB カメラを用いる手法 [9] では、対象者の顔や身体 of RGB 画像を使用するため高い推定精度を得ることができ、かつ、Web カメラのような持ち運び可

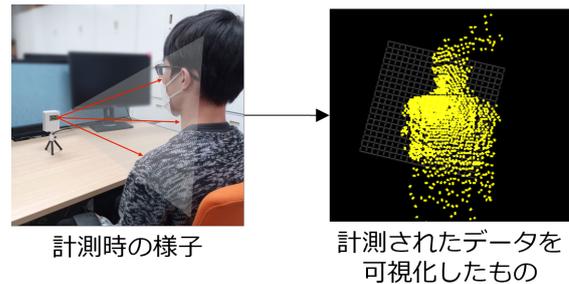


図 3 3 次元点群データの例

能なデバイスを用いるため、場所に依存しない。しかしながら、常に顔や身体を撮影し続ける必要があるため、対象者のプライバシーの侵襲性が高く、不快感を感じやすい。デプスカメラを用いる手法 [10, 11] では、深度画像 (RGB-D 画像) を使用するため、RGB 画像を用いる手法より、プライバシーの侵襲は少ないが、デバイス自体が RGB カメラのレンズも備えており、手法によっては、精度向上のために RGB 画像も併用しているため、カメラのレンズで顔を撮影され続けるという不快感は大きい。

近年、自動運転や考古学、農業など様々な領域において、レーザー光により計測した 3 次元点群データを基に物体の形状や距離を測定する LiDAR センサが利用されている。LiDAR センサを姿勢推定に利用する研究も行われており、Giovanni Diraco ら [12] は、対象人物の 3 次元点群データから、人体の中心軸からの距離と体積、骨格データを取得し、それらのデータからサポートベクタマシンを用いて 19 種類の姿勢を推定している。しかしながら、LiDAR センサは、一般的に広範囲、長距離を対象としており、上記の手法においても、部屋単位で、立ち、かがみ、座り、寝る (あおむけ) など、人の大まかな姿勢のみを推定している。このような、広範囲にわたって計測を行う LiDAR センサは、非常に高価で、持ち運びも困難であり、着座姿勢など範囲の狭く細かな姿勢の推定にも向かない。

