

超低解像度遠赤外線画像からの人物骨格推定の検討

岩田 紗希† 川西 康友† 出口 大輔† 井手 一郎† 村瀬 洋† 相澤 知禎‡

†名古屋大学 ‡オムロン株式会社

E-mail: iwatas@murase.is.i.nagoya-u.ac.jp

1 背景・目的

近年，社会の高齢化が問題となっている．高齢者の健康で安全な暮らしのためには，身体機能の維持と緊急時への対応が必要である．そこで独居高齢者を対象とした見守りシステムが注目されている．その実現のためには，屋内に可視光カメラを設置して撮影した画像を用いて，人物の行動を認識することが一般的である．しかし，可視光カメラでは夜間など暗闇での人物の認識が難しいほか，高解像度で日常生活の様子を撮影することには，プライバシー上の懸念がある．

これらの問題の解決策として，赤外線センサアレイの利用が考えられる．赤外線センサアレイとは複数の赤外線センサを格子状に集約したもので，特定の領域内の温度分布を計測することができる．非常に安価であることから，人物の位置を特定したり，体温を測定するセンサとして，空調などの家電製品にも用いられている．取得した画像（図 1 (ii)）は低解像度であるため，個人の特定は困難であり，プライバシー上の懸念を回避できる．また，暗闇でも熱源を感知できるため，昼夜を問わない見守りシステムに適している．

川島らはこのセンサを用いた人物行動認識手法を提案している [1]．しかし，この手法はあらかじめ定義されたいくつかの行動の分類のみ想定しており，それ以外の行動を認識できないことに加え，同じ行動クラス内の細かな違いを知ることもできない．もし，行動を認識するだけでなく，図 1 (iii) のように各フレームで人物の関節位置を推定し，骨格を可視化することができれば，これをもとに任意の行動を詳細に理解できると考えられる．そこで，本研究では赤外線センサを用いて撮影をした FIR (Far InfraRed) 画像からの骨格推定を目的とする．

本研究で扱う FIR 画像は低解像度であり，可視光画像とは性質が異なるため，OpenPose [3] のような既存の複雑な人体骨格推定手法をそのまま適用して骨格推定をすることは難しい．そこで，FIR 画像から回帰により骨格を推定するモデルを提案する．このモデルを学習するためには，大量の関節位置の教師信号が必要であるが，低解像度である FIR 画像に人手で関節位置を付与することは困難である．そこで，本研究では，赤外

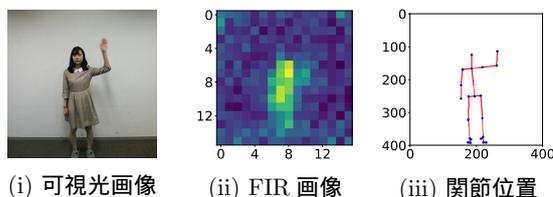


図 1 赤外線センサアレイにより人体を撮影した例と関節位置

線センサアレイと可視光カメラを同期させて学習データを撮影し，可視光画像に対して OpenPose を適用した推定結果から自動的に教師信号を取得して学習する手法を提案する．また，赤外線センサアレイ固有の，室温や被写体との距離に応じて観測画素値が変化する問題や雑音の問題を軽減するために，FIR 画像に事前処理を施す．以上の処理により，FIR 画像を入力とする関節点の直接回帰によって，人物の骨格位置を 2 次元座標で出力する骨格推定器を構築する．なお，本発表は文献 [2] を拡張したものである．

2 骨格推定法

本研究では，骨格推定のために FIR 画像系列から 16×16 画素の画像 N フレームを入力，その最終フレームにおける骨格として 21 個の関節点の 2 次元座標を並べた 42 次元ベクトルを出力とする畳み込みニューラルネットワーク (CNN) を設計した．このネットワークでは，特徴抽出部は各フレームで共有する．図 2 に提案手法の処理手順を示す．提案手法は，学習段階と推定段階に大きく分かれる．以下，各々について述べる．

2.1 学習段階

2.1.1 教師信号の自動取得

骨格推定器を構築するためには，FIR 画像と人物の各関節位置の情報が必要である．しかし，低解像度な FIR 画像の各フレームに関節位置を手により付与することは困難である．そこで本研究では，FIR 画像と同時に撮影した高解像度可視光画像中の関節位置に基づいて自動的に教師信号を付与することにした．そのために構築した赤外線センサアレイと可視光カメラで

