

在宅高齢者の見守りのための画像に基づくエッジ AI を活用した 人間中心のコンテキスト認識手法の考察

陳 思楠[†] 中村 匡秀^{†,††} 安田 清[†]

[†] 神戸大学 〒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 日本を含む世界人口の高齢化が進行する中、介護の施設や人手不足によって従来の施設介護から在宅介護への転換が大きな傾向となっている。施設より在宅生活に慣れやすい反面、高齢者の日常生活の介護や見守りは、家族介護者に大きな負担がかかっている。先行研究では、音声対話システムに基づく言葉による高齢者の「こころ」センシングや、骨格センシング技術を活用した高齢者の宅内活動の品質付け手法等を提案している。しかしながら、高齢者の顔表情や身体姿勢・行動（コンテキストと呼ぶ）の変化等の認識は、在宅高齢者の見守りに不可欠な技術であり、まだ実現できていない。そこで本研究では、非言語的な特徴量をめぐる状況、特に人間中心のコンテキスト認識手法を考察することを目的とする。我々のキーアイデアは、画像に基づく複数の事前学習済みモデルをエッジ環境で統合し、人間中心の特徴量を抽出して、コンテキストとして性質づけることである。アプローチとして、複数の事前学習済みモデルに基づき、ローカルで実行可能な画像認識技術を統合し、ライブ画像から人間中心のコンテキスト認識を行う。本手法に従って、汎用的なパソコンと定点 USB カメラのみを利用し、一般家庭に導入しやすい在宅見守りシステムとして期待できる。

キーワード 高齢者の見守り、画像認識、エッジ AI、人間中心、コンテキスト認識

Recognizing Human-Centered Contexts for In-Home Elderly Monitoring Using Vision-Based Edge AI

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

[†] 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 As the global population ages, including Japan, there is a significant trend toward transitioning from facility-based care to home-based care due to the shortage of caregiving facilities and personnel. While elderly individuals may find it easier to adapt to living at home, the daily caregiving and monitoring of their activities significantly burden family caregivers. Previous studies have proposed methods such as “mind” sensing for elderly people based on voice dialogue systems and quality assessment of elderly individuals’ in-home activities using skeleton sensing technology. However, recognizing changes in elderly individuals’ facial expressions, body postures, and behaviors (called context) is essential for monitoring elderly individuals at home and has yet to be fully realized. Therefore, this study examines a situation-centric context recognition method revolving around nonverbal features. Our key idea is to integrate multiple pre-trained models based on images in an edge environment, extract human-centric features, and characterize them as context. As an approach, we integrate locally executable image recognition technologies based on multiple pre-trained models to perform human-centric context recognition from live images. Following this method, we can expect to create a home monitoring system that is easy to implement in ordinary households using a general-purpose computer and a stationary USB camera.

Key words Elderly monitoring, Image recognition, Edge AI, Human-centered, Context recognition

1. はじめに

世界人口の高齢化が進んでいる中、在宅高齢者の見守りに関する問題が深刻となっている。特に家族と離れて暮らしている高齢者に対し、家族からの心配や不安はいつでもあって、高齢者の在宅日常生活状況を如何に効果よく見守るかが研究課題となっている。独居高齢者や、老老介護、認認介護のような世帯に対し、24時間の見守りは家族メンバーにとって、肉体的・心理的に大きな負担となっている。このような背景において、我々は工学技術を利用して、少しだけでも可能な範囲内、高齢者の見守り内容をより細分化し、世帯ごとに事情が異なっても、高齢者の様々な状況をモニタリングできるように目指している。

そこで本稿では、在宅高齢者の様々な状況を**宅内コンテキスト**（文脈）と定義している。我々の研究グループでは、Internet of Things (IoT) 技術と物理的なセンサを組み合わせた環境（くらし）コンテキストをセンシングする研究をはじめ、近年では**人間中心のコンテキスト**認識に焦点を当たっている。心理的なコンテキストを外化するための高齢者とバーチャル・エージェントとの対話インタラクション [1] や、画像に基づく骨格センシングを活用した高齢者の活動品質付け [2] など、多様なアプローチから在宅見守り支援技術の研究開発を進めている。

