5S-03

顔画像を教師データとした視線追跡によるマウス入力システムの研究

丹羽　夏紀†　　張　善俊‡

神奈川大学 大学院 理学研究科 理学専攻†　　神奈川大学 情報学部 計算機科学科‡

1 はじめに

現在、視線追跡には専用の複眼カメラやアイトラッカーなどの高価な機械が用いられることが多く、世間一般にはなかなか馴染みがない技術となっている。そこで本研究では、多くの人が利用できる比較的安価な単眼ウェブカメラ１台のみを使用した高性能な視線追跡手法を提案する。ユーザーがパソコンを操作する際にマウスカーソルの位置を注視していると仮定して、単眼ウェブカメラ１台から得られる顔画像とその時点でのマウスカーソル座標が結びついた独自のデータセットを収集する。このデータセットを学習データとして、顔画像を画像処理して得られる特徴量を入力、その時点でのマウスカーソル座標をラベルとして、機械学習を用いてモデルを構築する。学習モデルの構築過程で、学習データの数量や入力特徴量の組み合わせを変えて実験を行い、各モデルのマウスカーソル座標予測の性能を評価する。また、実際にモデルを使用してリアルタイムで顔画像の入力からマウスカーソル座標予測を行い、視線でマウスカーソルを動かせるか確認する。

2 先行研究

Sugano ら[1]は、単眼カメラから得た目の外観画像と頭部ポーズを入力、クリック点をラベルとして、頭部ポーズ依存のマッピング関数を学習するクラスタリングベース手法を提案した。また、クリック点を注視位置と仮定することで、リアルタイムに個人モデルの学習を行った。Krafka ら[2]は、スマートフォンやタブレットの画面上にドットを表示して被験者に注視させたときの内側の単眼カメラから得た顔画像と画像内の顔の位置グリッド情報と目画像を入力、ドット座標をラベルとして、畳み込みニューラルネットワークを用いてモデルを学習させる手法を提案した。

A Research of Mouse Input System by Eye Tracking with Face Image as Training Data

†Natsuki Niwa : Science Major, Graduate School of Science, Kanagawa University

‡Zhang Shanjun : Department of Computer Science, Faculty of Informatics, Kanagawa University

3 提案手法

3.1 ラベル付き顔画像データの収集

顔画像とマウスカーソル座標が結びついたデータセットを収集するために、画面内を反射し続ける円にマウスカーソルを合わせると、その時点の顔画像とマウスカーソル座標が記録されるようなプログラムを作成した。円は絶えず高速で移動しているため、円にマウスカーソルを合わせる行為は、本研究では円を注視できていると仮定する。マウスカーソル座標はパソコン側の解像度に依存しており、本研究では 2,560 * 1,440 pix の解像度である。また、単眼ウェブカメラから得られる顔画像の画素数は 640 * 480 pix である。今回作成したプログラムでは、１分間に約 300 組のデータセットが収集できる。

3.2 画像処理・特徴量抽出

データセットの顔画像から顔部分・右目・左目を抽出するために、今回は Dlib ライブラリの 68 点ランドマーク検出を使用した。68 点ランドマーク検出では、顔の各ランドマークに対応した番号が割り振られている。この性質を生かし、顔部分・右目・左目に対応した番号を取り出し、各座標群から x 座標同士および y 座標同士の最大距離を求め、大きい方の距離を用いて正方形に顔部分・右目・左目をそれぞれ切り抜く処理を行った。

ただし、特徴量として目や顔を使用する関係上、目のランドマークの関係性から瞬きをしたと判定された顔画像や 68 点ランドマーク検出ができなかった顔画像は本研究では採用しない。

次に、68 点ランドマーク検出を用いて数値的な特徴量を抽出する。今回使用した数値的な特徴量は、以下の９種類である。

- 顔方向 yaw、pitch、roll
- 右虹彩（黒目）中心座標 x、y
- 左虹彩（黒目）中心座標 x、y
- 右の目尻から目頭までの距離に対する、右虹彩の中心から目頭までの距離の割合
- 左の目尻から目頭までの距離に対する、左虹彩の中心から目頭までの距離の割合

