

修士論文

顕著性マップと視線情報に基づくVR酔い推定手法に関する研究

令和7年1月23日 提出

指導教員 小川 剛史 准教授

東京大学大学院
工学系研究科 電気系工学専攻
37-236471 岡本 一真

要旨

VR 酔いの対策では、画面周辺部を隠すことで周辺視野への刺激を低減するなど、ユーザ体験を損なう課題がある。本研究では、VR 酔いの程度に応じた対策を可能とするため、VR 体験中の酔い度推定手法を提案した。シーンの顕著性マップと視線情報から、ユーザがどこをどのように見ているのかという情報を取得し、構築した推定モデルの有効性を検証した。

Abstract

Mitigation strategies for VR sickness, such as occluding the screen periphery to diminish peripheral vision stimulation, have posed challenges by detracting from the user experience. In this study, we explored a method to estimate the level of VR sickness during VR experiences, facilitating adaptive countermeasures based on the sickness intensity. Specifically, we acquired data on where and how the user is looking from the scene saliency map and gaze information, and validated the efficacy of the estimation model we developed.

目次

第1章 序論	1
1.1 研究背景	1
1.2 研究目的	1
1.3 本論文の構成	2
第2章 関連研究	3
2.1 VR 酔いのメカニズム	3
2.2 VR 酔いに影響するコンテンツの要因	4
2.3 VR 酔い推定	4
2.4 コンテンツにおける VR 酔いリスクの推定	7
2.5 顕著性マップ	7
第3章 注視点周りの顕著性マップに着目した VR 酔い推定	9
3.1 目的	9
3.2 提案手法	9
3.2.1 顕著性マップの生成方法	9
3.2.2 注視点周りの顕著性マップ統計量	11
3.2.3 使用する特徴量	11
3.2.4 VR 酔い推定モデル	12
3.3 実験方法	13
3.3.1 データ収集システム	14
3.4 実験結果	15
3.4.1 データ分析	15
3.4.2 VR 酔い推定結果	16
3.5 考察	16

第4章 顕著性マップとオプティカルフローに着目したVR酔い推定	18
4.1 目的	18
4.2 使用するデータセット	19
4.3 提案手法	20
4.3.1 オプティカルフローの生成と分類方法	20
4.3.2 注視点画像	21
4.3.3 X3D+Crossformer によるVR酔い推定モデル	22
4.3.4 ハイパーパラメータ	24
4.3.5 学習設定	24
4.4 データ分析	25
4.4.1 視線情報の分析	25
4.4.2 頭部姿勢の分析	26
4.4.3 顕著性マップの分析	26
4.5 実験結果	28
4.5.1 モジュール間の比較	28
4.5.2 特徴量間の比較	28
4.5.3 先行研究との比較	29
4.6 考察	30
第5章 結論	32
5.1 結論	32
5.2 今後の課題	32
謝辞	34
発表文献	35
付録A	43

目 次

2.1 感覚矛盾説 (文献 [1] より引用)	3
2.2 VR 酔い状態と非 VR 酔い状態での視線ヒートマップの比較 (文献 [2] より引用)	8
3.1 UNISAL (文献 [3] より引用)	10
3.2 元画像と 1/16 サイズの顕著性マップ比較	10
3.3 顕著性マップの注視点の 2.5° 範囲	11
3.4 VR 酔い推定モデルの概要	13
3.5 シューティングゲーム (HyperDash[4])	13
3.6 レースゲーム (KartKraft[5])	13
3.7 データ収集システムの概要	14
3.8 VR 酔いアンケート	15
3.9 VR 酔い度別の注視点周りにおける顕著性マップ最大値	16
4.1 視野の非制限, 制限状態でのオプティカルフローと k-means 法による分類	21
4.2 生成された注視点画像	22
4.3 X3D+Crossformer による VR 酔い推定モデル	23
4.4 VR 酔い度別の視線のヨー, ピッチ方向の角速度の絶対値	25
4.5 VR 酔い度別の頭部の位置速度 ($ v_{hx} , v_{hy} , v_{hz} $)	27
4.6 VR 酔い度別の頭部の角速度 ($ \omega_{hx} , \omega_{hy} , \omega_{hz} $)	27
4.7 VR 酔い度別の注視点周りにおける顕著性マップ最大値	28
4.8 オプティカルフローによるクラスタリング別の注視点周りにおける顕著性 マップの最大値	29

A.1	k-means による非視野制限状態でのオプティカルフローの分類結果	43
A.2	k-means による視野制限状態でのオプティカルフローの分類結果	44
A.3	非視野制限状態のオプティカルフローの分類結果に対応する注視点周り (< 2.5°) の顕著性マップ画素値の最大値	45
A.4	視野制限状態のオプティカルフローの分類結果に対応する注視点周り (< 2.5°) の顕著性マップ画素値の最大値	46

表 目 次

2.1 先行研究の比較 (V: ビデオ、E: 視線移動、H: 頭部移動、P: 生理信号、U: ユーザープロフィール情報)	5
3.1 特徴量一覧	12
3.2 各特徴量毎の推定結果	16
4.1 VR.net の利用可能なデータ一覧	19
4.2 主要なハイパーパラメーター一覧	24
4.3 各モジュールの性能評価	28
4.4 各特徴量ごとの推定結果	30
4.5 先行研究との比較	30

第1章 序論

1.1 研究背景

VR 酔いとは、メタバースや VR ゲームなど VR 体験中に感じる吐き気やめまい、不快感などといった乗り物酔いと類似した症状のことを指す。VR 酔いが引き起こされる主な要因のひとつが、視覚と平衡感覚が一致していないことだとされており、これを感覚矛盾説 [6] という。また、HMD(Head Mounted Display) に表示されている映像の遅延や低フレームレートなどによっても酔いが誘発されることが知られている [7]。文献 [7] では、被験者にいくつかの VR 空間内で自由に行動させた結果、80%の被験者が VR 酔いの症状を感じたと報告し、VR 酔いを感じた被験者のうち 5%は途中で VR 体験を中止している。したがって、VR 酔いはユーザ体験の質を大きく低下させる原因のひとつだといえる。VR 酔いの対策として、視界内で固定されたマーカやグリッドを表示しておいたり [8]、視野角を狭くする方法 [9] などが提案されている。しかし、これらの方法は視覚情報の変化を伴うため、VR 環境への没入感を低下させる恐れがある。ユーザの VR 酔いの状態をリアルタイムに推定できれば、VR 酔いの対策を適応的に行ったり、体験の中断を勧めることができ、VR 空間への没入感を維持しつつ VR 体験の質を向上できると考えられる。

1.2 研究目的

VR 酔いは主に SSQ(Simulator Sickness Questionnaire)[10] などのアンケートによって評価されることが多く、リアルタイムな評価ができなかったり、評価の仕方に個人差があるといった課題がある。そのため、アンケートに代わる手法として心拍数や皮膚電気反応、脳波といった生理学的信号を利用する手法 [11, 12] や、HMD から得られる頭部姿勢や視線情報を用いるもの [13, 14] が研究されている。生理学的信号を利用した推定手法は HMD とは別にセンサが必要であったり、緊張や運動などの状態によりセンサデータが大きく変化するため、実用化が困難である。一方で、視線や頭部姿勢といった HMD から得られた情報のみを利用した手法は、追加のデバイスが必要ないという利点がある。先行研究 [13, 14]

において、視線や頭部姿勢がVR酔いを推定する上で重要な特徴量であると示された。一方、ユーザが見ている動画データはほとんど推定に寄与しないことが報告されている。しかし、動画データによってユーザが意図したオブジェクトに対し、視線を正確に素早く合わせることであったかを確認することができるため、動画データも重要な特徴量となると推測される。そのため、先行研究では動画データの特徴量を効果的に抽出できていなかった可能性がある。本研究では視線と頭部姿勢とともに動画データの要素をVR酔い推定に取り入れるため、顕著性マップを利用する。顕著性マップとは画像に対して、人がどこを注視しやすいのかを画像で表したものであり、動画データと視線の関係を表す特徴量として、VR酔い推定に有効な可能性がある。

第3章では、被験者1名を対象として、VRゲームをプレイ中の動画データ上の注視点座標付近の顕著性マップの画素値を取得し、その平均値や標準偏差値、最大値、最小値を新たな特徴量として導入し、データ分析やVR酔い推定モデルの構築を通じて、特徴量の有効性を検証した。