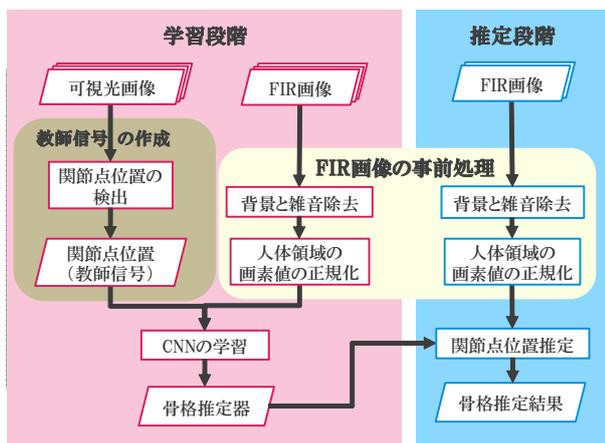


図2 提案手法の処理手順



図3 同期撮影システム

同時に撮影を行なう同期撮影システム（図3）を用いて撮影した可視光画像に OpenPose [3] を適用して得られ骨格を、対応する FIR 画像に対する骨格の教師信号とする。なお、CNN による学習のため、検出した関節点の画素位置は $[-1, 1]$ の範囲に収まるように設定する。

2.1.2 FIR 画像への事前処理

本研究で用いた赤外線センサアレイには以下のような特性がある。

1. 室温に応じて背景の観測画素値が変化
赤外線センサアレイは温度分布を計測するセンサであるため、背景温度も計測結果に含まれる。
2. 観測画素値に雑音が重畳
空気の流れや人物の動きが周りの空気の温度に影響する。また、センサの S/N 比も低い。
3. センサと人体との距離によって観測画素値が変化
センサと熱源となる人体の距離の変化により、観測範囲に対する占有面積率が変化するため、センサの観測熱量が変化する。

上記の特性による影響を軽減するため、提案手法では FIR 画像に事前処理を施す。まず、特性 1 による影響を軽減するために、各 FIR 画像において、最大画素値、最小画素値、平均画素値から背景領域と人体領域を分離するしきい値を算出し、背景除去を行なう。背景除去後も、背景に人体領域であるとみなされた部分が雑音として残る場合があるので、孤立領域を除去す



図4 FIR 画像の背景除去の例

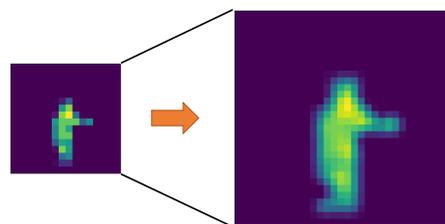


図5 FIR 画像における観測画素値を用いた正規化後の拡大例

ることで、これらを除く。次に、特性 2 及び 3 より、同じ状態の人物を観測していても、フレームごとに画素値が変化することを軽減するため、人体領域の画素値を正規化する。最後に CNN の入力層の大きさに合わせるため、バイキュービック法を用いて FIR 画像を 32×32 画素に拡大する。

2.1.3 CNN の学習

学習では図6に示す CNN を構築し、FIR 画像 N フレームを入力し、同時に撮影した可視光画像系列から検出した関節位置を教師信号として学習をする。学習には Adam [4] を用い、学習データから無作為に N 個の標本を抽出し、それを全体の近似として学習を進めるミニバッチ学習を行なう。また、損失関数には、この教師信号と推定骨格の平均 2 乗誤差 (MSE) を用いる。

2.2 推定段階

推定段階では、学習した骨格推定器を用いて、入力された N フレームの 16×16 画素の FIR 画像の最終フレームにおける人物の関節位置を推定し、42 次元ベクトルとして出力する。

3 実験および考察

3.1 データセット

赤外線センサアレイ（オムロン製 D6T-1616L）と可視光カメラ（バッファロー製 BSW20KM11BK）で同期撮影を行ない、データセットを作成した。FIR 画像及び可視光画像において、1 人の人物の体全体が画像の中心に写るようにした。行動のクラスは表 1 に示す日常生活で行なう行動 A、行動 B、手の動きが分かりや

