スマートスペース実現に向けた小型デバイスを用いた動作判定のための研究

代表研究者　　力　丸　彩　奈　　長野工業高等専門学校 工学科 准教授

1 研究目的と背景

人の動きが機器を操作する。空間が知能を持つことでこのようなスマートシステムが実現する。本研究の目的は，動きで機器を制御する空間「スマートスペース」の実現に向け，映像から得られる骨格情報を用いて動作の認識を行うことである。

近年の IoT 機器の普及により多くのデータが入手可能となった。それらは解析をされることで価値が見出される。しかし，監視カメラ等の映像は人による目視確認以外では有効的な活用がされていない。収集されるデータは大量になるため限られた人手ですべてを監視することは不可能である。特に，映像に映る人の動きを判断することは非常に困難である。人の動作には個体差があること，状況よって同じ動作でも姿勢が異なることなど，様々な要因から動作の指標を作ることが難しいからである。

その解析を可能にする技術として行動認識技術がある。行動認識は，カメラやセンサなどから得られるデータから人の行動の特徴を抽出し，それをもとに対象者の行動を分析する認識技術である。従来の行動認識の手法として深度カメラを用いての行動認識[1]やオプティカルフローを用いた手法[2]も提案されている。これらの手法では装置が大型であることや，素早い動きに対応できないという問題点がある。また，近年では機械学習アルゴリズムである Deep Learning が多くの分野で成果を上げており，行動認識の分野でも，そうした報告が上がっている[3-6]。本研究でも機械学習による行動認識を行う。近年では教師あり機械学習アルゴリズムを用いた行動認識も注目され，特定の行動に対して高い認識精度を実現している。教師あり機械学習では行動一つ一つに教師ラベルと呼ばれる名前を付ける必要があるが，行動の種類は多く，また，名前の付け難い曖昧な行動も存在するため，日常の行動すべてに名前を付けることは不可能に近い。そこで，本研究では教師ラベルを必要としない教師なし機械学習アルゴリズムを用いて行動の分類を行う。最終的な目的は行動種類を認識し，それを機器の遠隔操作に使用することだが，その前段階として，複数の行動が教師なし学習によって行動の種類ごとに分類可能かを検討する。

2 研究内容

2-1 自己組織化マップによる行動分類

本研究では自己組織化マップ(Self-Organizing Map: SOM)[7]による行動分類を行う。教師なし機械学習アルゴリズムである SOM は，教師あり学習では必須となる行動の名前付けを必要としないため，様々な種類の行動にも対応できると考える。ニューラルネットワークの一種である SOM は複数の入力データから類似度

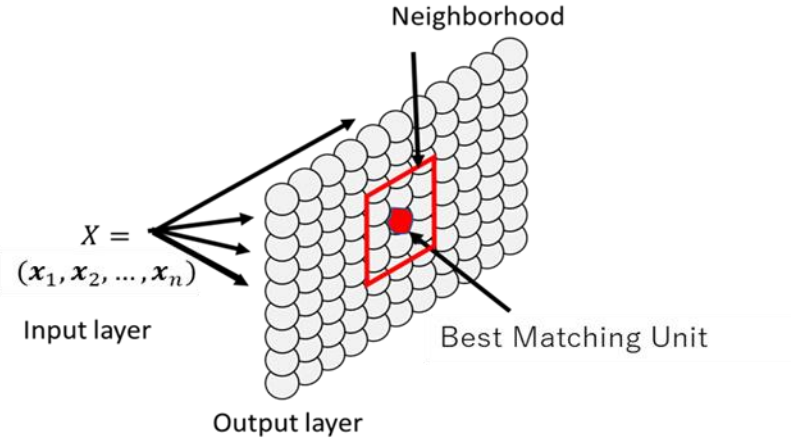

図 1 SOM のネットワーク構造

公益財団法人電気通信普及財団
研究調査助成報告書　第39号　2024年度

を算出しマップ上で表現するデータ解析手法であり，高次元データを低次元に写像することで，入力データ群に存在する傾向を視覚化することができる。ネットワーク構造は図 1 に示すように入力層と出力層を有している。 SOM の出力層はユニットの集合によって形成され，各ユニットは入力ベクトルと同じ次元の参照ベクトルを保持している。ある距離において入力に対して近い参照ベクトルを持つ勝者ユニット(BMU, Best Matching Unit)とその近傍を更新していくことで，類似の入力が近い領域に集まるように学習を行っていく。