本稿では、安価かつ軽量の LiDAR センサを備えた、持ち運び可能な卓上型センサデバイスを用いて、着座姿勢の推定を行える手法を提案する。

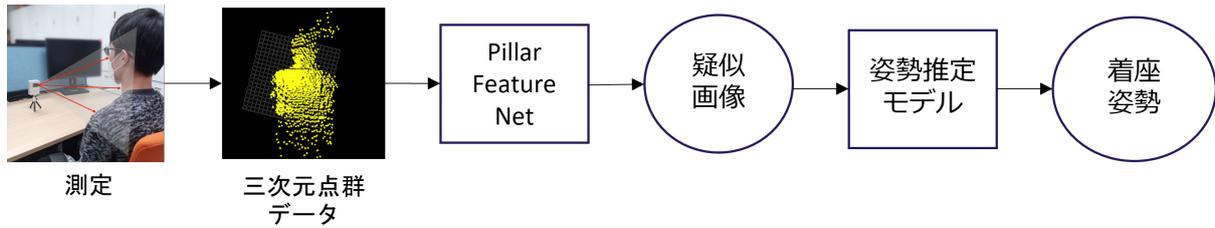


図 4 提案手法の概要図

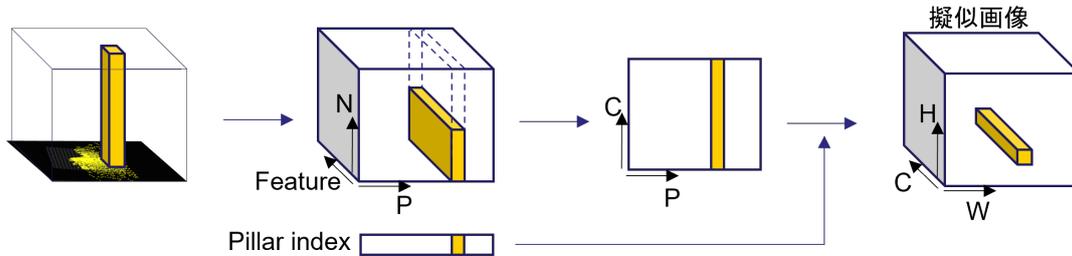


図 5 Pillar Feature Net の概要図

3. 提案手法

本章では、3次元点群データを計測するための卓上設置型センサデバイスについて述べたのち、着座姿勢推定手法について述べる。

3.1 卓上設置型センサデバイス

一般的な LiDAR センサは、放射したレーザパルスが物体の表面で反射して戻ってくるまでの時間を計測し、広範囲に渡って物体検知および距離を計測する [13]。しかし、市場に存在する一般的な LiDAR デバイスは高価、かつ、サイズが大きいため、持ち運びに適しておらず、また、広範囲、かつ、高精度な計測は、個々人の着座姿勢を推定するという目的にも合致していない。そこで本研究では、我々の研究室で開発している小型 LiDAR センサを搭載したセンサデバイス Hitonavi- μ [4] を用いて 3次元点群データを計測する。

図 2 に、Hitonavi- μ [4] の外観を示す。このデバイスは、LiDAR センサとして、小型かつ軽量の DK-ILT001 を搭載しており、重さは 16 g と軽量で、最高測距速度も 120 fps と高速に点群データを収集できる。消費電力も平均 0.6 W と低消費電力であるため、バッテリーによる駆動も比較的容易であり、持ち運びにも適している。また、低消費電力ゆえにレーザ出力が低く、測距距離 150~800mm、測距有効視野角 (FOV) 約 55 度×約 43 度と、100 m 以上先の物体を検出することを目的とした一般的な LiDAR と比較して測定範囲は狭い。しかし、着座姿勢を測定するという観点からは、不要な点群が計測されない分、こちらのセンサの方が適していると言える。図 3 は、正面にデバイスを設置して、計測された 3次元点群データを可視化したもので

ある。点群内にある各 3次元点は x, y, z 座標の値のみで表現され、色などの情報は含まれず、粒度も粗いためプライバシーの侵襲性は低くなっている。

3.2 着座姿勢推定手法

図 4 に提案する着座姿勢推定手法の概要を示す。提案手法は、3次元点群から物体の検出とそのカテゴリを予測する物体検出アルゴリズムである PointPillars [14] を参考に、Pillar Feature Net と姿勢推定モデルで構成されている。PointPillars [14] は、点群を箱状に分割する物体検出アルゴリズムとは異なり、点群を柱状に分割することでモデルの変数や層数を減らし、高速、かつ、高精度に物体検出を可能にしている。本研究では、物体検出は必要ないため、点群を柱上に分割し、効率的に 3次元点群を 2次元特徴量 (疑似画像) へ変換する特徴量エンコーダである Pillar Feature Net の部分を提案手法に取り入れた。

Pillar Feature Net の概要図を図 5 に示す。Pillar Feature Net では、まず点群を x - y 平面上で等間隔な格子状に分割する。これにより、点群は柱状に分割される。以下ではこれをピラーと呼ぶ。その後、各ピラーに含まれる 3次元点 (x, y, z) の次元を 8次元に拡張する。これは、所属する 1つのピラーにある全ての点の座標の平均からある点までの距離 x_c, y_c, z_c と、所属するピラーの中心からのオフセット x_p, y_p を加えることで行われる。PointPillars [14] では、反射率のデータを含んでいるが、本研究で使用するセンサデバイスでは、反射率データを得ることができないため、すべての点において反射率は 1 とし、学習には使用しない要素とした。これより得られるデータは、存在するピラーの数を P 、一つのピラーの中にある点群の最大数を N とすると $(P, N, 8)$ である。次に、活性化関数に線