本研究では、在宅高齢者の見守りのための人間中心のコンテキスト認識手法を考察することを目的とする。特に身体的なコンテキストとして、非言語的情報と言われる動作、行動、表情、視線などの特徴を画像データから捉え、高齢者の在宅日常生活における見守りデータの豊富性を向上することを目指している。我々のキーアイデアとして、*TensorFlow.js* を代表とする画像に基づく複数の事前学習済みモデルを統合し、人間中心のコンテキストに性質をつけることである。アプローチとして、定点カメラと汎用コンピュータを用い、リアルタイムに撮られた画像を、エッジ環境でモデルを呼び出して出力結果を分析する。

本稿で考察する全体フレームワークはまた実装しておらず適用性が不明であるものの、我々は高齢者の見守りに必要な技術や役に立つ可能性がある技術の利活用には価値が大きく存在すると確信している。今後の課題も含め、実用性の高い一般家庭向けの高齢者の見守り支援技術の研究開発をしつつ、コストやプライバシー、セキュリティ、豊富性、粒度、侵襲性を考慮する新たなテクニックを検証したいと考える。さらに、既存の関連研究にも色々比較を行い、本手法の不足点や制限を改善しながら、高齢者の自宅でより長い期間中に役に立つ見守りサービスとして、我々の研究進展をもう一歩進めたいと強く思っている。

2. 関連研究

本章では、在宅高齢者を対象として、「身体的な状況」と「心理的な状態」、「生理的な状況」という3つの見守りの焦点から、工学技術を用いたそれぞれの最新の関連研究を紹介する。

これまでに Internet of Things (IoT) センサや USB 定点カメラを用いて、在宅高齢者の身体的な動作や、行動、姿勢、視線、表情、顔の色など、様々な特徴量を取得・分析するような見守りのアプローチが存在する。Marin et al. [3] は空気組成に関

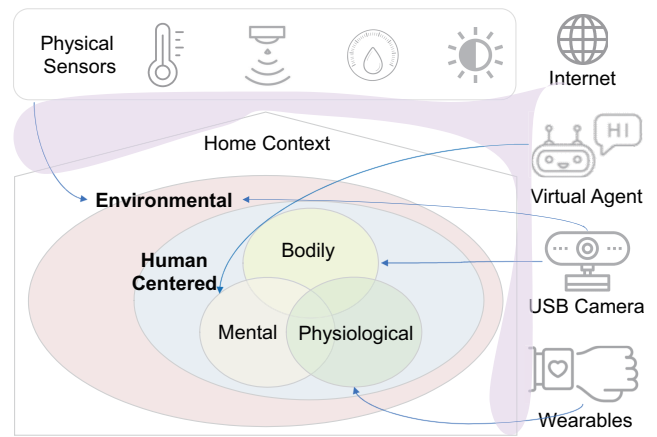


図1 工学技術を活用した宅内コンテキスト認識の分類例。

するデータを収集するガスセンサーをベースとした低コストで非侵襲的なシステムを利用し、高齢者の在宅活動を見守るシステムを提案している。また、在宅高齢者のプライバシーを保護した遠隔ケア方法として、Regress all meshes in a One-stage fashion for Multiple 3D People (ROMP) を活用して監視ビデオ内の高齢者の3D人体モデルを抽出し、k-Nearest Neighbor method (kNN) 姿勢推定アルゴリズムを利用して潜在的な異常行動を検出する提案手法の研究 [4] もある。