SOM は出力層の可視化を行った上で目視による解析をする利用法が一般的であるが，ここでは行動の認識を行うために分類手法として運用する。

2-2 行動データの取得

行動データは深度カメラやセンサを介して取得する方法が多いが，この方法では大型機器の設置やセンサ装着が事前に必要となるため，設置コストの増加や対象者が限定されるなどのデメリットもある。そのため本研究では行動データとして，撮影済みの映像からでも得られる骨格情報を用いる。骨格情報はすでに撮影された映像からの取得が可能である。

本研究では数秒の行動の記録映像から得られる骨格情報を行動データとして使用する。骨格情報の取得には機械学習を用いた骨格推定手法を用いる。機械学習による骨格位置推定手法には Top-down 型と Bottom-up 型の 2 種のアプローチが存在する。単一画像に複数人が含まれている場合，Top-down 型では撮影した画像に対して人間の検出を行い，各人物ごとに骨格推定を行う。一方 Bottom-up 型では，画像に存在する骨格点すべてに対し推定を行い，得られた骨格点について各人物ごとに結合させる。Top-down 型と比較して Bottom-up 型では複数人の骨格抽出時に処理速度が落ちにくいという利点が存在することから，本論文では，Bottom-up 型に属する Cao らの人物姿勢推定手法[8]を用いて骨格の抽出を行う。

Cao らの方法では VGG-19[9]の一部を用いて特徴量の抽出を行っている。VGG-19 は物体認識に用いられる畳み込みニューラルネットワークであり，畳み込み層が 16 層，全結合層が 3 層，プーリング層が 5 層で構成されている。Cao らの人物姿勢推定手法ではまず，図 2 に示す構造の CNN により解像度を 1/8 にまで圧縮した特徴マップを生成する。図中の C は畳み込み層，P はプーリング層を表す。その後の骨格抽出部分では図 3 に示すような二つに分岐するネットワーク構造を持つ。Stage 1, Stage t(t≥2) で示されるネットワークでの学習を繰り返し，精度を向上させていく。ここで図の Branch 1 は骨格推定を行うネットワークであり，関節位置の予測結果を得る。Branch 2 は骨格結合の可能性推定を行うネットワークであり，関節間の領域の全ピクセルに対し，方向ベクトルが定義される。これらを繰り返すことで複数人の同一骨格に対して，互い違いにならないように骨格点とその結合を考慮し，マッチングに基づいた複数人の骨格点の推定を行っている。取得する骨格点の種類について，実際に取得した骨格データを画像に重ねた結果を図 4 に示す。使用したモデルは COCO2016 のデータセットに基づく学習モデルである。このモデルにおいては，目，鼻，耳，首の付け根，肩，ひじ，手首，臀部，膝，足首の計 18 個の骨格位置について x,y 座標と尤度を取得することができる。

本研究では 1 秒程度の行動の記録映像から得られる骨格情報を行動データとして使用する。まず，分類の対象とする行動映像から連続する静止画を取得し，10 フレーム分の静止画となるよう均等に間引きを行う。次に，フレームごとに骨格点を抽出し，(x,y)座標を取得する。同様に，均等に間引きした 10 フレームを用いて，隣接する 2 フレーム間の位置変化から 9 個の各骨格点の x 方向の速度，y 方向の速度を取得する。そのため，一つの行動データは 648 次元の数値データとなる

