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 *

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††

Action Recognition Using a Far-Infrared Sensor Array by Convolutional RNN

Takayuki KAWASHIMA, Yasutomo KAWANISHI, Daisuke DEGUCHI, Ichiro IDE,
Hiroshi MURASE, Tomoyoshi AIZAWA and Masato KAWADE

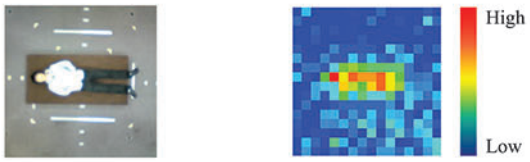
This paper proposes a Deep Learning-based action recognition method from an extremely low-resolution FIR image sequence. The method recognizes daily actions by humans (e.g. walking, sitting down, standing up, and so on) and abnormal actions (e.g. falling down) without privacy concerns. While privacy concerns can be ignored, it is difficult to compute feature points and to obtain a clear edge of the human body from an extremely low-resolution FIR image. To address these problems, this paper proposes a Deep Learning-based action recognition method whose inputs are the FIR images and their frame differences cropped by the gravity center of human regions.

Key words: far-infrared sensor array, action recognition, deep learning, extremely low-resolution image

1. はじめに

近年，高齢化社会の進展に伴い，独居高齢者の数が増加している[1]．このような状況において，高齢者への訪問介護や，健康な高齢者の身体機能の維持が重要となっている．高齢者の身体機能の低下を察知するためには，日常生活における行動（ADL：Activity of Daily Living[2]）を把握し，その変化を見つけることが重要である．ADL とは，食事や入浴，椅子とベッド間の移乗など，日常生活を営む上で不可欠な基本的行動を指す．各行動について自立度を評価することで，その人に必要な介護の程度を判断できる．ADL 遂行能力は，介護士の監督下でテスト項目を実施することで判断されるが，テスト実施日の体調の良し悪しだけで判断が下されるおそれがあり，実際には介護が必要な状況であるにもかかわらず，適切な介護を受けられない状況が起き得る[3]．この問題を解決するためには，高齢者の日常行動を常に見守り，身体機能の状態を把握する必要があるが，昼夜を問わず家族や介護士などが見守り続けることは困難であるし，大きな負担になる．また，高齢者は住宅内で転倒・転落することが多いが，独居高齢者の場合は，そのような場合に自力で助けを呼ぶことができないおそれがある．そのため，高齢者の日常生活を常に見守り，歩行や着席・起立などの日常行動や，転倒といった異常行動を自動的に認識するシステムへの関心が高まっている．

このような背景のもと，現在までに様々なセンサを用いた行動認識手法が提案されている．その中でも近年，赤外線センサアレイへの期待が高まっている．赤外線センサアレイは複数の赤外線センサを格子状に集約した非常に安価なセンサであり，



(a) Visible-light image　(b) FIR image

Fig. 1 Example of images captured by a visible-light camera and a 16 × 16 far-infrared sensor array

(a) Visible-light image　(b) FIR image

Fig. 2 Example of images captured at night-time

ある領域内の温度分布の計測が可能である[4]．人体を可視光カメラで撮影した場合の画像と，縦横 16 × 16 に配置された赤外線センサアレイで撮影して得られた 16 × 16 画素の熱画像の例を図 1 に示す．図 1 から分かるように，赤外線センサアレイで撮影した熱画像は超低解像度であるため，個人の特定が困難であり，プライバシー上の問題を回避できる．以降，本論文ではこのような画像のことを超低解像度 FIR 画像と呼ぶ．ここで，FIR は Far-InfraRed の略であり，遠赤外線を意味する．図 2 に示すように，赤外線センサアレイは暗所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．

そこで本研究では，超低解像度 FIR 画像系列から人物の日常行動（歩行，着席，起立）や異常行動（転倒）を認識することを目的とする．表 1 に本研究で認識対象とする行動の一覧を示す．本研究では，人物行動認識を行なう際に，赤外線センサアレイを天井に設置し，鉛直下向きに撮影することを想定する．

* 原稿受付　平成 30 年 5 月 7 日
掲載決定　平成 30 年 8 月 2 日
** 名古屋大学 大学院情報科学研究科（愛知県名古屋市千種区不老町）
*** 名古屋大学 大学院情報学研究科
† 名古屋大学 情報連携統括本部
†† オムロン株式会社 技術・知財本部（京都府木津川市木津川台 9-1）

Table 1 Definition of actions of recognition targets

Action	Definition
Walking	Moving in a standing posture
Sitting down	Motion from standing to sitting
Standing up	Motion from sitting to standing
Falling down	Motion from standing/sitting to lying

さらに，独居世帯の生活を想定し，センサの観測範囲内に入る人物は 1 人であると仮定する．また，本論文において，行動認識とは行動検出と行動分類という 2 つのタスクから構成されるものと定義する．前者は，入力系列から認識対象の行動が生起している区間を検出するタスクである．一方，後者は，入力系列に対して行動の種類を分類するタスクである．

行動を認識するためには，人物の姿勢やその時間的変化に着目することが重要である．可視光カメラを用いた人物行動認識では，それらの時空間的特徴を学習するために深層学習を用いた様々な手法が提案されており，高い性能を発揮している [5]．そこで，本研究では Convolution 層と Recurrent 層を組み合わせたニューラルネットワークを用いて人物行動認識を行なう．行動認識手法では，オプティカルフロー [6] を用いることで動き特徴を学習することが一般的である [5]．しかし，超低解像度 FIR 画像はノイズを多く含むため，オプティカルフローを正しく算出することは困難である．また，可視光画像系列からの行動認識では，特徴を抽出するために画像をそのまま用いることが一般的である [5]．しかし，天井から撮影された画像をそのまま特徴抽出に用いる場合，人物の動きではなく家具の配置や人物の位置を学習してしまい，学習データに含まれていない位置において生起する行動を認識できない可能性がある．

