

映像からの姿勢推定処理を利用した子どものリスク評価の定量化

Quantification of risk assessment of children using posture estimation processing from Videos

高尾 哲 康

Takao Tetsuyasu

現在、子どものアレルギー疾患が増加しており、その大きな要因として室内環境汚染物質の存在が挙げられている[1]。このため、多くの時間を過ごす屋内環境中の真菌（カビ）や有害化学物質を子どもがどれくらい体内に取り込むのかを見積もるリスク評価が求められている。子どもは単位体重あたりの呼吸量や吸収量、食事量が大人と比較してその割合が大きいこと、化学物質に対する感受性が大人と比較して敏感なこと[2]、身体特性や行動特性が大人と異なることなどの理由から、正確なリスク評価をするためには子どもの行動特性に合った評価をしていく必要がある。現在はコロナ感染防止のために屋内で過ごすことが多くなってきたので、リスク評価の必要性はますます大きくなってきている。本研究では、コンピュータによる画像処理手法として人工知能（AI）を利用した姿勢推定処理を用いて子どもの行動特性を定量的に評価する手法を開発し、子どもの行動特性を定量化することを目的とする。これにより、子どもの活動強度別に呼吸量を推定するとともに、子どもの環境汚染物質への曝露経路を推定する可能性が開ける。さらに、子どもの行動特性を考慮した室内環境汚染物質の曝露評価のための資料を得ることができ、リスク回避や行動見守りなどの応用も期待できる。

キーワード：子ども、環境汚染物質、リスク評価、画像処理、姿勢推定処理

1. はじめに

現在、子どものアレルギー性鼻炎、気管支ぜんそく、過敏性肺炎、アトピー性皮膚炎、じんましんなどのアレルギー性疾患が増加しており、これらの原因のひとつに室内の真菌（カビ）や環境汚染物質がアレルゲンとしてこれらの疾患を引き起こす可能性が指摘されている[3]。そのため、これらのアレルゲンを子どもがどれくらい体内に取り込むかを測定することが重要になる。アレルゲンを体内に取り込む可能性がある子どもの行動としては、呼吸や飲食とともに、おしゃぶりなどを含め、特に顔の目や鼻、口などの部位を触る・ひっかく・なめる・いじるなどの動作が考えられる。このような動作は本人が意識しているか、意識していない（くせなど）かを含めて個

人差も大きいと考えられる。いずれにしてもこのような動作を定量的に測定することでリスク評価が可能となり、リスク回避や行動見守りなどの応用ができるようになる。

前述のような可能性がある行動を自動的に検出するために、幼稚園・保育園・子ども園にて定点撮影した映像をもとに、画像処理分野で開発されつつある、人工知能を利用した姿勢推定処理手法を導入することで定量的な観測データを得ることを考えた。さらに気温や湿度、行動に伴う呼吸量の変化などのデータを組み合わせることでこのような行動を起こしやすい環境・状況を見つけ出すことも可能になる。また、どのような活動をしているときか、他者との接近・接触、どの物体に接近・接触した後にこのような行動を起こしやすいのかなど、さまざまな要因を見つけ出す手がかりとなり、リスク回避や行動見守り手法の開発につながる可能性がある。特に子どもの行動は多彩であり、事故防止の面でもリスク回避や行動見守りはより重要である。これまでの研究では、自由行動における子どもの行動特性の実態を定量的に評価することが困難であった。本研究では、子どもの自由行動を定点撮影した映像データをもとに姿勢推定処理を適用して動作特性を解析することを試みた。

2. 定点撮影映像データ

今回解析対象にした、子どもの自由行動の定点撮影映像データの詳細を表1に示す。天気、気温、湿度は気象庁による過去の気象データ[4]から得た。撮影日時、時間、映像解像度、フレームレートなどは映像データ内のメタデータ部分から抽出した。定点撮影が1カ所のみのため、被写体のサイズを計算しやすいように起立した自由行動が多い映像を解析対象とした。定点撮影映像のうち、あるフレーム（コマに相当）の画像例を図1（フレーム番号9603）に示す。この画像では個人識別しにくいように人物に姿勢推定処理を適用している。画像内の各人物についている番号は個人識別用の仮番号であり、撮影映像全体について通し番号を付けることにした。被写体が映像内で動き回っても、他人とすれ違ったり、何かの物体の陰に隠れてしまうことがあっても、追従できる限りは同一の通番とした。

