赤外線グリッドセンサを用いた深層学習での人の位置推定手法の検討

戸出 悠太[1] 片山 晋[1] 浦野 健太[1] 青木 俊介[2] 米澤 拓郎[1] 河口 信夫[2,1] 泉 智博[3] 荻野 淳[3]
小山 信之[3] 太田 智浩[3]

概要：本研究では, 赤外線センサから得られた 2 次元温度情報を基にする CNN を用いた人の位置推定手法の検討を行う. 深層学習を用いることによりマニュアル設計のアルゴリズムでは取り除かれてしまう場合のある入力される温度データの潜在的な特徴を学習できるため, 推定精度の向上, 対象に取る人数の増加, 使用環境, 推定対象による差異や人以外の熱源等ノイズに対してのロバスト性の向上などが期待される. 本提案の検証として室内で取得した温度情報を基に提案手法と代表的な分類器により学習を行い, 推定精度の比較を行った. 結果, 高い粒度での位置推定においては提案手法の方が高精度であることを確認できた. 今後, 学習モデルの改良を行い位置推定対象人数の増加, 位置推定粒度の向上, 様々な場所での検証を行っていく予定である.

A Study of Human Location Estimation Method Using Infrared Grid Sensors with Deep Learning

YUTA TOIDE[1] SHIN KATAYAMA[1] KENTA URANO[1] SHUNSUKE AOKI[2] TAKURO YONEZAWA[1]
NOBUO KAWAGUCHI[2,1] TOMOHIRO IZUMI[3] JUN OGINO[3] NOBUYUKI KOYAMA[3]
TOMOHIRO OTA[3]

1. はじめに

人の位置情報・移動情報はナビゲーション, マーケティング, 災害時の対策など多様な目的で利用されており, 需要が増加している. 中でも屋内における展示会等の人の集まるイベントにおいては, 来場者の移動経路情報を需要の分析や感染症対策といった目的で利用することが可能である.

しかし, 屋内や地下空間においては, 屋外での位置推定に一般的に利用されている GPS の精度が低下するため正確な位置情報の取得ができない. そのような背景から, 屋内における位置推定手法に関する様々な研究が行われている [1][2][3]. その一つとして赤外線センサを用いた位置推定手法の研究が行われており, その中でも我々は低解像度の赤外線センサを用いた手法に注目する. 低解像度の赤外線センサを用いた位置推定では Wi-Fi や BLE ビーコンを用いた手法のように推定対象がスマートフォン等の端末を持っている必要がない. また, カメラや高解像度の赤外線センサを用いた画像による位置推定手法のようにプライバシー上の問題が発生しづらいという特徴がある. 現在までにマニュアル設計のアルゴリズムや分類器を用いた位置, 移動方向の推定手法が研究されているが, 推定粒度の低さやノイズ, 推定使用環境による差異に対してのロバスト性といった課題があった [4][5][6].

本稿では, 深層学習を用いて低解像度の熱画像を元にした位置推定手法を検討する. 入力として赤外線センサから得られた熱画像を与え, 人の位置情報を End-to-end で出力する CNN ベースの深層学習モデルを構築した. システムの概要を図 1 に示す. また, 推定する位置の粒度によって精度がどれほど異なるのかを確認するために, 出力する位置の粒度を 3 段階用意し, それぞれ学習を行い精度の比較

1 名古屋大学大学院 工学研究科
Graduate School of Engineering, Nagoya University
2 名古屋大学 未来社会創造機構
Institutes of Innovation for Future Society, Nagoya University
3 パナソニック株式会社 プロダクト解析センター
Product Analysis Center, Panasonic Corporation

© 2021 Information Processing Society of Japan

を行った．また, 提案手法が他の手法と比較したときに有用性を持っているのか調べる目的で, 同様の入出力を行う分類器を SVM, ランダムフォレスト, ロジスティック回帰を用いて構築し精度の比較を行った．

2. 関連研究

2.1 屋内位置推定手法

屋内位置推定に関する研究は現在までに数多く行われている．Bluetooth を用いた手法では位置推定の対象が持つ BLE タグやスマートフォンと多数設置された Bluetooth 端末の間で通信を行い位置の推定を行うが, Bluetooth 端末からの通信要求に対しスマートフォンが応答を返さない設定になっている等の理由により推定精度が低くなるケースも存在した [7].