一方で音声対話技術や Virtual Agent (VA) 技術を活用して、在宅高齢者の考えていることや、思っていること、気づくことなどを、機械との音声会話を通してインタラクションを行い、「こころのうち」を外化するアプローチも存在する。Lima et al. [5] は会話技術 (Amazon Alexa) を活用した音声インタラクションを通し、高齢化と認知症のサポートのためのデジタル健康モニタリングにおける会話型 AI の有望性を実証する研究を行っている。また、Mild Cognitive Impairment (MCI) のある高齢者の在宅ケアに個人適応支援のためのソーシャルロボットシステムの研究 [6] もあり、適用効果を示された。

さらにウェアラブルヘルスセンサや専用機器を利用して、在宅高齢者の血圧や、ストレス、心拍数、脈拍数、睡眠の質などをリアルタイムに計測し、気づかない生理的な状況を把握するアプローチも存在する。Jachymek et al. [7] は2つの手首モニター (Fitbit Charge 4 と Xiaomi Mi Band 5) によって実行される Heart Rate (HR) 測定の精度を評価する研究を行っている。また、IoT を活用したウェアラブル健康監視システムが提案され、呼吸数、体温などの重要なパラメータを測定するためのセンサで構成される統合デバイスの研究 [8] もある。

3. 準備

本章では、本稿で考察する在宅高齢者の見守りの技術手法の前提知識として、在宅高齢者の見守りや、先行研究、技術的なチャレンジ、画像に基づくエッジ AI を展開して述べる。

3.1 状況や目的を合わせた在宅高齢者の見守り

家族と離れて自宅に住み慣れた高齢者に対し、「体調や食事が大丈夫か」、「何かあった際にどうしよう」という、家族からの心配の気持ちが増えている。在宅高齢者の心身状態をリアル

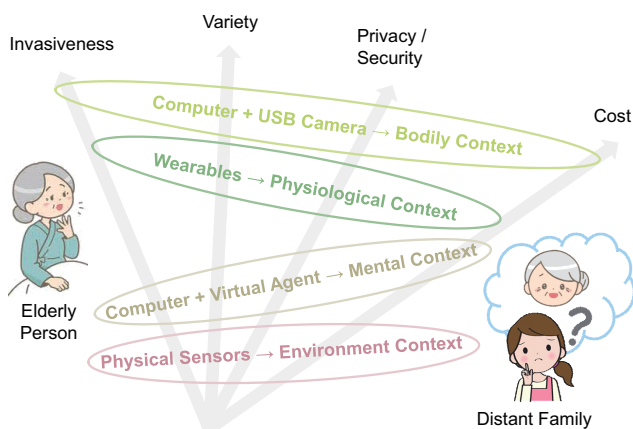


図 2 先行研究に基づく高齢者見守り技術の比較イメージの例。

タイムに把握し、家族の心配の気持ちを軽減するために、様々な見守りを考慮する必要となる。見守りに抵抗がある高齢者も少なくないが、家族の心配という気持ちを伝えることが大切であり、高齢者とよく話し合うことが不可欠である。

そこで世帯ごとにおいて、在宅高齢者のニーズや目的によって、見守りの時間軸やアプローチも異なる。例えば、「老老介護」や「認認介護」という、高齢者が互いにケアする世帯に対し、**24時間の見守りが必要か、必要なタイミングのみ**に見守りを行うかを柔軟に選別する必要となる。また、見守りの焦点も考える必要となる。在宅生活において、身体・動作行動の変化や心理的な変化、生理的な変化などが含まれている。

3.2 宅内コンテキスト認識に関わる先行研究

本稿では、我々は在宅高齢者の見守りの焦点となる様々な変化の状況を **宅内コンテキスト**と呼ぶ。図 1 に工学技術を活用した宅内コンテキストの分類例を示す。我々が考える宅内コンテキストの中には、環境（くらし）のコンテキストがメインとして存在する。環境コンテキストの中には、人間中心のコンテキストがあり、その中に身体的、心理的、生理的という3種類のコンテキストが含まれている。つまり、我々は人間中心のコンテキストが環境コンテキストの一部と見なし、そこで無人の場合にも人間中心のコンテキストの1種とされている。

