

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

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

1 はじめに

近年、高齢化社会の進展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。このようなシステムを実現するためには、屋内に設置した可視光カメラで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は、プライバシー保護上の問題がある。これらの問題を回避するために、川島らは図 1 (ii) に示すような 256 画素 (16 × 16) 画素の超低解像度遠赤外線画像 (FIR 画像) 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、それらに属さない行動は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ができない。もし、図 1 (iii) のように各フレームで人がどのような姿勢をしているのかを知ることができれば、これをもとに任意の行動を詳細に理解することができると考えられる。そこで、本研究では FIR 画像からの姿勢推定を目的とする。

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、既存の人体姿勢推定法をそのまま適用することはできない。この問題に対し、高齢者が日常的にとりうる姿勢は限定的であることに着目し、FIR 画像系列から関節点の座標をニューラルネットワークを用いて回帰し、直接推定する手法を提案する。

2 提案手法

提案手法では、姿勢推定のために 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姿勢として 21 個の関節点の 2 次元座標位

A Preliminary Study on Human Pose Estimation from an Extremely Low-Resolution FIR image

Saki IWATA[†], Yasutomo KAWANISHI[†], Daisuke DEGUCHI[†], Ichiro IDE[†], Hiroshi MURASE and Tomoyoshi AIZAWA[‡]

[†]Nagoya University

[‡]OMRON Corporation

iwatas@murase.is.i.nagoya-u.ac.jp

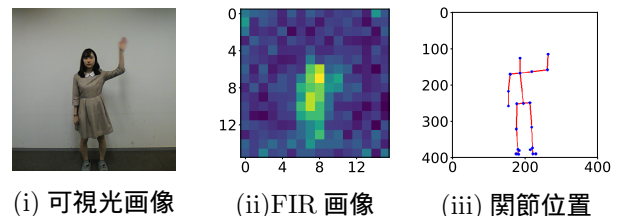


図 1: 人体を撮影した例とその関節位置

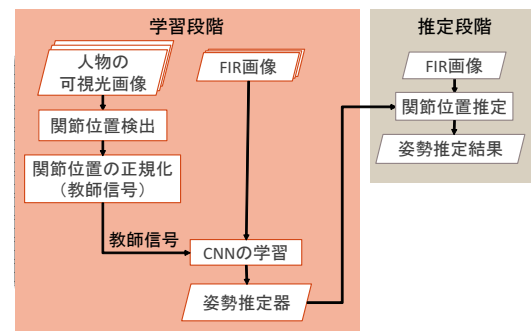


図 2: 提案手法の処理手順

置を並べた 42 次元の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。 N が 2 以上のとき、人体の動きの情報も利用できることが期待される。なお、このネットワークでは、CNN の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。

2.1 学習段階

学習段階では、教師信号には FIR 画像と同時に撮影した可視光画像から OpenPose[2] を用いて検出した関節位置を用いる。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正規化する。CNN の損失関数には、この教師信号と推定姿勢の平均 2 乗誤差 (MSE) を用いる。

2.2 推定段階

推定段階では、学習した姿勢推定器を用いて、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 次元のベクトルとして出力する。

表 1: 真値と推定結果の平均 2 乗誤差 (MSE)

行動	A	B	C
提案手法 1	0.0084	0.0139	0.0097
提案手法 2	0.0072	0.0262	0.0090

3 実験および考察

3.1 データセット

実験では赤外線センサアレイ (OMRON Thermal sensor D6T-1616L) と可視光カメラ (iBAFFALO BSW20KM11BK) で同期撮影を行ない、データセットを作成した。FIR 画像と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、日常生活で行なう、起立と着席をする (行動 A)、床に置いてあるものを拾って置く (行動 B)、さらに関節の動きが分かりやすい、右手と左手を交互に振る (行動 C) の 3 クラスとした。それぞれの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、行動 A は 7 人 (男性 6 人、女性 1 人) で合計 2,830 フレーム、行動 B, C は 9 人 (男性 7 人、女性 2 人) で、各々 2,026 フレーム、2,029 フレームからなる。

3.2 実験概要

時系列情報の有効性を確認するため、フレーム数 $N = 1$ (提案手法 1) と $N = 2$ (提案手法 2) の比較を行なった。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、交差検定を行なった。評価指標は可視光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 MSE とした。

3.3 実験結果

表 1 に各行動における真値と推定結果の MSE を、図 3 に行動 C に対する提案手法 1 による姿勢推定結果の例を示す。表 1 より行動 A, C においては提案手法 2 が、行動 B においては提案手法 1 の方が精度が高いことが分かる。さらに、図 4 より、推定対象の人物によっても精度が大きく異なることが分かる。

3.4 考察

人物ごとに推定精度が異なった理由として、身長差や男女の振る舞いの違いが挙げられる。特に女性被験者のデータと比較して、男性被験者のデータが多かったため、女性被験者の推定姿勢が実際よりも大きくなる傾向が見受けられた。また行動 B において提案手法 2 が有効でなかった理由として、人物ごとの動きの特徴が大きいためと考えられる。

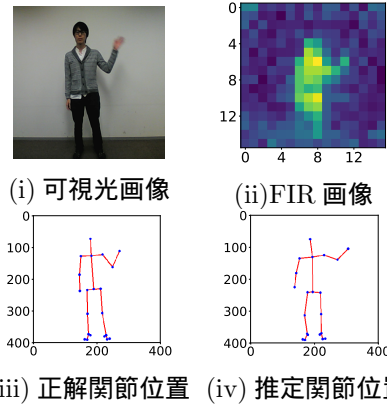


図 3: 行動 C に関する提案手法 1 の推定結果の例

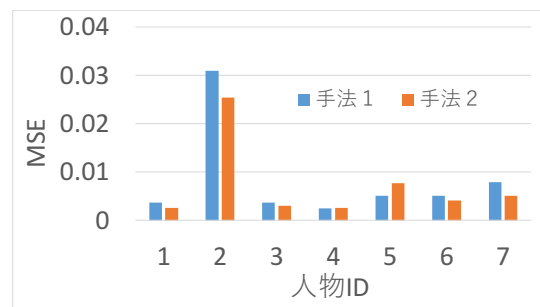


図 4: 行動 A に関する人物ごとの MSE

4 むすび

本研究では、低解像度である FIR 画像から人物の姿勢推定を行なう手法を提案した。今後の課題として、人物の身長差や振る舞いの違いに対応することや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技術研究費補助金の助成を受けたものである。

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 (2018).
- [2] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299 (2017).