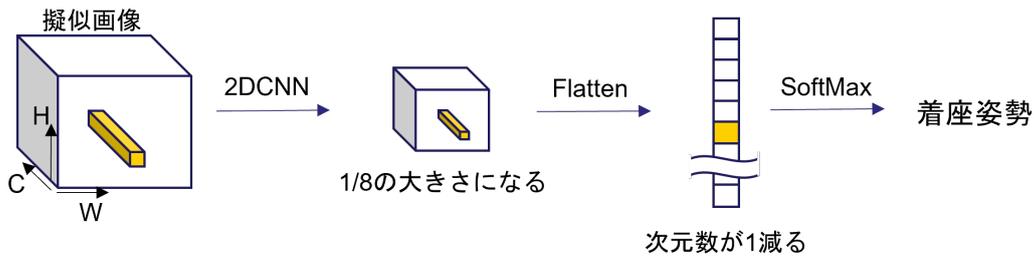


図 6 姿勢推定モデルの概要図

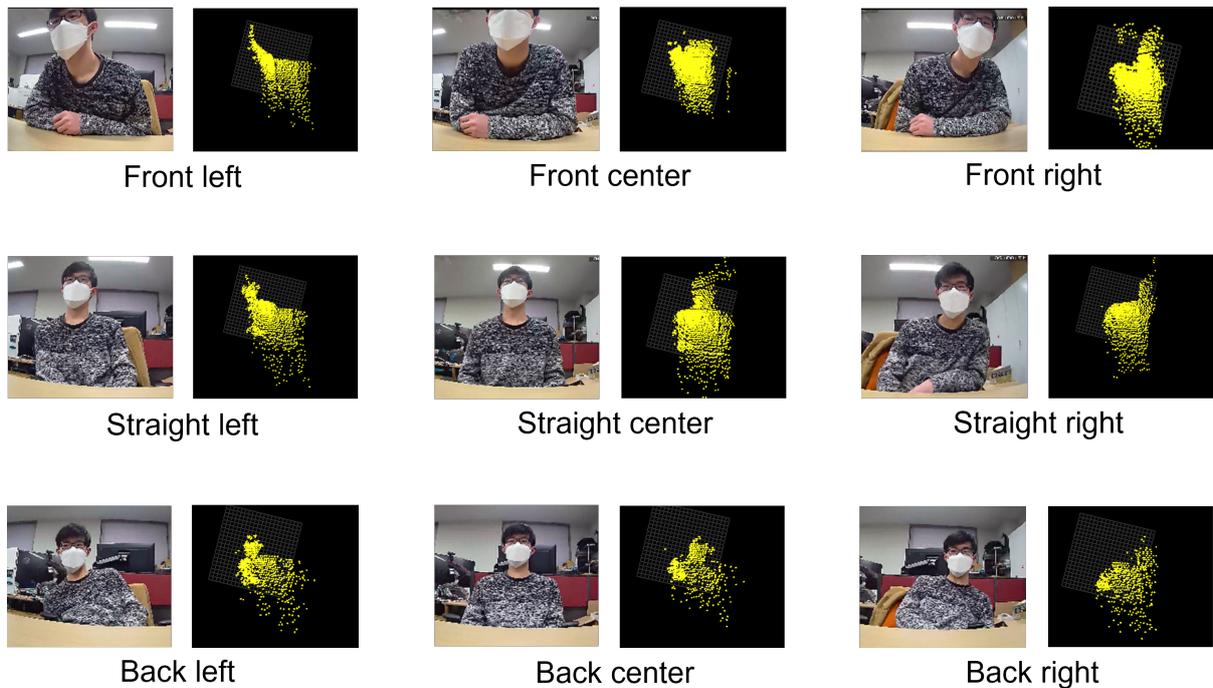


図 7 9 種類の着座姿勢と対応する 3 次元点群の例

形関数を用いた 2 次元畳み込み層、バッチ正規化レイヤ、ReLU 関数、Max Pooling 層の順に、先ほど得られたデータを入力することで、特徴を抽出した (C, P) テンソルを得る。最後に得られたテンソルとそれぞれのピラーの位置から、 (C, H, W) の疑似画像を得る。この疑似画像は一般的な RGB 画像と同じように使用することができるため、画像分類などで使用される 2 次元畳み込みニューラルネットワーク (2DCNN) を活用することができる。

抽出された疑似画像は、姿勢を分類するための姿勢推定モデルに送られる。PointPillars [14] では、物体検出のためにバウンディングボックスを必要とするため、物体検出を行う Detection Head として SSD (Single Shot Detector) を使用していたが、本研究の目的は姿勢推定であるため、姿勢推定モデルとして、図 6 のような、2DCNN 層と flatten 層、Softmax 層から構成されるネットワークを使用する。2DCNN 層と Flatten 層は、疑似画像の大きさや次元を落としていながら特徴を抽出する層である。また Softmax 層は、対象とする姿勢それぞれについて、取得した点群データがその着座姿勢である確率を出力する。推定結果は

出力値が最も高いラベルに対応する姿勢とする。

4. 性能評価

提案手法では、椅子や壁にデバイスを固定設置するのではなく、持ち運びが可能な卓上設置型センサデバイスを用いるため、実利用においては、人や日ごとに設置位置が異なることが予想される。そこで、異なる位置にデバイスを設置して 3 次元点群データを収集し、収集したデータを用いて姿勢を正しく推定できるか検証する。本実験では、図 8 のように、デバイスの設置位置を、着座者に対して、正面右手、正面中央、正面左手の 3 ヶ所に変えながら、1 人の被験者から着座姿勢ごとに 3 次元点群データを収集した。また、本実験では、前傾、直立、後傾の前後方向と、左、中央、右の左右方向の組み合わせた 9 種類を推定対象として、1 ヶ所ごとに、全ての姿勢について、それぞれ 3 分間ずつ 3 次元点群データを収集した。推定対象の 9 種類の姿勢と、正面中央にセンサデバイスを設置した際に収集された 3 次元点群データの例を図 7 に示す。