浦野らの研究 [8] では推定対象のもつ BLE タグの位置を BLE スキャナを用いて推定する研究が行われた．ハードウェア面ではパケットロスを防ぐためタンデム BLE スキャナを導入する, アルゴリズム面では受信強度の低さに対処するため尤度マップ生成やパーティクルフィルタによる推定を行うなどの工夫により推定精度の向上の可能性を示した．

2.2 高解像度の赤外線カメラを用いた人の検出

Jisso Park らの研究 [9] では赤外線カメラを用いて, 夜間に危険区域に侵入する人の自動検出が試みられた．赤外線カメラから得られた熱画像に対してピクセル単位のセグメンテーションを行い, 物体の近傍を細かく予測できる CNN ベースの人物検出アプローチが開発された．この手法は従来の物体検出手法を上回る性能を示した．

2.3 Synerex

センサデータの送受信, 保存, 生データの可視化を行う上で Synerex[10] を利用したシステムを構築し利用した．システムの概要を図 2 に示す．Synerex は名古屋大学河口研究室が開発している情報交換基盤であり, 図 2 に示すシステムの中では RaspberryPi からデータ保存用サーバ, データ可視化プロバイダへの熱画像データの受け渡しをする役割を担っている．移動ロボットのデジタルツインなどにも利用されている [11].

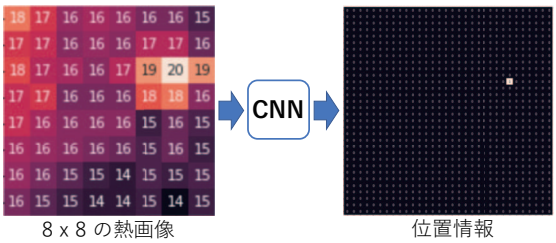

図 1 位置推定の概要

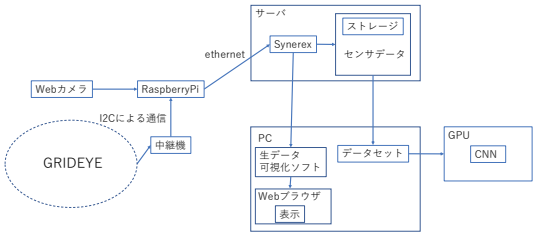

図 2 Synerex を利用したデータ収集システム全体概要

3. 提案手法

本章では学習に用いた熱画像データ, 位置データの収集方法と深層学習を用いて熱画像から人の位置を推定する手法に関する説明を行う．

3.1 データ収集に用いた機器

高い粒度で人の位置を推定するためには正確な正解位置データが必要となる．そこで, 今回の実験では正解位置データの収集にレーザー測量機 (Geodetic Total Station GT1205 | TOPCON) を用いた．GT1205 は高精度な 3 次元座標が取得可能なセンサである．外観を図 3 に示す．0.001m 単位で位置の測定が可能であり, 正確な正解位置データという目的に即していると考えられる．サンプリングレートはカタログ上は 10Hz であるが今回の実験で得られたデータは平均しておよそ 3Hz ほどであった．今回は 2 次元空間での位置推定を行うため, 得られたデータのうち 2 次元座標のみを学習に用いた．

赤外線センサには市販の赤外線グリッドセンサである GridEYE (Panasonic) を用いた．GridEYE は 8×8 の 2 次元熱画像を取得することができるセンサである．GridEYE の外観を図 5 に示す．サンプリングレートはカタログ上は 10Hz であるが今回の実験において得られたデータは平均

図 3 GT1205 の外観

図 4 実際のデータ収集の様子

© 2021 Information Processing Society of Japan

して 20Hz ほどであった．温度の分解能は 0.25 ℃, 視野角は 60° , 検知可能な範囲は 7m 以内となっている．GridEYE の検知範囲及び出力イメージを図 6 に示す．位置推定, 行動推定の分野における研究にも使用されている [6].