次に第4章では、多くの被験者と多様なVRゲームからなる、最新のVR酔い評価のためのデータセットであるVR.net[15]を使用して検証をおこなった。また、第3章で提案された特徴量に加えて、オプティカルフローの導入を検討した。オプティカルフローは視覚的な刺激の量を定量化することができるため、特定の視覚刺激に対して顕著な領域に視線を合わせられているかといったユーザの反応を分析することができる。さらに、より複雑な動画データと注視点の関係を捉えるために、注視点から中心視・周辺視野といった人の視覚システムを表すような注視点画像を生成し、動画データや顕著性マップとともに、3DCNNベースのモデルであるX3D[16]を用いて、直接処理することで、VR酔い推定の精度を向上できるかについても検証した。

1.3 本論文の構成

本稿の構成は以下の通りである。まず、第2章ではVR酔いを引き起こすメカニズムや要因、VR酔いの推定手法について紹介した後、顕著性マップについて紹介する。第3章で注視点周りの顕著性マップに着目したVR酔い推定、第4章で顕著性マップとオプティカルフローに着目したVR酔い推定の概要、実験方法、結果、考察を述べる。最後に、5章で結論及び今後の課題について述べる。

第2章 関連研究

2.1 VR酔いのメカニズム

VR酔いは乗り物酔いなどと同じ動揺病の一種とみなされており、その発生メカニズムについても同じ仮説で説明されることが多い。これまで多くの仮説が提案されているが、感覚矛盾説 [6] は最も広く受け入れられている。この仮説は身体の姿勢や運動の感覚と視覚や前庭感覚が矛盾することが、酔いを引き起こすとするものである。また、感覚矛盾説を発展させた感覚再配置説 [17] では、過去の記憶から予測される情報と一致しないことが酔いの原因になると主張している。そして、Subjective Vertical Conflict 仮説 [18] は感覚再配置説を更に発展させたものであり、予測された主観的な垂直方向が、知覚された垂直方向との不一致だけが動揺病を引き起こすというものである。

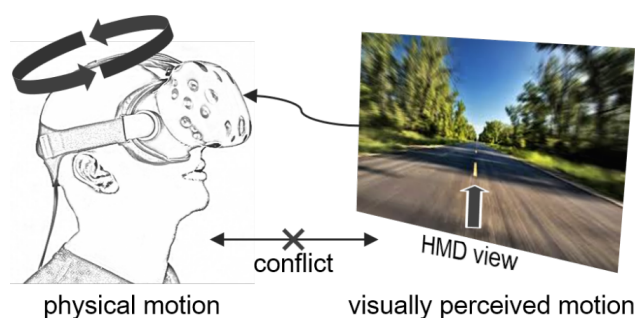


図 2.1 感覚矛盾説 (文献 [1] より引用)

上述の他にも様々な仮説が提案されているが、依然として動揺病のメカニズムは完全には解明されていない。これらの仮説はいずれも、視覚、前庭感覚、体性感覚といった複数の感覚モダリティ間の不一致が動揺病に寄与している可能性を示唆している。そこで本研究では視線情報、頭部姿勢、動画データといった複数の感覚からなるデータを統合することで、VR酔いの予測精度の向上を図った。

2.2 VR酔いに影響するコンテンツの要因

VR酔いを引き起こす要因の一つに考えられているものとして、ベクシオン(視覚性自己運動感覚)[19, 20]がある。ベクシオンとは視覚刺激が与えられた際に、静止しているのにも関わらず身体が運動しているかのように錯覚されることを指す。ベクシオンによる錯覚情報と静止した身体の情報とが矛盾しているため、感覚矛盾説によってベクシオンがVR酔いを引き起こす要因だと説明される。Soら[21]は被験者に、前後方向で8つの速度条件(3.3, 4.3, 5.9, 7.9, 9.5, 23.6, 29.6, 59.2 m/s)でVR体験をさせた。その結果、速度が10 m/sになるまでは、速度が上昇するにつれてベクシオンとVR酔いを強く感じるようになったが、それ以降の速度上昇はベクシオン、VR酔い共に10 m/sの場合とほとんど変化しなかったと報告している。文献[22][23]ではオプティカルフローを用いて、VR酔いを感じ始める速度及び加速度を調査した。その結果、並進運動よりも回転運動の方がVR酔いを感じ始める速度が有意に低く、一定速度よりも加速度条件の方がVR酔いを感じやすい傾向がみられた。

その他に、VR酔いに影響するコンテンツの要因として、グラフィックの品質[24, 25, 26]や、遅延の大きさ[7, 27]、受動的な運動の有無[28, 29]などがあげられる。Davisら[24]はグラフィックの品質が異なる2つのローラーコースターアプリケーションを使用し、その際のVR酔いを収集した。その結果、グラフィックの品質が高い方がVR酔いを引き起こしやすいとしている。Poukeら[25]は同一のアプリケーションを用いて、光源やモーションブレンダーなどを用いてグラフィックの品質を変化させ、VR酔いとの関連を調査した。その結果、有意差は認められなかったもののグラフィックの品質が高い方がVR酔いを引き起こしやすい傾向があると報告している。Dongら[28]は市販のレースゲームを使用して、被験者を運転手と乗客に分け、受動的な運動の有無によって生じるVR酔いの違いについて調査した。その結果、乗客の方がVR酔いを引き起こしやすいことが確認された。運転手は視覚刺激を予測できるため、感覚再配置理論などから乗客よりもVR酔いが少なくなったと説明することができる。

2.3 VR酔い推定

本研究と同様に主にHMDから得られる頭部姿勢、視線情報、動画データなどを使用してVR酔いの推定を行った研究について述べる。

Jinら[30]は5つのVRゲームを使用して、24名の被験者からプレイ中の頭部姿勢、動画データを収集した。動画データからコントラストやエントロピーといった特徴量を抽出

表 2.1 先行研究の比較 (V: ビデオ、E: 視線移動、H: 頭部移動、P: 生理信号、U: ユーザープロフィール情報)

文献	被験者数	データ	予測対象	性能
Jin et al. 2018[30]	24	H/V	SSQ スコア	R^2 0.868
Islam et al. 2021[13]	26	E/H/V	4 段階 VR 酔い度	RMSE 0.51
Lee et al. 2021[31]	25	V/P	SSQ スコア	RMSE 20.35
Shimada et al. 2023[32]	30	E	4 段階 VR 酔い度	精度 0.71
Jeong et al. 2023[14]	27	E/H/V/P	4 段階 VR 酔い度	精度 0.87
Kundu et al. 2023[13]	30	E/H	4 段階 VR 酔い度	RMSE 0.30
Zhao et al. 2023[33]	96	V/U	11 段階 VR 酔い度	MSE 5.67
Jeong et al. 2024[34]	45	E/H/P	20 段階 VR 酔い度	RMSE 0.25

し、LSTM-RNN を用いて VR 酔いを推定した。ランダム反復サブサンプリングによる評価の結果、決定係数が 0.868 となった。

島田ら [35] はリアルタイムで動作可能な VR 酔い推定に焦点をあて、視線情報データのみからの推定を試みている。島田らは 30 名の被験者にジェットコースターや車などの VR 動画を見せた際の視線情報データを収集した。そして注意機構を持つ LSTM と FCN を組み合わせたモデルの学習を行い、4 段階での推定を行った結果、71.09%の精度を達成した。

Islam ら [13] はジェットコースターなどの仮想環境を 5 つ作成し、26 名の被験者から動画、頭部姿勢、視線情報データを収集した。動画を 3D-CNN、その他の特徴量を CNN-LSTM に入力することで特徴抽出を行い、最後に全結合層を通して VR 酔いの度合いを 4 段階で出力するモデルを作成した。結果として頭部姿勢と視線情報データのみを用いた場合が最高で 87.7%で推定する事ができ、0~10 の段階で回帰を行った場合の RMSE は 0.51 だった。しかしながら、動画データを加えた場合は 52.2%となっており、動画データは VR 酔いの推定精度を低下させた。Kundu ら [36] は Islam らの研究を発展させ、SHAP(SHapley Additive exPlanations) により各特徴量の重要度を評価し、それを元に特徴量削減を行い、