本研究では，これらの問題に対して，以下に述べる 2 つの事前処理を超低解像度 FIR 画像系列に施すことによって解決を図る．

- ネットワークへの入力として，超低解像度 FIR 画像だけでなく，フレーム間差分画像も利用する．これにより，見えの特徴だけでなく，動きの特徴も学習できるようにする．
- ネットワークへの入力として，超低解像度 FIR 画像中の人体領域を切り出した画像を利用する．これにより，室内の家具の配置や人物の位置情報に依存しない特徴を学習できるようにする．

これらの事前処理と深層学習を組み合わせることにより，超低解像度 FIR 画像系列からの高精度な人物行動認識を実現する．なお，本論文で紹介する研究は，著者らによる先行研究 [7] を発展させたものである．この先行研究は，行動認識を構成する行動検出タスクと行動分類タスクのうち，行動検出タスクには未着手であったのに対し，本論文では，行動検出タスク及び行動分類タスクの両方を対象としている．

以降，2. では，超低解像度 FIR 画像の特性について，3. では関連研究を紹介する．4. では，超低解像度 FIR 画像系列から人物行動を認識する手法を提案する．5. では，提案手法の有効性を確認するために行なった実験について報告し，6. では，その考察を述べる．最後に 7. で，本論文のまとめと今後の課題について述べる．

Table 2 Specifications of 16 × 16 far-infrared sensor array

Spatial resolution	16 × 16
Temperature range	5–50℃
Temperature resolution	0.15℃
Angle of view	(horizontal) 100° × (vertical) 98°
Size	20.0 × 37.0 × 10.7 mm

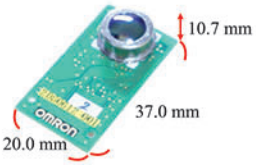


Fig. 3 16 × 16 far-infrared sensor array

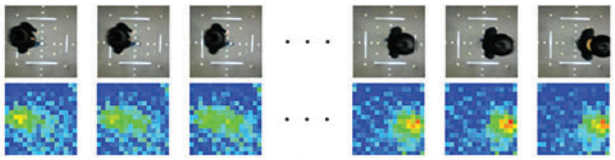


Fig. 4 Example of an extremely low-resolution FIR image sequence

2. 超低解像度 FIR 画像の特性

本研究では，オムロン株式会社製の縦横 16×16 に配置された赤外線センサアレイ（OMRON Thermal sensor D6T-1616L）を使用する．図 3 に本研究で使用する赤外線センサアレイの外観を示す．また，表 2 に，この赤外線センサアレイの仕様をまとめる．

図 4 に超低解像度 FIR 画像系列の例を示す．これは，人物が歩行する様子を赤外線センサアレイで撮影したものであり，基本的には人体表面の温度を計測したものである．超低解像度 FIR 画像には以下のような特性がある．

- 人体の正確な輪郭を算出困難である．
- 人物が動くと人体とその周辺領域の画素値が変化する．
- センサと人体との距離が変化すると画素値が変化する．
- 室温が変化すると画素値が変化する．

超低解像度 FIR 画像系列から人物行動を認識するためには，これらの特性を考慮した特徴を抽出することが必要と考えられる．

3. 関 連 研 究

3.1 赤外線センサアレイを用いた行動認識手法

赤外線センサアレイを用いた人物行動認識に関する研究としては，Toriyama ら [8] がセンサから得られる画像に対して温度と空間の絞り込みを適用することで手振り動作の認識を行なっている．この手法では，温度の絞り込みによって人体を強調し，空間の絞り込みによってノイズの影響を抑えることで認識性能の向上を図っている．しかし，認識対象としているのは事前に登録した特定の動作のみであり，一般的な行動については検討されていない．

Hevesi ら [9] は，センサ出力値の変化から人物の存否の判定と行動認識を行なっている．この手法では，センサ出力値が急激に変化した時間をとらえることで，台所における 5 種類の行

動と居間における 9 種類の行動を高精度に認識している．しかし，認識結果はセンサから得られる画像中の位置に依存しており，例えば，冷蔵庫付近の画素値の変化が大きい場合には冷蔵庫を使用していると判別する．そのため，転倒など観測範囲内の様々な場所で起こり得る（位置に依らない）行動を認識することは想定していない．

増山ら [10] は，縦横 32 × 31 に配置された赤外線センサアレイを用いた認識手法を提案している．この手法では，背景差分によって得られる人体領域の画素数の変化を用いて行動を検出し，その人体領域から抽出した特徴量を用いて歩行・着席・転倒の人物行動を認識している．また，検出された行動区間を 2 分割することで，歩行中に転倒するといった連続した行動にも対応している．しかし，背景差分の結果から行動検出を行なうため，その性能に大きく依存する．

3.2 可視光カメラを用いた行動認識手法

可視光カメラを用いた行動認識に関する研究において，深層学習を用いる手法が高い性能を発揮している [11–15]．

行動認識手法は，そのアプローチの違いにより大きく 3 つに分けられる．1 つ目は，入力系列に対して解析窓をずらしながら，各区間の行動を分類するものである [11]．このアプローチは，解析窓に対して行動分類手法を適用すればよいため，最も単純である．しかし，解析窓の幅を固定した場合，データごとの入力長の違いに柔軟に対応できないという問題がある．2 つ目は，入力系列に対して最初に行動検出を行ない，その後，検出された区間に対して行動分類を適用するものである [12) 13)]．このアプローチは，行動検出と行動分類を別の処理で行なうため，最も再現率（Recall）が高くなると言われている [14]．しかし，入力系列に対して行動検出と行動分類の 2 つの処理を施す必要があるため，計算量が多くなる．3 つ目は，行動検出と行動分類を同時に行なうものである [14) 15)]．このアプローチは，入力系列に対して 1 つの処理を行なえば良いため，計算量が少なく抑えられる．また，入力系列と正解ラベルとの対応を直接学習する end-to-end 学習をすることができる．

