

## 直感的に操作可能な音声認識リモコンの検討

T120414 荒木 恭平

指導教員 三好 力 教授

### 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する。

### 2. 提案手法

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する位置にカメラを配置して、使用者が家電に視線を向けるとその家電が識別され、指定される手法である。なお、家電の識別には深層学習とパターンマッチングを検討した。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、視線を検出してその家電が指定される手法である。なお、視線検出は OpenCV の顔認識や顔のパーツ認識を用いる。これら 3 つの手法を比較し、どれが適しているのかを検討する。

### 3. 実験結果

画像識別の実験はテレビ、エアコン、その他の 3 値分類で識別した。テレビの画像を深層学習とパターンマッチングに与えた結果を以下に示す。図 1 および図 2 はテレビとの距離や角度を変えた画像を深層学習に与えた結果である。表 1 はテレビとの距離や角度を変えた画像をパターンマッチングに与えた結果である。

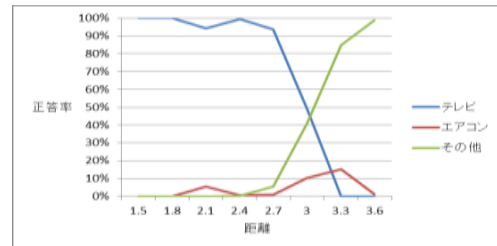


図 1 : 距離 1.5m~3.6m (深層学習)

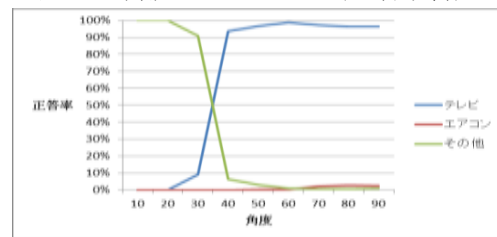


図 2 : 角度 10~90 度 (深層学習)

表 1 : パターンマッチングの結果

距離	正答	角度	正答
1.5m	○	10 度	×
1.8m	○	20 度	×
2.1m	×	30 度	×
2.4m	×	40 度	○
2.7m	×	50 度	○
3.0m	×	90 度	○

視線検出の実験結果では、100 枚の画像を用いた結果、37 枚正しく判別できたので正答率は 37%となった。

### 4. まとめ

画像認識では、図 1 の通り距離は 1.5m から 3m まで正確に識別することができ、図 2 の通り 90 度から 30 度までは正確に識別が可能であった。距離の変化には画像の中央を切り出すことで解決でき、角度は 30 度の位置で見ることがほぼないので本システムには利用可能である。視線認識では、認識率が 37%と低く、本システムでの利用に不十分であった。よって、本システム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が適していることが判明した。