

機械学習を用いた視線方向推定と仮想文字盤への応用 Gaze Direction Estimation Method using Machine Learning and Application

磯部 佑太[†]
Yuuta Isobe

山地 秀美[†]
Hidemi Yamachi

1. はじめに

筋萎縮側索硬化症(ALS)や発達性筋ジストロフィー等の進行性の難病患者は、症状が進行するにつれて声によるコミュニケーションが困難になる。そのため意思を伝達する手段として残った運動機能を使ったコミュニケーション支援ツールが不可欠となる。

既存の支援ツールとして、目の動きを利用したものが広く用いられている。例えばひらがな表の書いてある透明な文字盤を用いる手法やパソコンと視線入力デバイスを併用する手法等が挙げられる。前者は介護者が肢体不自由者の視線の動きを文字盤越しに読み取り、文字をメモしていく。最後にメモした文字列を読み上げることでコミュニケーションをとる手法である。これは他者の手を借りる必要があり、視線を読み取るには訓練が必要なためコミュニケーションの手段として誰でもすぐに使うことはできない。後者は視線入力デバイスを用いて視線を推定し、文字盤をディスプレイに表示して文字の入力を行う。他者の手を借りずとも使用することができるためコミュニケーションの手段としては比較的容易に使うことができる。しかし、モニターやノートパソコンを使用者の顔の正面に設置するためにモニターアームを使う必要がある。装置自体が大掛かりなものになりがちで設置コストも少なくない。落下の危険性もあり、外出先での使用も難しいなど問題は多い。

また、症状の進行によっては眼球運動にも影響が始め動かせる範囲が狭くなることもある。そのような場合、従来の文字入力方法では画面端の文字を見ることができず入力できない問題がある。オリイ研究所が開発した支援ツールの OriHime eye^[1]ではディスプレイに表示した文字盤を使用者の目の注視点方向に動かすことで少ない目の動きにより全ての文字を入力可能にしている。このようにユーザーインターフェースの観点からも改善が必要とされる。

これらの問題を踏まえ本研究はディスプレイによる情報の提示が不要で、目の移動量を抑えて入力が行えるコミュニケーション支援ツールの提案を目的とする。

2. システムの構成

2.1 概要・仕様

提案する支援ツールの外観が図 1 である。眼鏡のフレームに取り付けた Web カメラで装着者の左目を撮影する。その画像を入力画像とし機械学習にて視線方向の推定を行う。方向は正面と目を閉じた状態、また上下左右斜めの各方向、計 10 パターンに分類し、特定の方向を一定時間見つめることで対応した文字を入力する。入力が一定時間無ければ文字列を読み上げる。簡便な操作で入力が可能であり従来の様にディスプレイへのインターフェースの表示やモニターアームの設置を必要とせず仮想の文字盤にてコミュニケーションを図ることができる。

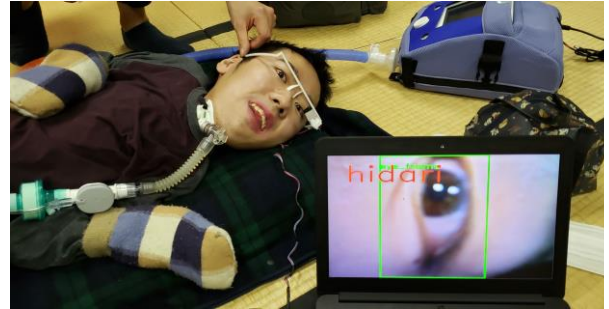


図 1 提案する支援ツールを装着した様子

2.2 視線方向推定

カメラを使う視線計測手法は角膜反射や強膜反射等を用いて視線方向を推定するモデルベース手法^[2]と、目の画像自体を特徴量として推定を行うアピランスペース手法^[3]の大きく 2 つに別けられる。モデルベース手法では赤外線などの外部光源を用いた計測が必要になる。今回の場合では装置の単純化の観点から外部光源が不要でカメラ画像のみで推定が可能なアピランスペース手法を用いる。

アピランスペース手法にはサポートベクターマシン(SVM)やニューラルネット(NN)などの様々な手法があるが本手法では畳み込みニューラルネットワーク(CNN)を用いる。環境は Python3.6 と Keras2.2.4 に TensorFlow1.14.0 を使用しネットワークには Keras^[4]が公開している一般物体認識用のものをファインチューニングして活用する。学習に使うデータセットは、装置を装着した際の左目の動きを撮影した 300×480 ピクセルの 12000 枚のカラー画像を使用する。

2.3 インターフェース

提案するインターフェースの遷移が図 2 である。9 つのブロックに分かれた設計となっており、識別方向を取得しその方向に応じた処理を実行する。図 2 の例では入力したい文字の「あ」行を選択し、2 回目の操作で入力したい文字(この場合「う」)を選択する。同じ方向の識別が 2 秒以上続くと入力として保持され、使用者に音声で伝える。目を閉じた状態が 2 秒以上続くと保持されている文章を読み上げ、意思を伝えることができる。従来の文字盤と大きく違う点は注視点による操作ではなく、視線の方向のみで入力を行うため操作が簡便になり、視線の可動域に捕らわれることなく扱えることにある。

