4G-05

直感的に操作可能な音声認識リモコンの研究

荒木恭平　三好力
龍谷大学

1　はじめに

現在、多くの家電が普及することで生活は便利になっている。それらの家電を操作するために、赤外線リモコンなどを用いて操作している。しかし、家電が多くなると、リモコンの数も同様に増えてしまう。その問題は、リモコンを一括で管理することで解決できる。一例として、音声認識リモコンというシステムがある。しかし、音声認識リモコンは家電を一括で操作するために、家電の種類を音声で指定する必要があった。そこで、人の視線と音声認識リモコンを組み合わせ、人の視線で家電を指定し、ボイスコマンド 1 単語で家電を操作するシステムの開発を提案した。

2　提案手法

従来のリモコンと音声認識リモコンを組み合わせ、音声認識リモコンに指向性をもたせることで、誤作動が少なく、直感的に操作可能なリモコンができると考える。

本研究では、指向性をもたせた音声認識リモコンを開発するのに、2 通りの手法を提案する。

1 つは、使用者本人の視線と同様の画像を取得できる位置にカメラを配置することで、人が家電に視線を向けるとカメラに家電が写り、その家電が指定される手法である。もう 1 つは、家電にカメラを置き、使用者がカメラの方へ視線を向けていると、その家電が指定されているという手法である。この 2 つの手法を比較し、どちらが適しているのかを検討する。

これらの手法を図 1.1 および 1.2 に示す。

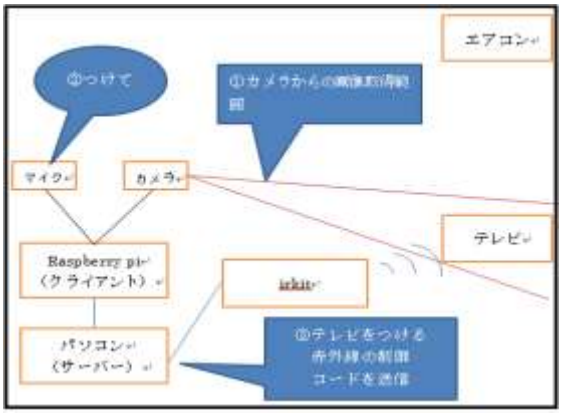

図 1.1：システムのブロック図（家電認識）

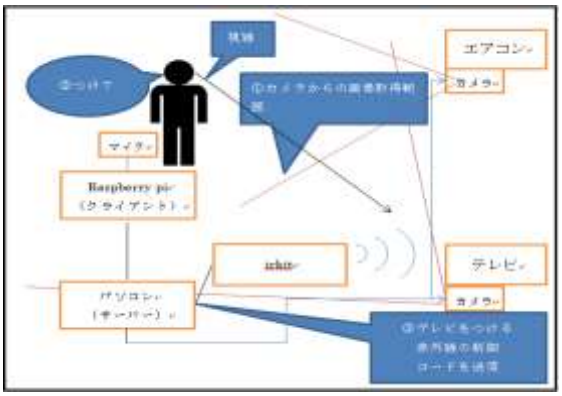

図 1.2：システムのブロック図（視線検出）

図 1.1 および 1.2 に示した図はそれぞれの手法でのシステムのブロック図である。

システムを構築するのにサーバークライアントシステムを用いる。このシステムを用いることで、処理を役割分担させる。サーバー側は家電の認識を行い、家庭のエアコンやテレビ、ライトなど、赤外線で操作できる家電を、一括で操作する赤外線リモコンデバイスである“irkit”を用いて、家電を操作するという処理を行わせる。クライアント側は web カメラやマイクから画像や音声を取得し、サーバーサイドに常時送信する処理を行わせる。サーバーは移動する必要がないので、据え置きのパソコンを流用する。クライアントは人が装着し、人がみている画像を取得する場合と家電に設置する場合を考えるので、設置が簡単にでき、軽量でなければならない。今回は Raspberry pi を用いて実験した。

Title:Study of intuitive speech recognition remote control
†Araki　Kyohei　Miyoshi Tsutomu
Factulty of Science and Tecnology , Ryukoku University

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.