顔方向は、３D 座標点と各点が対応した画像上の

２D 座標点からカメラの位置と姿勢を求める Perspenctive_n_Point 問題を用いて算出した。本提案手法ではカメラはパソコンに固定されているため、顔の位置と姿勢が逆算される。

また、顔画像は基本的に正面から撮影されているため、注視している虹彩が楕円形になることはほぼないものと判断した。

3.3 学習モデルの構築

顔部分画像、右目画像、左目画像それぞれの入力に対して、３つの ResNet-50 モデルが適用され、各特徴マップが生成される。各特徴マップは連結・平滑化・ReLU 活性化関数を介した全結合が行われ、１つの特徴マップとなる。そこに数値的な特徴量の入力が結合され、ReLU 活性化関数を介した全結合が行われ、最後に linear 活性化関数を用いて出力が２次元（x,y）となるようなモデルを構築した（図１）。

4 実験

本研究では、以下の２つの実験を行う。性能の評価方法として、モデルに対してテストデータセットのマウスカーソル座標予測を行い、実際の値と予測値の平均二乗誤差（MSE）と平均絶対誤差（MAE）の２つを算出する。

4.1 データセットの規模による性能評価

訓練データセットの枚数を 5,000 枚、15,000 枚、30,000 枚、45,000 枚の４つのケースで図１のモデルを作成し、それぞれの性能を比較する。

4.2 入力特徴量の違いによる性能評価

45,000 枚のデータセットで、入力する特徴量を「顔部分画像のみ」、「両目画像のみ」、「顔部分,両目,数値」の 3 つのケースで異なるモデルを作成し、性能を比較する。他の入力がないとしても、基本的なモデル構造は図１と同じである。

5 評価結果

4.1 の評価結果を表１、4.2 の評価結果を表２に示す。単位は pix である。

表 1

	5,000	15,000	30,000	45,000
MSE	15,900	12,400	8,930	9,110
MAE	95.7	84.5	69.5	72.7

表 2

	顔部分のみ	両目のみ	顔,両目,数値
MSE	9,000	11,400	9,110
MAE	73.4	78.9	72.7

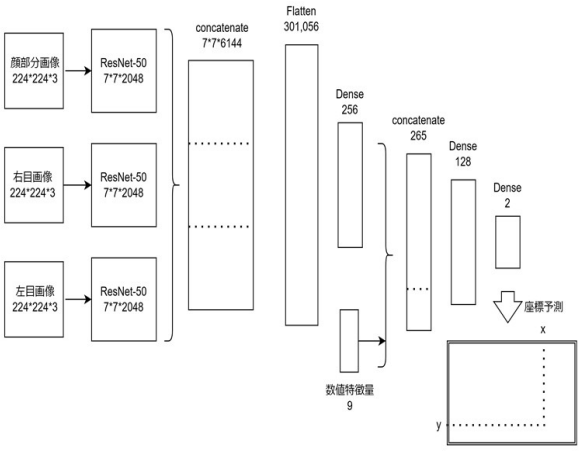

図 1　使用した学習モデル

6 考察

訓練データセットの枚数が増えるほど精度は上がっていったが、45,000 枚の精度が 30,000 枚よりも悪いことから、訓練データセットの枚数に関係なく 70 pix 辺りに単眼ウェブカメラによる座標予測の限界がありそうだと考えた。

また、顔部分画像のみのモデルでも精度がよく、頭部の向きと目以外からも有用な情報が含まれているのではないかと考えた。

7 まとめ

本研究では、単眼ウェブカメラを用いて顔の視線予測を行うモデルを提案した。異なる訓練データセットの枚数および入力特徴量の条件下での実験を通じて、性能の向上や有益な情報源について洞察を得ることができた。今後は、環境条件や様々なユーザーにおける応用可能性について検討し、モデルの拡張や改善を行いたい。

参考文献

[1] Yusuke Sugano, Yasuyuki Matsushita, Yoichi Sato, and Hideki Koike :“Appearance-Based Gaze Estimation With Online Calibration From Mouse Operations” IEEE Transactions on Human-Machine Systems, Vol. 45, No. 6, December 2015, pp.750-760.

[2] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, Antonio Torralba :“Eye Tracking for Everyone” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27-30 June 2016.