

低解像度遠赤外線センサアレイを用いた 表面熱分布付き 3 次元人体モデル推定

玉木 太耀^{1,2,a)} 藺頭 元春^{2,1} 北野 和哉¹ 藤村 友貴¹ 向川 康博¹ 川西 康友^{2,1}

概要

本研究では 32×32 画素の低解像度遠赤外線 (LFIR) センサアレイから得られる LFIR 画像を元に表面熱分布付き 3 次元人体モデルを推定する手法を提案する。LFIR 画像はプライバシーに配慮できる一方、高ノイズかつ低解像度である。そのため既存の可視光画像向け手法の流用は難しく、また手動でのアノテーションも困難であり、適切なデータセットが未整備である。これに対し、本研究では時空間情報の活用とデータセット自動構築の両面からアプローチする。推定では、複数フレームの時間的文脈を統合することで観測ノイズの影響を抑え、人体の形状・表面の隣接関係に基づく空間的制約を導入することで、表面熱分布付き 3 次元人体モデルを推定する。加えて、自動アノテーションシステムを用いてデータセットを構築し、その有効性を評価した。

1. はじめに

近年、IoT 技術の急速な普及により、居住空間における人物センシングの重要性が高まっている。特にスマートホームや高齢者の見守り支援システムにおいて、居住者の姿勢や動作、および温熱状態を非接触かつ継続的に把握することは、空調制御による省エネルギー化や、転倒検知・体調急変の早期発見といった安全・安心な生活環境の構築に不可欠である [6]。従来、こうしたシステムでは可視光画像や深度情報を入力とする深層学習ベースの推定手法が成果を挙げているが [8]、これらは外見情報の取得を伴うため、プライベート空間への導入には心理的・倫理的な懸念が強い。

これに対し近年、低解像度遠赤外線 (low-resolution far-infrared; LFIR) センサアレイが注目されている。LFIR センサアレイは、 32×32 画素程度の LFIR 画像を出力する。LFIR 画像は顔の特徴や衣服の模様といった外見情報を含まないため匿名性が極めて高く、暗闇や逆光環境でも人体

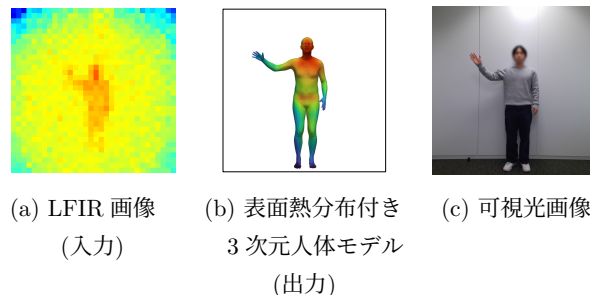


図 1: LFIR 画像, 表面熱分布付き 3 次元人体モデル, 対応する可視光画像の例。

を安定して検知できる。既存の LFIR 画像に基づく研究は、主に行動認識 [6, 9] や骨格推定 [11, 13] に焦点を当てている。しかし、高度な見守りや精密な空調制御には、骨格情報だけではなく、人体全表面の形状および温熱状態（以下、表面熱分布）の把握が重要となる。表面熱分布は、着衣状態や体温を反映する極めて重要な指標であり、これを 3 次元人体モデル上で可視化できれば、周囲環境との相互作用を含めた居住者の状態をより直接的に把握できる。

そこで本研究では、プライバシーを保護しつつ詳細な人体情報の取得を実現するため、LFIR 画像から表面熱分布付き 3 次元人体モデルを推定することを目的とする。図 1 に、(a) 入力として用いる LFIR 画像の例、(b) その出力である表面熱分布付き 3 次元人体モデルの例を示し、参考として (c) 対応する可視光画像を示す。