表1. 定点撮影映像データの詳細

撮影日時	2016年6月17日9～11時
撮影時間	1時間7分34秒
場所	M幼稚園保育室
天気	くもり
気温	24.2～25.6℃
湿度	59～69%
映像解像度(ピクセル)	1920×1080
フレームレート	29.97 fps
フレーム総数	121,559
ファイルサイズ	約4.3 GB

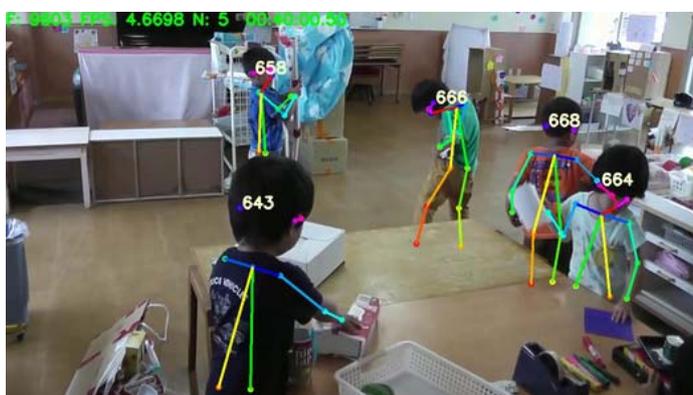


図1. 定点撮影映像データの例

姿勢推定処理はフレームごとに処理するため、フレーム数が多いと映像データ全体を処理するのに非常に時間がかかる。例えば、CPU: Core i5-6400 2.7GHz のコンピュータでは1フレームにかかる処理は画像解像度や画像の複雑度にもよるが、CPUのみでの処理で数秒～20秒かかる。表1の映像データ全体について処理する場合、単純比例計算で8日～30日かかることになる。画像処理をビデオカード上のGPU (Graphics Processing Unit) を利用することにすれば、GPUの種類にもよるが、約20倍の高速化が図れる。今回はGPU搭載のNVIDIA製ビデオカードを利用することにした。さらに処理を高速化・効率化するために、フレームの間引きも行った。スポーツ分野などのスピードが重要な映像に見合うフレーム数ほどの必要はなく、通常の身体活動を追従できる程度のフレーム数であれば問題ないと考え、0.25秒間隔のフレーム(4fps)とした。さらに姿勢推定処理後の映像を確認しやすいようにタイムラプス(デフォルト再生で早送り状態)処理も適用した。映像をフレームごとに間引いたり、再生速度を調整したりするなどの映像変換ツールとしてFFmpeg[6]を利用した。その結果、表1の映像データ(1時間7分)について姿勢推定処理にかかる時間が1時間以内となり、姿勢推定処理後の映像を5倍速14分程度で再生・確認ができるようになった。これにより、映像変換を適切に行うことでリアルタイムでの姿勢推定処理も可能となり、AIエッジ型コンピューティングを活用することでコンパクトなリスク回避や自動行動見守りシステムへの期待がもてる。

3. 姿勢推定処理

映像データの姿勢推定処理のために利用したハードウェア・ソフトウェアを表2に示す。高速化で重要な部分は、前述

表2. 利用したハードウェア・ソフトウェア

CPU	Intel Core i5-6400 2.7GHz
メモリ	16GB
ビデオカード	NVIDIA GTX-1080Ti (GPUメモリ 11GB)
OS	Linux (Ubuntu 16.04 LTS)
GPU向けプラットフォーム	CUDA(Compute Unified Device Architecture)11.0 [5]
映像変換ソフトウェア	FFmpeg 4.2.3 [6]
開発言語	Python3, Perl, bash, 他
AIシステム開発環境	Anaconda3 [7]
画像処理ライブラリ	OpenCV 4.4 [8]
Deep Learningモデル開発	Keras TensorFlow [9]
姿勢推定処理	OpenPose [10]
ビデオ再生	VLC [11], 他

のNVIDIA製ビデオカードとGPUを利用するためのプラットフォームCUDA[5]である。姿勢推定処理には機械学習を利用して開発された技術であるOpenPose[12]を利用した。姿勢推定のための他の方法としては、オンラインゲーム向けに開発された、Microsoft Azure Kinect[13]などのモーションキャプチャーを利用する方法がある。関節マーカー付きのウェアなどを必要とせず、深度センサを搭載していることで、ゲームプレイヤーの動きを3Dトラッキングした精度の高い姿勢データを得ることができ

