

在宅高齢者の外的状況における特徴収集とコンテキスト分析手法の検討 ～ 時系列画像と姿勢検出モデルを統合して ～

陳 思楠^{†,††} 中村 匡秀^{†,††,†††} 安田 清^{††}

[†] 神戸大学 数理・データサイエンスセンター 〒657-8501 神戸市灘区六甲台町 1-1

^{††} 神戸大学 大学院工学研究科 〒657-8501 神戸市灘区六甲台町 1-1

^{†††} 理化学研究所・革新知能統合研究センター 〒103-0027 東京都中央区日本橋 1-4-1

E-mail: [†]chensinan@gold.kobe-u.ac.jp, ^{††}masa-n@cs.kobe-u.ac.jp, ^{†††}yasukiyo.12@outlook.jp

あらまし 在宅高齢者の見守りにおける研究分野では、IoT センサーやケアロボットの利用が注目されてきた。我々の研究グループでは、これまでに画像ベースの事前学習済みモデルを用いた高齢者見守り手法に取り組んできた。しかし、単一画像に基づく認識 AI から得られる情報は、見守りに必要な情報としては不十分である。本研究では、時系列画像に焦点を当て、認識 AI の出力結果における文脈の前後関係を解明する新たな認識手法の開発に取り組む。本研究の成果により、時系列画像に基づく動的なコンテキスト認識が可能となり、高齢者に対するより柔軟で個別化された見守りサービスの提供が期待される。

キーワード 時系列画像, コンテキスト, 外的特徴, 姿勢検出, スマートヘルスケア

Retrieving and Analyzing External Features of Elderly Individuals at Home – Integrating Time-Series Images and Pose Detection –

Sinan CHEN^{†,††}, Masahide NAKAMURA^{†,††,†††}, and Kiyoshi YASUDA^{††}

[†] Center of Mathematical and Data Sciences, Kobe University

Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan

^{††} Graduate School of Engineering Faculty of Engineering, Kobe University

Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan

^{†††} Riken AIP 1-4-1 Nihon-bashi, Chuo-ku, Tokyo 103-0027 Japan

E-mail: [†]chensinan@gold.kobe-u.ac.jp, ^{††}masa-n@cs.kobe-u.ac.jp, ^{†††}yasukiyo.12@outlook.jp

Abstract In the field of research focused on monitoring elderly individuals at home, the use of IoT sensors and care robots has garnered attention. Our research group has previously worked on monitoring methods for the elderly using image-based pre-trained models. However, the information obtained from AI recognition based on single images has proven insufficient for monitoring needs. In this study, we focus on time-series images and embark on developing a new recognition method that elucidates the contextual relationships in the output of recognition AI. The results of this research enable dynamic context recognition based on time-series images, leading to the provision of more flexible and personalized monitoring services for the elderly.

Key words Time-series images, Contexts, External features, Pose detection, Smart healthcare

1. はじめに

世界人口の高齢化の進行に伴い、日本を代表とする超高齢化社会の中、在宅で生活する高齢者、特に認知症高齢者の増加が問題となっている [1]。在宅高齢者の安全と健康を確保するため、工学技術を活用し、高齢者の行動や生活パターンの変化、

緊急時の対応などの外的状況を適切に把握し、必要な支援を行うことが重要である。関連研究では、IoT センサやバーチャルエージェントを活用し、高齢者の「暮らし」と「こころ」の両面をセンシングする研究が行われている。本稿では、これらの外的状況を「コンテキスト」と称し、時系列画像からコンテキスト認識・分析する研究に取り組む。