図 5　赤外線センサ GridEYE

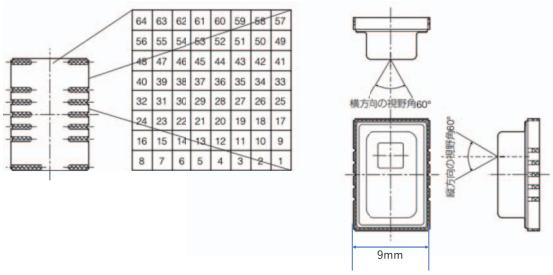

図 6　GridEYE の検知範囲及び出力イメージ

3.2　データ収集

図 7 のように高さ 2.7m の天井に垂直下向きに設置した赤外線センサの下を被験者が無作為に歩行し, その際の位置データ及び熱画像データを収集した．データ収集のイメージを図 8 に示す．赤外線センサの検知範囲内の位置推定のみを行うため, 赤外線センサの真下を中心とした 1.95m 四方の正方形エリア内のデータのみをデータセットとして使用した．ここで,GT1205 は時刻情報の記録を行わないため, データ収集段階で GridEYE との時刻同期を取る必要がある．今回は GT1205 と GridEYE 間でデータ収集の開始と終了を可能な限り同時に行うことで開始, 終了時刻の同期をとった．また, 一度の測定で多くのデータを収集すると正解位置データに対してリサンプリングを行う際に熱画像データとの時刻ずれが大きくなると考えられるためデータ収集は 3,4 分ごとに分けて行い, 計 11 回分のデータを収集した．被験者の人数は 1 名である．今回の推定では時系列情報を用いないため, データの偏りを防ぐために収集したデータセットに対して時系列方向にランダムな並べ替えを行った．60 %をトレーニング用,20 %をバリデーション用, 残りの 20 %をテスト用として提案手法モデルの学習, 評価に用いた．ここで, 提案手法と分類器でデータセットに差が出ないようにする必要がある．従って, 各分類器では提案手法モデルに使用したものと同じトレーニング用データ, テスト用データで学習, 評価を行った．実際のデータ収集の様子を図 4 に示す．

3.3　深層学習を用いた人の位置の推定

提案手法では深層学習を用いて熱画像から人の位置を推定する．入力データとして扱うのは赤外線センサを用いて観測した 8×8 の熱画像である．時刻 t における熱画像 I^t は各画素における温度を i^t_{mn} とすると式 (1) で表される．

$$\boldsymbol{I^t} = \begin{pmatrix} i^t_{11} & i^t_{12} & \cdots & i^t_{18} \\ i^t_{21} & i^t_{22} & \cdots & i^t_{28} \\ \vdots & \vdots & \ddots & \vdots \\ i^t_{81} & i^t_{82} & \cdots & i^t_{88} \end{pmatrix} \tag{1}$$

$$P^t = M \times m + n \tag{2}$$

また, 収集した 2 次元座標は図 9 に示すように指定の粒度へ変換を行っている．時刻 t における正解位置 P^t は, これを 1 次元の整数値に変換したものである．式 (2) は収集した正解位置データを M*M の座標に変換した際の座標 (n, m) の点に対する 1 次元への変換を示している．今回はセンサの検知範囲全体を 1.95m 四方の正方形として扱い, それを元に 8×8, 16×16, 32×32 の座標に変換した 3 つの正解位置データを生成し使用した．

本稿では I^t を入力として,P^t を出力とする深層学習モデルを考える．学習に用いるデータセットを作るために入力データと正解位置データの同じ時間幅におけるサンプル数

