

[研究助成 (C)]

意図しない操作に対して堅牢な視線に基づくコンピュータの操作手法

Gaze-based interaction robust for unwante interaction

2207001



研究代表者

筑波大学 大学院
システム情報工学研究群

博士後期課程

礪 本 俊 弥

[研究の目的]

本研究の目的は、視線追跡機器から得られるユーザの視線にまつわる情報を用いた入力インタフェースである、視線入力インタフェースにおける課題である意図しない操作に対して堅牢なコンピュータの操作手法を確立することである。

背景. 人間は日常的に意識的・無意識的に眼を動かし、ある対象に注意を向けることや、ある対象の情報を得るといったことを行っている。また、周囲の明るさや、見ている対象への興味によって瞳孔径の大きさも変化する。こういった眼の動きや、眼周辺の状態をコンピュータの操作に用いる操作を視線入力インタフェースと呼ぶ。中でも、操作したい対象をじーっと見続ける操作があり、これは凝視入力[a]と呼ばれ、マウスやタッチパネルを用いたコンピュータの操作における、左クリックやタップにより行う選択に相当する。凝視入力とは視線入力インタフェースのなかでも、簡単な操作としてこれまで多くの研究が行われてきた。凝視入力は操作が簡単である一方で、人間が無意識的に動かし、眼の動きを使用するため、意図しない操作が発生することが知られている。この意図しない操作の解決は長く研究されてきたが未だ解決されていない課題である。

意図しない操作の発生は凝視入力のアプローチに起因している。コンピュータはユーザがある対象をじーっと見ていることを認識するために、「ユーザの視線座標が一定時間以上ある対象の中に入り続けている」かどうかを調べる。一定時間以上入り続けている場合に、ユーザがその対象を選択したくて見ていると判断し、選択を実行する。この時間閾値が短い場合には、ぱっと見た・見てしまった時にも選択が実行され、意図しない操作が発生する。一方で2秒といった長い時間の閾値を用いることで意図しない操作の解決は可能であるが、選択するまでに必要以上に見続ける必要があり、操作性が減少してしまう。先行研究では、意図しない操作が発生しない、かつ操作性を減少させない時間閾値の探索が行われてきた[b]。

[研究の内容、成果]

本研究では、意図しない操作を解決する方法として、先行研究とは異なり、時間閾値に加えて、見続けている時および見続ける前に、ユーザの眼にまつわる情報がどのように変化をしたかを用いる。これらの情報からユーザが操作をしようとしているのか、そうでないかの意図を検出し、凝視入力に適応する。本研究において眼にまつわる情報とは、サッカード、フィクセーションと呼ばれる人間の認知処理にまつわ

る眼の動き、瞳孔径、および輻輳反応のことを指し、いずれも市販の視線追跡機器から得られる情報である。

本研究では、凝視入力時に生じる意図しない操作を防ぐ手法として、2つの凝視入力手法を示した。

まず一つめが、DTD (Dwell-Time-Dispersion) 手法である。これは、従来の凝視入力では、ある対象の中に一定時間以上視線が入っていることをもとにユーザの操作意図を検出していたのに対して「一定時間内の視線座標のばらつきが一定範囲未満である」かどうかをもとに操作意図を検出する。つまり、時間閾値に加え、ばらつき (dispersion) の閾値を用いて、ユーザの意図を検出する手法である[1, 2]。

2つめが、DTD-ML 手法である。これは DTD 手法に、機械学習 (machine-learning: ML) による、意図検出を組み合わせたものである。具体的には、DTD 手法により、意図を検出したのちに、その意図が選択意図であれば、さらに DTD 手法を行う以前の眼にまつわる情報を用いて機械学習による2回目の意図検出を行う。そして機械学習の結果も選択意図がある

場合にのみ、選択を実行する手法である (図1) [2]

実験1 DTD 手法を実装するためのユーザ実験を行った。実験では14名を対象とし、実験参加者には、2秒間画面に表示された点 (図2a) を見つめるよう指示した。表示された点を見つめている際の各時間閾値での視線のばらつきを調査した (図2b, c)。この結果を基に DTD 手法に用いる時間閾値とばらつき閾値を、それぞれ 600 ms および 0.3 度とした。

実験2 DTD-ML 手法における機械学習部分を実装するための実験を実施した。実験では、文字、単語、画像、動画を DTD 手法にて選択するタスク、および動画を視聴するタスクを採用した。これは、DTD-ML 手法が DTD 手法において、検出できなかった意図しない操作を防ぐことを目的としているためである。また、眼にまつわる情報は、何を見ているかや、タスクによって異なるため、複数タスクにおける眼にまつわる情報を収集した。実験は24人を対象として行った。また、機械学習に必要な