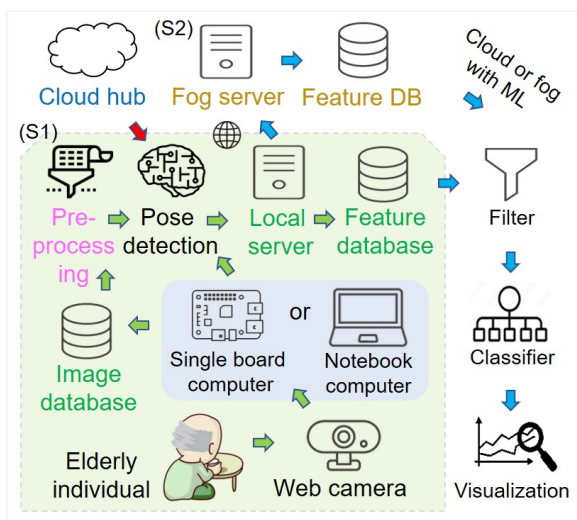


図1 提案手法の全体アーキテクチャ。

我々の研究グループでは、認識AIを活用して、環境をめぐるコンテキストと、人間をめぐるコンテキストのセンシングサービスの実現を目指している。先行研究では、単一画像に基づくコグニティブAPI (i.e., クラウドサービス) に基づき、軽量な機械学習を用いた環境コンテキスト (i.e., 活動内容) の認識技術の研究開発を行った [2]。また、単一画像とエッジ環境で実行可能な学習済みモデルを駆使して、人間コンテキスト (i.e., 細粒度な動き) の特徴付け手法を提案した [3]。

本研究では、在宅高齢者の外的状況における特徴収集とコンテキスト分析手法を検討することを目的とする。我々のキーアイデアは、時系列画像と姿勢検出モデルを統合することである。アプローチとして、(S1) 時系列画像に基づく特徴収集基盤の構築と、(S2) 連続的な変化を用いたコンテキスト分析手法の検討という2つの項目から構成される。図1に提案手法の全体アーキテクチャを示す。

2. 準備

2.1 技術的なチャレンジ

我々は画像に基づくより高性能なコンテキスト認識サービスの実現には、以下の2つのチャレンジ課題があると考えられる。

(C1) **特徴収集基盤の構築**: 宅内の環境や人間に対し、認識AIから出力された特徴の収集がチャレンジだと考える。特に一定時間間隔を持つ連続的な特徴収集基盤の構築に様々な注意点が存在する。特徴収集のプロセスでは、対象者の在宅生活に侵襲性やプライバシー影響を最小限にしたい点がある。また、導入コストやメンテナンスに対し、対象者のご負担にならない点も含む。さらに、ライブ画像のような生データから特徴を収集する場合、全てのデータを認識AIに入力することでなく、生データそのものの変化差分を予め検出する必要となる。

(C2) **連続的な変化による分析**: 収集基盤から得られた特徴を時系列上に並べ、連続的な変化による分析がチャレンジだと考える。認識結果におけるそれぞれのスコア (i.e., 正解率) によって、特徴の一部をフィルター処理する必要となる。また、特定な目的を明確し、連続的な変化の可視化からデータの密度や

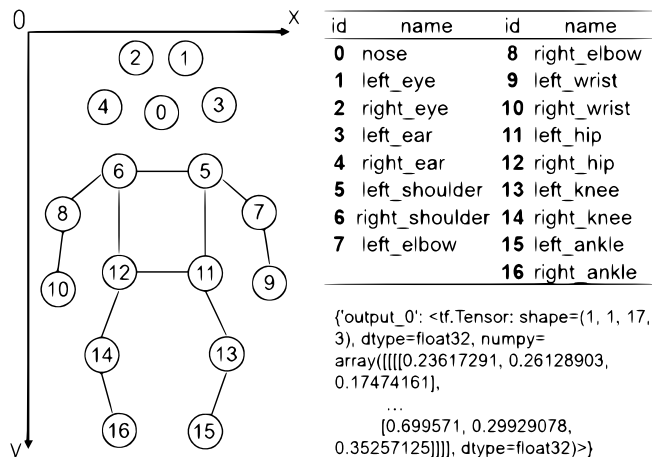


図2 姿勢検出モデル MoveNet のキーポイントの位置分布及び出力例。

粒度を調整しながら、コンテキストを分類・予測する。さらに、タイムリーな支援の兆候を検出するために、同期的な分析も大きなチャレンジとなる。動的な個人適応を実現するには、コンテキスト認識モデルのファインチューニングも不可欠である。