実際の実験に使用機材を図2に示す。

図2：実験機材

3　画像認識を用いる家電指定

家電認識には、既存の深層学習フレームワークである、caffeを用いた。
多数の画像を与え自動で特徴量を抽出した分類機を作成する。この分類機を用いて、カメラからの画像で家電を判定する。

4　視線検出を用いる家電指定

Webカメラから人の顔画像を取得することで視線検出を行う。詳細な視線検出は行うことが難しいが、家電側を向いているのか向いていないのかを検出することは可能だと考える。視線検出を行う際、OpenCV の顔の認識、目の認識を行う機能を用いて視線を向けているのかを判別する。

5　実験結果および考察

5.1　画像認識を用いる家電指定の結果

本実験の深層学習の学習モデルはテレビ、エアコン、その他の 3 値分類機である。以下がテレビに対して、距離および角度と正答率のグラフである。

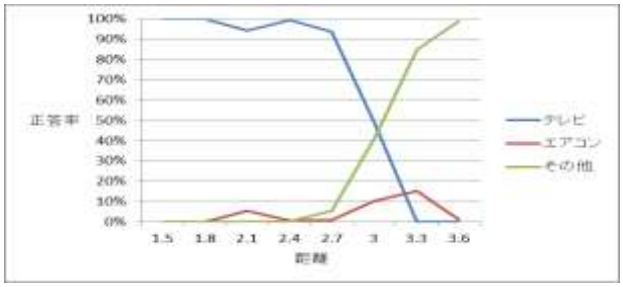

図3：距離1.5m～3.6m（テレビ）

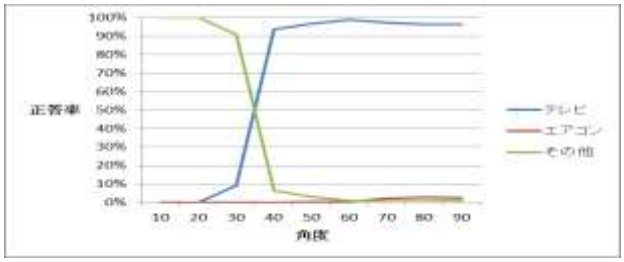

図4：角度10～90度、距離2m（テレビ）

図3はテレビから距離150cm～360cmまでを間隔30cmおきでテレビの画像を取得し、分類機に与え、分類した結果である。
図4は距離200cmにおいて、カメラとテレビのなす角度を10～90度まで間隔10度ずつのテレビの画像を取得し、学習モデル1の分類機に与え、分類した結果である。

5.2　視線認識を用いる家電指定の結果

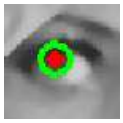

図5：左目の認識結果

図6：右目の認識結果

表1：目の座標

	W(pixcel)	H(pixcel)
左目	323	266
右目	265	262

図5、6及び表1はある画像を用いて、視線検出をおこなった結果である。他の実験画像で行った結果、100枚中37枚が正しく、残りは誤りであった。

5.3 考察

画像認識で今回作成した学習モデルを用いた場合、30度までは正確に識別ができ、距離は300cmまで正確に識別することが可能であった。しかし、家電を識別するのに2秒ほどの時間を要した。

視線認識では、処理時間が早いが、顔認識、目の認識、口の認識、のどれか1つでも失敗すれば、正しい視線は検出されない。今回用いたOpenCV の分類機を用いて、これらの認識を行ったが、画像により認識の失敗などが63%であった。また、瞳孔の座標検出では、Hough 変換を用いたがあらかじめに与えておいた円の直系の範囲では、全ての画像で正しい検出することがなく、画像によって、円サイズの変更が必要であった。

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.