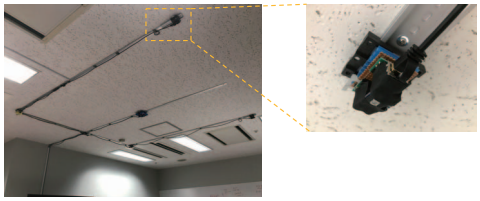

図 7　天井に設置した GridEYE

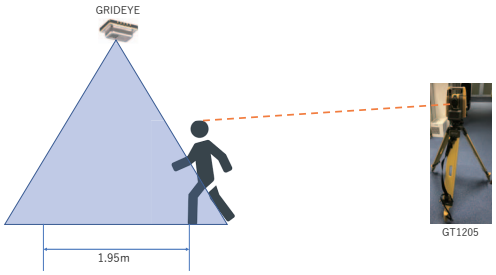

図 8　データ収集のイメージ

© 2021 Information Processing Society of Japan

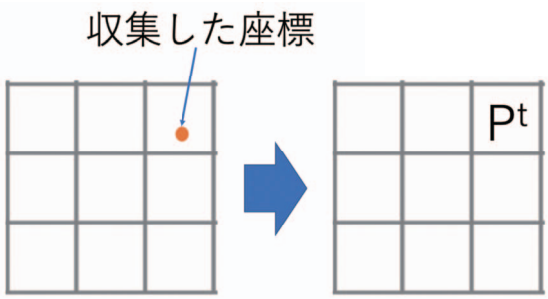

図 9　正解位置データの整形（3x3 に整形した例）

を揃える必要があった．そこで, 入力データよりサンプル数の少ない正解位置データに対して線形的にリサンプリングを行い, サンプル数を入力データと揃える処理を行った．時刻のずれが大きくなるのを防ぐため, リサンプリングを行う時間幅は数分毎となっている．リサンプリング後に入力データと正解位置データの時刻同期が出来ているのかをデータをそれぞれ座標, ヒートマップとして描画し目視により確認した．

今回使用したデータセットの総サンプル数は 61636 であった．また, その位置に人がいるデータの数をグリッドごとに示した図を図 10,11,12 に示した．

図 13 は今回使用した深層学習モデルである．推定精度を向上させる目的で入力データに対して正規化を施した後, CNN を用いて I^t の特徴量を学習し, 全結合層で P^t の出力を行った．CNN ベースのモデルを使用したのは,CNN が画像の特徴を学習するタスクにおける有用性が認められており低解像度の熱画像においてもその性能は発揮されると

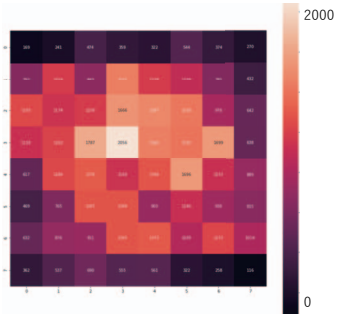

図 10　8 × 8 のグリッド毎のその位置に人がいるデータの数

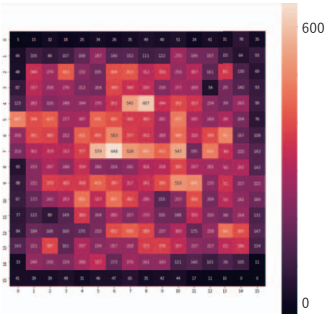

図 11　16 × 16 のグリッド毎のその位置に人がいるデータの数

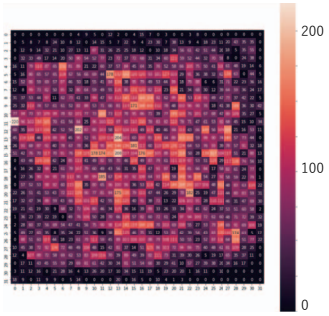

図 12　32 × 32 のグリッド毎のその位置に人がいるデータの数

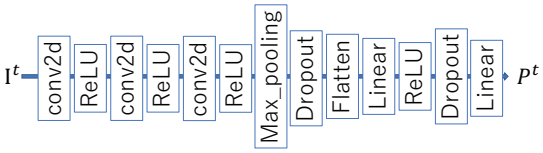

図 13　使用した深層学習モデルの構成

考えたためである．各層の数やパラメータの調整は学習後の推定精度を見て手作業で行った．提案手法モデルで使用した各層におけるパラメータは表 1 に示した．最適化法には Adam を, 損失関数には NLLLoss を用いた．学習に使った epoch 数は 8 × 8 の推定においては 500,16 × 16 の推定においては 500, 32 × 32 の推定においては 300 であった．