より軽量かつ高精度な LSTM によるモデルを作成した。その結果、4 段階分類では 94%, 0~10 の段階で回帰では RMSE が 0.30 となった。

Jeong ら [14] は 27 名の被験者に VR 動画を視聴させ、その際の頭部姿勢、視線情報、EDA(Electro Dermal Activity: 皮膚電気活動)、動画データ、を収集した。注意機構をもった CNN と BiLSTM を組み合わせ、4 段階で VR 酔いの程度を推定するモデルを作成した。なお、動画データは ResNet による特徴量抽出の後に入力された。その結果、最高で 87% の精度で推定することができている。また注意機構の重みから各モダリティの重要度を知ることができ、視線情報が 44%, 頭部姿勢が 36%, EDA16%, 動画 4%となり、動画データは殆ど推定結果に寄与しなかったとしている。Jeong らはその後の研究 [34] で、45 名の被験者に VR 動画を視聴させ、頭部姿勢、視線情報、EDA、血圧、皮膚温を収集し、空間と時間それぞれの次元に対して Transformer を適用するモデルを作成した。その結果、20 段階での VR 酔い推定で RMSE が 0.25 となった。上述の文献 [13][36][14][34] では、モデルを評価する際に分割交差検証を用いているが、被験者数が十分でないといった理由からユーザや提示した刺激を区別しない分割交差検証を行っているため、モデルの汎化性能については十分に評価できていない可能性がある。

Lee ら [31] は 96 名の被験者に VR 動画を視聴させ EEG(Electroencephalogram: 脳波), ECG (Electrocardiogram: 心電図), EDA、動画データを収集した。生理学的情報から特徴量を抽出するネットワークと、動画データから特徴量を抽出するネットワークを別々に作成し、それらの特徴量から SSQ Score を推定するモデルを作成した。特に動画データからは、次のフレームを予測し実際のフレームとの差分を見ることで、どれだけ視覚刺激とユーザの予測との間に不一致があるのかという特徴量の抽出を図った。動画データをもとにした分割交差検証を行い、RMSE は全ての特徴量を用いた場合が最も低く、20.35 となった。動画データを用いなかった場合は 23.83 であったため、VR 酔い推定における動画データの有効性を示している。

Zhao ら [33] は、124 名の被験者に仮想空間内でコイン収集やシューティングタスクをさせ、その際の 3 次元オプティカルフローとユーザ情報を収集した。3 次元オプティカルフローはゲームエンジンから直接収集され、ユーザ情報は性別や年齢、日常生活で映像酔いを経験する頻度などが収集された。3 次元オプティカルフローを 3D CNN、ユーザ情報を MLP で特徴抽出し VR 酔いを推定するモデルを作成した。94 名を学習、30 名をテストデータに分割し推定を行った結果、10 段階の回帰で MSE が 5.67, 2 値分類では 86.5%の精度となり、通常のオプティカルフローを用いる場合と比べて、3 次元オプティカルフローによって推定精度を向上させられたとしている。

以上のように、視線情報や頭部姿勢が重要な特徴量であることが示されている。しかし、一部の研究では動画データはほとんど VR 酔いの推定に寄与せず、効果的な特徴量を抽出できなかったことが報告されている。また、動画データに対して視線がどのように動いているのかという関係を明示的に VR 酔い推定に取り入れた研究は、著者の知るところ未だない。そこで、本研究では動画と視線の関係を表している顕著性マップを利用することで VR 酔い推定の精度が向上できないかを検討する。

2.4 コンテンツにおける VR 酔いリスクの推定

VR コンテンツ利用時のユーザ体験を向上させるために、予めコンテンツの VR 酔いのリスクをユーザに提示する手法の研究が行われている。Quan ら [37] は、VR 動画の特徴量としてオプティカルフローの総和や回転、輝度値に加え、ResNet50 を用いた深層学習による特徴量抽出を行った。そして、これらの特徴量を用いて SVR (Support Vector Regression: サポートベクター回帰) モデルを構築し、VR 酔いのしやすさを予測した。その結果、PLCC (Pearson Linear Correlation Coefficient) が 0.942 を達成した。一方、Kuo ら [38] は、ガウス分布を用いたオプティカルフローの重み付き和を特徴量として抽出し、ランダムフォレストを用いた VR 酔いのしやすさを予測を行った。これにより、PLCC が 0.91 を達成したことが報告されている。Lee ら [39] は動画データからオプティカルフロー、視差マップ、顕著性マップを入力とし、VR コンテンツ自体の酔いやすさを推定する 3DCNN モデルを作成した。顕著性マップを入力に含めることで、推定精度を上げられる事を示した。Du ら [40] は Lee らと同様にオプティカルフロー、視差マップ、顕著性マップを使用して、3DCNN と Attention 層を組み合わせたモデルを提案した。その結果、全ての特徴量を用いた場合の PLCC が最も高く、0.963 だった。

以上のように、動画データからベクションを生じる箇所を特定するために、オプティカルフローを用いた研究が多い。また、顕著性マップについてもコンテンツの VR 酔いのリスクを推定する上で、有効な特徴量であることが示されている。

2.5 顕著性マップ

顕著性マップとは、画像を見る際に人がどこを注視しやすいのかをグレースケールの画像で表したものである。顕著性マップを生成するモデルとしては、ボトムアップ型とトップダウン型の手法がある。ボトムアップ型のモデルとはコントラスト、色、エッジといっ

た画像の低レベルな特徴量から顕著性マップを推定するモデルであり、神経回路に見られる中心-周辺抑制メカニズムを参考にしたもの [41] や、画像のスペクトルに関するスケール不変性を用いたもの [42] 等がある。そして近年では、顕著性マップのための大規模データセットの登場 [43, 44] により、深層学習によるトップダウン型のアプローチ [3] が主流となっている。ボトムアップ型の手法は計算コストが少なく済むが、精度はトップダウン型には及ばない。

VR 酔いと顕著性マップの関連について調べた研究として、Yang ら [2] によるものがある。Yang らは VR 酔い状態における顕著性マップを生成する方法を研究した。VR 酔いを誘発する VR 動画を被験者に提示した後、すぐに 360° 静止画像を見せることで VR 酔い状態での視線データの収集を行った。そして、既存の顕著性マップを生成する幾つかの深層学習モデルをファインチューニングする事で VR 酔い状態での顕著性マップを生成した。また、図 2.2 に示すように、VR 酔いでない状態に比べて VR 酔い状態では明らかに視覚的注意が分散することを報告していた。但し、Yang らは静止画像を提示した場合のみでしか実験をしておらず、VR ゲームなどのユーザが自由に VR 空間内を動き回れるような場合についても同様の結果が得られるのかについては調査が必要である。

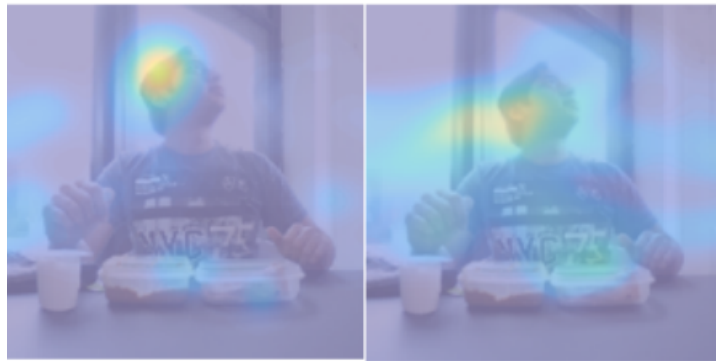


図 2.2 VR 酔い状態と非 VR 酔い状態での視線ヒートマップの比較 (文献 [2] より引用)

第3章 注視点周りの顕著性マップに着目したVR酔い推定

3.1 目的

VR酔い状態の人は、注意力が散漫になる [2, 45], あるいはVR酔いを低減するために特定の場所を注視する [46] など、普段とは異なる行動をとっていると考えられる。これらの行動を検出し、VR酔いの推定に活用するためには、動画データと注視点との関係が重要であると推測できる。そこで、視線と頭部姿勢とともに動画データとの関係をVR酔い推定に取り入れるため、顕著性マップを利用する。顕著性マップとは画像に対して、人がどこを注視しやすいのかを画像で表したものであり、動画データと視線の関係を表す特徴量として、VR酔い推定に有効な可能性がある。具体的には、画像上の注視点座標付近の顕著性マップの画素値を取得し、その平均値や標準偏差値、最大値、最小値を新たな特徴量として導入し、その有効性について検証した。

3.2 提案手法

本研究では、画像データと視線との関係を示す特徴量として、注視点座標付近の顕著性マップの画素値を取得し、その特徴量を考慮したVR酔いの程度を推定するモデルを作成する。

