

動画を入力として自発的に動作するホームAI

—sumica：操作のいらぬホームAI—

1. 背景

近年、Alexa や Google Assistant などの対話型AIが搭載された Amazon Echo、Google home などのデバイスが出現し、徐々に我々の生活の中に浸透し始めている。これらのデバイスは、人間からの発話指示に応じて様々なサービスを提供してくれる。例えば、外部WebサービスのAPIやIoTデバイスと連携することで、照明を点灯させたり、ピザの注文がしたりする事が可能である。しかしながら、これらのデバイスは少なくとも以下の二点の問題があるため、本来求められるホームAIとしての要件を十分に満たしているとは言えない。

1. 対話指示や事前設定によってのみ動作を行い、自発的に動作を行わない。
2. パーソナライズが不十分である。

前者の理由の一つとして、基本的な入力にマイクが使われることが挙げられる。これらのデバイスは音声入力を必要とするので、基本的に人間が能動的に操作しなければならない。ユーザが寝ている間や不在時などにおいては、AIが自発的に適切な行動を行うことはできないし、人間の指示がない状態で気を利かせて行動することもない。また、音声コマンドに応じたホームAIの動作は予め決められており、ユーザごとのニーズや行動習慣に合わせたパーソナライズを行うことも出来ない。

2. 目的

本プロジェクトではユーザの行動や部屋内の状況を認識し、それに依って自発的に動作するホームAIを開発することを目的とした。

本ホームAIは部屋に設置されたマイクとカメラを用いて常時センシングを行い、撮影された画像に対して人物検出、姿勢推定等の画像処理を行う。また、ユーザが自分で行動認識器を学習させたり、システムの動作を直感的に設定できるUIを提供する。このUIにより、ユーザの習慣や部屋の違いを考慮してパーソナライズされた行動認識器を学習することができる。また、各ユーザのニーズに合わせたホームAIの動作を柔軟、かつ容易に設定することが可能である。

3. 開発の内容

本システムは、マイクとカメラを用いてユーザの発話内容や、行動をセンシングする。また、ユーザは本システムの提供するUI上からパーソナライズされた行動認識器の学習、部屋内の状況に応じたシステムの動作の設定を行うことができる。

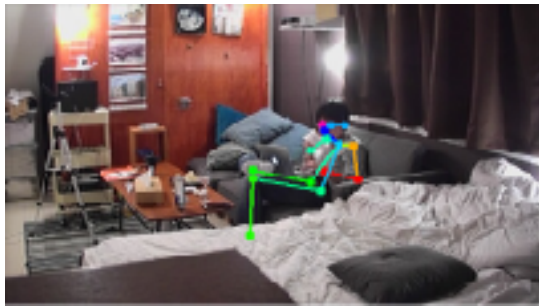


図1 姿勢推定



図2 人物検出

本システムはカメラによって部屋内で撮影された画像を解析し、ユーザの行動を認識する。ここでは画像解析処理について詳細に説明する。

本システムでは、画像に対して大きく3つの処理を行っている。

- ・ 姿勢推定
- ・ 人物検出
- ・ 学習済み行動認識器を用いた特徴抽出

カメラから撮影された画像に対して人物検出を行った後、人物を含む領域でクロップした画像を学習済み行動認識器を用いて特徴抽出を行う。また、姿勢推定を行い、推定されたユーザの関節位置のベクトルを特徴として扱う。このようにして得られた二つの特徴ベクトルを連結したものを最終的な画像の特徴ベクトルとして用いる。

次に本システムで提供するユーザインタフェース（UI）について説明する。ユーザは図3に示すUIを通じて以下の2つの操作を行うことができる。



図3 ユーザインタフェース

- ・ 行動認識器に入力する学習データのラベル付け
- ・ システム動作の設定

ユーザはUI上のタイムラインから、過去に撮影された画像を確認し、適切な行動のラベル(睡眠、食事等)を付与することができる。画像に付与されるラベルが更新されたことを検知し、行動認識器は自動でアップデートされる。

また、部屋内の状況に応じたシステムの動作の設定はUI上のグラフから行うことができる。部屋内の状況を表す入力ノード、システムの動作を表す出力ノードが存在し、対応する入力ノードと出力ノードをエッジで紐付けることで、部屋内の状況に応じたシステムの動作を直感的に設定することができる。