これらの手法では，比較的高解像度の画像を用いており，画像をそのまま直接ネットワークへ入力して学習を行なっている．しかし，前述のように天井から撮影した画像を直接ネットワークへ入力する場合，動きの特徴よりも行動の位置を学習してしまう可能性がある．この問題を解決するためには，膨大な数の学習データが必要となる．また，超低解像度画像に対して Pooling や大きなカーネルのフィルタを繰り返し適用することは困難である．そのため，従来の CNN 構造は超低解像度画像には適していないと考えられる．

4. 超低解像度 FIR 画像系列からの人物行動認識手法

2. で述べたように，超低解像度 FIR 画像にはいくつかの特性がある．本論文では，深層学習を用いることで，それらの特性を考慮した特徴を自動的に学習させる手法を提案する．その際に以下の点を考慮する．

- ネットワーク構造
  時空間的な特徴を学習するために，見えの特徴を学習するための Convolution 層と時間的依存関係を学習するための Recurrent 層を組み合わせる．また，超低解像度の入力から行動認識に有効な特徴を得るために，Pooling の回数を少なくする．

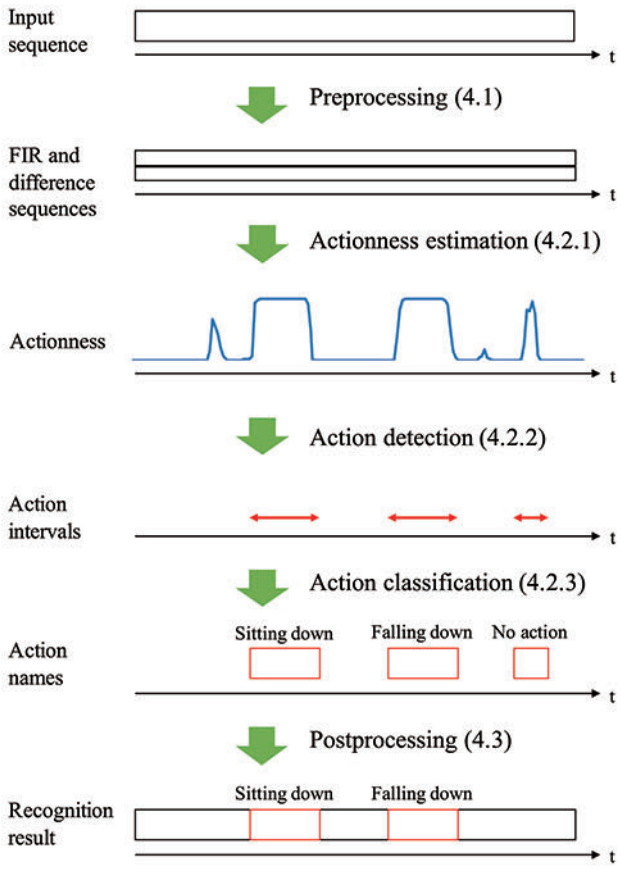


Fig. 5 Overview of the process

- ネットワークへの入力
  動きの特徴を強調するために，超低解像度 FIR 画像だけでなく，そのフレーム間差分を追加チャネルとしてネットワークへ入力する．また，位置不変な特徴を抽出するために，人体領域の重心位置を中心として画像を切り出す．

3. で述べたように，行動認識手法は大きく 3 つのアプローチに分類される．本論文では，行動検出と行動分類を別の処理で行なうアプローチにより，行動を認識する．これは，見守りシステムにおいて未検出を防ぐことが重要であると考えられるためである．そのため，Recall が最も高くなるアプローチを採用する．図 5 に提案手法の処理を示す．まず，入力系列から各フレームの行動尤度（Actionness）を推定する．そして，推定した行動尤度をもとに行動が生起している区間を検出する．これにより，行動検出結果が得られる．最後に，検出された区間に対して行動分類を適用し，最終的な行動認識結果を得る．

以下，具体的な処理手順について述べる．

4.1 超低解像度 FIR 画像系列に対する事前処理

行動認識の性能を向上させるために，ネットワークに入力する前の超低解像度 FIR 画像系列に対して事前処理を施す．

まず，入力 FIR 画像系列 $\{I_t\}_{t=0}^{T}$ に対して背景差分を適用し，2 値画像系列 $\{B_t\}_{t=0}^{T}$ を抽出する．ここで，T は系列のフレーム数（系列ごとに異なる）である．背景差分には GMM（Gaussian Mixture Model）を用いた手法 [16] を利用する．

そして，背景差分によって得られた各 2 値画像から前景領域の重心位置 $\{g_t\}_{t=0}^{T}$ を算出し，これらを各画像における人体位

置とみなす．

次に，以下のように入力 FIR 画像系列 $\{I_t\}_{t=0}^{T}$ を正規化する．

$$\widehat{I}_t(i,j) = \frac{I_t(i,j) - m}{M - m} \tag{1}$$

ここで，$I_t(i,j)$ は画像 I_t における画素 (i,j) の画素値であり，M，m は各系列における画素値の最大，最小値である．M，m はそれぞれ以下の式で表される．

$$M = \max_{t,i,j} I_t(i,j) \tag{2}$$

$$m = \min_{t,i,j} I_t(i,j) \tag{3}$$

この正規化により，室温の違いに頑健になる．そして，以下のように正規化後の系列からフレーム間差分 $\{D_t\}_{t=1}^{T}$ を求める．

$$D_t(i,j) = |\widehat{I}_t(i,j) - \widehat{I}_{t-1}(i,j)| \tag{4}$$

最後に，系列中の各画像から人体領域の重心位置を中心として R 画素四方の画像を切り出す．フレーム間差分についても同様に切り出しを行なう．ここでは，切り出された各画像に人体領域が完全に含まれるように $R = 10$ とする．画像の切り出しにより，人物の位置に依存しない特徴抽出が可能になる．これらの切り出された FIR 画像系列 $\{\widehat{I}'_t\}_{t=1}^{T}$ とフレーム間差分画像系列 $\{D'_t\}_{t=1}^{T}$ を人物行動認識の入力にする．なお，$t = 0$ のときはフレーム間差分が求められないため，$t = 1$ 以降の画像を人物行動認識の入力にする．図 6 に各系列の例を示す．なお，参考のために 1 段目に可視光画像系列を載せたが，提案手法では可視光画像は用いない．