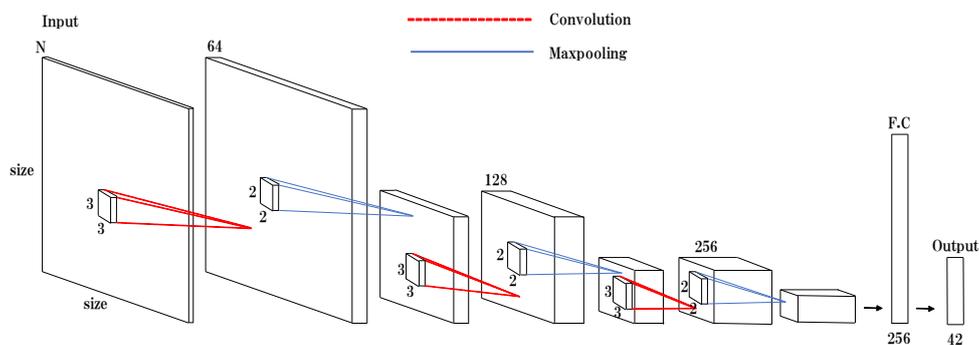


図 6 CNN のネットワーク構造 (図中の数値はカーネルの大きさ, チャンネル数, ユニット数を示す)

表 1 データセット

行動	行動内容	被験者数 (男性/女性)	フレーム数
A	着座状態から立って座る	7 (6/1)	2,830
B	床に置いてある物を拾って置く	9 (7/2)	2,026
C	右手と左手を交互に上げる	9 (7/2)	2,029
D	右足と左足を交互に上げる	12 (11/1)	2,716

表 3 真値と推定結果の RMSE ($\times 10^{-2}$)

行動	比較手法 1	比較手法 2	提案手法 1	提案手法 2
A	8.24	5.62	4.83	4.72
B	11.2	5.92	4.94	4.69
C	8.96	5.23	4.34	4.14
D	4.77	5.80	4.91	4.75
平均	8.29	5.42	4.75	4.58

表 2 実験に用いた手法

手法	比較手法 1	比較手法 2	提案手法 1	提案手法 2
事前処理	なし	あり		
正規化方法	—	2 値化	観測画素値	
入力枚数		1	2	
FIR 画像例				

すい行動 C, 足の動きが分かりやすい行動 D の 4 クラスとした。また, 各行動クラスに関して, 被験者 1 人につき 1 本の映像を撮影した。

3.2 実験概要

表 2 に示す 4 つの手法で実験を行なった。比較手法 1 では FIR 画像に事前処理を施さず, 1 枚だけ入力 ($N = 1$) する。比較手法 2 では FIR 画像に事前処理を施すが, 人物領域に観測画素値を用いず, 人物の輪郭のみが分かるよう, 背景と人物を 2 値化した画像を 1 枚だけ入力する。提案手法 1 では FIR 画像に事前処理を施し, 1 枚だけ入力する。提案手法 2 では時系列情報の有効性を確認するために FIR 画像に事前処理を施し, 2 枚入力 ($N = 2$) した。ただしこのとき, ネットワークは 2 ストリームになっており, パラメータは共有している。各行動の画像系列のうち, 1 人の系列を評価, その他を学習に用いて, 交差検証を行なった。評価指標は可視画像から推定した真値と提案手法による推定結果の各関節位置の RMSE (Root Mean Square Error) とした。

3.3 実験結果

表 3 に各行動における真値と推定結果の RMSE を示す。実験結果よりすべての行動において提案手法 2 が最も高精度であった。

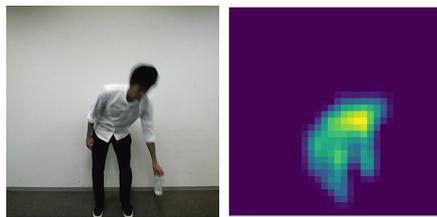
3.4 考察

3.4.1 RMSE に基づく関節位置の推定誤差

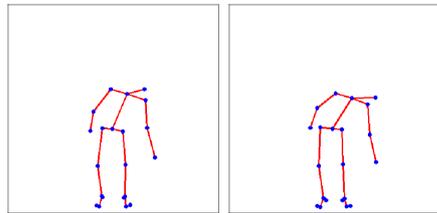
提案手法では CNN における学習のために, 図 8 に示すような座標の正規化を行なった。正規化前の画像の大きさは 400×400 画素, 正規化後の画像の大きさは 2×2 画素なので, 正規化後に 1 画素の大きさを表す最小単位は 0.005 となる。これをふまえると, 提案手法における関節 1 つあたりの平均誤差 (0.05 前後) は, 図 8 (ii) に示す程度の大きさである。