表 1　提案手法モデルで使用した各層におけるパラメータ

層	8 × 8 での パラメータ	16 × 16 での パラメータ	32 × 32 での パラメータ
Conv2d	入力チャネル: 1 出力チャネル: 16 kernel: 3 stride: 1 padding: 1	入力チャネル: 1 出力チャネル: 16 kernel: 3 stride: 1 padding: 1	入力チャネル: 1 出力チャネル: 16 kernel: 3 stride: 1 padding: 1
Conv2d	入力チャネル: 1 出力チャネル: 64 kernel: 3 stride: 1 padding: 1	入力チャネル: 1 出力チャネル: 64 kernel: 3 stride: 1 padding: 1	入力チャネル: 1 出力チャネル: 128 kernel: 3 stride: 1 padding: 1
Conv2d	入力チャネル: 1 出力チャネル: 128 kernel: 2 stride: 1 padding: 1	入力チャネル: 1 出力チャネル: 256 kernel: 2 stride: 1 padding: 1	入力チャネル: 1 出力チャネル: 512 kernel: 2 stride: 1 padding: 1
Dropout	0.25	0.25	0.25
Linear	出力次元数: 128	出力次元数: 512	出力次元数: 2048
Dropout	0.5	0.5	0.5
Linear	出力次元数: 64	出力次元数: 256	出力次元数: 1024

© 2021 Information Processing Society of Japan

4. 評価

本章では提案手法と代表的な分類器の精度評価を行う．提案手法で用いる深層学習モデルは CNN をベースとした構造であり, 実装は Pytorch 1.8.1 で行った．評価はテスト用データから推定した位置と正解位置データが一致した割合を求め比較を行う．また, 提案手法についてのみ推定結果が 1 グリッドずれていた場合も正解とした場合の正解率も算出し, 本来の正解率との比較を行った．

4.1 比較手法

提案手法の比較手法として, 代表的な分類器である SVM, ランダムフォレスト, ロジスティック回帰を用いた位置推定手法を用いる．提案手法モデルと同様に I^t を入力として,P^t を出力する用に学習を行った．提案手法と条件を揃えるためにトレーニング用, テスト用のデータは提案手法モデルと同じものを使用した．実装は Scikit-learn 0.24.2 を用いて行った．SVM のパラメータ調整には GridSearchCV を用いた．GridSearchCV は指定した範囲内で最適なパラメータの探索を行う．今回は,C は 0.1 から 100 の範囲,gamma は 0.01 から 0.06 の範囲, kernel は linear,rbf,poly,sigmoid の中から最適なパラメータの探索を行った．学習は提案手法と同様に入力データに対して正規化を施した後に行った．

4.2 推定精度の比較

推定精度の評価にはテスト用のデータに対する正解率を用いる．これを各手法において求め, 推定粒度ごとの比較を行う．推定精度の評価結果は表 2 に示した．表 2 より全ての粒度での推定において提案手法の推定精度が最も高くなっていることが分かった．

また, 提案手法における epoch 毎の訓練用データと評価用データに対する損失, 正解率 (%) の遷移を図 14,15,16,17,18,19 に示した．

表 2 推定精度の比較

推定粒度	CNN (提案手法)	SVM	ランダムフォレスト	ロジスティック回帰
8 × 8	**91.12 %**	91.04 %	87.05 %	52.77 %
16 × 16	**87.71 %**	87.09 %	82.29 %	37.46 %
32 × 32	**81.41 %**	80.51 %	76.17 %	38.95 %

4.3 1 グリッドの誤差を許容した場合の正解率

提案手法の出力した位置と正解位置データとの誤差が縦横 1 グリッド以内であれば正しい推定が出来たと判断する場合, 各粒度における推定精度は表 3 のようになった．