る。また、モーションキャプチャーを複数台利用して同期をとることで被写体全身のより正確な位置検出が可能である。この方法では撮影機器が限定され、ハイエンドな PC も必要とする。今回は一般的なビデオカメラで撮影した映像を利用することを考え、OpenPose を利用することにした。OpenPose では1枚の画像(フレーム)上で人体骨格の関節、両手足の関節、顔の各部位のキーポイント(最大130個)を位置情報(2Dデータ、X-Y座標)と検出の信頼度スコア(0.0~1.0)を得ることができる。2Dデータをもとに3Dデータ化する3D-Pose-Baseline[14]などの試みもあるが、解析対象人数の制限や計算コスト、解析精度などの問題があり、今回は2Dデータをもとに実施した。本研究の目的は子どもの顔の目や鼻の部位と両手との位置関係を検出することであり、検出できた多くの場合で定点撮影しているビデオカメラ方向側に向いているためである。顔が横向きになっている場合や手をカメラ視線方向に伸ばしている場合などは深度データ(奥行きZ座標)が必要となるが、2Dデータからは可能性として検出するにとどまる。後ろ向きの場合を含め、高精度な3Dデータ化には定点撮影カメラを複数台配置することで可能となる。

姿勢推定処理では、画像フレーム内の被写体の身体部位18カ所(両目、両耳、鼻、首、両肩、左右腰、両手肘、両手首、両足膝、両足首)の位置情報と検出の信頼度スコアが得られる。学習モデルを変更することで顔面や指などの細かい部位の位置情報も得られる。今回は、定点撮影映像のためにひとり分の解像度が低いこともあり、顔や指の細かい部位のデータはうまく得られなかった。画像からの姿勢推定処理の一例を図2に示す。各キーポイントの2D座標と座標ごとの信頼度スコアをもとに可視化したものである。キーポイント位置は丸印で表示し、キーポイント間を実線で接続している。また、検出信頼度が高い部位ほど明瞭な画像表示となっている。

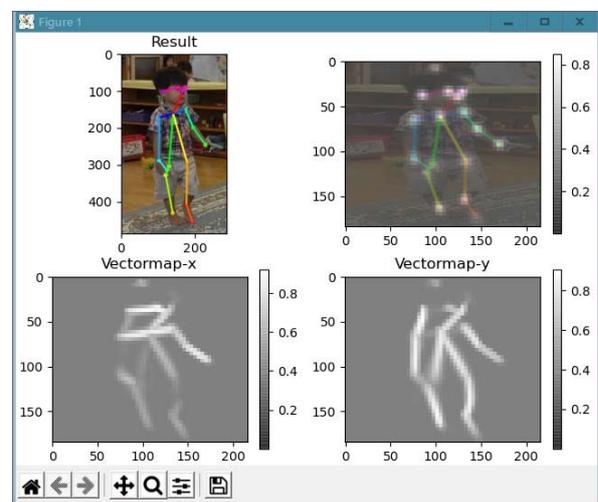


図2. 画像からの姿勢推定処理の例

姿勢推定処理はフレーム単位で行うため、複数の被写体が映っているフレーム間で被写体の同定処理(同一人物判定)は後処理として行う。今回は被写体の身体部位の位置変化と動き方向をもとに連続するフレーム間で補完することで同一人物判定を行った。しかし、被写体が定点撮影カメラの視界からはずれてしまうと、その被写体についてはそこでリセット状態となり、同一人物が再度カメラ視界内に入ってきたとしても以前とは別の人物として扱う(人物に割り当てる通番が更新される)ことになった。今回の解析対象映像は自由行動場面なのでこのようなことが頻発している。解決策としては室内に複数の定点撮影カメラを配置するか、服装や色など別の情報で同一人物判定することになる。また、被写体が定点撮影カメラの視界からはずれなくても、別の人物や物体の陰になることがあり、その場合はフレームから一時的に消えた状態になる(オクルージョン occlusion)。この場合でもフレーム間補完処理によりなるべく同一人物判定するようにした。その結果、ある程度の効果はあったものの、長時間のかくれんぼや子どもどうしで固まってだんご状態になった場合などはうまく判定できなかった。一方、短時間(時間的に近いフレー

ム間)の身体全体または身体の一部(今回の解析対象映像では起立動作が多いので下半身部分が多い)が隠れてしまったり、保育士(大人)とのすれ違いなどで陰になった場合(図3に141番の子どもの例を示す)はうまく判定できた場合があった。同一人物判定の精度を上げることは今後の課題であり、これにより個人ごとのリスク評価の定量化ができるようになる。