我々の先行研究は**環境のコンテキスト認識**をはじめ、IoT技術と複数の**物理的なセンサ**部品と組み合わせ、温度、人感など7種類を含む環境センシングボックスを開発している。我々は実装した環境センシングボックスを、高齢者の自宅に置くことで、環境データから高齢者の生活習慣やリズムを可視化・分析する研究も行っている[9]。また、生活習慣の分析に関連する他の研究[10]も存在する。さらに、**USB 定点カメラ**で撮影されたライブ画像を、クラウド・コグニティブ Application Programming Interface (API) に入力し、画像内に存在するオブジェクトやシーンのコンセプトを単語タグの集合で出力される。それらを軽量な教師あり機械学習にかけて、環境コンテキストの認識モデルを構築・評価する手法も提案している[11]。

一方で我々の近年の先行研究では、主に**人間中心のコンテキスト認識**に関する研究を絞っている。代表的な成果として、高齢者とバーチャル・エージェントとのインタラクションを通し

て、「心のうち」を言葉で外化するところセンシングサービス[1]や、高齢者向けの音声で制御可能なマイクロサービス[12]（例、Web 検索、動画視聴、カレンダーなど）を提案している。また、我々は USB 定点カメラで撮られたライブ画像を骨格センシング技術で特徴抽出を行い、高齢者の身体活動の品質付け[2]や位置ロケーションセンシング手法[13]も開発している。さらに、**ウェアラブルヘルスデバイス**とVAとの連携[14]や、マルチモーダル統合分析[15]の研究も進んでいる。

3.3 高齢者見守りのための技術的なチャレンジ

本稿では、我々は在宅高齢者向けの見守り技術のチャレンジポイントに関し、主に次の4つを考えている。図2に先行研究に基づく高齢者見守りアプローチの比較イメージの例を示す。

まず、1番目のチャレンジポイントは、利用機器から日常生活への**侵襲性**である。例えば、USBカメラや、身体装着型デバイスの利用に対し、抵抗がある高齢者は少なくないが、利用目的や今後期待できる効果などを、高齢者へ納得できることまで事前に説明する必要である。また、在宅高齢者とバーチャル・エージェントとの対話に関し、高齢者はアプローチに慣れる時間がかかる場合もある。さらに、対話シナリオの設計に関わる話題が個人の嗜好に合うかを、アンケートやインタビューで予め調査することもチャレンジである。

次のチャレンジポイントは、特徴データの**粒度と豊富性**である。例えば、対話ログとするテキストデータでは、高齢者が対話するタイミングのみに記録され、連続性のないデータや欠損するデータが多く存在する場合がある。それに対し、ウェアラブルヘルスデバイスを装着する場合、デバイスの電気が切れるまでの1~2週間間に、装着し忘れることが無ければ、生理的なデータを時系列上でリアルタイムに取得することが可能となる。また、物理的なセンサを通して収集されたデータの粒度は良いかもしれないが、それに基づく人間の状況への推定は難しく、データの豊富性の不足が存在すると考える。

また、**プライバシーとセキュリティ**保護の必要性も不可欠なチャレンジポイントである。特に生画像をクラウド送信・処理する場合、高齢者の顔画像や宅内の私物などのプライバシー損害の恐れが存在する。そこで多くのクラウドサービスでは、データ処理・認識の手法やプロセスが不透明であり、ブラックボックスとなる場合がある。また、ウェアラブルヘルスデバイスを利用して取得したデータも個人情報を含むデータとして、2時利用時にセキュリティ保護が必要となる。さらに、普段に高齢者がバーチャル・エージェントとの会話内容の中に、プライバシーに関わる部分が存在するかという点に対し、対話ログの保存・蓄積場所を検討することも不可欠である。