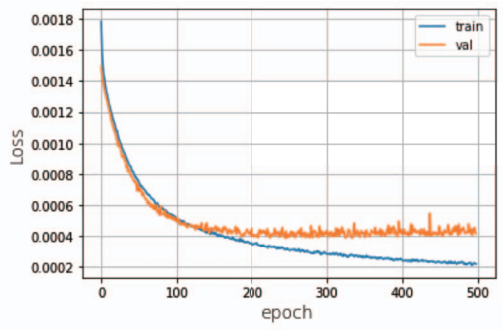

図 14 8 × 8 の推定における訓練用と検証用データの損失の推移

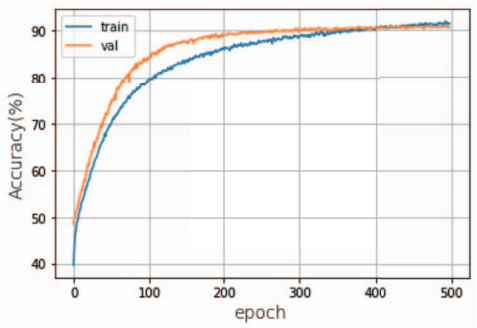

図 15 8 × 8 の推定における訓練用と検証用データの正解率の推移

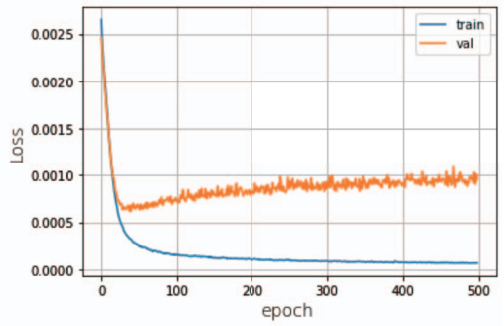

図 16 16 × 16 の推定における訓練用と検証用データの損失の推移

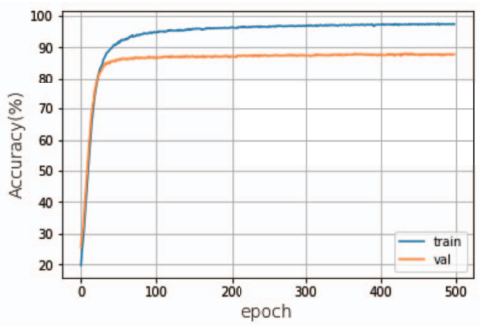

図 17 16 × 16 の推定における訓練用と検証用データの正解率の推移

4.4 考察

提案手法により学習を行った際の学習曲線について考察する．図 17,19 から, 学習の途中からトレーニング用データに対する精度は上昇しているが評価用のデータに対する

© 2021 Information Processing Society of Japan

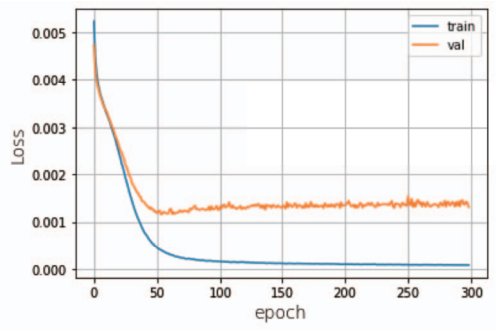

図 18 32 × 32 の推定における訓練用と検証用データの損失の推移

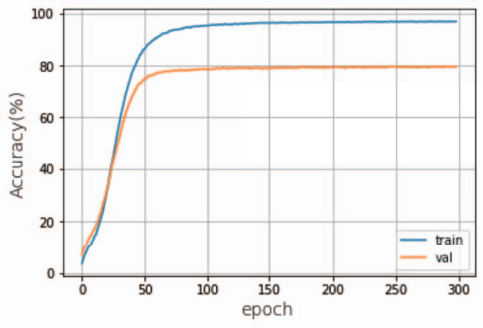

図 19 32 × 32 の推定における訓練用と検証用データの正解率の推移

表 3 1 グリッドの誤差を許容した場合の正解率

推定粒度	正解率
8 × 8	99.83 %
16 × 16	97.82 %
32 × 32	93.22 %

精度はほとんど変化していないことが読み取れる．これはトレーニング用データが持っている特徴と同じ特徴を持つデータが評価用データに存在していないために発生したと考えられ, 全体のデータ量の不足が原因と考えられる．

全ての粒度での推定において提案手法が最も高精度な推定ができた要因を考察する．提案手法が画像の特徴抽出において有効性が確認されている CNN で古典的な分類器では見つけられなかった特徴を捉えることが出来たと考えられる．

今回行った位置推定において提案手法と古典的分類器の精度にそれほど差がつかなかった要因を考察する．8 × 8 の入力データを元に 8 × 8,16 × 16,32 × 32 の位置推定を行うタスクは機械学習モデルを使って行うタスクとしてはそれほど複雑なものでは無いと考えられるため, 古典的な分類器においてもその性能を十分に発揮することができたと考えられる．また,SVM はライブラリを用いてパラメータの最適化を入念に行ったことで有利になり高い推定精度が得られたと考えられる．

推定の粒度が上がるにつれてに提案手法と他の分類器との間の推定精度差が少しずつ広がっている要因を考察する．推定の粒度が上がりタスクの複雑さが増したことで画像の特徴の抽出に優れる CNN ベースの提案手法の精度が古典的な分類器を上回ったと考えられる．