3.2.1 顕著性マップの生成方法

本研究では、深層学習による手法の中で、Droste らが提案した UNISAL [3] という顕著性予測モデルを利用して顕著性マップを生成した。UNISAL の概観を図 3.1 に示す。UNISAL は、スマートフォンなどに搭載されるような軽量なモデルである MobileNet V2 を用いて、転移学習をおこなうことで高精度かつ軽量の顕著性予測を可能にしている。UNISAL では MobileNet V2 による特徴量抽出をおこなうため、非常に低い解像度の画像で動作するように設計されている。しかし、HMD に表示されるような高解像度の画像をダウンサンプリ

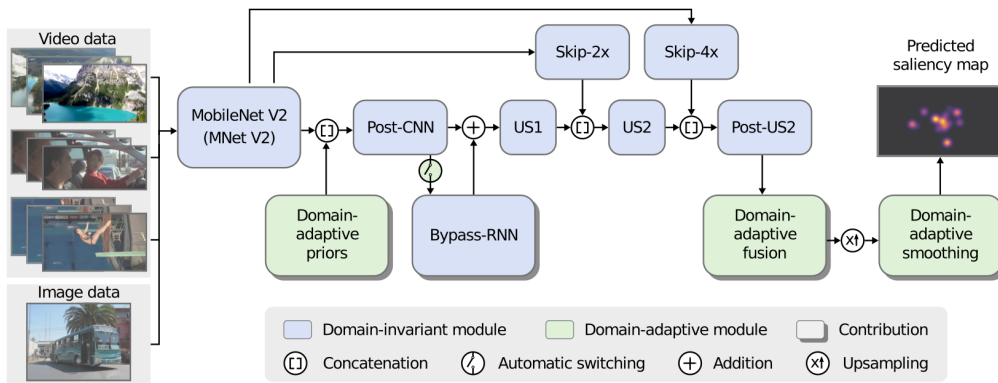


図 3.1 UNISAL (文献 [3] より引用)

ングすると、遠くにあるオブジェクトや人物を認識できず、顕著性マップが正しく予測されない可能性がある。そこで、本研究ではHMDに表示されている元画像に加えて、注視点を中心とした1/4、1/16の大きさで切り抜いた画像に対して顕著性予測モデルを適用した。図3.2は全体及び1/16サイズの画像と、それらの顕著性マップを示している。元画像

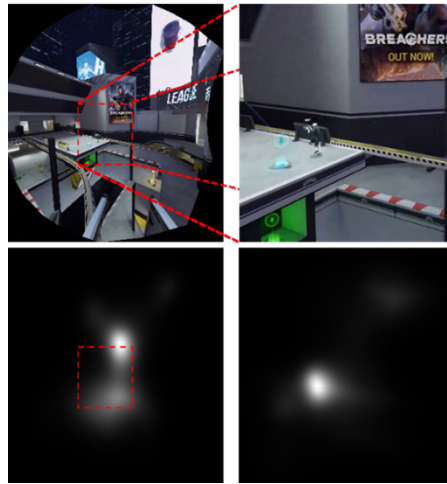


図 3.2 元画像と1/16サイズの顕著性マップ比較

では、奥にある看板が最も顕著性の高い領域で、注視点からの距離がやや大きくなってしまうのに対し、1/16サイズの画像では、人が最も顕著性の高い領域であり、注視点もその近くに存在することが分かる。

3.2.2 注視点周りの顕著性マップ統計量

顕著な領域に視線が合わせられているかを新たな特徴量とするため、図 3.3 に示されている様な注視点周りの 2.5° の範囲内にある顕著性マップの値を抽出し、最大値・最小値・平均値・標準偏差といった統計量を用いる。この 2.5° という範囲は、注視点と比べて視力がおよそ半分になる偏心度を表しており [47, 43]、人は興味のある物体に視線を合わせる場合やテキストを読む際に、この領域で細かい形状を認識している。なお、図は $1/16$ サイズの画像を使用している。

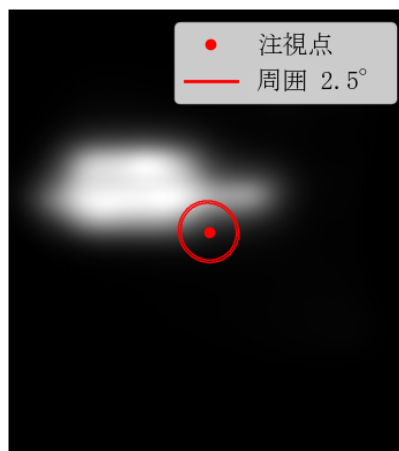


図 3.3 顕著性マップの注視点の 2.5° 範囲

3.2.3 使用する特徴量

使用する特徴量の一覧を表 3.1 に示す。視線情報については 8 次元の特徴量 (画像上の注視点座標, 目の開き具合, 瞳孔の直径), 頭部姿勢については 13 次元の特徴量 (位置座標, クォータニオン, 速度ベクトル, 角速度ベクトル) を使用した。顕著性マップについては左眼に表示された全体, $1/4$, $1/16$ サイズの画像, それぞれで注視点周り ($< 2.5^\circ$) の顕著性マップの画素値から平均値, 標準偏差値, 最大値, 最小値を計算し, 合計で 12 次元の特徴量を得る。これらのデータは 30Hz の時間解像度で収集され, 前処理として標準化がされる。VR 酔い状態は 10 段階 (0: 全く感じない, 9: 極めて強く感じる) の主観的な評価からなる。このような評価指標を FMS(Fast Motion Sickness Scale)[48] といい, リアルタイムな VR 酔いの評価が必要な際に多く用いられている。本研究では, VR 体験中に 1 分間隔で収集する。

表 3.1 特徴量一覧

モダリティ	データ
視線情報 (左, 右)	注視点座標 [px] 瞳孔直径 [mm] 目の開き具合 (0-1)
頭部姿勢	位置座標 [m] クォータニオン 速度ベクトル [m/s] 角速度ベクトル [rad/s]
顕著性マップ (左目用画像) 全体, 1/4, 1/16 サイズ	注視点周りの画素値の 平均, 標準偏差, 最大, 最小
VR 酔い度	FMS (0-9)

3.2.4 VR 酔い推定モデル

図 3.4 に本章で実装した VR 酔い推定モデルの概要を示す。このモデルは画像認識などで広く使われている ResNet[49] のアーキテクチャを基に、1 次元畳み込みを用いることで時系列予測に適応できるようにした ResNet1D というアーキテクチャである。ResNet は残差ブロックという構成要素を複数個繋げたモデルになっており、残差ブロックの入力を出力に足し合わせるという残差接続をおこなうことによって、より深いネットワークの学習を可能にしている。このモデルは 8 個の残差ブロックからなり、600 フレーム (20 秒×30Hz) 分の特徴量ベクトルから、その直後の VR 酔い状態を 3 段階 (低:0-2, 中:3-5, 高:6-9) で分類した結果を出力する。

エポック数 100, バッチサイズ 256, 損失関数は不均衡なデータに対応するために、重み付き交差エントロピー誤差を用いて学習させた。重みに各正解クラスのデータ数の逆数を掛けることで、クラスの不均衡を補正した。オプティマイザには Adam を使用し、学習率は 0.001 を初期値として各エポック毎に 0.95 倍に減衰させた。モデルの評価は層化 10 分割交差検証により行った。

データ量不足を補うため、フレームシフトによるデータ拡張を行った。具体的には、酔いの状態を回答した直前の 10 秒間は、VR 酔い状態が同じだと考え、その中でフレームシフトを行った。

