

視線動作に内在する行為意図の 動的特徴を捉える機械学習モデルの構築

Statistical Analysis of Gaze Movements with Different Intentions and Their Classification Using Machine Learning

岡田 来波生¹, 大谷 一真², 田中 康平², 金野 武司³

Ranai Okada, Kazuma Otani, Kohei Tanaka, Takeshi Konno

金沢工業大学¹ 情報工学科, ² 電気電子工学科, ³ 知能情報システム学科

Kanazawa Institute of Technology, ¹Information and Computer Science,

²Electrical and Electronic Engineering, ³Artificial Intelligence

c1204100@st.kanazawa-it.ac.jp

概要

画面上で特定の領域を注視することで操作が成立する視線のインターフェースは、人間と機械との間の有力なインタラクション手段の一つとして研究が進められている。しかしその多くは、注視点の滞在場所や時間から人間の関心対象を同定する手法が取られている。視線動作には異なる行為意図が潜在している可能性があり、より直感的なインターフェースの実現には、視線が持つ動的な情報からその意図を読み解くことが不可欠であると考えられる。そこで本研究では、異なる意図を持った視線の動きを取得することのできる実験課題を開発・実施し、その動作特徴を分析すると共に、機械学習による識別可能性を検証した。結果、オフラインでの識別において、全体でおよそ6割以上の、そして参加者によっては9割を超える識別率が得られることを確認した。これは、視線動作による行為意図の推定がリアルタイム環境において十分に機能する可能性を示唆するものであると考えられる。

キーワード：視線計測, 機械学習, 意図の検知

1. はじめに

近年、視線計測技術は補助代替コミュニケーション (Augmentative and Alternative Communication: AAC) 機器のとして実用化が進んでいる。代表的な製品には Tobii Dynavox 社の TD I-Series が挙げられる [1]。TD I-Series は最新の屋外対応アイトラックを組み込み、ALS (筋萎縮性側索硬化症) や脳性まひ、脊髄損傷といった重度運動障害の利用者が、視線だけでパソコンや家電の操作を行なうことができる。また、視線入力を車いすやロボットのナビゲーションに応用する研究も進んでいる。Subramanian ら [2] は、視線

で注視した対象に向かって車椅子を移動させることのできる半自律型の移動システムを提案している。しかし、人間の視線には注視する対象に向かう意図のない自然注視も含まれるため、「見ただけで操作が発生してしまう」問題 (Midas Touch 問題) が生じ、これを回避する高度な意図識別が不可欠であると報告されている。

既存の視線インターフェースは、注視点の位置やその滞在時間といった静的情報に基づいて注視対象を特定する方式が主流である。しかし人間の視線動作には、意図が異なることによる動的パターンの違いが潜在している可能性があり、これを識別できれば、意図しない操作を抑制し、より自然かつ直感的なインタラクションを実現できる可能性がある。

そこで本研究では、まず異なる意図の視線動作を取得することのできる環境を構築し、得られた視線動作の特徴を分析した上で、機械学習によってその異なる行為意図の識別が可能であるか検証することを目的とした。

2. 視線データを取得する実験課題

異なる意図を持つ視線動作を取得できるようにするために、多数のオブジェクトが描かれた画像から特定の対象を探索する課題を開発した。この課題では、画面上に 41 種類のオブジェクトを計 1930 個配置した Fig.1 に示すような画像¹を 5 枚用意し、実験参加者には各画像につき指定された 5 体のターゲットとなるオブジェクト (人型ロボット、犬、カブトムシ、カエル、ペットボトル) を 3 分以内に探し出すことが課された。