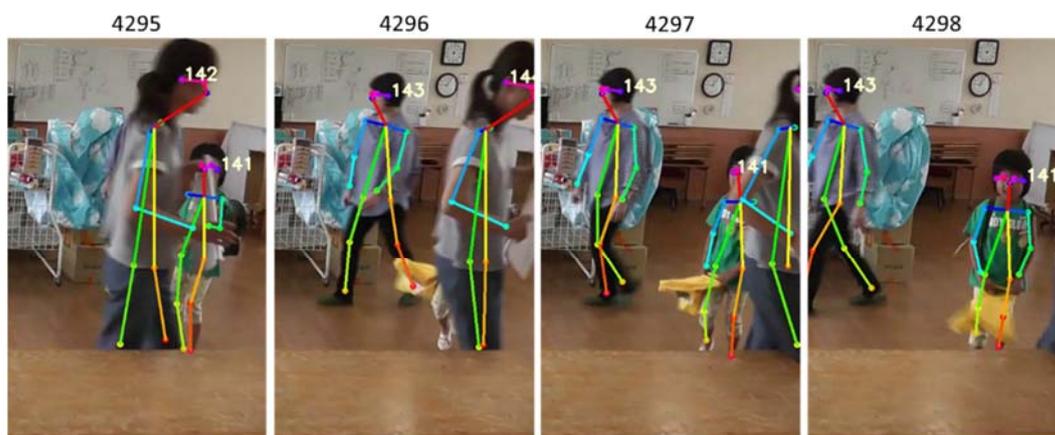


図3. オクルージョン対応処理結果の例(上部の数値はフレーム番号)

4. 姿勢推定結果からのリスク評価の定量化

姿勢推定処理結果のデータをもとにリスク評価の定量化を行う。今回は同一人物判定処理が不十分なために個人ごとのリスク評価はできなかったが、集団としてのリスク評価の定量化を行なうことができた。顔の部位(両目、鼻、口)と両手首との位置関係(合計8箇所)から距離を算出して集計した結果を図4に示す。距離を算出する基準として5歳児の座高平均値61cm(標準偏差2.8)を利用した[15]。子どもの場合は大人と比較して標準偏差が小さいので、算出精度がそれほど悪くなることはないと考えた。座高は尻の下部から頭の先端までの距離なので、姿勢推定処理で得られる両腰と首との距離を約50cmと見積もり、これを基準にして顔の部位と両手との距離を算出した。今回の定点撮影映像では、しばしば下半身が机などで隠れてしまうことが多いため、比較的安定している座高を基準とした。姿勢推定処理では机などに隠れた足の膝や足首の位置も一部推定はするが、精度は悪くなる。また、着席や伏臥、仰向け姿勢などが少ない、起立姿勢の自由行動なので、定点撮影カメラからはほとんど実測に近い状態となり、位置関係の計算誤りになることが少ないと考えられる。体の向きについては、両肩、左右腰位置から体を大きくねじっていない限りはある程度の角度計算ができるので算出しやすいと考えた。しかし、両腕の動きの自由度は子どもの場合は特に高いので、今回は複雑な計算はしないことにした。

図4からは、自由行動のほとんどの場合で顔の部位と両手が接近することはないが、たまに5cm以内に接近するケースが100件程度発生している。10cm以内であれば640件となった。例えば、鼻と接近した場合は口とも接近することが多いので、同じフレーム内で同時発生するケースが多い。顔の部位と両手首との間の距離が5cm以下となった人物が映像内に映っているフレームの番号から発生時刻と経過時間(例えば、手で鼻や目をこすっている時間)がわかる。図5に例を

示す。多くの場合、走り回るなどの全身運動が少ない時間帯（人物の移動量がフレーム間で小さい）に起こりやすいように見受けられた。より高解像度の画像であれば指の位置情報をもとにより精密な距離計算ができるが、多人数が映る定点撮影映像では計算コストの制限もあり、被写体数を限定した映像を利用するなどが今後の課題である。