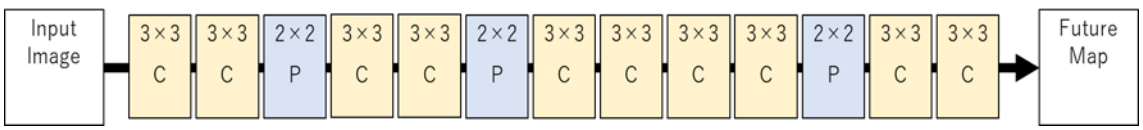

図 2 人物姿勢推定手法で用いられる VGG-19 を用いたネットワーク構造

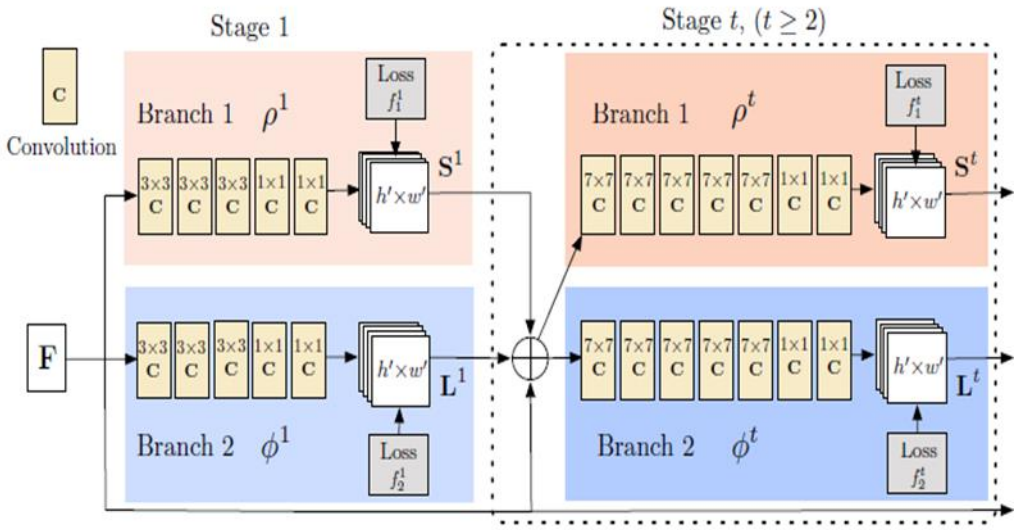

図 3 骨格抽出のネットワーク構造

図 4 映像から抽出した 18 骨格点

2-3 データ数による行動分類の比較

本研究ではまず，次の 4 つの行動についてデータを取得した。

(0)　歩く
(1)　走る
(2)　スマホ歩き
(3)　転倒

各行動，1 秒程度の動画から分類用のデータを作成した。(0)歩く では 2，3 歩程度，(1)走る は 3，4 歩程度の行動であり，(2)スマホ歩き は片手でスマートフォンを持ち，もう片方の手でそれを操作している状

況を想定した。(3)転倒は 1 歩足を踏み出した後に倒れる動作とした。

これらの行動映像から骨格情報を抽出し，SOM による分類を行い，分類後の出力マップは U-matrix 法[10]を用いて色付けした。

各行動データ 1 つを用いて分類を行った結果を図 5 に示す。マップ上の番号は行動データの番号に対応している。図 5 では転倒行動である 3 番の行動データを囲むように赤い線が現れている。この線は他のデータとの境界線を示している。線の色は青に近いほど隣接データと似ていることを示し，赤に近いほど隣接データとは異なることを示している。このことから，転倒行動は他の 3 つと大きく異なる行動であることを示している。

また，各行動データ数を 3 つに増やしての分類を行った結果を図 6 に示す。図 6 では 0～2 が行動(0)，3～5 が行動(1)，6～8 が行動(2)，9～11 が行動(3)のデータである。図 6 の左上にある 9,10,11 のデータが近くに配置されたことから，3 つのデータが似ている動きであると判断できる。また，境界線も赤く他の行動とは関連性の低いデータである事を表す結果となり，転倒は他と大きく異なる行動であることを示す結果となった。

一方で他のデータを行動の種類ごとに見ると 3，4，5 など離れた配置となった。これは転倒以外の行動が比較的似ていたために正しく分類されなかったと考えられる。データ数や行動数を増やすことで改善できる可能性がある。