¹画像のピクセルサイズは 1896 × 1030 とした。

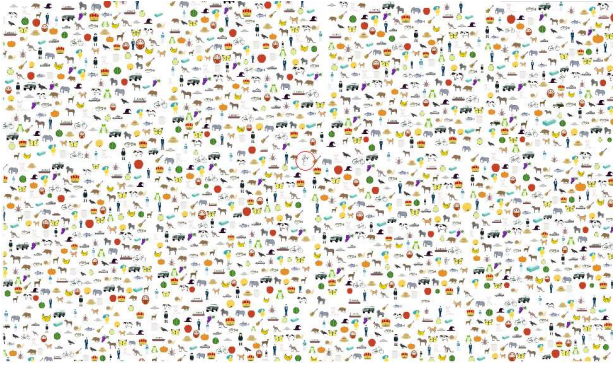


Fig 1 実験で提示されたオブジェクト探索画像



Fig 2 Experiment environment

この中で参加者は、視線によって画像の拡大・縮小、および平行移動できる「操作モード」と、そういった画面操作を停止させた状態にする「探索モード」が、手元のキーボードのZキーを押すことによって切り替えられるようになっていた。操作モードにおける画像の拡大（ズームイン）は、視線の移動速度が閾値以下に下がった状態で起こり、左目を閉じることによって縮小（ズームアウト）されるようになっていた。また、視線を画像の端に移動させると、その方向に移動するようになっていた。これによって、画面を操作しようとする意図と、対象を探そうとする意図という2つの異なる意図が、異なるモードとして取得できるようになっていた。

視線計測には Tobii 社製 X2-30 Compact アイトラッカを用い、これを PC (iMac 4.5K Retina ディスプレイモデル) のモニタ下部に設置した (Fig.2)。

実験には金沢工業大学の学生 10 人 (全員男性、平均年齢 21.6, $SD = 0.966$) が参加し、すべての実験は同大学の研究室内で行われた。また、実験は同大学の倫理審査委員会の承認を得て実施された。

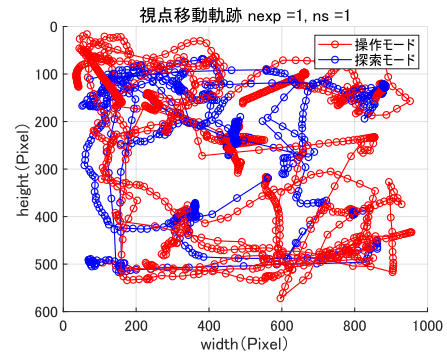


Fig 3 視点の画像上の移動軌跡. 参加者 1 の 1 回目の試行

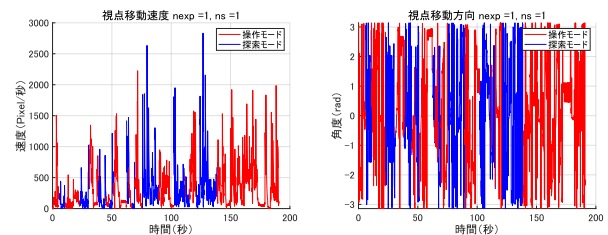


Fig 4 視点の移動速度 (左) と移動方向 (右). 参加者 1 の 1 回目の試行

3. 視線データの分析

一回の実験で参加者は、用意された 5 つの画像に対してそれぞれ 3 分間、特定の対象を探索した。これにより、モードの識別の対象となる視線のデータは 10 名 $\times$ 5 枚 $\times$ 180 (秒) であった。一例として、画面上を移動した視点の軌道を Fig.3 に示す。

我々はまず、各参加者の各試行ごとに視線の移動速度および移動方向を確認した (Fig.4)。全てのデータを確認すると、参加者ごとに視線の動かし方は一様ではなく、探索モードと操作モードの期間も長短入り乱れていたため、一見して両モードを判別できる特徴は見当たらなかった。

そこで次に、視線移動の速度と方向に頻度的な違いがないかを調べるべく、データを探索モードと操作モードに分け、それぞれの移動速度および方向を確率密度分布としてプロットした (Fig.5)。速度分布を見ると、探索モードのほうがやや高速域での頻度が高い傾向があったものの、両モード間で明確に状態を識別できるほどの差異があるとは言い難い。同様に方向分布においても、十字方向 (0, 90, 180, 360 度) 付近にピークが認められる程度であり、2 つのモードで大きく異なる様相は見られなかった。個別試行ごとに分布を確認しても、速度・方向ともに試行間で変動が