4.2 人物行動認識

事前処理を施した FIR 画像系列 $\{\widehat{I}'_t\}_{t=1}^{T}$ とフレーム間差分画像系列 $\{D'_t\}_{t=1}^{T}$ を用いて行動認識を行なう．行動認識は行動尤度推定，行動検出，行動分類の処理からなる．以下に，各処理の内容について述べる．

4.2.1 行動尤度推定

行動尤度推定では，時刻 $t = \ell$ において認識対象の行動が生起しているか否かを識別することを目的とする．そのため，「認識対象の行動」クラスと「その他」クラスの 2 クラス分類問題として定式化する．時刻 $t = \ell$ における行動尤度推定を考え，$\{\widehat{I}'_t, D'_t\}_{t=\ell-L+1}^{\ell}$ を入力したときの各クラス（$a = 1, 2$）の応答関数を $f^a_{t=\ell}\left(\{\widehat{I}'_t, D'_t\}_{t=\ell-L+1}^{\ell}\right)$ とする．ここで，L は入力画像の枚数を表すパラメータである．このとき，各クラスの尤度 $P^a_{t=\ell}$ は以下のように求められる．

$$P^a_{t=\ell} = \frac{\exp\left(f^a_{t=\ell}\left(\{\widehat{I}'_t, D'_t\}_{t=\ell-L+1}^{\ell}\right)\right)}{\sum_{a=1}^{2}\exp\left(f^a_{t=\ell}\left(\{\widehat{I}'_t, D'_t\}_{t=\ell-L+1}^{\ell}\right)\right)} \tag{5}$$

ここで，$P^a_{t=\ell}$ は $P^a_{t=\ell} \in [0,1]$，$\sum_{a=1}^{2} P^a_{t=\ell} = 1$ を満たす．提案手法では，「認識対象の行動」クラスの尤度 $P^1_{t=\ell}$ を行動尤度推定結果として用いる．

4.2.2 行動検出

推定された行動尤度を用いて行動検出を行なう．検出には TAG（Temporal Actionness Grouping [13]）を用いる．TAG では行動尤度に対して複数のしきい値 γ_ξ を設定することにより，区間の重なりを許した複数の区間を検出できる．これにより，行動区間 $\Phi = \{\phi_n = [s_n, e_n]\}_{n=1}^{N}$ が得られる．ここで，N は検出された区間の数，s_n，e_n はそれぞれ行動の開始時刻と終了時刻を表す．

4.2.3 行動分類

検出されたすべての区間 $\Phi = \{\phi_n = [s_n, e_n]\}_{n=1}^{N}$ に対して行動分類を行なう．ここでは，行動クラスの数を K 個とする．K 個のクラスの中には「着席」や「起立」，「転倒」など認識対象の行動クラスだけでなく，ここで対象とする行動以外を行なっていることを表す「認識対象以外の行動」や，何も行動を行っていない「その他」というクラスも含める．これにより，行動検出で誤検出した区間を行動分類で除去できる．

区間 ϕ_n の行動を分類することを考える．$\{\widehat{I}'_t, D'_t\}_{t=s_n}^{e_n}$ を入力したときの各クラス（$k = 1, 2, ..., K$）の応答関数を $f^k_n\left(\{\widehat{I}'_t, D'_t\}_{t=s_n}^{e_n}\right)$ とする．このとき，各クラスの尤度 P^k_n は以下のように求められる．

$$P^k_n = \frac{\exp\left(f^k_n\left(\{\widehat{I}'_t, D'_t\}_{t=s_n}^{e_n}\right)\right)}{\sum_{k=1}^{K}\exp\left(f^k_n\left(\{\widehat{I}'_t, D'_t\}_{t=s_n}^{e_n}\right)\right)} \tag{6}$$

ここで，P^k_n は $P^k_n \in [0,1]$，$\sum_{k=1}^{K} P^k_n = 1$ を満たす．各クラスの尤度 P^k_n を用いて，以下のように行動を分類する．

$$c_n = \arg\max_k P^k_n \tag{7}$$

$$p_n = \max_k P^k_n \tag{8}$$

ここで，c_n は分類されたクラスのラベル，p_n は分類されたクラスの尤度である．これにより，行動分類結果 $\Psi = \{\psi = [\phi_n, c_n, p_n]\}_{n=1}^{N}$（行動区間，行動クラスのラベル，行動クラスの尤度）が得られる．

4.3 行動分類結果に対する事後処理

行動検出では，区間の重なりを許して複数の区間を検出する．そのため，行動分類結果 $\Psi = \{\psi = [\phi_n, c_n, p_n]\}_{n=1}^{N}$ にも区間の重なりが含まれる．そこで，行動分類で出力された行動クラスの尤度をもとにして NMS（Non-Maximum Suppression）を適用する．これにより，重複して検出された区間を最終的な行動認識結果として出力しないようにする．このとき，異なるクラスが重複した場合には，尤度が最も高いクラスのみを出力する．また，行動分類で「認識対象以外の行動」や「その他」というクラスに分類された区間も出力しないようにする．これにより，最終的な行動認識結果 $\Psi' = \{\psi' = [\phi_{n'}, c_{n'}, p_{n'}]\}_{n'=1}^{N'}$（$N' \leq N$）が得られる．

4.4 深層学習による実装

提案手法では，行動尤度推定と行動分類を深層学習で実装する．提案手法では，時空間的な特徴を学習するために，Convolution 層と Recurrent 層を組み合わせる．その際，Recurrent 層として BLSTM（Bidirectional Long Short Term Memory [17]）を用いる．BLSTM は LSTM から派生した再帰型深層学習モデルの一つである．LSTM は過去から未来への順方向の情報だけを学習するのに対して，BLSTM はそれに加えて未来から過去の逆方向の情報も学習できる．提案手法では，ネットワーク構造を 3 つの Convolution 層，2 つの Fully-connected 層，1 つの BLSTM 層とする．各 Convolution 層では，3×3 のカーネルを移動幅 1 で適用する．そして，Convolution 層 1 と 2 では Max-pooling を移動幅 2 で適用する．Convolution 層と Fully-connected 層の活性化関数には ReLU [18] を用い，BLSTM 層の活性化関数にはシグモイド関数を用いる．出力層にはソフトマックス関数を用いる．表 3 にネットワークの構成をまとめる．