評価方法として、まず、1 ヶ所で収集された 3 次元点群

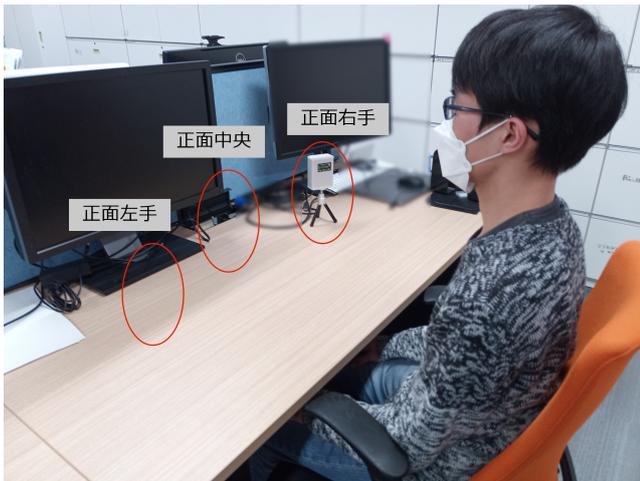


図 8 センサデバイスの設置位置

データのみを用いて構築された推定モデルで、他の位置に設置した際にも推定が行えるか検証する。各設置位置において収集した3次元点群データを用いて、それぞれ姿勢推定モデルを構築し、モデル構築に使われなかった2ヶ所の3次元点群データを検証データとした場合の、Precision, Recall, F 値を表 1 に示す。正面中央に設置したデータのみを用いた推定モデルは、正面右手および正面左手に設置して収集したデータを用いたモデルより推定精度が高いことが分かる。正面中央が高くなったとはいえ、F 値 0.22 では推定ができるとは言い難く、単純に1ヶ所のデータを用いただけでは、設置位置を変えた場合に対応できないと言える。

前述の通り、1ヶ所から収集した3次元点群データのみでは、設置位置を変えた場合に対応できないため、複数の設置位置で収集したデータを学習に用いた場合を検証する。ここでは、各設置位置で収集したデータのうち、80%を学習データ、20%を検証データとしてそれぞれ統合し、推定モデルの構築および評価を行った。表 2 は、3ヶ所のデータを用いた場合の Precision, Recall, F 値を示しており、0.87 以上の F 値で9種類の姿勢について推定が行えていることが分かる。本実験では、3ヶ所に設置して収集した3次元点群データを用いたが、実利用では、より様々な位置にセンサデバイスが設置され、また、様々な体格の人が利用することとなる。しかしながら、設置位置や体格の全てを網羅してデータを収集することは困難である。そのため、代表的な位置や体格の人からデータを収集するとともに、例えば、Pillar Feature Net で疑似画像を作成する前に、3次元点群データを回転、縮小、拡大するなどデータ拡張を行い、様々な設置位置や体格などに対応できるように学習データを増やしていく必要があると考えられる。

5. おわりに

本稿では、健康的な職場環境を実現するために、卓上設

表 1 設置位置毎にモデルを構築した場合の評価結果

設置位置	Precision	Recall	F 値
正面右手	0.16	0.22	0.11
正面中央	0.29	0.25	0.22
正面左手	0.27	0.15	0.14

表 2 3ヶ所のデータを用いて構築したモデルの評価結果

Precision	Recall	F 値
0.92	0.88	0.87

FrontLeft	1464	0	0	0	0	0	559	1316	30
FrontCenter	0	1885	2	0	0	0	786	618	65
FrontRight	0	4	3079	0	0	0	2	202	78
StraightLeft	0	0	0	3377	0	0	2	0	0
StraightCenter	0	3	1	1	3495	0	5	7	1
StraightRight	0	2	0	0	0	0	3334	1	0
BackLeft	0	0	0	6	0	0	0	3342	2
BackCenter	0	4	2	8	7	1	2	3343	1
BackRight	0	5	0	0	0	1	2	1	6
									3319
	FrontLeft	FrontCenter	FrontRight	StraightLeft	StraightCenter	StraightRight	BackLeft	BackCenter	BackRight