最後に、一般家庭への**費用負担と導入難しさ**も考慮しなければならない。例えば、汎用的なパソコンの購入が負担可能となっても、USBカメラの導入に納得できない高齢者もいる。また、大量の物理的なセンサ部品を宅内に導入する際に、メンテナンスが難しくなる場合もある。つまり、負担や導入の難しさは、初期にしか存在しないものではかく、今後の長い間での問題点への対応や対策を、全面的に考慮する必要となる。さらに、世帯ごとのニーズや事情が異なり、高齢者を如何に効

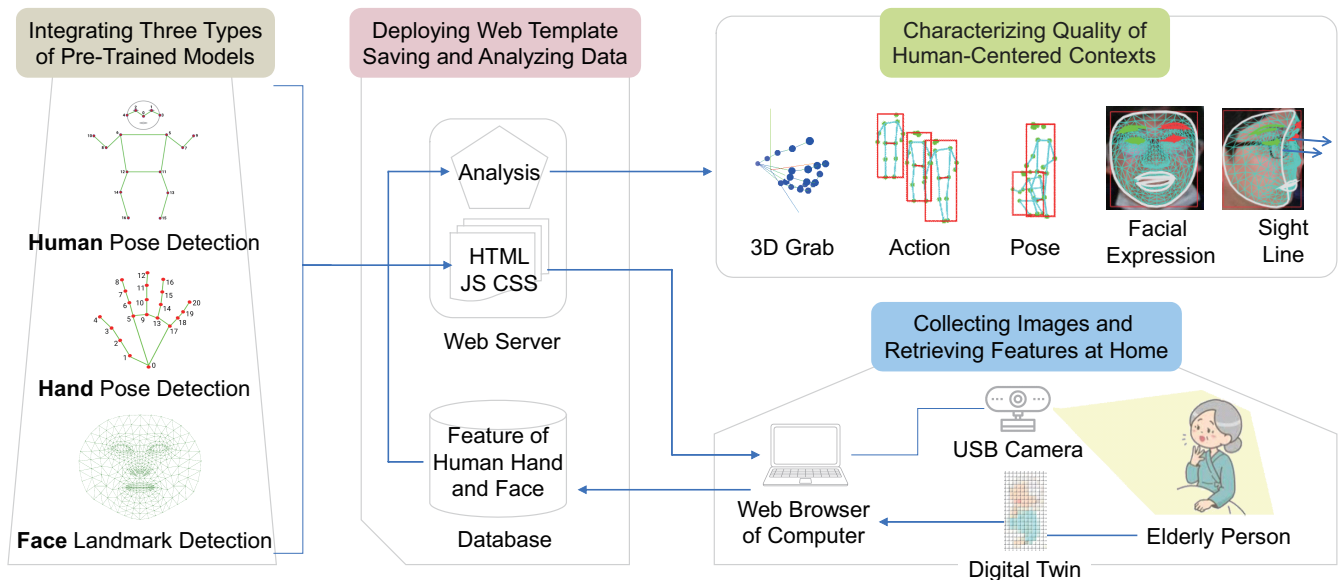


図3 本稿で考察する人間中心のコンテキスト認識手法の全体フレームワーク。

率よく負担よく見守るかに関し、サービス導入者は家族メンバーからのご意見も聞くことが重要である。

4. 宅内における人間中心のコンテキスト認識

本章では、在宅高齢者の見守りのための人間中心のコンテキスト手法を考察する。また、本稿で目指す手法に必要なコンポーネントやモジュールを整理し、以下に具体的に述べる。

4.1 目的とキーアイデア

本研究の目的は、在宅高齢者の顔表情や身体姿勢・行動などの非言語的な特徴量をめぐる状況、特に人間中心のコンテキスト認識手法を考察することである。我々は画像に基づく複数の事前学習済みモデルを統合し、エッジ環境で特徴量抽出とコンテキスト性質付けることをキーアイデアとする。近年、TensorFlow.jsを代表とする事前学習済みモデルが登場している。画像や音声、テキストなど、マルチメディアデータに対し、Webブラウザまたは汎用的なアプリケーション開発言語Node.jsで簡単に必要なモデルをローカルに導入できる。また、導入したモデルの一部では、テスト試行や特定なラベリングデータを用い、再学習することも可能となる。