3.4.2 FIR 画像における事前処理の有効性

提案手法では, FIR 画像において人物の輪郭を明確化するために事前処理を施した。図 9 に示す行動 C における人物ごとの RMSE を見ると, 比較手法 1 においては人物によって大きな差があることが分かる。理由として, 人物 1 と 9 は他の被験者と比較して身長が低い, 本来の大きさより大きく推定され, 精度が低下したと考えられる。その他の手法では, 全般的に精度が低かった人物の精度が上がったため, 事前処理の有効性が示唆される。また, 比較手法 2 と提案手法 1 を比較して, 後者の方が精度が高かったことから, 人体領域の観測画素値を用いた方が良かったことが分かった。こ

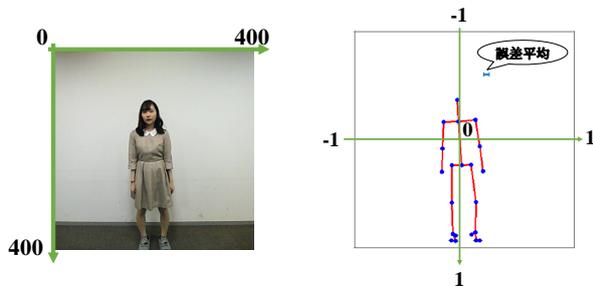


(i) 可視光画像 (ii) FIR 画像



(iii) 正解関節位置 (iv) 推定関節位置

図 7 行動 B に関する提案手法 2 による推定結果の例



(i) 正規化前 (ii) 正規化後

図 8 正規化前後の座標関係

これは、2 値化してしまうと人物領域の輪郭しか分からず、本来は胴体と比較して温度が低い手足などの部位を区別できなかったためと考えられる。このことから、人物の観測画素値は人体の部位をより明確にしておき、学習において有用な情報であることが分かった。

4 むすび

本研究では、低解像度 FIR 画像から人物の骨格を推定する手法を提案した。実験結果より、簡単な行動においては、おおむね正確に推定できることを確認した。今後の課題として、FIR 画像の事前処理の改良やネットワークを工夫することで時系列情報をより効果的に用いることが考えられる。

謝辞 本研究の一部は科学研究費補助金の助成を受けたものである。

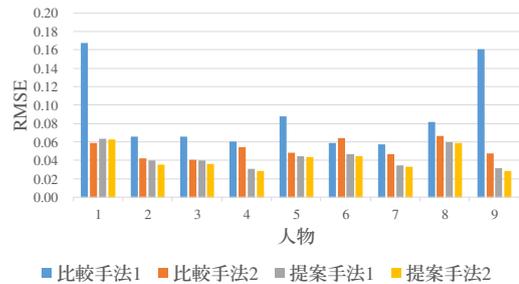


図 9 行動 C における人物ごとの RMSE

参考文献

- [1] 川島 昂之, 川西 康友, 出口 大輔, 井手 一郎, 村瀬 洋, 相澤 知禎, 川出 雅人, “赤外線センサアレイを用いた畳み込み RNN による人物行動認識,” 精密工学会誌, vol.84, no.12, pp.1025–1032, Dec. 2018.
- [2] 岩田 紗希, 川西 康友, 出口 大輔, 井手 一郎, 村瀬 洋, 相澤 知禎, “超低解像度遠赤外線画像からの人物姿勢推定の初期検討,” 情報処理学会第 81 回全国大会, no.5R-05, Mar. 2019.
- [3] Cao, Zhe and Simon, Tomas and Wei, Shih-En and Sheikh, Yaser, “Realtime multi-person 2D pose estimation using part affinity fields,” Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition, pp.7291–7299, July. 2017.
- [4] Kingma, Diederik P and Ba, Jimmy, “Adam: A method for stochastic optimization,” Computing Reserch Repository arXiv preprint, arXiv:1412.6980, Dec. 2014.