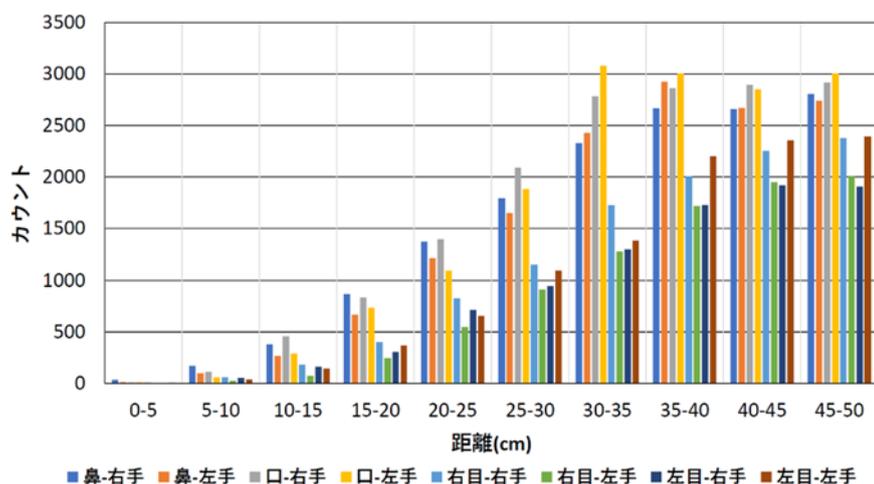


図4. 顔の部位と両手との位置関係（距離換算）

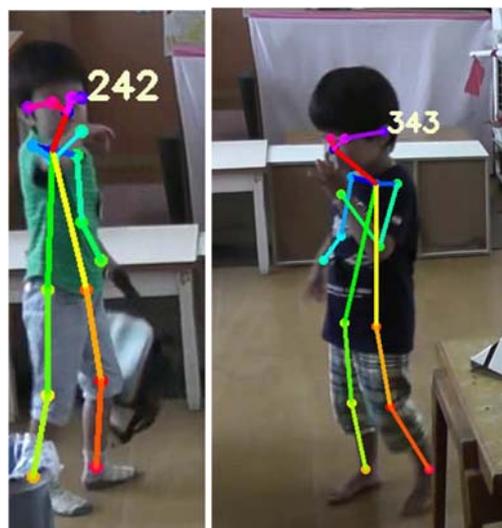


図5. 鼻-手が接近している例

5. まとめ

映像からの姿勢推定処理を利用した子どものリスク評価の定量化を行った結果について報告した。定点撮影映像データがあることで集団についてどのくらいのリスクがあるかを常にモニターできることがわかった。気温や湿度、二酸化炭素量などのデータや子どもの活動内容ごとの呼吸量などと組み合わせることでより詳しい状況判断や原因の特定ができるようになる。画像からの物体検出（画像に映っている物体が何であるかを識別）と組み合わせることでリスクを引き起こす要因の自動検出も可能となる。また、室内汚染物質の吸入や皮膚をひっかく・なめるなどの経皮によるリスクの回避にも応用できるツールとなることが期待される。映像変換を適切に行えばリアルタイムでモニターできることから、今後は性能向上が著しい AI チップ搭載小型マイコンなどのエッジ型コンピューティングを活用し、存在を感じさせないコンパクトな自動行動見守りシステムへの発展も期待できる。

謝辞

本研究は 2019 年度学長裁量経費第 3 号および JSPS 科研費 JP15K00616 の助成にて実施した。幼稚園・保育園・こども園での定点撮影映像データを提供していただいた高橋ゆかり先生に感謝します。

参考文献

- [1] 高橋ゆかり、李剛、石倉卓子、末永美雪、保育園および幼稚園における室内外浮遊真菌の実態調査、富山国際大学紀要、第9巻、pp.127-132、(2017).
- [2] 化学物質評価研究機構、化学物質のリスク評価がわかる本、丸善出版、(2012).
- [3] 室内環境学会、室内環境学概論、東京電機大学出版局、(2010).
- [4] 気象庁、過去の気象データ検索、<http://www.data.jma.go.jp/obd/stats/etrn/index.php> (2020.9 閲覧)
- [5] CUDA: <https://developer.nvidia.com/cuda-toolkit> (2019.9 閲覧)
- [6] FFmpeg: <https://ffmpeg.org/> (2019.9 閲覧)
- [7] Anaconda3: <https://www.anaconda.com/> (2019.4 閲覧)
- [8] OpenCV: <https://opencv.org/> (2019.4 閲覧)
- [9] Keras TensorFlow: <https://www.tensorflow.org/guide/keras?hl=ja> (2019.9 閲覧)
- [10] OpenPose: <https://github.com/CMU-Perceptual-Computing-Lab/openpose> (2019.9 閲覧)
- [11] VLC: <https://www.videolan.org/> (2019.4 閲覧)
- [12] Z. Cao, T. Simon, S.E. Wei, and Y. Sheikh, "Realtime multiperson 2d pose estimation using part affinity fields," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit, (2017).
- [13] Microsoft Azure Kinect: <https://azure.microsoft.com/ja-jp/services/kinect-dk/> (2020.3 閲覧)
- [14] 3D-Pose-Baseline: <https://github.com/una-dinosauria/3d-pose-baseline> (2020.3 閲覧)
- [15] 子供の年齢別・平均体重と平均身長・座高データ (年度別平均表) : スクスクのっぼくん: https://www.suku-noppo.jp/data/average_weight.html (2020.3 閲覧)