インターフェースには Processing3.3.7 を使用し、文章の読み上げには音声合成システムの OpenJTalk を用いる。

3. 評価実験の方法

提案する支援ツールでは設置を容易にするため装着型と

[†] 日本工業大学 情報工学科, Nippon Institute of Technology

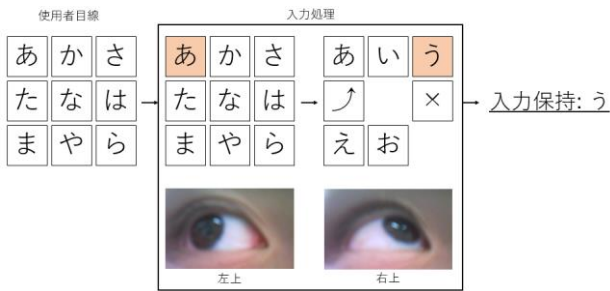


図 2 インターフェースの遷移

し、ディスプレイを取り除いて使用する。それによる有益性を示すためにインターフェースとデバイスのそれぞれに対して評価実験を行う。インターフェースでは従来の入力方法と比較し入力速度や精度、目の移動量の観点にて調査を行う。デバイスではディスプレイの有無による影響の違いを見るため、インターフェースを表示して入力を行った場合と表示しないで入力を行った場合での入力速度や精度、使用感を実験とアンケートから調査し評価する。

以下、二つの評価実験において入力速度は一文字目の入力開始から最後の文字の入力を完了するまでの時間を文字数で割ったものとし、精度は指定する文字列を入力する際の誤入力の平均回数とする。目の移動量は入力時の注視点の移動量の総和とする。入力する文章は「こんにちは」「おはよう」「おやすみ」「たすけて」「しせいなおして」の5つとし、表 1 の数値は同じ文字数である3つの文章に対して算出したものである。

3.1 インターフェースの評価実験

比較には従来の手法で一般的な入力方法である Tobii^[5]を使用したデジタル文字盤を用いる。インターフェースとしてひらがな表タイプの文字盤と提案する文字盤を扱う。最初にひらがな表タイプの文字盤を表示し、入力速度と精度、目の移動量を計測する。その後、提案する文字盤にて同じ項目を計測する。視線入力の経験の有無を考慮せず計 5 名の被験者に対して、操作・注意事項の説明を 2 分、試用時間を 5 分、実験時間を 10 分とし実験を行った。

3.2 デバイスの評価実験

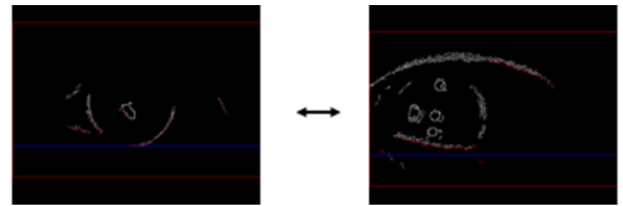
始めにディスプレイにインターフェースを表示した状態で入力を行う。入力中の文字や選択中のブロック、入力課題を表示した状態での入力速度や精度を計測する。次にインターフェースの表示を行わずに支援ツールのみで入力を行う。この際、ディスプレイには入力課題のみを表示し、入力中の文字や文字盤を表示しないものとする。

4. 実験結果と考察

インターフェースの評価実験の結果を示したものが表 1 である。表より提案するインターフェースを用いた方が入力速度・精度ともに改善が見られた。対してデバイスでは入力精度の影響から正確性を欠く結果となり実験を正しく行うことが叶わなかった。識別精度の低さの原因として考えられるのは環境光の影響である。図 1 にも見られるように使用者は横になった状態で目の画像を撮影する。そのため天井の照明が瞳孔に反射しエッジ検出の段階で学習時のデータセットと大きく異なる特徴が出た(図 3)。その差に

表 1 インターフェースの実験結果

| | 従来の文字盤 | | | 提案する文字盤 | | |
|-------|--------|--------|--------|---------|--------|-------|
| | 速度 (s) | 精度 (回) | 距離 | 速度 (s) | 精度 (回) | 距離 |
| 被験者 1 | 5.9 | 21.8 | 90814 | 2 | 1 | 28463 |
| 被験者 2 | 5.3 | 6.4 | 28251 | 4.4 | 0 | 23846 |
| 被験者 3 | 8.25 | 37.8 | 144514 | 8 | 12.6 | 96298 |
| 被験者 4 | 6.91 | 9.8 | 50369 | 5 | 1 | 35384 |
| 被験者 5 | 6 | 12 | 43868 | 5 | 2 | 39498 |



影響の少ないエッジ検出

影響があったエッジ検出

図 3 照明によるエッジ検出への影響

より識別がうまくいかなかったと考えられる。アンケートからは提案するインターフェースを用いた方がミスは減るが入力までの工数が増えるためストレスに感じたという意見が上がった。

5. まとめ

提案する支援ツールは従来のインターフェースを簡便なものに改善し、装着型のデバイスとした。それにより設置が容易になり扱いやすくなった。有用性を示すために評価実験を行った結果、インターフェースに関しては従来のデジタル文字盤に比べて、入力速度や精度、目の移動量において改善が見られた。デバイスに関しては識別精度が想定よりも低かったため有効性を示すことはできなかった。データセットを改善しつつ、機械学習以外の識別方法も併せて検討していく必要がある。また、入力では最低 2 回の操作を必要とするため、少ない工数で入力できる予測変換の機能や一文字当たりの入力時間を調整できる機能があるとストレス軽減につながるのではないかと考え今後の課題としていく。

参考文献

- [1]株式会社オリイ研究所 OriHime eye デジタル透明文字盤 <https://orihime.orylab/eye>(閲覧日 2019 年 12 月)
- [2] T. Ishikawa, "Passive driver gaze tracking with active appearance models," 2004, technical Report.
- [3] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Adaptive linear regression for appearance-based gaze estimation," IEEE transactions analysis and machine intelligence, vol. 36, no. 10, pp. 2033-2046, 2014
- [4] Keras https://github.com/keras-team/keras/blob/master/examples/cifar10_cnn.py(閲覧日 2020 年 1 月)
- [5] Tobii Pro フュージョン <https://www.tobii.com/ja/product-listing/tobii-pro-fusion>(閲覧日 2020 年 1 月)