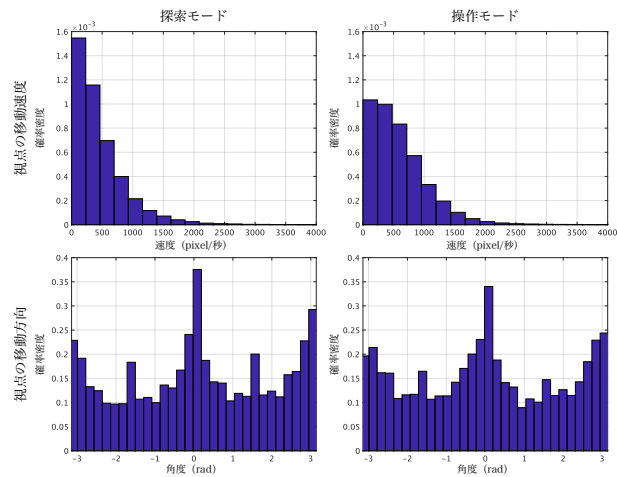


Fig 5 探索モード（左列）と操作モード（右列）における視点の移動速度（上段）と移動方向（下段）の確率密度分布

大きく、探索モードと操作モードを統計量だけで区別することは困難であることが確認された。

4. 視線データの識別方法と結果

視線の停留点ではなく、その動きの特徴を把握できるようにするために、我々は視線の移動方向と速度に着目したが、定量的な特徴としては探索モードと操作モードの間に違いを見出すことはできなかった。しかし、人間の視線動作には非線形かつ微細なパターンが含まれている可能性があることから、単純な統計的手法では捉えきれない差異が機械学習によって抽出できるのではないかと我々は考えた。

そこで我々は、視線の移動速度と方向を二次元平面上で散布図に描き、その頻度でヒートマップ化した画像を畳み込みニューラルネットワーク（CNN）に識別させる手法を採用した。時系列データの識別にはリカレントニューラルネットワーク（RNN）が広く用いられるが、RNNは学習が不安定になりやすくハイパーパラメータの調整も煩雑である。本研究のデータ規模と実装コストを考慮し、チューニングの自由度が低くとも高い識別性能が期待できるCNNを用いる方がより実践的であると判断した。CNNの構造は特殊なものを用いずに、Mathworks社のMATLABで提供されているサンプルネットワーク [3] を利用し、3層の畳み込み層と2層のプーリング層を組み合わせた構成とした。

CNNに入力する画像は、まず各時間ウィンドウ内の視線移動を速度と方向の散布図にプロットし (Fig.6 左)、その点群をもとにヒートマップへ変換した (Fig.6 右)。外れ値の影響を抑えるため、速度軸は全データ

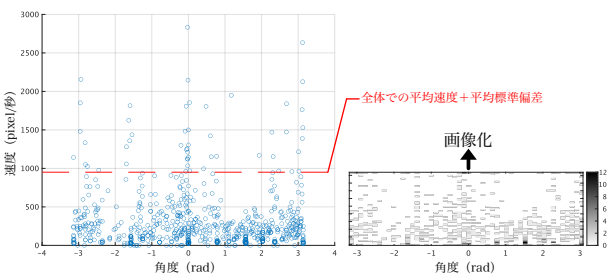


Fig 6 視点の移動方向と速度の散布図とその画像化。参加者1の1回目の試行

の平均値に標準偏差を加えた位置で上限を設け、それを超える点を上限値に置き換えてから集計した。速度と方向の両軸をそれぞれ50分割した50×50の格子に頻度を集計し、最小値と最大値の幅で線形スケリングした256階調のグレースケールで画像化した。操作モードと探索モードの時間間隔にはばらつきがあるため、5秒を最小窓幅とし、10秒まで1秒刻みでデータを切り出して画像を作成した。窓幅に満たない区間が生じた場合は、データを繰り返して長さを補完し、すべてのウィンドウが同一長になるよう整形した。