本研究で扱う LFIR 画像はノイズが多く解像度も限られるため、可視光画像向けの 3 次元人体モデル推定やテキスト推定の手法を LFIR 観測へ流用し、表面熱分布付き 3 次元人体モデルを高精度に推定することは難しい。そこで本研究では、時空間的な情報と人体表面構造を考慮した学習を導入し、LFIR 観測のみから表面熱分布付き 3 次元人体モデルを推定する手法を提案する。また、本手法の学習には詳細な 3 次元人体モデルと熱分布の真値が必要となるが、LFIR 画像は低解像度かつ高ノイズで幾何学の手掛かりに乏しいため、手動アノテーションは現実的ではなく、そのような真値を備えた適切なデータセットも存在しない。そこで本研究では、LFIR と同期して撮影した高解像

¹ 奈良先端科学技術大学院大学

² 理化学研究所, ガーディアンロボットプロジェクト

^{a)} taiyo.tamaki@riken.jp

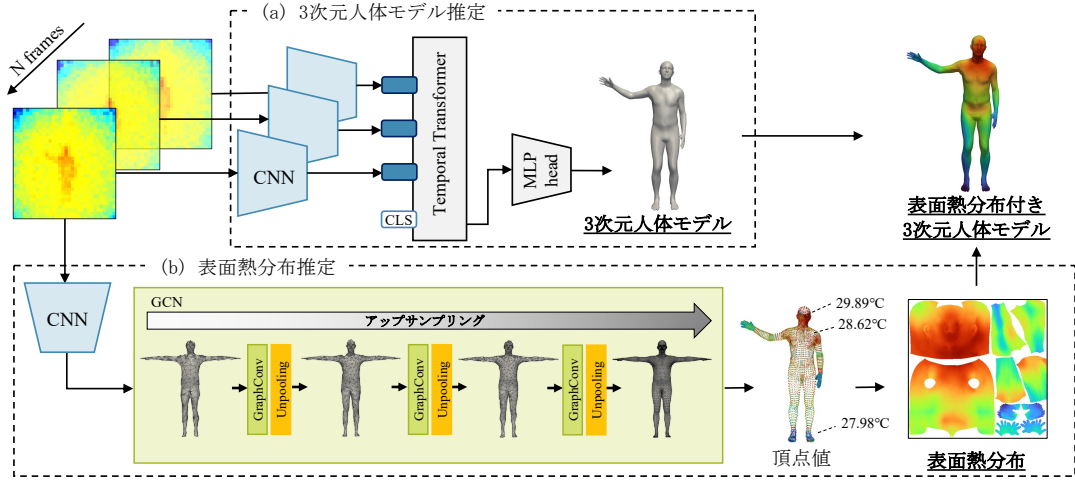


図 2: 提案手法の全体構成. (a) LFIR 画像系列から 3 次元人体モデルを推定し, (b) 単一フレームから頂点値を推定して表面熱分布へ変換する. 最後に, 推定した 3 次元人体モデルに表面熱分布を対応付け, 表面熱分布付き 3 次元人体モデルとして出力する.

度可視光画像から真値を推定し, LFIR 画像へ対応付けることで自動アノテーションを行う. これにより, 人手では付与困難な真値を大量に生成し, 提案手法の学習および定量評価を可能にする.

以上より, 本研究は LFIR 観測に基づく表面熱分布付き 3 次元人体モデル推定を対象とし, そのための推定手法と自動アノテーションに基づくデータセット構築手法を提案する. 評価実験により提案手法の有効性を示し, プライバシーに配慮した居住空間内の人物状態把握に向けた基盤技術としての有用性を明らかにする.

2. 提案手法

本研究では, LFIR 画像から 3 次元人体モデルと人体の表面熱分布をそれぞれ推定し, 両者を統合することで表面熱分布付き 3 次元人体モデルとして出力する手法を提案する (図 2).

2.1 3 次元人体モデル推定

LFIR 画像は低解像度かつノイズも多く含むため, 単一フレームから幾何学的手掛かりを安定して抽出することが難しい. そこで本研究では, 連続フレームに含まれる姿勢の遷移過程を手掛かりとして時空間特徴を学習する構成を採用し, 情報の乏しい観測下でも安定した推定を可能にする. 3 次元人体モデルの表現には, Skinned Multi-Person Linear model (SMPL) [4] を用い, LFIR 画像系列 $X = (x_1, \dots, x_N)$ から最終フレーム x_N に対する姿勢パラメータ $\theta \in \mathbb{R}^{72}$ および形状パラメータ $\beta \in \mathbb{R}^{10}$ を推定する.

