

ディープラーニングによる視線方向推定

菅谷 裕一[†] 阿部 清彦^{††}

† 東京電機大学情報環境学部 †† 東京電機大学システムデザイン工学部

1. はじめに

肢体不自由者の支援のために、視線のみでコンピュータを操作できる視線入力システムが提案されている。今までに提案されてきた多くのシステムでは、特殊な赤外線カメラやセンサーを用いている。本研究では、ニューラルネットワークを用い、一台のカメラから得た画像のみで視線方向を推定する手法について検討した。

2. 視線方向の推定のための畳み込みニューラルネットワーク

CNNは、画像認識によく使用されるディープラーニングの代表的な手法の一つである。この方法を用いて、上下左右正面の5方向の画像サンプルを使用した分類学習を行い、5種類の視線方向を推定した。

使用するCNNは中間層が13層(畳み込み層9層、Max-pooling層3層、全結合層1層)で、畳み込み層3層おきにMax-pooling層を挟んでいる。畳み込み層のノード数は全て64、フィルタサイズは 1×1 、 3×3 、 5×5 の3種類を使用し、Max-pooling層のプーリングサイズは 2×2 とした。図1に使用したCNNの模式図を示す。

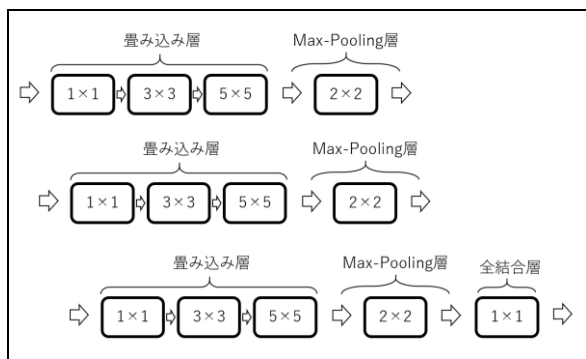


図1. 視線方向推定 CNN の中間層

3. 評価実験

3.1 学習用画像の分類学習

被験者1名が正面を向き上下左右正面の5方向に視線を向け、スマートフォンのカメラで動画を撮影した。撮影フレームを抽出することにより、各方向3200枚、テスト用に各50枚、計16250枚の学習用画像を作成した。作成した画像からHaar-like特徴量を用いたカスケード分類器を用いて顔部分をトリミングし、 64×64 (画素)に縮小した。学習の際には、目の周辺だけを学習させるために、 12×64 (画素)にトリミングを行った。バッチ数は128とし、エポック数は最大値を20とし、Early

Stoppingを設定した。学習の結果、テスト画像の推定精度は99.9%となった。

3.2 学習モデルを用いた評価実験

学習により得られたモデルを用い、Webカメラで動画を撮影しながらフレームごとに視線方向を推定するプログラムを作成した。評価実験の結果、65~89%の精度で視線方向を推定できた。これは、被験者1名による限定的な結果ではあるが、CNNを用いることによって視線の方向を画像から分類、推定できることを示している。

3.3 Grad-CAMによる特徴量の可視化

学習により得られたモデルが画像のどこに着目して視線方向を推定しているのかを確認するため、Grad-CAMという手法を用い[1]、学習用画像の内15枚について、特徴量の可視化を行った。図2に作成したヒートマップを示す。この結果から、CNNが黒目の部分ではなく目の周辺の部位を特徴として着目していることがわかった。これは、虹彩など、眼球の動きから視線を計測する従来の手法とは異なる着目点である。

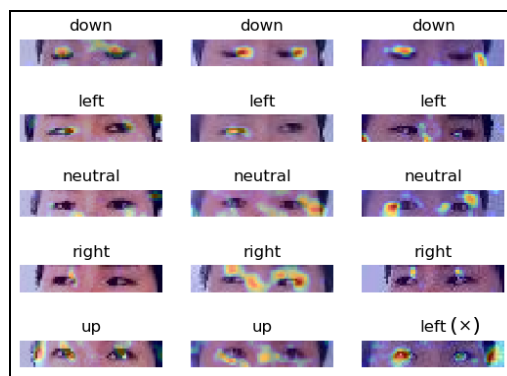


図2. 作成したヒートマップ

4. 今後の課題

今後は、さらに被験者を集め、学習モデルを汎用化させていく予定である。その際、特徴量の可視化も行ない研究を進めたい。また、5方向のみの視線方向推定だけでなく、より多くの注視位置を推定する手法について検討したい。

参考文献

- [1] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, 2017 IEEE International Conference on Computer Vision, pp. 618-626, 2017.