2.2 姿勢検出モデル MoveNet

近年、機械学習や深層学習モデルを用いて、画像や映像から人物の姿勢を推定・処理する技術が登場している。先行研究では、PoseNet という前の世代の姿勢推定モデルを利用した [4]。本稿では、精度とレイテンシのより良い MoveNet を活用して、時系列画像から特徴収集を行い、連続的な変化から在宅高齢者のコンテキスト分析に進むことを狙っている。姿勢検出モデルに MoveNet では、主な体の関節 (i.e., 17 のキーポイント) の空間的な位置をリアルタイム (i.e., 30+ FPS) より高性能かつ高速に検出することができる [5]。図2に姿勢検出モデル MoveNet のキーポイントの位置分布及び出力例を示す。

3. 提案手法

章2.1で述べたチャレンジ課題に向け、我々は以下の項目を提案し、在宅高齢者の外的状況把握を最終目的とする。

3.1 (S1) 時系列画像に基づく特徴収集基盤

Step 1: 利用機器と技術・主要な環境の選択

表1に特徴収集基盤を構築するための利用機器と技術・主要な環境を示す。一般的にWebカメラをノートブックコンピュータとつなぐ場合、Webブラウザ上でライブ画像を自動的に取得し、姿勢検出モデルを呼び出すことまでには、主にPythonまたはJavaScriptを用いたプログラミングからシステム開発を始める。また、Webカメラとシングルボードコンピュータ (i.e., Raspberry Pi or Orange Pi) とつなぐ場合、主にNode.jsを用いてプログラミングを行い、時系列画像を収集・認識するための特定かつ独立なプログラムを作ることが必要となる。

Step 2: Webカメラの規格選択と解像度設定

表2にWebカメラの選択可能な規格とオリジナルな解像度の種類を示す。時系列画像を収集する前、Webカメラの規格 (i.e., 視角の広さ) 選択と解像度の設定を実施することが不可欠である。現在で登場しているWebカメラには狭角 (60度) と一般

表 1 特徴収集基盤を構築するための利用機器と技術・主要な環境.

| Hardware | Software | Main environment |
|-----------------------|-----------------|------------------|
| Notebook computer | Web browser | JavaScript |
| Notebook computer | Web browser | Python |
| Single board computer | Special program | Node.js |

表 2 Web カメラの選択可能な規格とオリジナルな解像度の種類.

| Camera name | Angle | Original resolution type |
|-------------------------|-------|--------------------------|
| Narrow angle camera | 60 | 3840 × 2160 |
| General camera | 90 | 2040 × 1080 |
| Wide angle camera | 120 | 1920 × 1080 |
| Super-wide angle camera | 150 | 1280 × 720 |

(90 度), 広角 (120 度) および超広角 (150 度) の 4 種類がある. また, Web カメラの解像度に関し, 低解像度 (1280 × 720 ピクセル以下) と標準解像度 (1920 × 1080 ピクセル) が最も多かった. Web カメラの規格選択と解像度設定は, 画像データの受信や, 認識精度, 記録容量などの項目と緊密な関係がある.

Step 3: ライブ画像の前処理と特徴取得

ライブ画像の全てを姿勢検出モデルに入力することではなく, 取得した画像において, 前後変化差分の有無をチェックしたり, 姿勢検出の確率を向上するために画像の照度や角度を変えたりする前処理が不可欠だと考える. ライブ画像の前処理が達成したに基づき, *MoveNet* を代表とする高性能な姿勢検出モデルに入力し, 体の関節特徴点の 2 次元座標と検出確率スコアが自動的に出力される. 機械学習や深層学習の訓練済みモデルとして, 姿勢検出モデルをローカルにダウンロードし, インターネットのない場合でも, ライブ画像から特徴が検出できる.

Step 4: 特徴データの蓄積保存と可視化

姿勢検出モデルから取得された特徴データを一定周期で蓄積してから, データベースに送信して保存するという流れで進むことを考える. また, 特徴データを記録する際に, 取得日時の情報も時系列データとして不可欠である. さらに, 体の関節特徴を Web ブラウザの *Canvas* [6] や *Python OpenCV* [7] で機械から描画され, 可視化する機能もできる. 生画像のプライバシー保護などに対し, 特徴は座標の 2 次元データと検出確率スコアのみテキストファイルのため, より安全性の高い形式だと思う. 姿勢特徴の可視化によって, 遠方の家族や福祉施設・病院からプライバシー保護を遠慮なく, リアルタイムに高齢者の在宅様子を見守ることが期待できる.