2.1.1 推定モデルの構成

推定モデルは, 各フレームの空間特徴を抽出する畳み込

みニューラルネットワーク (CNN), 時系列情報を統合する Temporal Transformer, およびパラメータを推定する回帰ヘッドから構成する (図 2 中の (a) 参照).

まず, 各フレーム x_t に対して 2 次元 CNN を適用し空間特徴を抽出する. この際, 情報欠落を防ぐため中間層での pooling を避け, 最終段の global max pooling により特徴ベクトル f_t を得る [3]. 次に, 得られた特徴系列 $F = (f_1, \dots, f_N)$ の先頭に学習可能な CLS トークン f_{cls} を付加して Transformer へ入力する. 自己注意機構により, ノイズや自己遮蔽を含む不安定なフレームの影響を動的に抑えつつ, 過去の推移と整合する時空間表現を獲得する. 最後に, MLP head により SMPL パラメータ (θ, β) を回帰し, 真値との L_1 損失を最小化するように学習する.

2.2 表面熱分布推定

本研究では, 推定した表面熱分布を 3 次元人体モデル上で扱うため, コンピュータグラフィックスで標準的に用いられるテクスチャ形式を採用する. 具体的には, 3 次元表面を 2 次元平面座標 (u, v) に対応付ける UV 展開を行い, 熱分布を画像形式 (UV テクスチャ) として保持する. この形式は汎用的な 3 次元レンダリングパイプラインとの親和性が高く, 本手法では SMPL の UV 展開に従う 2 次元画像を最終的な出力形式とする. 可視光画像を対象とした従来研究 [2, 10] では, UV テクスチャを CNN により直接回帰する手法が一般的である. しかし, UV 展開には切れ目が存在し, 背中や腕・脚の付け根など 3 次元的に隣接する部位が 2 次元画像上では離れて配置される. そのため, 2 次元畳み込みの局所近傍のみに基づく直接回帰では, 3 次元表面としての連続性を十分に考慮できず, 切れ目近傍で不連続やにじみが生じやすい.

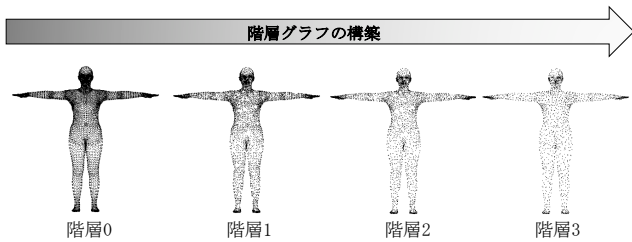


図 3: メッシュ表面グラフの階層化の概念図. 粗い階層で大域的整合性を扱い, 細かい階層で局所の変化を復元する.

そこで本研究では, 2 次元画像としての直接推定を避け, 人体メッシュ各頂点に対応する熱分布の値 (以下, 頂点値) をまず推定し, UV 展開に基づいて UV テクスチャへ変換する方針をとる. これにより, 3 次元表面として連続な部位間で整合性を保った推定が可能となる.

2.2.1 推定モデルの構成

推定モデルは, 3 段の畳み込み層により最終フレーム x_N から特徴抽出を行う CNN エンコーダと, メッシュトポロジに沿って頂点値を回帰するグラフ畳み込みニューラルネットワーク (GCN) デコーダから構成する (図 2 中の (b) 参照). GCN はメッシュの隣接関係に沿って情報を集約できるため, 2 次元画像上で離れて配置される領域間でも 3 次元表面上の隣接関係を保った推定が可能であり, 切れ目近傍の不連続を抑制できる.