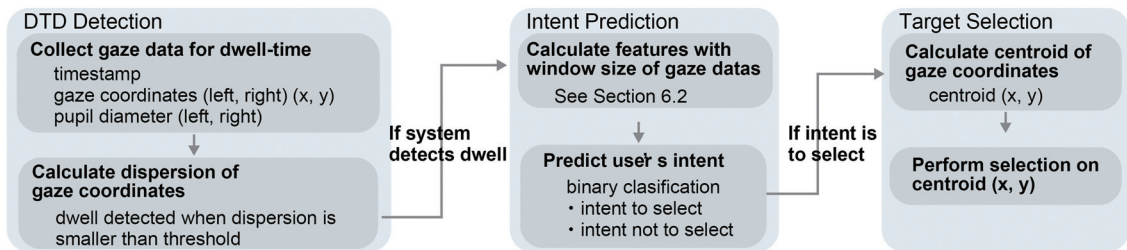


図1 本研究にて示す凝視入力 (DTD-ML) の実行の流れ [2]

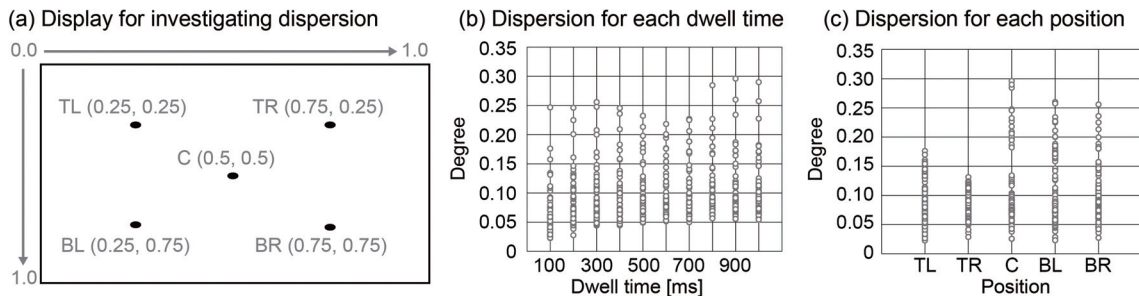


図2 DTD 手法を実装するために行った実験の結果 [2]

正解ラベルは、各選択が行われた際に実験参加者に、その選択は意図した者であったかを質問する形で収集した。

機械学習モデルの作成 実験2にて収集した眼にまつわる情報および正解ラベルを用いて意図検出を行う機械学習モデルを作成した。なお眼にまつわる情報は、視線追跡時に生じるノイズが含まれ、個人によっても異なるため、ノイズおよび個人差を取り除くための以下の工夫を行った。

- 1) まず、視線追跡機から得られた情報を20サンプルにダウンサンプルした
- 2) その後、DTD手法により選択が実行された時間を基準とした変化値を計算した。
- 3) 変化値に対して、表1に示す127の特徴量を計算した。

特徴量には、視線のx座標およびy座標の変化を示す **x**, **y**, 瞳孔径の変化を示す **pupil**, 輻輳距離の変化を示す **diffx**, およびサッカード、フィクセーションを示す **saccade**, **fixation** の統計量を用いた。

上記の特徴量と正解ラベルを用いて機械学習モデルを作製した。機械学習アルゴリズムには Gradient Boosting Decision Tree のアルゴリズムである LightGBM を使用した。本機械学習モデルによって検出する意図は「操作を行う」「操作を行わない」の2値である。

機械学習モデルのテスト結果を以下に示す。なお性能指標には AUC-ROC (以下 AUC) を用いた。

・学習データ全体：0.903

・Leave-one-participant-out 交差検証：0.890

・Leave-one-task-out 交差検証：0.601

本結果より、作製した機械学習モデルは、ユーザに依存しないが、タスクに依存することがわかった。

実験3 作製した機械学習モデルを使用した DTD-ML 手法が実際のインタフェースにおいてどの程度の性能であるかを示すための実験を実施した。実験に使用したインタフェースを図3に示す。なお、このインタフェースは実験2に用いたものとは異なるインタフェースである。20名の実験参加者に、このインタフェースを従来の凝視入力手法 (DT 手法), DTD 手法, および DTD-ML 手法を用いて10分間自由に操作してもらった。実験2と同様に、各選択が行われた際に、その選択が意図したものであったかの質問を行った。