3.2 (S2) 連続的变化を用いたコンテキスト分析

Step 5: 特徴データのフィルター処理

Step 4 で記録・保存された特徴データの全てを利用することではなく, 我々はまず, 各特徴データの検出確率スコアをチェックし, フィルター処理を行う必要と考える. 実際のライブ画像を取得する際には, 例えば体の一部しか画像内にない場合は,

表 3 本実験の環境設定.

| Item | Contents |
|------------------------|------------------|
| Main machine | Raspberry Pi 3B |
| Environment | Node.js |
| Camera type | Super wide angle |
| Image resolution | 320 × 240 |
| Subject info | An elderly (70s) |
| Experiment place | Living room |
| Experiment period | One day |
| Shooting time interval | One second |
| Pose detection model | MoveNet |
| Analysis | Python |
| Database | MongoDB |

検出結果自体が出力できず, あるいは検出確率が低い特徴が出力結果の中に混ざっている場合もある. また, 部屋中の照度やカメラの設置場所, 高さ, および視角など, 様々なパラメータから検討する必要だと考える.

Step 6: 特定期間におけるコンテキスト分類

Step 5 で特徴データのフィルター処理ことから, 我々は検出確率スコアのより高いデータのみを着目し, 出力時刻の差分や出力結果の差分から, 高齢者の部屋内で活動する時間帯と動きの幅を可視化することができる. これらの結果に基づき, 生活パターンの検出や不規則な生活リズム, 異常検出なども期待できる. また, 出力データの時間的な連続性がより高い期間のみをさらに注目し, 軽量の機械学習アルゴリズムにかけて, コンテキスト内容の推測や分類もできる.

4. 予備実験

4.1 実験目的と環境設定

本実験では, 時系列画像と姿勢検出モデルを統合し, 在宅高齢者の外的特徴の収集基盤の構築とコンテキスト分析手法の検討を目的とする. 表 3 に本実験の環境設定を示す. 我々はシングルボードコンピュータ *Raspberry Pi* と超広角 Web カメラをつなぎ, 70 代の在宅高齢者 1 名を対象とし, 実際のリビングルームに実験機器を設置した. 実験期間は 1 日であり, 我々が開発した *Node.js* [8] に基づく特定なプログラムから 1 秒間隔でライブ画像を 320 × 240 ピクセルの解像度で収集した. 我々は姿勢検出モデル *MoveNet* を利用し, 出力結果が存在する場合のみ画像のローカルへの記録・保存を *MongoDB* で行った. また, コンテキスト分析手法の検討には, 我々は汎用的なプログラミング言語 Python を駆使して, 特徴フィルターリングや教師なし学習のクラスタリングなどを進めた.

4.2 評価方法・尺度および結果

我々はまず, *MoveNet* から出力された 17 個の XY 座標およ



図 3 本実験で時系列画像と姿勢検出モデルを統合した連続的な出力関節の可視化結果例。

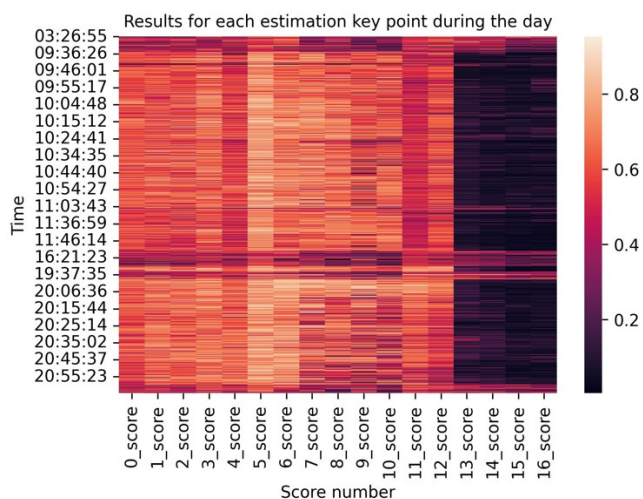


図 4 本実験で体の各キーポイントの検出確率スコアの分布。

びスコア値を Python で日時情報も含めて一つの表にまとめ、その一部に OpenCV ライブラリと連動して出力関節の可視化を行った。図 3 に本実験で時系列画像と姿勢検出モデルを統合した連続的な出力関節の可視化結果例を示す。本結果から高齢