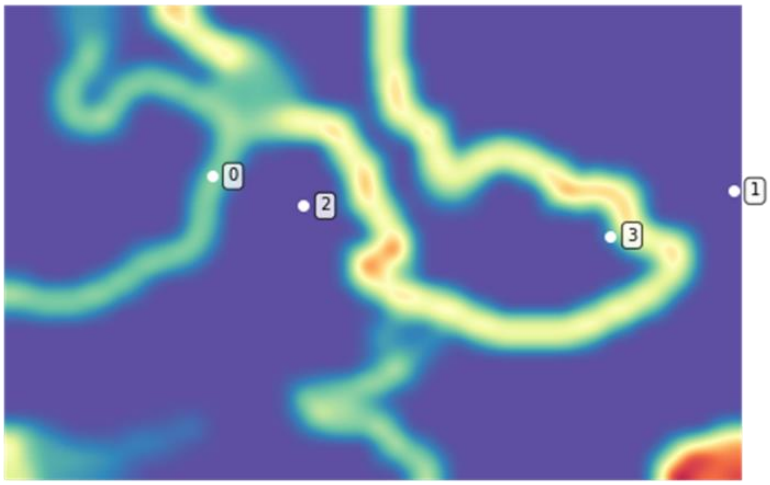

図 5 各行動データ 1 つの U-matrix 出力結果

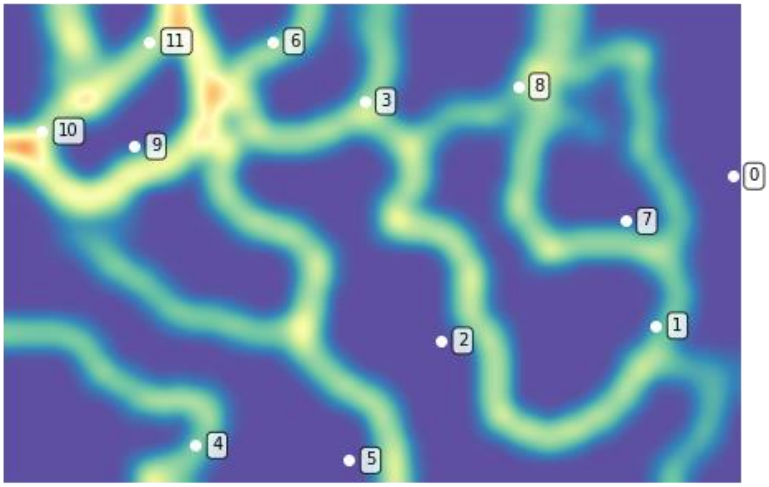

図 6 各行動データ 3 つの U-matrix 出力結果

2-3 10 種類の行動分類

次に，行動の種類を増やし，10 種類の行動について分類を行った。行動種類を表 1 に示す。1 種類の行動につき，3 回データを取得した。行動の種類とデータ番号の対応を表 1 に示す。映像取得方向は正面・背面・側面とし，「歩く」の正面はカメラに向かって歩き，背面はカメラを背にして遠ざかる行動を意味する。

作成した行動データを用いてマップの作成を行った。学習回数 10 回の分類マップを図 7，学習回数 200 回の分類マップを図 8 に示す。

学習回数 10 回の分類結果では行動種類が同じデータはマップ上の近くに配置されていることがわかる。しかし，データ番号 0〜5 付近境界線は赤に近く，他とのデータに差異があることを示している。これは「走る」を背面と正面から撮影した行動であり，他のデータと比較すると骨格の座標値の変化が速いことが原因と考えられる。しかし，15〜17 の側面方向から撮影した「走る」については付近の境界線は薄い結果となった。この理由として，データ作成時に，位置変化の影響を受けないよう各フレームにおいて常に首の骨格点が(0,0)となるよう補正したことが考えられる。

学習回数を 200 回にすると，同じ種類の行動が離れたところに配置されるデータも増加したが，境界線の強さが全体的に均一になりつつある。これは学習が進むにつれ，10 種類の行動の差異が小さくなったことを示している。

表 1 日常行動種類とデータ番号

データ番号	行動種類	映像取得方向
0，1，2	走る	背面
3，4，5	走る	正面
6，7，8	歩く	背面
9，10，11	歩く	正面
12，13，14	歩く	側面
15，16，17	走る	側面
18，19，20	椅子に座る	側面
21，22，23	椅子から立つ	側面
24，25，26	椅子に座る	正面
27，28，29	椅子から立つ	正面