また、複雑な部屋内の状態を記述するために中間ノードを用いることも可能である。中間ノードは、「ある状態が一定時間以上続く」「ある二つの状態が同時に起こる」などの条件を記述することができる。これらのノードを組み合わせることで、システムの多様な動作を柔軟に設定することが可能である。

図4は「夜間に、スマホを累積1時間以上使用した場合に、注意する」という動作を設定したときのグラフの例である。

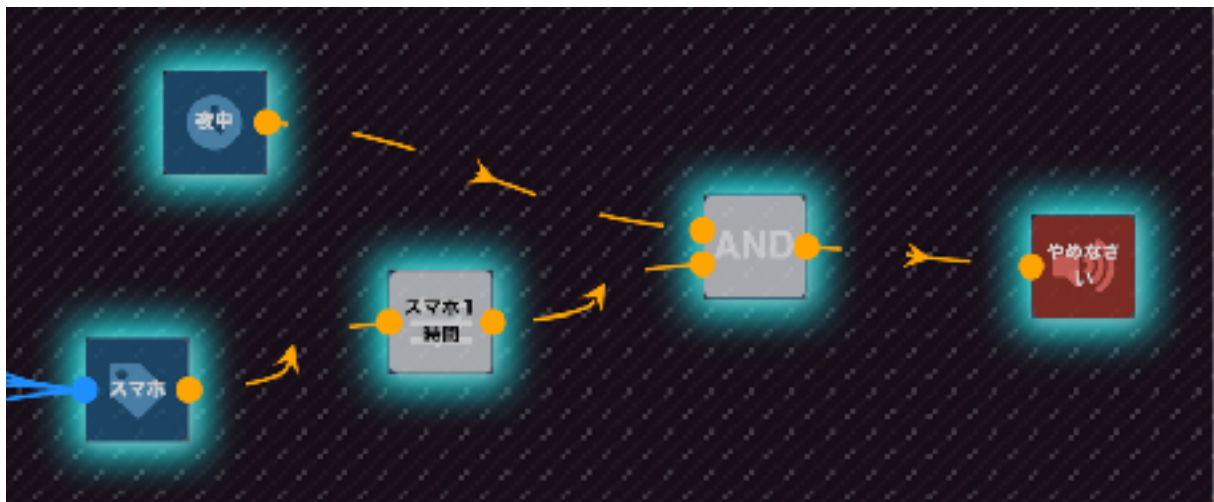


図4 グラフによるシステムの動作設定

4. 従来の技術（または機能）との相違

従来のホームAIは、スマートスピーカーと呼ばれる音声操作可能なホームアシスタントデバイスであった。これらのデバイスは既存の一般的な家電操作方法である、スイッチやリモコンによる操作を音声操作で置き換えた。しかしながら、これらのデバイスは依然操作の必要があり、音声認識の精度もまだ完全とはいえない。また、システムの動作にユーザによる操作を必要とするため、システムが動作できるシーンに限られるという問題点も存在する。

本プロジェクトで提案するホームAIはユーザの行動認識を行い、それに応じて自動的に動作する。そのため、ユーザによる明示的な操作なしに、システムはテレビ

を消したり、照明をつけたりといった機器操作を行うことが可能である。また、従来のシステムでは、ユーザによる操作が必要であるため動作ができなかった、ユーザの睡眠時や不在時、また何かに集中していてシステムを意識していないような場面でも、システムが動作することが可能になった。

5. 期待される効果

本システムは従来の受動的に動作するホームAIとは異なり、自発的に動作する。既存のホームAIは、「操作」が必要であるという点において、従来の機器操作と変わらず、そのため既存のホームAIに対し利便性を感じないユーザも一定数存在した。従って、本システムはそのようなユーザのニーズを掴み、今後のホームAIの普及に貢献することが期待される。

6. 普及（または活用）の見通し

本システムは画像解析にGPUを使用している。そのため、本システムを利用するユーザは自分でGPUを用意する必要がある。そこで、今後はGPUを必要としない軽量な特徴抽出手法の導入を行うことで、多くのユーザに使ってもらえるように努める。普及についてはオープンソースプロジェクトとして公開することを考えている。

7. クリエータ名（所属）

中村 晃貴（慶應義塾大学 大学院 理工学研究科）

野口 裕貴（慶應義塾大学 環境情報学部）