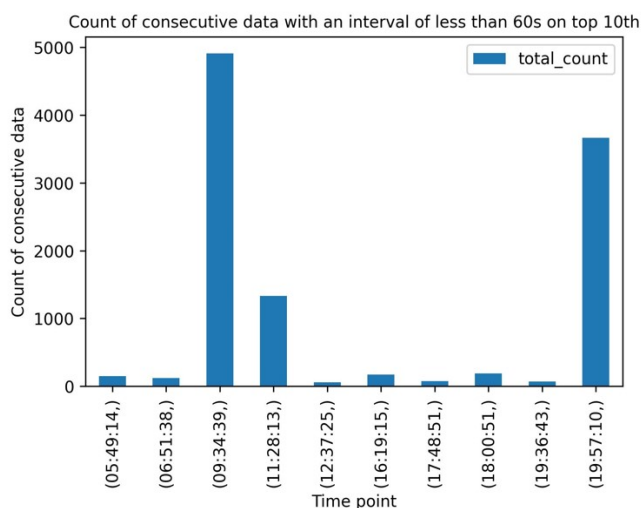


図 5 連続的かつ時間間隔 60s 内のデータ合計数の上 10 位。

者と Web カメラとの距離変化や行動・姿勢の連続性が一定程度に把握できるが、一日中の変化はデータ数が多いため、グラフ分析が必要となると考えた。

姿勢検出モデル MoveNet から出力された結果の確率を明確

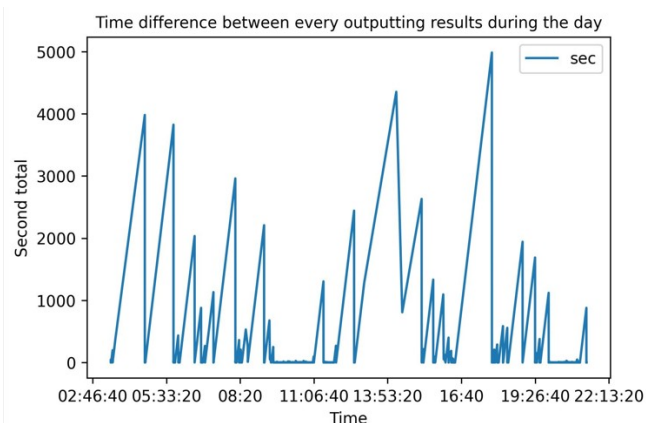


図 6 本実験における一日中に結果出力時刻の差分値の変化。

にしたいため、我々は一日中のそれぞれ特徴点のスコア結果を可視化した。図 4 に本実験で体の各キーポイントの検出確率スコアの分布を示す。本結果から特徴点 ID の 12 以降のスコアがかなり低いことがわかったため、この後の分析では特徴点 ID の 0-12 のみを着目することにした。

一方で一日中に高齢者がずっと部屋内にいるわけではなく、出力された結果の時間的な連続性も明らかにしたいため、60 秒間を閾値としてそれ以下の連続するデータの合計数を集計した。図 5 に連続的かつ時間間隔 60s 内のデータ合計数の上 10 位を示す。本結果から朝 9 時 34 分から約 2 時間内に出力結果の時間的な連続性が最も高かったことがわかった。

また、我々は姿勢検出モデル MoveNet から出力された結果ごとの出力時刻の前後の差分時間にも着目した。図 6 に本実験における一日中に結果出力時刻の差分値の変化を示す。本結果から、昼 12~15 時と午後 17~18 時の間に部屋にいない時間が最も長かったことがわかった。

以上のため、我々は時間的な連続性の高いデータの分析を優先し、午前中の出力結果から、姿勢検出モデル MoveNet から出力された 12 個特徴点の変化差分の合計も集計した。図 7 に朝の時間帯に 12 特徴点の出力座標差分の合計値の変化を示す。本結果から、朝 9 時半ごろの体の動きが最も大きく、部屋内に入ったり出かけたりする際の動き変化もわかった。