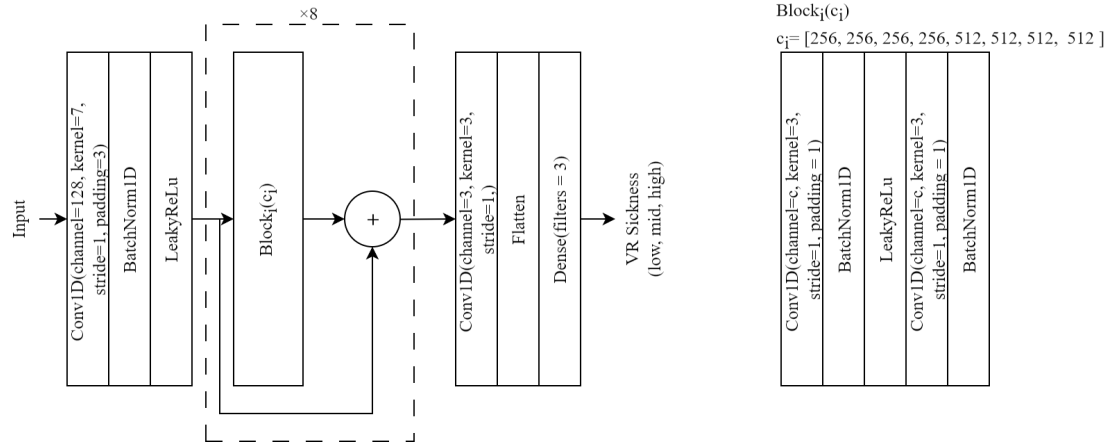


図 3.4 VR酔い推定モデルの概要

3.3 実験方法

実際のVR体験をデータセットに反映するため、市販のVRコンテンツを実行しながらバックグラウンドで動画データ、頭部姿勢、視線情報を収集することのできるデータ収集システムを構築した。そして、様々なインタラクションや移動方法(車、歩き、テレポーション)が可能なVRコンテンツとして、図3.8,3.6に示すシューティングゲーム、レースゲームの2つを選択した。実験は10分間のプレイの後、5分以上の休憩をするという一連の流れを繰り返すことで行われた。最終的に休憩を含めずに約12時間分のデータセットが集められた。なお、被験者は著者一名のみである。



図 3.5 シューティングゲーム
(HyperDash[4])



図 3.6 レースゲーム(KartKraft[5])

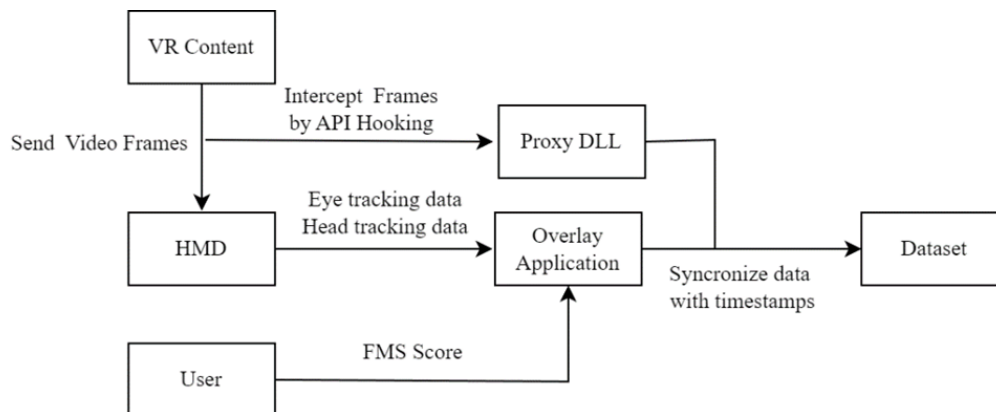


図 3.7 データ収集システムの概要

3.3.1 データ収集システム

HMD にはアイトラッキングが可能な HTC Vive Pro Eye を用いた。図 3.7 に実装したデータ収集システムの概要を示す。本実装で映像はプロキシ DLL を介した API フック、視線情報と頭部姿勢及び VR 酔いの度合いはオーバレイアプリケーションで収集する。OpenVR の `IVRCompositor::Submit` は HMD に表示されているシーンをアップデートする API であり、これをフックすることで HMD に表示される全ての動画フレームにアクセスし、それらが HMD に送信される正確なタイムスタンプを記録できる。本実装ではプロキシ DLL を介して API フックを行い、プレイ中の画面を 30Hz のフレームレートで保存した。続いて VR コンテンツに対して、バックグラウンドで実行可能なオーバレイアプリケーションを作成した。オーバレイアプリケーションの実装には Unity を使用しており、SteamVR を用いたアプリケーションであれば実行可能である。視線情報は VIVE Eye and Facial Tracking SDK、頭部姿勢は SteamVR の API により収集した。視線情報のセンサは 120Hz のサンプリングレートでデータの取得が可能であるため、オーバレイアプリケーションでは可能な限り多くのデータを収集し、画像データとタイムスタンプを合わせて 30Hz のサンプリングとなるように調整した。VR 酔いの状態は、図 5 に示すような画面を被験者に提示して 10 段階 (0: 全くなし, 9: 極めて強く感じる) で回答させた。この画面を 60 秒間隔で表示し、被験者はその時点での酔いの度合いを回答する。なお、回答中は VR 酔いの進行を防ぐため、被験者には VR コンテンツが見えないようにしている。

現在の酔いの度合いを吐き気、不快感、胃の状態
を考慮して0～9の数値で教えてください

全くなし 2 極めて強く感じる

図 3.8 VR酔いアンケート

3.4 実験結果

3.4.1 データ分析

注視点周りの顕著性マップ画素値の最大値について、各VR酔い度の回答の直前600フレームを平均し、それらを集計して分析を行う。注視点周りの顕著性マップ画素値の統計量として、最大値の他に最小値、平均値、標準偏差があるが、ここでは顕著な領域に視線を合わせることが出来ているかという点に着目して、最大値のみを取り扱う。なお、同一被験者の時系列データであるため、独立とはみなせず、自己相関を伴う可能性があるため、検定の信頼性については注意が必要である。図4.7に、VR酔い度別の注視点周りにおける顕著性マップの画素値の最大値について箱ひげ図を示す。図より、1/16サイズの画像ではVR酔い度が上昇するにつれて注視点周りの顕著性マップ最大値が減少する傾向にあり、VR酔いによって顕著な領域に視線を合わせること難しくなっていることが観察できる。元画像及び注視点を中心に1/4サイズ、1/16サイズで切り抜いた画像から顕著性マップを作成した場合のそれぞれについて、ボンフェロー二補正を適用したKruskal Wallis検定を行った結果、元画像と1/4サイズの画像については有意差が認められなかったものの、1/16サイズの画像では有意差が見られた。元画像： $(H(2) = 1.63, p_{corr} = 1.33)$ 、1/4サイズ： $(H(2) = 1.11, p_{corr} = 1.72)$ 、1/16サイズ： $(H(2) = 16.37, p_{corr} = 0.00083)$ 。1/16サイズの画像について、Mann-Whitney U検定を用い、ボンフェロー二補正を適用した多重比較検定を行った結果、低群と高群($p_{corr} = 0.00083$)、中群と高群($p_{corr} = 0.022$)で有意差が認められた。

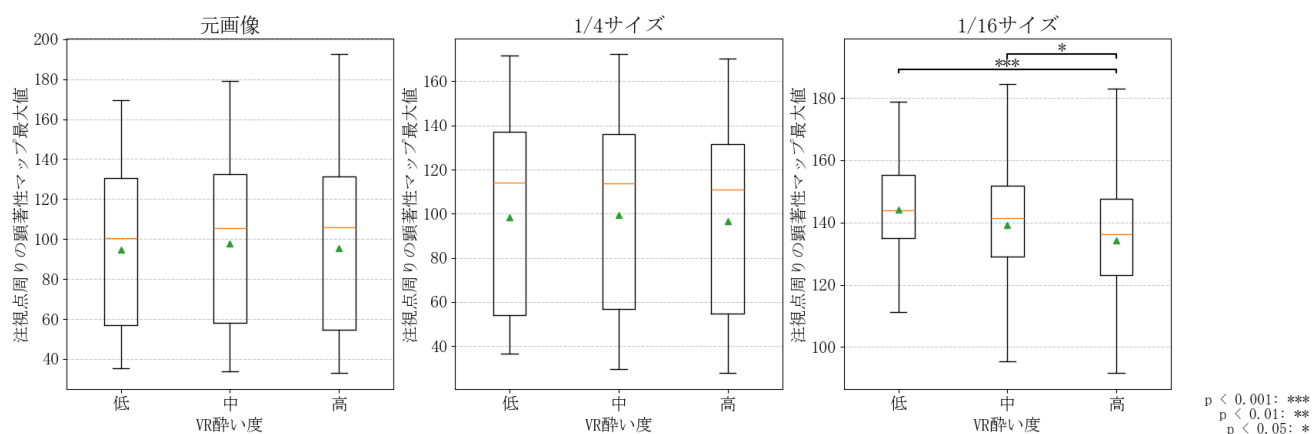


図 3.9 VR酔い度別の注視点周りにおける顕著性マップ最大値

3.4.2 VR酔い推定結果

表 3.2 に各特微量を用いて VR酔いの度合いを推定した結果を示す。全ての特微量を用いた場合が最も精度が高く 62.2%となった。視線情報と頭部姿勢のみを用いた場合と比べて、2.9%推定精度が向上した。そして視線情報、頭部姿勢、顕著性マップそれぞれのモダリティのみを用いて推定した場合の精度は 56.9%, 58.5%, 44.3%となった。