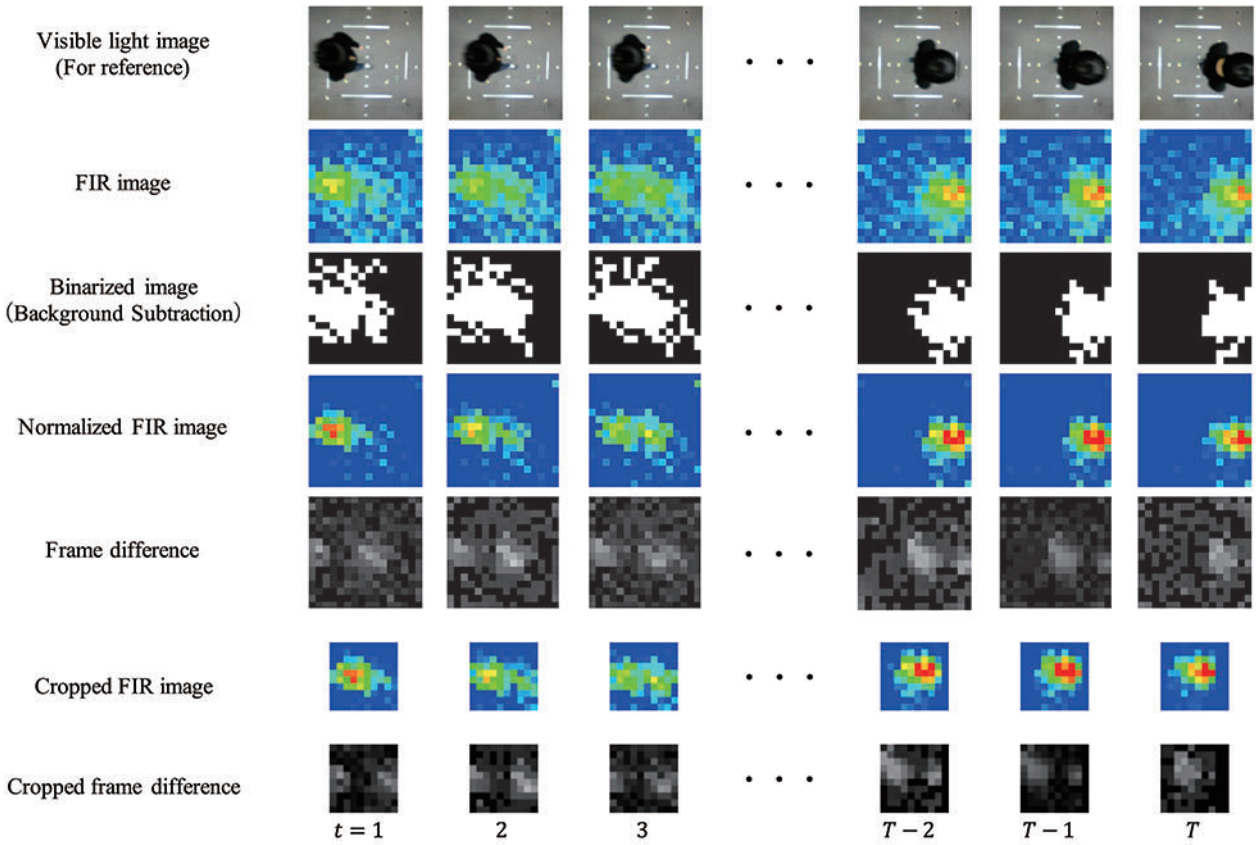


Fig. 6 Example of FIR and frame difference image sequences

Table 3 Network architecture

Input	Units: 10 × 10 × 2 (2 channels: FIR and diff.)	
Conv. 1	Kernel: 3 × 3 (Stride: 1) Channel: 32 Maxpooling: 2 × 2 (Stride: 2)	ReLU
Conv. 2	Kernel: 3 × 3 (Stride: 1) Channel: 64 Maxpooling: 2 × 2 (Stride: 2)	ReLU
Conv. 3	Kernel: 3 × 3 (Stride: 1) Channel: 128	ReLU
F.C. 4	Units: 256	ReLU
F.C. 5	Units: 64	ReLU
BLSTM 6	Units: 64	Sigmoid
Output	Units: #classes	Softmax

表 3 に示すように，ネットワークへの入力は FIR チャネルと差分チャネルの 2 チャネルである．FIR・フレーム間差分画像が各時刻で入力され，ネットワークは各時刻で各行動クラスの尤度を出力する．最終フレームまで入力した時の出力は系列全体の時間的変化を考慮できるため，最終フレームの出力を系列全体の認識結果として採用する．

行動尤度推定モデルでは出力層のユニット数を 2 とし，行動分類モデルでは K とする．この 2 つのモデルの違いは出力層のユニット数のみであり，その他の構造は同じである．

Table 4 Dataset contents

Group	1	2	3	4
Subject	A, B	C, D	E, F	G, H, I
# of sequences	100	100	100	100
# of frames	35,968	35,813	35,384	48,971

5. 実　　験

提案手法の有効性を確認するために，実際の環境で撮影した超低解像度 FIR 画像系列を用いて実験を行なった．以下，実験で使用したデータセット，実験条件，実験結果について報告する．

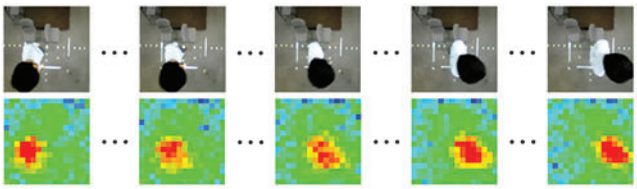
5.1 データセット

本実験で用いるデータセット[*1]は，縦横 16 × 16 に配置された赤外線センサアレイを天井に設置して収集した．センサのフレームレートは 10 fps とし，床面から 220 cm の高さに鉛直下向きに設置した．データ撮影時には，空調の温度を 19 °C に設定し，風が直接センサに当たらないようにした．