SMPL は頂点と面の接続関係に基づくグラフとして扱うことができる. しかし, SMPL は頂点数が多いため, 単一解像度の近傍集約のみでは全身の大域的な整合性を得ることが困難である. そこで本研究では, メッシュを段階的に縮約した階層グラフ構造を導入する (図 3). 階層構造は, 細かい階層から順に頂点をクラスタとしてグループ化し, 各クラスタを粗い階層の代表頂点へ対応付けることで構築する. ここで, 3 次元座標の距離のみに基づくクラスタリングでは, メッシュ表面上で隣接しない頂点が同一クラスタに混在する恐れがある. そこで本研究では, 各メッシュの接続関係を反映した隣接行列に基づくスペクトルクラスタリング [4] を採用する. これにより, 人体形状のトポロジを維持したまま頂点の階層構造を構築でき, 大域的な分布から局所的な詳細へと段階的に情報を復元する構成が可能となる. デコーダはこの階層上でグラフ畳み込みと unpooling を適用し, 元のグラフ解像度における頂点値を出力する. 学習時には, 出力頂点値と真値との $L1$ 損失を最小化するように学習する.

3. 実験設定

3.1 データセット

データ収集には, 32×32 解像度の FIR センサアレイ (Heimann Sensor 社製サーモパイル赤外線アレイ) と, 真

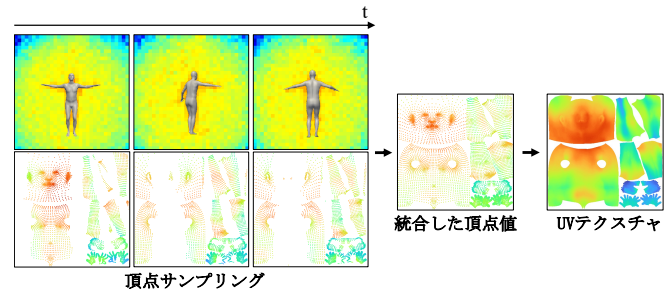


図 4: 表面熱分布の真値構築手順. 被験者の回転により複数視点の LFIR 観測を取得し, メッシュ頂点への画素値サンプリングと時間平均により, 真値 UV テクスチャを得る.

値生成のための RGB-D センサ (Microsoft Azure Kinect) を使用した. 両センサは被験者正面に設置して同期撮影を行い, 室温は $(21 \pm 0.5)^\circ\text{C}$ および $(24 \pm 0.5)^\circ\text{C}$ の 2 条件とした. 撮影においては, 被験者 10 名から約 4,000 フレーム, 合計約 40,000 フレームを取得し, 腕上げ (左右交互), 深呼吸, 拾い上げ, 片足上げ, 歩行等 (中間姿勢を含む) を収録した. 3 次元人体モデルの真値には可視光画像に ScoreHMR [8] を適用して得た各フレームの SMPL パラメータを用いた. また, 表面熱分布の真値は, 単一視点では背面などの死角が生じるため, 被験者にその場で全身回転してもらい, 複数フレームの観測を統合して構築した (図 4). 具体的には, メッシュ頂点を LFIR 画像へ投影して画素値をサンプリングし, 観測が得られたフレームで時間平均して統合頂点値を得た後, UV テクスチャへ展開して生成した. 以上の手順により, 各被験者・各室温条件について, 1 枚の真値表面熱分布を得た.

3.2 比較手法

LFIR 画像から表面熱分布付き 3 次元人体モデルを直接推定する既存手法は存在しないため, LFIR 画像から骨格を推定する既存研究 [9] を参考に比較手法を設定した. 3 次元人体モデル推定に対しては, 単一フレームの空間特徴のみに基づき SMPL パラメータを回帰する 2D-CNN と, 連続フレームを時空間特徴として扱い 3 次元畳み込みで時系列情報を統合して SMPL パラメータを回帰する 3D-CNN を比較対象とした. 表面熱分布推定に対しては, 最も直接的なベースラインとして, 単一フレームの LFIR 画像から UV テクスチャを画像として直接回帰する CNN を比較手法とする. この手法はメッシュトポロジを用いず, UV テクスチャ上の局所画素近傍に基づいて推定を行うため, UV 展開の切れ目に起因する不連続性を考慮できない.