表 3.2 各特微量毎の推定結果

使用した特徴量					評価指標			
視線情報	頭部姿勢	顕著性マップ			精度	適合率	再現率	F 値
		full	1/4	1/16				
○					0.569	0.502	0.509	0.501
	○				0.585	0.529	0.544	0.531
○	○				0.593	0.525	0.505	0.509
		○	○	○	0.443	0.411	0.448	0.406
○	○	○	○	○	0.622	0.557	0.532	0.539

3.5 考察

データ分析の結果より、VR酔いの症状が強くなるほど顕著な領域に対して視線が合わせられなくなっている傾向が観察できた。また異なる画像サイズ間では、画像サイズが小さくなるほど、それぞれ増加傾向にあることが観察できる。画像を切り取ることによって、

遠くの物体に対して顕著性を正しく予測できるようになったことに加えて、周囲の顕著性の高い領域がなくなり、相対的に視線を合わせようとしているオブジェクトの顕著性が高まったことが原因だと考えられる。

先行研究 [13, 14] では動画データは推定にほとんど寄与しないとされていたが、本研究では顕著性マップから新たな特徴量を導入することにより、推定精度を 2.9% 向上させることができた。これらの特徴量は比較的単純なものであり、顕著性予測モデルの精度にも依存することから、動画データを十分に活用できていなかった可能性が高い。それにもかかわらず、推定精度の改善が見られたことから、動画データが有効な特徴量である可能性が示唆された。

また、各モダリティのみを用いた推定結果を比較すると、頭部姿勢が最も重要な特徴量であることが観察できる。そして頭部姿勢のみを用いた場合と頭部姿勢と視線情報を用いた場合で比較すると、精度は 0.8% しか増加せず、F 値は 0.022 減少しているため、視線情報からは有用な情報が得られなかったと考えられる。

先行研究 [14] では、注意機構の重みから特徴量の重要度を計算した結果、頭部姿勢ではなく視線情報が最も重要な特徴量だとしており、本実験の結果とは異なる。また、全ての特徴量を用いた場合の推定精度に関しては、先行研究 [13, 14] の推定精度はそれぞれ 87.7%, 87% (4 段階分類) だったため、本実験の推定精度はかなり低いと言える。これらはデータセットのサイズや性質、被験者が一人であったこと等が影響している可能性がある。本来、被験者は少ないほど推定がしやすくなるはずである。しかし、VR 酔いは症状の出やすさに大きく個人差があり、先行研究 [13, 14] では被験者を区別しない分割交差検証によって評価していたため、症状が出やすい人と出にくい人を分類することによって、推定精度が上がっている可能性がある。

第4章 顕著性マップとオプティカルフローに着目したVR酔い推定

4.1 目的

本章では、前章で提案した特徴量に加えて、オプティカルフローを導入することを検討した。コントローラ入力による移動中など、ベクションが発生する状況では急激なVR酔いを感じる可能性がある。このような状況で、ユーザが顕著な領域に視線を合わせられているかといった反応を捉えることで、VR酔い推定の精度を上げられる可能性がある。オプティカルフローは、ユーザが受ける視覚的な刺激を定量化できるため、VR酔いを感じやすい状況を特定するために有効である。具体的には、k-meansによるオプティカルフローのクラスタリングを行い、VR空間内でのユーザの動きを大まかに分類し、さらに放射基底関数を用いて類似度計算を行い、データ分析やVR酔い推定のための特徴量とした。

前章では、被験者1名で実験を行い、顕著性マップと視線との関係をVR酔い推定に取り入れることで、僅かに推定精度を向上することが出来た。しかし、VR酔いは個人差やVRゲームの特性による影響が大きく、VR酔いを感じやすい人とそうでない人、またタスクによって視線のパターンが大きく異なることが予想されるため、より多数の被験者やVRゲームからなるデータセットで評価を行う必要がある。そこで、本章では多数の被験者とVRゲームを含むデータセットであるVR.net[15]を使用することにした。

前章では、注視点周りの顕著性マップの画素値の統計量という非常に単純な特徴量を使用していた。そこで、動画データと注視点の複雑な関係をより直接的に推定へ反映するため、注視点から中心視・周辺視野といった人の視覚システムを表すような画像を生成し、これを動画データと顕著性マップやオプティカルフローと統合させた。これらを効率的に処理するため、動画データの時空間的特徴を抽出する3DCNNベースのX3D[16]を採用した。その他の頭部姿勢、視線方向などの多変量時系列データについては、Crossformer[50]により特徴抽出を行った。このモデルは、時間軸と特徴量軸それぞれの依存関係を注意機構で学習するTSA層を持ち、VR酔い推定における視線情報、頭部姿勢、注視点周りの顕著性マップ、オプティカルフロー間の複雑な関係を効果的にモデル化できる可能性がある。

以上を踏まえ、本章ではオプティカルフローと顕著性マップを交えた、より詳細な分析を行う。そして、Crossformer と X3D による VR 酔い推定の有効性を検証する。

4.2 使用するデータセット

本章では、2024年に公開されたVR酔いのためのデータセットであるVR.netを使用する。VR.netは、アクション、レース、シミュレーション、スポーツ等の代表的なジャンルから多くのVRゲームを選定し、そのVRゲームを被験者に1回のセッションにつき最大15分間プレイさせることでデータを収集している。VR.netにおける全てのデータはMeta Quest Pro[51]を使用して収集されている。現在公開されているVR.netは、70個のVRゲーム、被験者数253名、493セッションからなるデータセットとなっており、利用可能なデータの一覧は表4.1の通りである。視線情報として、眼球の位置座標と、視線方向を表

表 4.1 VR.net の利用可能なデータ一覧

種類	特徴量
視線情報	位置座標 [m] クォータニオン 信頼度 (0-1)
頭部姿勢	位置座標 [m] クォータニオン 速度ベクトル [m/s] 角速度ベクトル [rad/s]
入力情報	ボタン ジョイスティック (X/Y, -1.0~1.0)
オブジェクト情報	名前 3D バウンディングボックス モデル行列
カメラ情報	ビュー行列 プロジェクション行列
動画データ	左目用画像 (解像度: 480 × 528)
VR 酔い度	FMS(1-5)

すクォータニオン、視線情報の信頼度がある。頭部姿勢としては位置座標、クォータニオン、速度ベクトル、角速度ベクトルがある。入力情報には、押下されたボタンとジョイスティックの動きを2次元で表した値がある。オブジェクト情報としては、オブジェクトの

名前と3Dバウンディングボックス、モデル行列がある。カメラ情報として、左目用カメラのビュー行列、プロジェクション行列がある。動画データは左目用画像のみがある。VR酔い度はFMSを使用して、約1分間隔で5段階でのVR酔い度が収集され、その他のデータは約14Hzまたは7Hzで収集されていた。そのため7Hzに合わせてダウンサンプリングを行った。カメラ情報とシーン情報に関しては188セッションしか収集されていなかったため、本研究では使用しないことにした。また、セッション単位で視線情報や動画データに欠損がみられたものは使用せず、最終的に373セッションを使用することとした。

4.3 提案手法

本章では、動画データから顕著性マップとオプティカルフローを生成し、新たな特徴量とした。オプティカルフローについては、kmeansによるクラスタリングを行い、ユーザがどのような視覚的刺激を受けているかを、大まかに表すことのできる特徴量に変換した。顕著性マップとその注視点周りの画素値の統計量に関しては、前章と概ね同じ方法で生成されたが、VR.netの動画データが 480×528 という低解像度であったことから、 $1/16$ サイズの画像に対しては顕著性マップを生成せず、元画像と $1/4$ サイズの画像から顕著性マップを生成した。