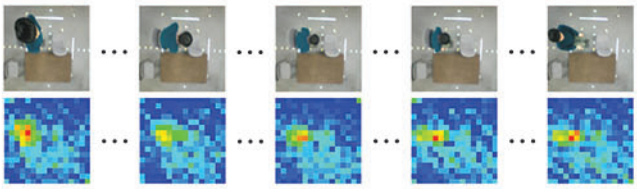
室内にはマットと椅子を設置し，人体以外の熱源は存在しない状況下でデータを撮影した．その際，マットや椅子の設置位置から行動が認識されることを防ぐため，様々な配置で撮影を行なった．

表 4 に撮影したデータの内容を示す．20 代の男女 9 名を 4 グループに分け，グループごとにデータを撮影した．空調の設定温度は 19 °C に統一したが，実際の室内温度はグループごと

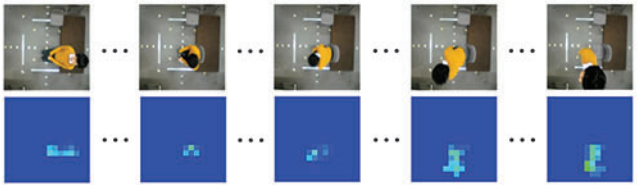
[*1] 公開予定



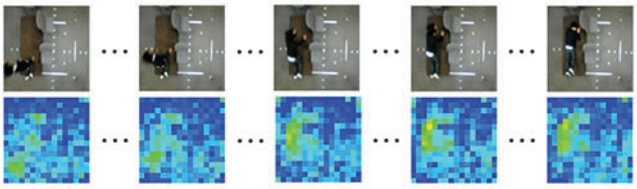
(a) Walking



(b) Sitting down



(c) Standing up



(d) Falling down

Fig. 7 Example of human actions from the dataset

に差が生じた．各被験者には，シナリオに沿って演じてもらった．各シナリオには，歩行や着席，起立，転倒といった動きがあるものや，立位，座位，臥位で静止しているものが含まれている．また，被験者には各シナリオを明所（照明を点けている状況）と暗所（照明を消した暗闇の状況）でそれぞれ演じてもらい，各グループ 100 系列，合計 400 系列のデータを撮影した．

各系列データに対して，行動開始フレームと行動終了フレームのアノテーションを人手で付与した．歩行，着席，起立，転倒は各被験者 40 本ずつ，合計 360 本のアノテーションが付与されている．静止は各被験者 120 本ずつ（立位，座位，臥位で各 40 本ずつ），合計 1,080 本のアノテーションが付与されている．データセット中の歩行，静止データの中には，アノテーションが付与されていないものが含まれているため，行動検出から分類までを含めた行動認識の評価実験では「歩行」や「静止」を評価対象の行動として含めない．行動分類単体の評価実験を行なう場合は，アノテーションが付与されたデータのみを用いることで，「歩行」や「静止」を評価対象として含めた評価を行なう．図 7 にデータセットに含まれる行動の例を，図 8 に暗所での転倒の例を示す．なお，図 7，図 8 には，超低解像度 FIR 画像に対応する可視光画像をそれぞれ載せたが，評価実験には超低解像度 FIR 画像のみを用いた．

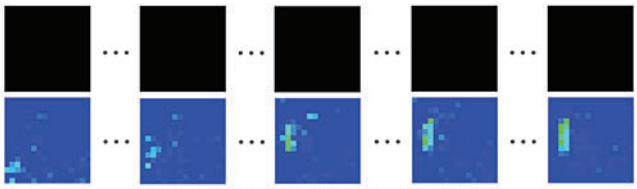


Fig. 8 Example of “Falling down” in the dark

5.2 実験条件

本実験では，「着席」，「起立」，「転倒」を認識対象の行動とした．評価は 4 分割交差検証で行なった．つまり，学習データと評価データを入れ替えて 3 グループ分のデータで学習，1 グループ分のデータで評価という試行を 4 回行ない，その平均により評価した．比較手法として，増山ら [10] の論文を参考に，我々が実装した手法を用いた．その際，背景差分手法は提案手法と同じものを用いた．

5.2.1 ネットワークの学習とパラメータ

行動分類モデルの出力層のクラス数は「歩行」，「着席」，「起立」，「転倒」，「静止」の 5 クラスとした．学習時には損失関数として交差エントロピーを用い，Fully-connected 層に対しては Dropout を 0.2 の割合で適用した．カーネル，重み，バイアス項は無作為に初期化し，最適化アルゴリズムとして Adam [19] を用いた．データセットに付与されたアノテーションをもとに「静止」，「歩行」，「着席」，「起立」，「転倒」データを人手で切り出し，学習データとして用いた．また，画像を x，y 軸方向にそれぞれ ±1 画素ずらすことでデータ拡張を行ない，600 エポックの学習を行なった．

行動尤度推定モデルの出力層は「行動（着席，起立，転倒）」と「その他（静止，歩行も含む）」とした．行動分類モデルと同様に，学習時には損失関数として交差エントロピーを用い，Fully-connected 層に対しては Dropout を 0.2 の割合で適用した．行動分類モデルで学習されたカーネル，重み，バイアス項を初期値として用い，最適化アルゴリズムとして Adam を用いた．データ拡張は行なわず，500 エポックの学習を行なった．また，行動尤度推定時の入力長は $L = 50$ とした．行動検出における TAG（Temporal Actionness Grouping）のしきい値 γ_ξ は [0,1] の範囲に対して 0.1 刻みで設定した．

5.2.2 評価指標

行動認識の性能を評価するため，評価指標として mAP（mean Average Precision）を用いた．また，クラスごとの性能を評価するために Precision と Recall も用いた．