4.2 S1: 複数の学習済みモデルのエッジでの統合

図3に本稿で考察する人間中心のコンテキスト認識手法の全体フレームワークを示す。我々は主に3種類の事前学習済みモデルのTensorFlow.jsモデルに注目している^(注1)：(1) 人間姿勢検出。(2) 手姿勢検出。(3) 顔ランドマーク検出。人間姿勢検出では、主に3つのモデル(即ち、MoveNet, BlazePose, PoseNet)が含まれ、WebブラウザまたはNode.jsで入力画像から身体特徴点の検出が実行可能となる。一方で、手姿勢検出も、顔ランドマーク検出も人間姿勢検出と実行方法が同じである。本稿では、我々は主に、前文の(1)~(3)のモデル呼び出しのためのJavaScriptコードを統合し、HTMLとCSSと含んでWeb

サーバへデプロイする。Webサーバでは、Webページのコードのほか、データ分析に必要なルーティングも用意する。さらに、Webサーバ稼働の同時にデータベース構築も行う。

4.3 S2: 人間中心のコンテキストへの性質付け

高齢者の自宅では、我々は汎用的なコンピュータとUSBカメラと繋ぎ、Webブラウザから複数の学習済みモデルが稼働可能なページへ読み込む。カメラの視点や視角の制限は存在かもしれないが、撮影できる範囲内の人間の関連特徴量を、データベースに自動的に保存・分析を行う。生のライブ画像は保存する必要はユーザに聞いていく。また、人間中心のコンテキストを性質に付けるためには、我々は次の5つのパラメータを考えている：(I) 手姿勢検出モデルを用いた3D手姿勢データの分析である。特に高齢者の手は物を持っているかを画像の特徴量から検討してみる。次に(II)行動と(III)姿勢のコンテキスト認識を人間姿勢検出技術を行っている。さらに(IV)表情と(V)視線に関し、顔ランドマーク検出技術を活用して、特徴量の変化を捉まえて分析する。

4.4 考察

本稿で考察した手法の利点として、画像処理の全てのプロセスはエッジ環境で行い、高齢者のプライバシーやセキュリティ保護に繋がられる。また、本研究で着目した3つのモデルとも24fpsの粒度で連続に特徴抽出可能となり、高齢者の身体、手、および顔のコンテキスト変化を時系列上で同時に取得することが期待でき、より豊富性の高いコンテキスト認識手法の実現が有望だと考える。しかしながら、本手法の制限にも次の通りが存在している。まずは、各モデルの精度の評価結果は不完全となり、部屋のレイアウトの異なる全ての世帯で適用できるとは言えない。次に、本稿では主に、1つの定点USBカメラの設置を考え、たとえ部屋の明るさや、撮影範囲外の場合では、モデルの認識結果が欠損することも想定されている。また、本稿で考察するフレームワークは、技術の実装や実際の高齢者の自宅での臨床実験はまだ行っていない。既存技術は家族のニーズと

(注1) : <https://github.com/tensorflow/tfjs-models>

如何にうまくつなぐかはまた様々な検証が不可欠となる。

5. ま と め

本稿では、在宅高齢者の見守りのための人間中心のコンテキスト認識手法を考察することに着目し、関連研究や既存技術に基づく全体フレームワークの詳細内容を述べた。我々は、人間の姿勢検出や、骨格検出、および顔表情、視線の検出に関する画像に基づく事前学習済みモデルを統合することをキーアイデアとした。また、高齢者の手周りの持ち物や触り物、行動、姿勢、表情、視線へのより豊富性の高いコンテキスト認識手法への考察も行った。今後の課題として、本稿で考察したフレームワークの実装を行い、高齢者の自宅で臨床実験を行いたいと考える。さらに、見守り現場の様子に合わせて、複数の定点カメラを設置するアプローチも検討し、異なる視点と視角範囲での実際の検証効果を明確に取得したいと思う。