4.3 考察

本実験で得られた知見として、一日中だけでも高齢者が部屋内の行動や動作に関する不規則な瞬間が多く、コンテキストの内容を推定するにはさらなる検討が必要と考える。本研究の利点として、時系列画像と姿勢検出モデル MoveNet を非同期に統合し、これらを基盤として今後の研究に重要な役割を当てると思う。一方で、本研究の制限として、同期的な特徴検出に基づくコンテキスト分析はまだ実現していません、タイムリーな支援には更なる努力が必要と考える。

5. おわりに

本稿では、在宅高齢者の外的状況における特徴収集基盤を構築し、それに基づくコンテキスト分析手法を検討するための予備実験を行った。先行研究で活用した姿勢検出モデル PoseNet

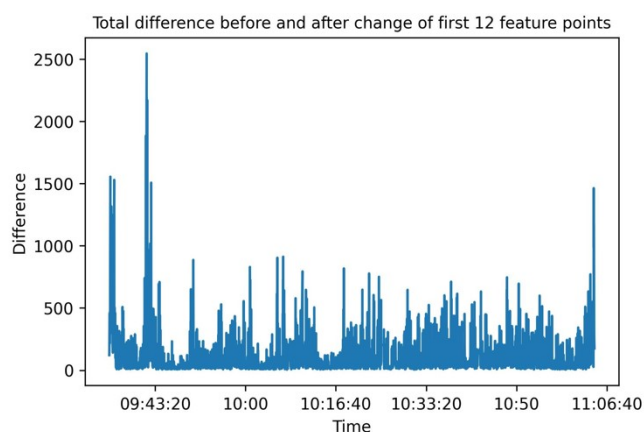


図 7 朝の時間帯に 12 特徴点の出力座標差分の合計値の変化。

より高性能なモデル MoveNet を利用し、時系列画像の連続的な変化からコンテキスト分析手法の検討に進めた。今後の課題として、同期的な特徴収集基盤の実装を行いたい。また、収集した特徴を機械学習のクラスタリングや分類アルゴリズムにかけることも検討したいと考える。

謝辞 本研究の一部は JSPS 科研費 JP20H05706, JP22H03699, JP22K19653, JP23H03401, JP23H03694, 24K02765, JP23K17006 の助成を受けて行われている。

文献

- [1] S. Hanaoka, K. Matsumoto, T. Kitazawa, S. Fujita, K. Seto, and T. Hasegawa, "Comprehensive cost of illness of dementia in japan: a time trend analysis based on japanese official statistics," *International Journal for Quality in Health Care*, vol.31, no.3, pp.231-237, 2019.
- [2] S. Chen, S. Saiki, and M. Nakamura, "Integrating multiple models using image-as-documents approach for recognizing fine-grained home contexts," *Sensors*, vol.20, no.3, p.666, 2020.
- [3] S. Chen, S. Saiki, and M. Nakamura, "Nonintrusive fine-grained home care monitoring: Characterizing quality of in-home postural changes using bone-based human sensing," *Sensors*, vol.20, no.20, p.5894, 2020.
- [4] S. Chen, M. Nakamura, and K. Yasuda, "Evaluating an in-home exercise program using vision-based edge ai for elderly healthcare," *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol.1.IEEE, pp.1017-1020 2023.
- [5] P. Kaushik, B.P. Lohani, A. Thakur, A. Gupta, A.K. Khan, and A. Kumar, "Body posture detection and comparison between openpose, movenet and posenet," *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, vol.6.IEEE, pp.234-238 2023.
- [6] M. Bajammal and A. Mesbah, "Web canvas testing through visual inference," *2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST)IEEE*, pp.193-203 2018.
- [7] S. Gollapudi and S. Gollapudi, "Opencv with python," *Learn Computer Vision Using OpenCV: With Deep Learning CNNs and RNNs*, pp.31-50, 2019.
- [8] L.P. Chitra and R. Satapathy, "Performance comparison and evaluation of node.js and traditional web server (iis)," *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAM-MAET)IEEE*, pp.1-4 2017.