5.3 実験結果

表 5 に tIoU のしきい値を 0.3 にしたときの各行動の Precision と Recall を示す．また，表 6 に，提案手法と，比較手法において，正解判定とする tIoU（temporal Intersection over Union）のしきい値を変えた場合の mAP の値を示す．フレーム間差分画像を入力に加え深層学習によって認識を行う提案手法の方が，それらを用いない増山らの手法に比べ，表 5，表 6 の全ての指標において高い値を示した．これらの結果から提案手法の有効性が確認された．

6. 考　　察

本章では，5. で報告した実験結果に対する考察を述べた後で，行動検出の単体性能を評価する．まず，6.1 で行動認識の

Table 5 Precision and Recall of each action when tIoU threshold is set to 0.3

	Sitting down		Standing up		Falling down		Average	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Mashiyama et al. [10]	0.110	0.600	0.225	0.203	0.125	0.644	0.153	0.482
Proposed method	**0.918**	**0.842**	**0.888**	**0.814**	**0.860**	**0.856**	**0.889**	**0.837**

Table 6 Action recognition result

Threshold of tIoU	0.3	0.5	0.7	0.9
Mashiyama et al. [10]	0.303	0.239	0.100	0.011
Proposed method	**0.839**	**0.649**	**0.199**	**0.018**

Table 7 Training condition of each actionness estimation model

	Input sequence	# of training sequences
Model A	All frames	100 (= 2 subjects)
Model B	Last 50 frames	100 (= 2 subjects)
Model C	Last 50 frames	300 (= 7 subjects)

Table 8 Action recognition result of each actionness estimation model (mAP)

Threshold of tIoU	0.3	0.5	0.7	0.9
Model A	0.679	0.566	0.271	**0.029**
Model B	0.874	0.619	0.219	0.015
Model C	**0.946**	**0.794**	**0.296**	0.004

Table 9 Action detection result

Threshold of tIoU	0.3	0.5	0.7	0.9
Precision	0.591	0.450	0.156	0.015
Recall	0.940	0.803	0.373	0.061

総合性能について考察する．次に，6.2 で行動検出の単体性能について評価する．なお，行動分類の単体性能については文献 [7] で評価した．

6.1 行動認識の総合性能評価

表 6 より，比較手法と比べて提案手法の方が高い mAP を示した．提案手法では，深層学習によって行動の認識に有効な時空間特徴を学習できたため，高精度な結果が得られたものと考えられる．なお，提案手法のネットワーク構造やネットワークへの入力の有効性については文献 [7] で述べた．

表 5 より，比較手法は Precision の値が低いため，誤認識が非常に多いことが分かる．これは，比較手法が背景差分の結果を利用して行動検出をしているためと考えられる．超低解像度 FIR 画像は，人体の輪郭が不明瞭でありノイズも多く含まれる．そのため，人物が静止している場合にも前景領域の画素数が変化し，行動が検出されてしまう．これにより，比較手法では誤検出が多く発生したものと思われる．提案手法も事前処理で背景差分を用いているが，その結果を人体の大まかな位置を取得するためだけに用いている．そのため，その性能が全体の性能に与える影響は小さく，誤認識につながらなかったものと考えられる．

6.1.1 行動尤度推定モデルのパラメータと性能

提案手法では，行動検出時の行動尤度推定と行動分類の両者で深層学習を用いる．そこで，行動尤度推定モデル学習時のパラメータを変えたときに，全体の性能に及ぼす影響を確認する実験を行なった．変化させたパラメータは，「ネットワークへの入力長」と「学習データ数」の 2 つである．実験では，変更するパラメータの組み合わせを変えた 3 つの学習条件を比較した．表 7 に各行動尤度推定モデルの学習条件を示す．Model A では，行動尤度推定時の入力長を $L = \ell$ とした．つまり，行動尤度推定の際に過去全フレームを入力とした．また，Model A と B では表 4 のグループ 1 の 100 系列を，Model C ではグループ 1，3，4 の 300 系列を学習データとして用いた．テストデータにはグループ 2 の 100 系列を用いた．なお，TAG（Temporal Actionness Grouping）のしきい値 γ_ξ の設定，行動分類モデルの学習は 5.2 と同様にした．

表 8 に各行動尤度推定モデルでの実験結果を示す．tIoU のしきい値を 0.3 にしたとき，Model A に比べて Model B が高精度であった．この結果から，行動尤度推定モデルの入力として過去全フレームを用いた場合，性能がむしろ低下することを確認した．これは，過去全フレームを入力とすることで，現在フレームには関係のない過去の情報を BLSTM 層が保持し続けたことが原因である．データセットは 10 fps で撮影されたため，50 フレームは 5 秒間に相当するデータである．したがって，少なくとも 5 秒程度過去の情報を保持していれば，現フレームで行動が生起しているか否か判定できると考えられる．今後の検討課題として，行動尤度推定モデルの入力長をさらに短くしたときの性能変化を確認することが挙げられる．

Model B と C を比較すると，tIoU のしきい値を 0.3 にしたときに Model C の方が高精度であった．したがって，学習データ数を増加させることで行動尤度推定の性能が向上し，全体の性能向上にもつながったものと考えられる．

6.2 行動検出の単体性能評価

5.1 で作成したデータセットを用いて，行動検出の単体性能評価実験を行なった．本実験では，行動分類は行なわず，行動検出の性能を単体で評価した．ここでは，行動区間が正しく検出されているか否かだけを評価し，行動クラスについては評価しない．

表 9 に実験結果を示す．全体の傾向として，Precision が低いことから誤検出が多く発生していたことが分かる．一方，Recall が高い値を示しており，未検出の行動はほとんどなかった．行動検出では未検出を減らすことが最も重要と考えられるため，提案手法は目的にかなっていると思われる．さらなる性能向上のためには，行動尤度推定モデルの改善を行ない，Recall の性能を維持したまま Precision を向上させる必要がある．