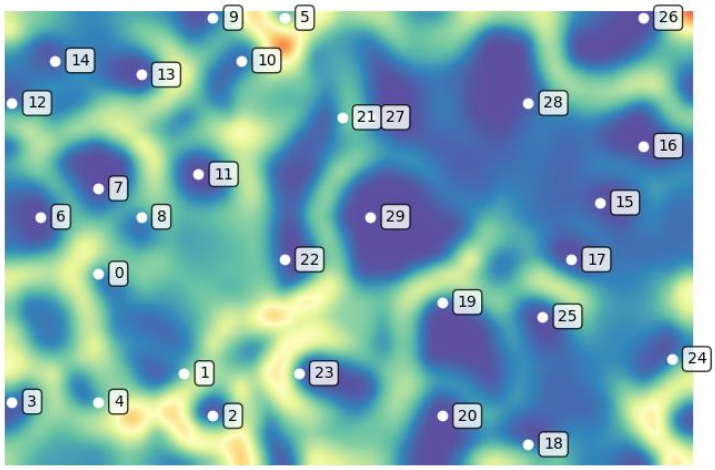

図 7　10 種類の行動の分類マップ（学習 50 回）

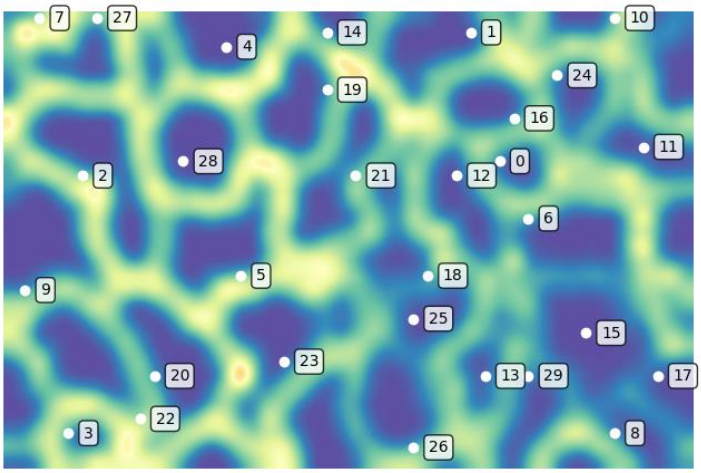

図 8　10 種類の行動の分類マップ（学習 200 回）

さらに，作成した分類マップに特徴的な行動である「転倒」を入力した。その結果を図 9，図 10 に示す。

図 9 は学習回数 50 回のマップであり，「転倒」行動である 30～32 付近の境界が強い結果となった。また，日常行動の出力結果で境界線が強く出ていた「走る」行動の境界線が弱くなっていることが分かる。これは新たに入力された「転倒」が他の行動と大きく異なっていたため「走る」の差異が弱まったと言える。また，行動種類が同じデータもマップ上での近い配置を保っている。

図 10 は学習回数 200 回のマップである。50 回の出力と比較すると「転倒」付近の境界が，より明確に表れている。しかし，他の行動種類同士の類似度は「転倒」入力前のマップと同様に低いことが分かる。

今回の実験では，日常行動と，特徴的な「転倒」行動について学習回数に依らず，分類可能な結果となった。

しかし日常行動のみの分類では学習回数による分類結果の違いも大きく，用途に合わせた適切な学習回数の設定が必要である。日常行動のみの分類マップで学習回数を増やすことで境界線が薄れたことから，検出したい行動の数が増えると，データ間の差異が弱まり検出が難しくなる可能性がある。

また，日常行動のデータについて映像取得方向の違いによって同じ行動でも境界線の強さに違いが表れた。今後は映像の取得方法や行動データの作成方法についても検討する必要がある。