3.3 評価指標

3 次元人体モデル推定の評価には, 関節位置誤差である MPJPE (Mean Per Joint Position Error) と, メッシュ頂点間のユークリッド距離を全頂点で平均した値である再

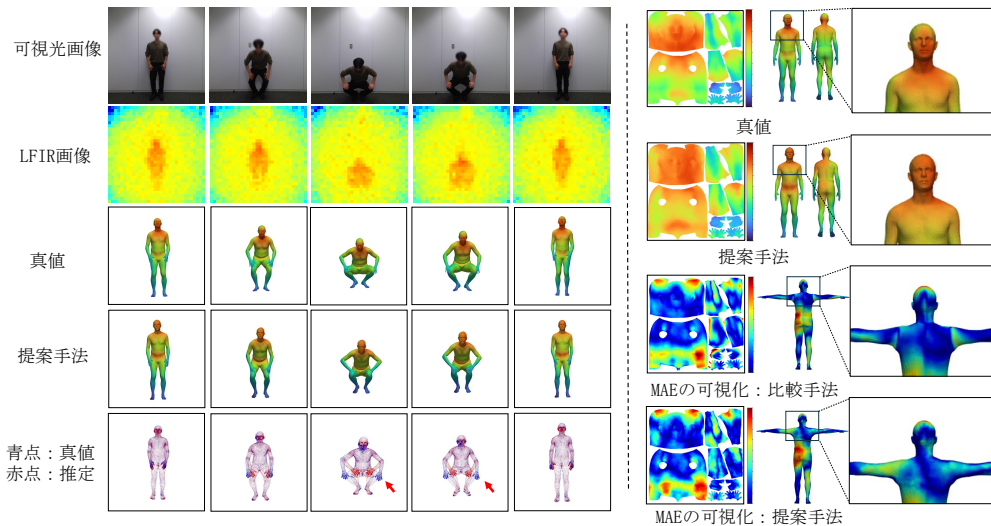


図 5: 左: 3 次元人体モデル推定の定性結果 (可視光画像, LFIR 画像 (入力), 真値, 提案手法, 頂点重ね合わせ). 右: 表面熱分布推定の定性結果 (真値, 提案手法) と, 誤差 (MAE) の可視化 (比較手法, 提案手法).

表 1: 3 次元人体モデル推定の定量結果.

	MPJPE [mm] ↓	再構成誤差 [mm] ↓
2D-CNN	121.6	106.9
3D-CNN	88.9	78.6
提案手法	70.3	61.9

構成誤差を用いる. 表面熱分布推定の評価には, 画素の平均誤差を表す MAE (Mean Absolute Error) と PSNR (Peak Signal-to-Noise Ratio) を用いる. 二乗誤差に基づく PSNR は局所的に大きな誤差を強く反映するため, MAE との併用により平均精度と局所的破綻の両面を評価できる. 評価は, 被験者単位でデータを学習用とテスト用に分割して行い, 未学習の人物に対する汎化性能を確認した.

4. 実験結果・考察

3 次元人体モデル推定

定量結果を表 1 に示す. 提案手法は両指標において最良の結果を示した. 2D-CNN は幾何学的手掛かりが乏しく誤差が大きいが, 時系列を利用する 3D-CNN では大幅に改善した. さらに提案手法は, Transformer による系列の選択的な統合により, 高ノイズなフレームの影響を抑えつつ最も高い精度を達成した. また, 定性的な結果を図 5 に示す. 提案手法は基本的な動作において体幹および四肢の配置を概ね自然に再現できている. 真値メッシュ頂点 (青) と推定メッシュ頂点 (赤) を重ねた可視化では, 全体の頂点分布が良く一致しており, 姿勢の概形を適切に推定できていることが確認できる. 一方で, 図中の赤矢印で示すように, 自己遮蔽や四肢が体幹に近接するフレームでは局所的なずれが増加する傾向が見られた.

表 2: 表面熱分布推定の定量結果.