上記の処理により、探索モードで9650枚、操作モードで19075枚の画像が得られた。CNNはクラス不均衡に弱いため、操作モードの画像からランダムに9600枚を抽出し、両クラスを9600枚ずつにそろえた。データセットの70%（各クラス6720枚）を訓練用、残りを検証用に割り当てた。

用意した画像群を探索モードと操作モードに分けて学習させた結果、識別率は0.97だった。この値は、人間の目にはほとんど差異が認識できない状態でも、CNNはモードの違いを高精度に識別できることを示している。ただし、この結果はあらかじめ切り出したデータの識別性能に過ぎず、連続する視線データに対して同等の精度でモードを推定できることを示すものではない。

そこで、学習済みモデルを180秒間の2つのモードが混在する連続した視線の移動データに適用し、5秒から10秒までの各窓幅でモード推定を行った。その結果をFig.7に示す。横軸はヒートマップ生成に用いた時間窓の長さを示す。最短窓長を5秒に固定しているため、それより短い区間は視線データを繰り返して5秒分に揃えた。縦軸には推定モードと実際のモードが一致した割合を示した。

結果を見ると、参加者ごとに一定の個人差はあるものの、一致率はいずれも6割を超え、中には9割を上回る参加者がいることも確認できる。この中でも、と

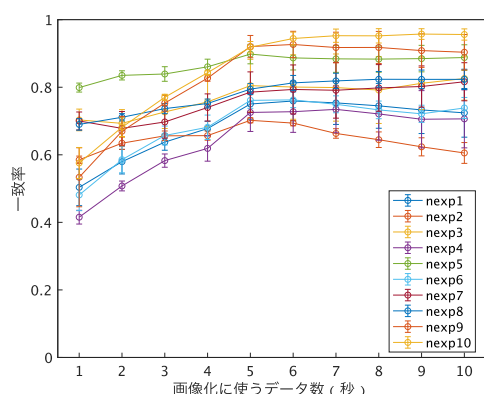


Fig 7 識別に使用するデータ時間幅に対するモード推定の平均一致率。エラーバーは標準誤差。

りわけ識別のための時間窓を5秒にした場合の識別精度が最も高く、視線に含まれる意図の変化を捉える上で有効な時間スケールがあることが示唆された。これらの結果は、視線を通じた行為意図の推定がリアルタイム環境においても十分に機能し得る可能性を示すものと我々は考えている。

5. 結論

本研究では、異なる意図を持つ視線動作を切り分けて取得することのできる実験環境を構築し、その違いを統計的な分析および機械学習によって識別できるかどうかを検証した。

我々は視線の停留点から注視対象を特定する従来の方法とは異なり、視線動作の移動速度と方向に着目してその統計量を調査したが、そこには異なる意図を定量的に識別できるような特徴を見出すことはできなかった。しかし、視線の動き（移動速度と方向）を画像化したデータに対して畳み込みニューラルネットワーク（CNN）を用いた識別を行なわせた結果、高い精度で識別可能であることが確認された。さらに、学習させたモデルを用いて異なる意図が混在する視線データを通時的に識別させてみた結果、その精度には参加者ごとのばらつきが見られたものの、全体としておよそ6割を超える識別率が得られ、参加者によっては9割を超えることが確認された。この結果は、モード推定が実時間において十分に機能する可能性を示唆していると考えられる。

文献

- [1] Tobii Dynavox, “TD I-Series,” <https://www.tobiidynavox.com/products/td-i-series/>, (2025.7.6 アクセス)
- [2] Subramanian, M., Park, S., Orlov, P., Shafti, A., & Faisal, A. A. (2021, May). Gaze-contingent de-

coding of human navigation intention on an autonomous wheelchair platform. In 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), pp. 335-338.

- [3] Mathworks, 分類用のシンプルな深層学習ニューラルネットワークの作成, <https://jp.mathworks.com/help/deeplearning/ug/create-simple-deep-learning-network-for-classification.html>, (2025.7.6 アクセス) .