表 9 より，tIoU のしきい値を 0.3 にしたときの行動検出結果の Recall は 0.940 であった．一方，表 5 より，行動認識結果の Recall の平均は 0.837 であった．行動検出結果に比べて行動認識結果では Recall が低下していることが分かる．これは，行動検出は高精度に行なわれていたにもかかわらず，行動分類で誤分類したものと考えられる．誤分類が発生する原因としては，学習時の入力データと認識時の入力データに違いがあることが挙げられる．行動分類モデルの学習は学習データとして，アノテーションに基づいて人手で正確に切り出された系列だけを用いた．一方，認識時には行動検出で検出された区間が入力

データとなるため，数フレームのずれが発生する．このフレームのずれにより，誤分類が発生したのものと考えられる．この問題を解決するためには，行動分類モデルを学習する際に，学習データとしてアノテーションに基づいて人手で正確に切り出された系列だけを用いるのではなく，真値とは数フレームずらした系列を切り出して用いると良いと思われる．

また，表 9 より，tIoU のしきい値を 0.3 にしたときの行動検出結果の Precision は 0.591 であった．一方，表 5 より，行動認識結果の Precision の平均は 0.889 であった．行動検出結果に比べて行動認識結果では Precision が大きく向上していることが分かる．これは，行動検出では行動が生起していない区間を誤検出したが，それらの区間を行動分類で除去することができたためと考えられる．この結果から，行動分類の出力層に認識対象以外のクラスを導入することの有効性が確認された．

7. むすび

本論文では，超低解像度 FIR 画像系列からの人物行動認識手法を提案した．本研究では，深層学習に基づく手法を提案し，ネットワークへの入力として，超低解像度 FIR 画像だけでなく，フレーム間差分画像も併せて利用した．これにより，見えの特徴だけでなく，動きの特徴も学習できるようにした．また，超低解像度 FIR 画像から人体領域を切り出して入力した．これにより，室内の家具の配置や人物の位置情報に依存しない特徴が学習できるようにした．評価実験の結果から提案手法の有効性を確認した．今後の課題として，様々な環境での実験や認識対象とする行動の種類の追加が挙げられる．

謝　辞

本研究の一部は，JSPS 科研費 JP17H00745 の助成を受けたものである．

参 考 文 献

1) 内閣府: 平成 29 年版高齢社会白書, http://www8.cao.go.jp/kourei/whitepaper/w-2017/zenbun/29pdf_index.html. （2018/1/5 参照）.
2) J.M. Wiener, R.J. Hanley, R. Clark, and J.F. Van Nostrand: Measuring the activities of daily living: Comparisons across national surveys, J. Gerontology, **45**, 6, (1990) S229.
3) 上田敏: 日常生活動作を再考する―「できる ADL」,「している ADL」から「する ADL」へ―, リハビリテーション医学, **30**, 8, (1993) 539.
4) M. Ohira, Y. Koyama, F. Aita, S. Sasaki, M. Oba, T. Takahata, I. Shimoyama, and M. Kimata: Micro mirror arrays for improved sensitivity of thermopile infrared sensors, Proc. 2011 24th IEEE Int. Conf. Micro Electro Mechanical Systems, (2011), 708.
5) S. Herath, M. Harandi, and F. Porikli: Going deeper into action recognition: A survey, Image and Vision Computing, **60**, (2017) 4.
6) T. Brox, A. Bwuhn, N. Papenberg, and J. Weickert: High accuracy optical flow estimation based on a theory for warping, Proc. 8th European Conf. Computer Vision, (2004) 25.
7) T. Kawashima, Y. Kawanishi, D. Deguchi, I. Ide, H. Murase, T. Aizawa, and M. Kawade: Action recognition from extremely low-resolution thermal image sequence, Proc. 14th IEEE Int. Conf. Advanced Video and Signal based Surveillance, (2017) 28.1.
8) C. Toriyama, Y. Kawanishi, T. Takahashi, D. Deguchi, I. Ide, H. Murase, T. Aizawa, and M. Kawade: Hand waving gesture detection using a far-infrared sensor array with thermo-spatial region of interest, Proc. 11th Int. Conf. Computer Vision Theory and Applications, **4**, (2016) 545.
9) P. Hevesi, S. Wille, G. Pirkl, N. Wehn, and P. Lukowicz: Monitoring household activities and user location with a cheap, unobtrusive thermal sensor array, Proc. 2014 ACM Int. Joint Conf. Pervasive and Ubiquitous Computing, (2014) 141.
10) 増山翔太，洪志勲，大槻知明: 低解像度赤外線センサアレーを用いた非接触行動識別法の識別精度改善，電子情報通信学会技術研究報告，ASN2015-109，2016.
11) Z. Shou, D. Wang, and S.F. Chang: Temporal action localization in untrimmed videos via multi-stage CNNs, Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition, (2016) 1049.
12) S. Buch, V. Escorcia, C. Shen, B. Ghanem, and J.C. Niebles: SST: Single-stream temporal action proposals, Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition, (2017) 2911.
13) Y. Zhao, Y. Xiong, L. Wang, Z. Wu, X. Tang, and D. Lin: Temporal action detection with structured segment networks, Proc. 16th IEEE Int. Conf. Computer Vision, (2017) 2914.
14) S. Buch, V. Escorcia, B. Ghanem, L. Fei-Fei, and J.C. Niebles: End-to-end, single-stream temporal action detection in untrimmed videos, Proc. 28th British Machine Vision Conf., (2017) 92.1.
15) T. Lin, X. Zhao, and Z. Shou: Single shot temporal action detection, Proc. 25th ACM Multimedia Conf., (2017) 988.
16) P. Kaewtrakulpong and R. Bowden: An improved adaptive background mixture model for real-time tracking with shadow detection, Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, (2001) 149.
17) M. Schuster and K.K. Paliwal: Bidirectional recurrent neural networks, IEEE Trans. Signal Processing, **45**, 11, (1997) 2673.
18) V. Nair and G.E. Hinton: Rectified linear units improve restricted Boltzmann machines, Proc. 27th Int. Conf. Machine Learning, (2010) 807.
19) D.P. Kingma and J. Ba: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, (2014).