1 グリッドの誤差を許容した場合の正解率について考察する．表 3 の結果から, 間違った位置を出力したデータのうち誤差が 1 グリッド以内であるものが半数以上あることが分かる．

4.5 今後の展望

今回の研究では入力データの時間情報を用いなかったが, これを学習に使用した場合に精度が向上すると考えられる．ノイズの抑制を行う層を追加するなどネットワークにも改善の余地が大きく残されている．

データ量の不足により推定精度が落ちていると思われるため, 今後同様の実験を行う際はデータセットの充実が必要である．また, 今回学習に用いたデータセットは単一の場所, 日時, 個人から得られたものであり, 提案手法を異なるデータセットに適用した場合に汎用性にかける可能性がある．そのため, より多様な場所, 時間から取得したデータセットを構築することが課題となる．

5. おわりに

低解像度の赤外線センサを用いた人の位置推定について,CNN を用いた手法を試みた．ネットワークは熱画像データを入力として受け取り, その時刻における推定した人の位置情報を出力する．古典的な分類器で同様の学習を行った場合の推定精度との比較を行うために,CNN（提案手法）,SVM, ランダムフォレスト, ロジスティック回帰に対して同じトレーニングデータ, テストデータを用いた学習, 評価を行った．

結果, 提案手法は全ての粒度での推定において古典的な分類器より高い精度を出すことを確認した．

今後は入力データの時系列情報を用いるなどの工夫が必要となると考えられる．

謝辞

本研究の一部は、パナソニック、JST CREST（JPMJCR1882)、AMED 委託研究（20he0622042h0001）の支援を受けています

参考文献

[1] Hui Liu, Houshang Darabi, Pat Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. 37, No. 6, pp. 1067–1080, 2007.

[2] Yanying Gu, Anthony Lo, and Ignas Niemegeers. A survey of indoor positioning systems for wireless personal networks. *IEEE Communications Surveys Tutorials*, Vol. 11, No. 1, pp. 13–32, 2009.

© 2021 Information Processing Society of Japan

[3] Robert Harle. A survey of indoor inertial positioning systems for pedestrians. *IEEE Communications Surveys Tutorials*, Vol. 15, No. 3, pp. 1281–1293, 2013.
[4] 岡部亮平, 久保田稔. 人感センサデータと機械学習による移動体の動線推定. 第 82 回全国大会講演論文集, 第 2020 巻, pp. 277–278, feb 2020.
[5] 河本満, 池田剛, 幸島明男, 車谷浩一. 焦電型赤外線センサアレイを用いた一測位方法. Technical Report 6, 産業技術総合研究所, may 2010.
[6] 岡田遼太郎, 矢入郁子. 赤外線グリッドセンサを用いた室内位置・行動推定の粒度及び精度の検討. 人工知能学会全国大会論文集, Vol. JSAI2014, pp. 2J14–2J14, 2014.
[7] Abdullah Kurkcu and Kaan Ozbay. Estimating pedestrian densities, wait times, and flows with wi-fi and bluetooth sensors. *Transportation Research Record*, Vol. 2644, No. 1, pp. 72–82, 2017.
[8] 浦野健太, 廣井慧, 梶克彦, 河口信夫. 配布型 ble タグとタンデムスキャナを用いた屋内位置推定手法. 情報処理学会論文誌, Vol. 60, No. 1, pp. 58–75, 2019.
[9] Jisso Park, Jingdao Chen, Yong K. Cho, and Dae Y Kang. Cnn-based person detection using infrared images for night-time intrusion warning systems. 2019.
[10] 河口信夫, 米澤拓郎, 廣井慧. Synerex: 超スマート社会を支える需給交換プラットフォームの設計コンセプトと機能. 情報処理学会研究報告 ユビキタスコンピューティングシステム, Vol. 2020-UBI-65, No. 49, pp. 1–6, 2020.
[11] Yuto Fukushima, Yusuke Asai, Shunsuke Aoki, Takuro Yonezawa, and Nobuo Kawaguchi. Digimobot: Digital twin for human-robot collaboration in indoor environments. In *IEEE International Conference on Intelligent Vehicles (IEEE IV)*. IEEE, 2021.

© 2021 Information Processing Society of Japan