謝辞 本研究の一部は JSPS 科研費 JP19H01138, JP20H05706, JP20H04014, JP20K11059, JP22H03699, JP19K02973, 若手研究 23K17006 の助成を受けて行われている。

文 献

- [1] C. Miura, S. Chen, S. Saiki, M. Nakamura, and K. Yasuda, "Assisting personalized healthcare of elderly people: Developing a rule-based virtual caregiver system using mobile chatbot," *Sensors*, vol.22, no.10, p.3829, 2022.
- [2] 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.
- [3] D. Marín, J. Llano-Viles, Z. Haddi, A. Perera-Lluna, and J. Fonollosa, "Home monitoring for older singles: A gas sensor array system," *Sensors and Actuators B: Chemical*, p.134036, 2023.
- [4] J. Shen and Y. Sun, "Privacy-preserved video monitoring method with 3d human pose estimation," 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)IEEE, pp.1502–1507 2023.
- [5] M.R. Lima, T. Su, M. Jouaiti, M. Wairagkar, P. Malhotra, E. Soreq, P. Barnaghi, and R. Vaidyanathan, "Discovering behavioural patterns using conversational technology for in-home health and well-being monitoring," *IEEE Internet of Things Journal*, 2023.
- [6] C. Di Napoli, G. Ercolano, and S. Rossi, "Personalized home-care support for the elderly: a field experience with a social robot at home," *User Modeling and User-Adapted Interaction*, vol.33, no.2, pp.405–440, 2023.
- [7] M. Jachymek, M.T. Jachymek, R.M. Kiedrowicz, J. Kaźmierczak, E. Płońska-Gościński, and M. Peregud-Pogorzelska, "Wristbands in home-based rehabilitation-validation of heart rate measurement," *Sensors*, vol.22, no.1, p.60, 2021.
- [8] B. Sumathy, S. Kavimullai, S. Shushmithaa, and S.S. Anusha, "Wearable non-invasive health monitoring device for elderly using iot," *IOP Conference Series: Materials Science and Engineering*, vol.1012IOP Publishing, p.012011 2021.
- [9] F. Tasaka, S. Chen, M. Nakamura, and S. Saiki, "Analysis and visualization of living habits of elderly people at home based on indoor environment sensing data," *IEICE Technical Report*, vol.122, pp.35–40, March 2023.
- [10] Y. Umetsu, W. Sasaki, H. Suwa, G. Shimoyama, M. Arahira, and K. Yasumoto, "Analyzing daily living patterns for optimizing anomaly detection conditions based on in-home sensor data," *JSAI Technical Report, Type 2 SIG*, vol.2019, no.SAI-034, p.02, 2019.
- [11] S. Chen, S. Saiki, and M. Nakamura, "Toward flexible and efficient home context sensing: Capability evaluation and verification of image-based cognitive apis," *Sensors*, vol.20, no.5, p.1442, 2020.
- [12] H. Ozono, S. Chen, and M. Nakamura, "Study of microservice execution framework using spoken dialogue agents," 2021 IEEE/ACIS 22nd International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)IEEE, pp.273–278 2021.
- [13] S. Chen, S. Saiki, and M. Nakamura, "Using human pose estimation for user-defined indoor location sensing," 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)IEEE, pp.495–501 2021.
- [14] H. Ozono, S. Chen, and M. Nakamura, "Encouraging elderly self-care by integrating speech dialogue agent and wearable device," *International Conference on Human-Computer InteractionSpringer*, pp.52–70 2022.
- [15] S. Chen, H. Ozono, M. Nakamura, and K. Yasuda, "Quantitative expression of elderly multi-modal emotions with spoken dialogue agent and edge ai," 2023 IEEE 6th Eurasian Conference on Educational Innovation (ECEI)IEEE, pp.219–221 2023.