	PSNR ↑	MAE ↓
CNN	25.54	0.0215
提案手法	29.36	0.0192

表面熱分布推定

定量結果を表 2 に示す. 提案手法は PSNR および MAE の両指標において比較手法を上回り, 有効性が確認された. PSNR の向上は, 画素単位の平均誤差 (MAE) の低減に加えて, 局所的に大きな誤差が生じる領域が抑制されたことを示唆する. このことから, 提案手法は平均的な再現性と局所的な破綻の抑制の両面で, 表面熱分布推定の品質を改善したといえる. また, 定性的な結果を図 5 に示す. 比較手法 (CNN) は UV テクスチャを 2 次元画像として直接回帰するため, UV 展開の切れ目近傍で領域間の連続性を十分に考慮できず, 不連続やにじみが生じやすいことが確認できる. 一方, 提案手法ではこれらの破綻が抑制され, 切れ目を跨いでも滑らかで整合的な熱分布が得られている. これは, メッシュ頂点値として回帰した上で, GCN によりメッシュトポロジに沿って情報を伝搬させることで, 3 次元表面としての整合性を推定過程に取り込めたためと考えられる.

5. おわりに

本研究では, LFIR 画像から表面熱分布付き 3 次元人体モデルを推定する枠組みを提案した. 実験により, 基本的な動作に対して提案手法が概ね妥当な推定結果を与えることを確認した. 今後の課題として, 室温条件・衣服・距離など観測条件の変動に対する頑健性の向上, ならびに遮蔽下における信頼性の担保が挙げられる.

参考文献

- [1] Arai, T., Isogawa, M., Sakurada, K. and Sugimoto, M.: 3D Human Pose Estimation Using Ultra-low Resolution Thermal Images, *SIGGRAPH Asia 2024* (2024).
- [2] Casas, D. and Comino-Trinidad, M.: SMPLitex: A Generative Model and Dataset for 3D Human Texture Estimation from Single Image, *Proceedings of the 34th British Machine Vision Conference* (2023).
- [3] Iwata, S., Kawanishi, Y., Deguchi, D., Ide, I., Murase, H. and Aizawa, T.: LFIR2Pose: Pose Estimation from an Extremely Low-resolution FIR image Sequence, *Proceedings of the 25th International Conference on Pattern Recognition*, pp. 2597–2603 (2021).
- [4] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G. and Black, M. J.: SMPL: A skinned multi-person linear model, *ACM Transactions on Graphics*, Vol. 34, No. 6, pp. 248:1–248:16 (2015).
- [5] Marikyan, D., Papagiannidis, S. and Alamanos, E.: A systematic review of the smart home literature, *Technological Forecasting and Social Change*, Vol. 138, pp. 139–154 (2019).
- [6] Morawski, I., Lie, W.-N., Aing, L., Chiang, J.-C. and Chen, K.-T.: Deep-Learning Technique for Risk-Based Action Prediction Using Extremely Low-Resolution Thermopile Sensor Array, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 33, No. 6, pp. 2852–2863 (2023).
- [7] Ng, A. Y., Jordan, M. I. and Weiss, Y.: On Spectral Clustering: Analysis and an Algorithm, *Advances in Neural Information Processing Systems 14* (2002).
- [8] Stathopoulos, A., Han, L. and Metaxas, D.: ScoreHMR: Score-Guided Diffusion for 3D Human Recovery, *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024).
- [9] Toriyama, C., Kawanishi, Y., Takahashi, T., Deguchi, D., Ide, I., Murase, H., Aizawa, T. and Kawade, M.: Hand-Waving Gesture Detection Using a Far-Infrared Sensor Array with Thermo-Spatial Region of Interest, *Proceedings of the 11th Joint Conference on Computer Vision and Computer Graphics Theory and Applications*, Vol. 4, pp. 545–551 (2016).
- [10] Wang, J., Zhong, Y., Li, Y., Zhang, C. and Wei, Y.: Re-Identification Supervised Texture Generation, *Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11846–11856 (2019).