実験3の結果を図4に示す。全実験参加者の意図しない操作の発生回数は、DT 手法が449回、DTD 手法が188回、DTD-ML 手法が43回であった。また、全選択に対する意図しない操作の発生割合は、それぞれ88.56%, 64.16%, 24.02%であった。実験参加者毎に見た場合でも、10分間の操作において、DT 手法、DTD 手法では、平均37.42回、15.67回の意図しない操作が発生していたが、DTD-ML 手法においては、3.58回と意図しない操作の発生回数が減少していることがわかった。また各選択が生じるまでにどの程度自由にインタフェースを見ることができていたかを示す時間は、DT 手法が4171ms、DTD 手法が6659ms、DTD-ML

表1 機械学習に用いた127の特徴量 [2]

Features		Numbers
plus, minus, absolute, and all (19) values of changes in x , y , diff_x , and pupil	average, standard deviation (SD), amplitude, skewness, kurtosis	80 (4×4×5)
durations of saccades, durations of fixations, distances of saccades, distances of fixations, velocities of saccades	average, first value, last value, last value minus first value, minimum value, max value, amplitude	35 (5×7)
Changes in x , y , diff_x , and pupil	1st value, 19th value, difference between 19th and 1st values	12 (4×3)

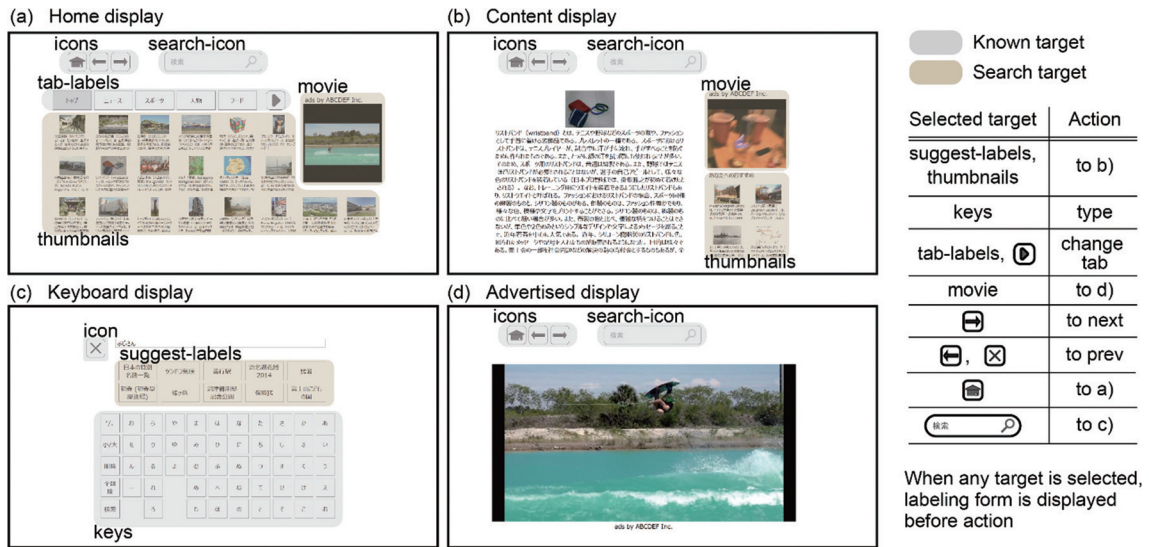


図3 実験3に用いたインタフェース [2]

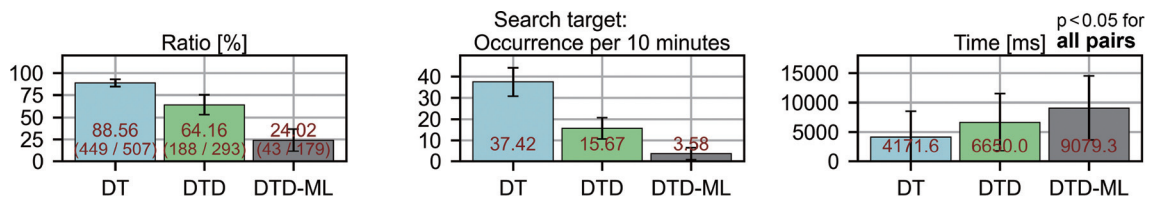


図4 実験3の結果 [2]

手法が 9079ms であった。これらから、DTD-ML 手法は、DT 手法と比べて意図しない操作を防ぐことができ、実験参加者は自由にインタフェースを操作できていたことがわかる。

また、実験 2 では、leave-one-task-out の AUC が 0.601 であることから、学習に用いたタスクとは異なるタスクにおいては十分な検出精度が得られないと考えられていたが、実験 3 では意図しない操作を防いでいることがわかった。これは、実験 2 において、文字、単語、文章、画像、動画と日常的に用いているインタフェースに対する選択タスクを網羅的行ったことが貢献していると考えられる。