VR酔いを推定するモデルとしては、多変量時系列データを処理するCrossformerと、動画データを直接処理するX3Dを用いる。また、注視点から動画データと注視点の複雑な関係をより直接的に推定へ反映するため、注視点から中心視・周辺視野といった人の視覚システムを表すような注視点画像を生成し、これを動画データと顕著性マップやオプティカルフローと統合させた上でX3Dに入力した。

4.3.1 オプティカルフローの生成と分類方法

本研究では、オプティカルフローを生成するために、高速なRLOF(Robust Local Optical Flow) アルゴリズム [52] を使用した。また、オプティカルフローに対して、k-means法を適用することで、ユーザがどのような視覚刺激を受けているのかの分類を行った。多くのVRコンテンツでは、図4.1のようにコントローラの操作などによる移動の際に視野を制限することでVR酔いを低減する機能がある。視野制限の有無があると、オプティカルフローの総量が大きく異なるため、まず視野が制限されている状態かどうかを二分し、それぞれの視野の条件に対してk-means法を適用することにした。また、オプティカルフローには外れ値が含まれる場合があり、極端に大きなオプティカルフローが観測されることがある。

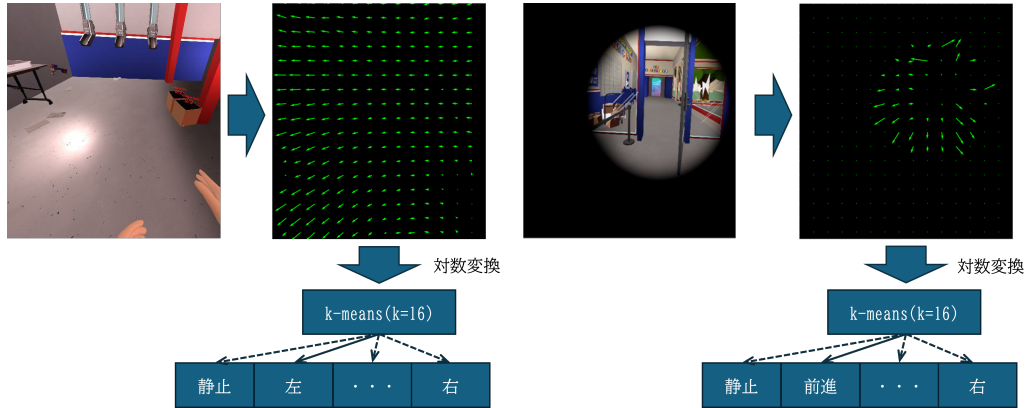


図 4.1 視野の非制限, 制限状態でのオプティカルフローと k-means 法による分類

このような外れ値は k-means 法のクラスタリング結果に大きな影響を及ぼす可能性があるため、オプティカルフローを対数変換してからクラスタリングを行った。これにより外れ値の影響を抑え、より適切なクラスタリングが可能となる。また、オプティカルフローは計算量を削減するために、 30×33 にリサイズされた。k-means のクラス数 k は、分類の多様さから $k = 16$ に設定し、合計で 32 クラスとなった。各クラスには静止、前進、左、右などのラベルが割り当てられた。詳細な各クラスのラベル割り当ては付録図 A.1, A.2 に示す。

さらに k-means 法によって求められた各クラス重心と類似度を計算することで、より柔軟な特徴量となるようにした。類似度計算には以下の式 4.1 で表される放射基底関数を用いた。ここで \mathbf{x}' はクラス重心を表し、 \mathbf{x} は各オプティカルフローを表す。調整可能なパラメータ σ はユークリッド距離で計算されたペアワイズ距離分布の下位 5% パーセンタイルに設定された。

$$K(\mathbf{x}, \mathbf{x}') = \exp \left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\sigma^2} \right) \quad (4.1)$$

4.3.2 注視点画像

3DCNN などを用いて、動画データや顕著性マップ、オプティカルフローと視線の関係を直接学習するためには、画像上に注視点をマッピングすることが有効な可能性がある。人の視覚システムは網膜偏心度に応じて、物体の輪郭や色の濃淡を識別できる感度が反比例的に低下することが知られている [53]。この特性に基づき、文献 [43, 47] では周辺視野で知覚できる解像度の低下を以下の式 4.2 で近似しており、本研究でも同様の手法を用いて人の視覚特性を模倣する注視点画像を生成した。

$$R(\theta) = \frac{\alpha}{\alpha + \theta}, \alpha = 2.5^\circ \quad (4.2)$$

$R(\theta)$ は網膜偏心度が θ の場所における相対的な解像度を表している。 $\theta = \alpha$ の時、 $R(\theta) = \frac{1}{2}$ となるため、 α は注視点と比べて解像度がちょうど $\frac{1}{2}$ になる偏心度を表し、 α は約 2.5° となることが知られている。この式を基に値を 1~255 に正規化し、図 4.2 のような注視点画像を生成した。



図 4.2 生成された注視点画像

4.3.3 X3D+Crossformer による VR 酔い推定モデル

本章で実装した VR 酔い推定モデルを図 4.3 に示す。このモデルは動画データ、顕著性マップ、オプティカルフロー、注視点画像を入力とし、時空間的な特徴を抽出する X3D、多変量時系列データである頭部姿勢、視線情報、入力情報、注視点周りの顕著性マップ画素値の統計量、オプティカルフローのクラスタ重心類似度を処理する Crossformer の 2 つのモジュールから構成される。これらのモデルの出力を全結合層により統合し、VR 酔い度を推定する。頭部姿勢については位置座標、クォータニオン、速度ベクトル、角速度ベクトルの計 13 次元の特徴量、視線情報については注視点座標、位置座標、クォータニオン、信頼度の計 12 次元の特徴量、入力情報についてはジョイスティックの動きのみの 4 次元の特徴量を使用した。入力情報として、ボタンの情報を含めなかったのは、VR ゲームによってボタンが異なる意味をもつためである。入力フレーム数は 300 フレームとし、画像サイズについては計算量を削減するため 60×66 にリサイズしている。1/4 サイズの画像から生成された顕著性マップは、 30×33 にリサイズし、 60×66 となるようにゼロパディングを行った。

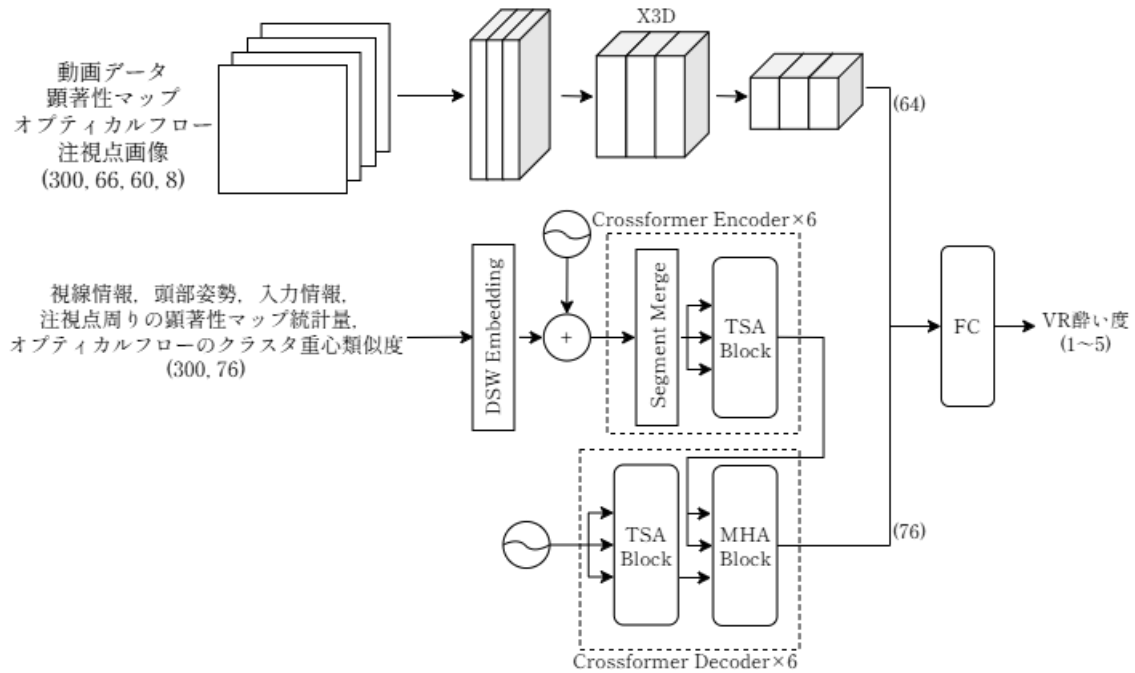


図 4.3 X3D+Crossformer による VR 酔い推定モデル

X3D は、動画データを効率的に処理する 3DCNN ベースのモデルの一つであり、2D CNN から 3 次元方向へと拡張を行うことで、動画データの時空間情報を効率的に捉えることを可能にしている。具体的には、入力動画のフレーム列に対して、ネットワークの深さやチャンネル数、さらに時間方向の畳み込みサイズを段階的に拡張していく設計が特徴であり、少ないパラメータでより高い性能をもつ。