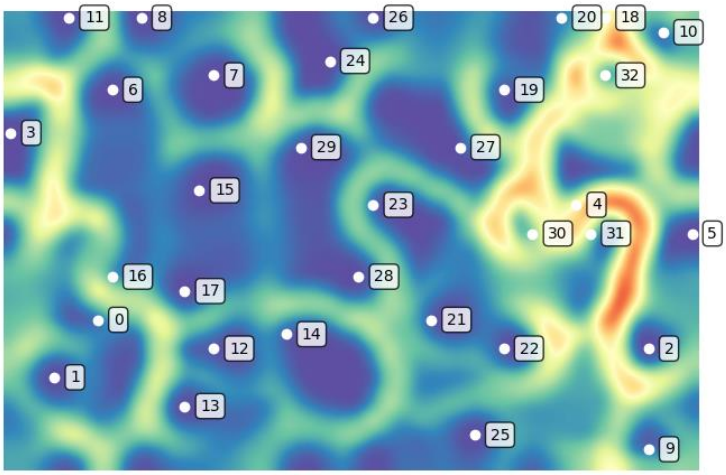

図 9 「転倒」行動入力後（学習 50 回）

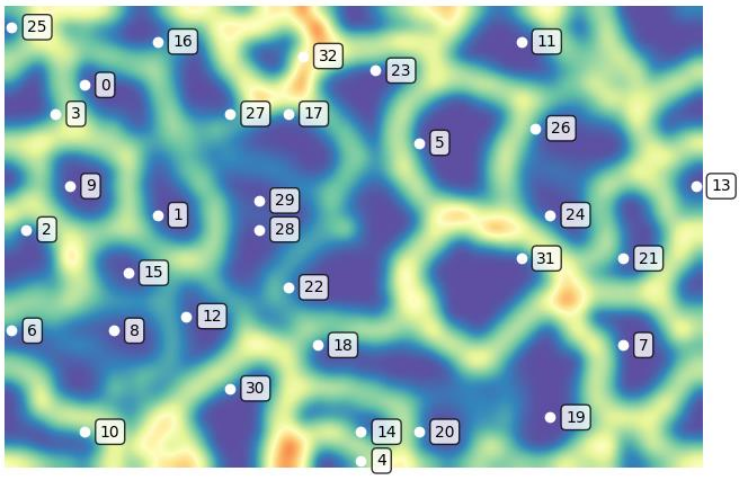

図 10 「転倒」行動入力後（学習 200 回）

3 まとめ

本研究ではスマートスペース実現に向け，行動の名前を必要としない教師なし学習の自己組織化マップを用いて行動の分類ができるか検討を行った。行動種類の分類を可視化するため，U-Matrix 法を用いて分類マップを作製した。分類マップの作成に使用する行動データは映像から得られる 18 骨格点の座標データを用いた。一つ目の実験として 4 種類の行動を対象とし，データ数の違いによる分類結果の比較を行った。各行動データ 1 つと 3 つではどちらも特徴的な行動である「転倒」が他の行動と大きく異なる行動であることを示す結果となった。

次に行動の種類を 10 種類にした実験を行った。行動の種類は走る・歩く・椅子に座る・椅子から立つ，とした。また，撮影方向を正面と側面の 2 方向からとした。歩く・走るについては背面からの撮影によるデータを取得し 10 種類の行動とした。学習回数は 50 回と 200 回として，学習回数によって結果がどのように変わるか比較を行った。また，日常行動のみによる分類マップ作成後，新規データとして「転倒」を入力し，検出ができることを確認した。しかし，学習回数による分類マップの出力差も大きく，適切な学習回数を設定する必要があることが分かった。また，日常行動のデータについて映像取得方向の違いによって同じ行動でも境界線の強さに違いが表れたことから，今後は映像の取得方法や行動データの作成方法についても検討

する必要がある。

また，小型のエッジデバイス機を用いて，映像からの骨格抽出や分類を行うことでより汎用的なシステムの実現が可能になるため，小型機への実装も今後の課題である。

【参考文献】

[1] 神園卓也, 高野茂, 馬場謙介, 村上和彰: “深度情報を含む映像からの行動認識に関する研究,” 情報処理学会研究報告, 4B-3 (2013)

[2] 玉木徹, 山村毅, 大西昇: “オプティカルフローを用いた複雑背景化における人物の腕領域の抽出と運動パラメータ推定,” 電気学会論文誌 C, Vol。120, No。12, pp。1-8 (2000)

[3] Minkin, V. A., Nikolaenko, N. N.: “Application of Vibraimage Technology and System for Analysis of Motor Activity and Study of Functional State of the Human Body,” Biomedical Engineering, Vol.42, No. 4, pp.196-200 (2008)

[4] 関 弘和, 堀 洋一: “高齢者モニタリングのためのカメラ画像を用いた異常動作検出,” 電気学会論文誌 D(産業応用部門誌), Vol. 122, No. 2, pp. 182-188 (2002)

[5] 三菱電機株式会社 情報技術総合研究所: “AI でカメラ映像から特定の動作を自動検出する「骨紋」を開発”, http://www.mitsubishielectric.co.jp/news/2019/pdf/1009.pdf, (Accessed 2024.6.12)

[6] Minh, T. L., Inoue, N. and Shinoda, K.: “A Fine-to-Coarse Convolutional Neural Network for 3D Human Action Recognition,” 29th British Machine Vision Conference (BMVC), p.227, BMVA Press (2018)

[7] T. Kohonen: ”Self-organized formation of topologically correct feature maps,” Biological Cybernetics, 1982, 43 (1) pp.59–69.

[8] Cao, Z., Simon, T., Wei, S. E. and Sheikh, Y.: “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.7291-7299 (2017)

[9] Simonyan, K. and Zisserman, A.: “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Published as a conference paper at ICLR, arXiv:1409.1556 (2015)

[10] Alfred Ultsch: “U*Matrix: a Tool to visualize Clusters in high dimensional Data,” University of Marburg, Department of Computer Science, Technical Report, Nr. 36, (2003)

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
機械学習による非日常行動検出手法の検討	映像情報メディア学会 2023 年年次大会	2023 年 8 月
教師なし学習を用いた動画からの行動分類手法の検討	映情学技報, vol. 48, no. 6,	2024 年 3 月