ユーザ評価もまた、DTD-ML 手法が他の手法と比べ優れていることもわかった。システムのユーザビリティを定量的に示す指標である SUS (system usability scale) の結果は、DT 手法が 31.88、DTD 手法が 50.62、DTD-ML 手法が 69.38 であった。これも意図しない操作

を防ぐことにより、実験参加者が自由にインタフェースを操作できたことに起因していると考えられる。

本研究の発展 本研究において、これまでの凝視入力手法において課題とされてきた意図しない操作の発生を、DTD 手法および DTD-ML 手法を実現することにより減少させることがわかった。しかし、10 分間の操作において 3.58 回の意図しない操作が発生している。まずは、特徴量選定、機械学習モデルの改良、およびさらなるデータ収集などを通して、意図しない操作の発生回数を減らすための改良を行っていく。

本研究は主に現実世界のディスプレイを対象とした内容である。一方、近年視線追跡が可能なヘッドマウントディスプレイの登場により、視線インタフェースを Augmented Reality, Virtual Reality に適応する研究が多く行われている。本研究も発展系としてこれらの環境にお

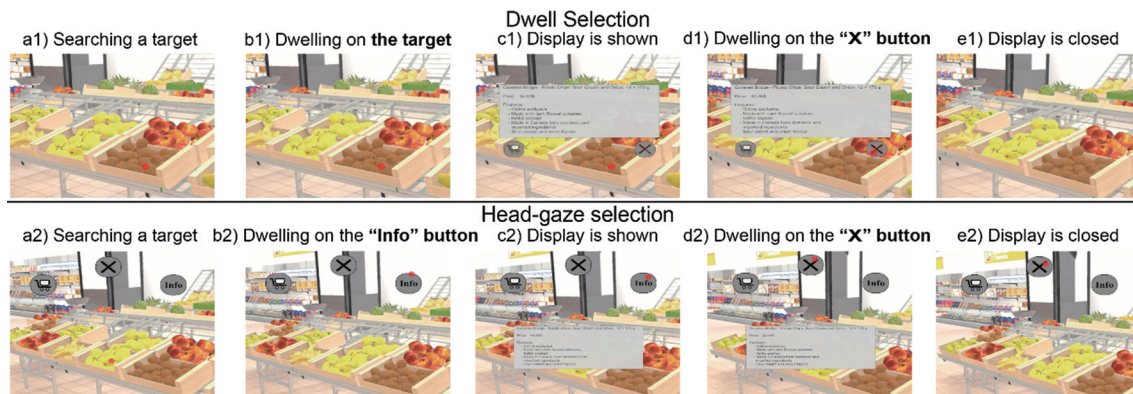


図5 DTD-ML 手法を Virtual Reality 環境に適応したインタフェース [3]

いて使用可能とするための改良を行っている (図5) [3]。

また、眼にまつわる情報を用いた意図検出技術に関しても、現状操作を行うか行わないかの2値の意図のみ検出可能であるが、例えば、これを3値以上の意図を検出するといった改良も検討する。このとき、操作意図だけではなく感情や興味の強さなどを検出対象とすることで、さらなるインタラクションの可能性が広がると考える。もちろんこれらの改良はチャレンジングな内容ではあるが、本研究における実験方法、特徴量選定などが今後の研究にも貢献すると考えている。

[関連研究]

- [a] Robert J.K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In Proceeding CHI 1990, 11-18.
- [b] Martez E. Mott et al. 2017. Improving Dwell-

Based Gaze Typing with Dynamic, Cascading Dwell Times. In Proceedings CHI 2017), 2558-2570.

[受給者による成果の発表, 論文など]

1. 磯本俊弥, 山中祥太, 志築文太郎. 時間および範囲をもとに認識する凝視に基づく操作手法. 第28回インタラクティブシステムとソフトウェアに関するワークショップ予稿集, Article 15, 6 pages. ソフトウェア科学会, 2020年12月(査読付き, 口頭発表)
2. Toshiya Isomoto, Shota Yamanaka, Buntarou Shizuki. Dwell Selection with ML-based Intent Prediction Using Only Gaze Data. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 6, No. 3, Article 120, September 2022 (査読付き, 論文誌)
3. Toshiya Isomoto, Shota Yamanaka, Buntarou Shizuki. Interaction Design of Dwell Selection Toward Gaze-based AR/VR Interaction. Proceedings of the 2022 ACM Symposium on Eye Tracking Research & Applications, Article No.: 39, 2 pages, June 2022 (査読付き, ポスタ発表)