図 9 3ヶ所のデータを用いて構築したモデルの混同行列

置可能な小型 LiDAR センサデバイスを用いて計測した3次元点群データに基づき着座姿勢を認識する手法を提案した。提案手法では、3次元点群データから Pillar Feature Net を利用して2次元特徴量に変換し、深層学習モデルにより着座姿勢を推定する。実利用を想定し、9種類の基本的な着座姿勢について、卓上の3ヶ所にデバイスを設置して収集した3次元点群データを用いて、推定モデルを構築した結果、F 値 0.87 以上で推定可能なことを確認した。今後は、様々な設置位置や体格が異なる複数人からデータを収集するとともに、データ拡張により学習データを増やすことで、提案手法の精度向上および検証を行う。また、労働者や学生に長期的に利用してもらい提案手法の有効性の検証を行う予定である。

謝辞

本研究は JSPS 科研費 19K11924、および JST A-STEP JPMJCR21M5 の支援を受けたものです。

参考文献

- [1] Nyssa T Hadgraft, Genevieve N Healy, Neville Owen, Elisabeth AH Winkler, Brigid M Lynch, Parneet Sethi, Elizabeth G Eakin, Marj Moodie, Anthony D LaMontagne, Glen Wiesner, et al. Office workers' objectively assessed total and prolonged sitting time: individual-level

- correlates and worksite variations. *Preventive medicine reports*, Vol. 4, pp. 184–191, 2016.
- [2] Hadi Daneshmandi, Alireza Choobineh, Haleh Ghaem, and Mehran Karimi. Adverse effects of prolonged sitting behavior on the general health of office workers. *Journal of lifestyle medicine*, Vol. 7, No. 2, p. 69, 2017.
- [3] Adrian Bauman, Barbara E Ainsworth, James F Sallis, Maria Hagströmer, Cora L Craig, Fiona C Bull, Michael Pratt, Kamalesh Venugopal, Josephine Chau, Michael Sjöström, et al. The descriptive epidemiology of sitting: a 20-country comparison using the international physical activity questionnaire (ipaq). *American journal of preventive medicine*, Vol. 41, No. 2, pp. 228–235, 2011.
- [4] Shota Yamada, Hamada Rizk, and Hirozumi Yamaguchi. An accurate point cloud-based human identification using micro-size lidar. In *International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 2022.
- [5] Jawad Ahmad, Henrik Andersson, and Johan Sidén. Sitting posture recognition using screen printed large area pressure sensors. In *IEEE Sensors Journal*, pp. 1–3, 2017.
- [6] Jianquan Wang, Basim Hafidh, Haiwei Dong, and Abdulmotaleb El Saddik. Sitting posture recognition using a spiking neural network. *IEEE Sensors Journal*, Vol. 21, No. 2, pp. 1779–1786, 2020.
- [7] Teruhiro Mizumoto, Yasuhiro Otoda, Chihiro Nakajima, Mitsuhiro Kohana, Motohiro Uenishi, Keiichi Yasumoto, and Yutaka Arakawa. Design and implementation of sensor-embedded chair for continuous sitting posture recognition. *IEICE TRANSACTIONS on Information and Systems*, Vol. 103, No. 5, pp. 1067–1077, 2020.
- [8] Sangyong Ma, Woo-Hyeong Cho, Cheng-Hao Quan, and Sangmin Lee. A sitting posture recognition system based on 3 axis accelerometer. In *2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp. 1–3. IEEE, 2016.
- [9] Kehan Chen. Sitting posture recognition based on open-pose. In *IOP Conference Series: Materials Science and Engineering*, Vol. 677, p. 032057. IOP Publishing, 2019.
- [10] Weiyan Ren, Ou Ma, Hongxin Ji, and Xinyuan Liu. Human posture recognition using a hybrid of fuzzy logic and machine learning approaches. *IEEE Access*, Vol. 8, pp. 135628–135639, 2020.
- [11] Audrius Kulikajevs, Rytis Maskeliunas, and Robertas Damaševičius. Detection of sitting posture using hierarchical image composition and deep learning. *PeerJ computer science*, Vol. 7, p. e442, 2021.
- [12] Giovanni Diraco, Alessandro Leone, and Pietro Siciliano. Human posture recognition with a time-of-flight 3d sensor for in-home applications. *Expert Systems with Applications*, Vol. 40, No. 2, pp. 744–751, 2013.
- [13] Hamada Rizk, Hirozumi Yamaguchi, Moustafa Yousssef, and Teruo Higashino. Gain without pain: Enabling fingerprinting-based indoor localization using tracking scanners. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, pp. 550–559, 2020.
- [14] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12697–12705, 2019.