Crossformer は多変量時系列予測のための Transformer ベースのモデルの一つである。従来の Transformer では、同じ時間の複数の特徴量を 1 つのベクトルに埋め込み、注意機構によって時間の依存関係を学習していた。それに対して、Crossformer は一定の時間窓で各特徴量毎に埋め込みを行う DSW(Dimension-Segment-Wise) Embedding、時間軸と特徴次元軸それぞれの依存関係を注意機構で学習する TSA(Two Stage Attention) 層を導入している。そのため、時間に加えて特徴量間の関係をより捉えやすいという特長がある。VR 酔い推定では、視線情報や頭部姿勢、顕著性マップとオプティカルフローなどの間にある複雑な関係を捉えることが有効であると考えられるため、VR 酔い推定に適したモデルとなる可能性がある。

4.3.4 ハイパーパラメータ

本実装における, X3D および Crossformer の主要パラメータについて表 4.2 に示す. Crossformer では埋め込みの次元数 (d_{model}) を 128 に, エンコーダの層数 (e_{layers}) を 6 に, フィードフォワード層の中間次元数 (d_{ff}) を 128, 埋め込みの時間窓の長さ (seg_len) を 12 に, アテンション層におけるヘッド数 (n_heads) を 16, 計算量を削減するためのアテンション層のスケーリングファクター ($factor$) を 12 に設定し, ドロップアウト率 ($dropout$) は 0.2 に設定した. また X3D については PyTorchVideo[54] による実装を使用して, パラメータは X3D-S の初期設定を基に, 時間方向の計算量を削減するために $stem_conv_stride, stage_temporal_stride$ を $(2, 2, 2), (2, 2, 2, 2)$ に調節した.

表 4.2 主要なハイパーパラメータ一覧

モジュール	パラメータ	値
Crossformer	d_{model}	128
	e_{layers}	6
	d_{ff}	128
	seg_len	12
	n_heads	16
	$factor$	12
	$dropout$	0.2
X3D	$stem_conv_stride$	$(2, 2, 2)$
	$stage_spatial_stride$	$(2, 2, 2, 2)$
	$stage_temporal_stride$	$(2, 2, 2, 2)$
	out_dim	64

4.3.5 学習設定

VR酔い度 1~5 を -1~1 の範囲に正規化し, 損失関数には MSE を用いて学習させた. エポック数 64, バッチサイズは 16 で 8 回の勾配累積の後にパラメータを更新する方式を採用した. オプティマイザには Adam を使用し, 学習率は初期値として 0.001 を設定した. また, 過学習を防ぐための L2 正則化の係数として $1.0e^{-10}$ を設定した.

そして学習率は cosine annealing によるスケジューリングを行い, 各エポックにおいて初期学習率 0.01 から 0.0001 までを徐々に減少させる設定とした. スケジューリングの周期は, 全エポック数 64 に対応させた.

入力データについてはすべて標準化を行った上で入力された。データ量不足を補うため、フレームシフトによるデータ拡張を行った。具体的には、酔いの状態を回答した前後10秒間は、VR酔い状態が同じだと考え、その中でフレームシフトを行った。モデルの評価には、ユーザを区別した5分割交差検証を用いた。

4.4 データ分析

視線情報と頭部姿勢及び、注視点周りの顕著性マップ画素値の最大値について、VR酔い度を低(1,2)、中(3)、高(4,5)の3群に再分類し、各VR酔い度でセッション毎に集計して分析を行った。また、3群いずれかのカテゴリが欠損している場合には除外した。本分析では、全373セッションのうち89セッションのみが使用された。データ除外率が高い理由は、全ての酔いカテゴリを網羅していない被験者・セッションが多かったためである。

4.4.1 視線情報の分析

図4.4には、VR酔い度別の視線のピッチおよびヨー方向の角速度の絶対値 ($|\omega_{ex}|, |\omega_{ey}|$) に関する箱ひげ図を示す。図4.4から、VR酔い度が上昇するにつれて $|\omega_{ex}|$ はやや増加する傾向がある一方で、 $|\omega_{ey}|$ については、明確な変化はみられなかった。これは図2.2の結果と一致しており、VR酔い度が上昇するほど視線が分散することを示唆している。また、 $|\omega_{ex}|, |\omega_{ey}|$ のそれぞれについてFriedman検定を独立に行い、ボンフェローニ補正をした結果、いずれも統計的に有意差が認められなかった ($\chi^2(2) = 6.13, p_{corr} = 0.092$, $\chi^2(2) = 1.17, p_{corr} = 1.12$)。

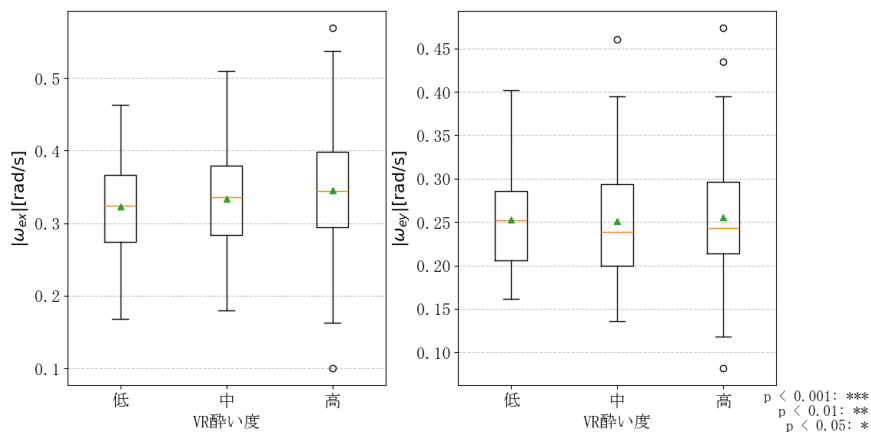


図 4.4 VR酔い度別の視線のヨー、ピッチ方向の角速度の絶対値

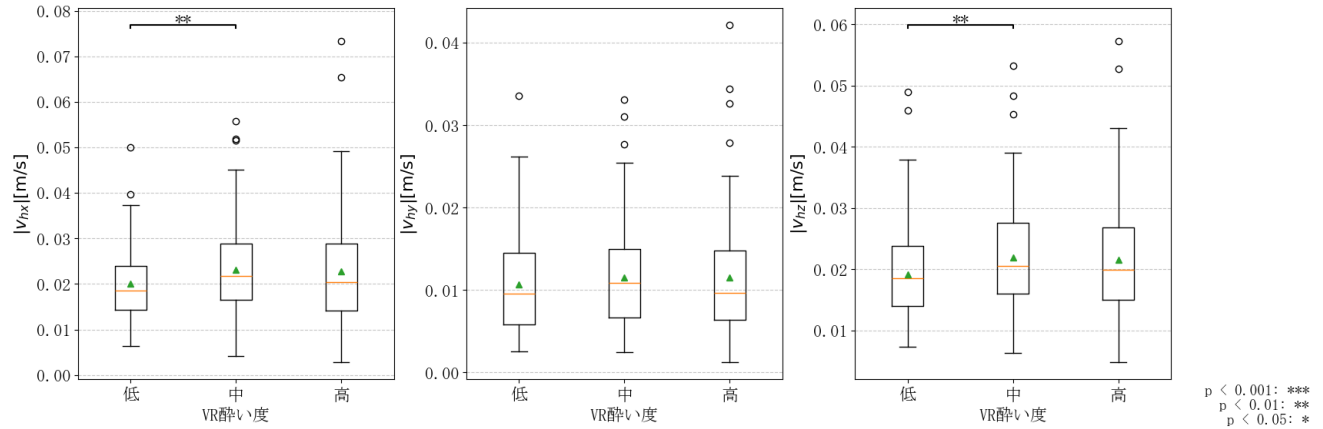
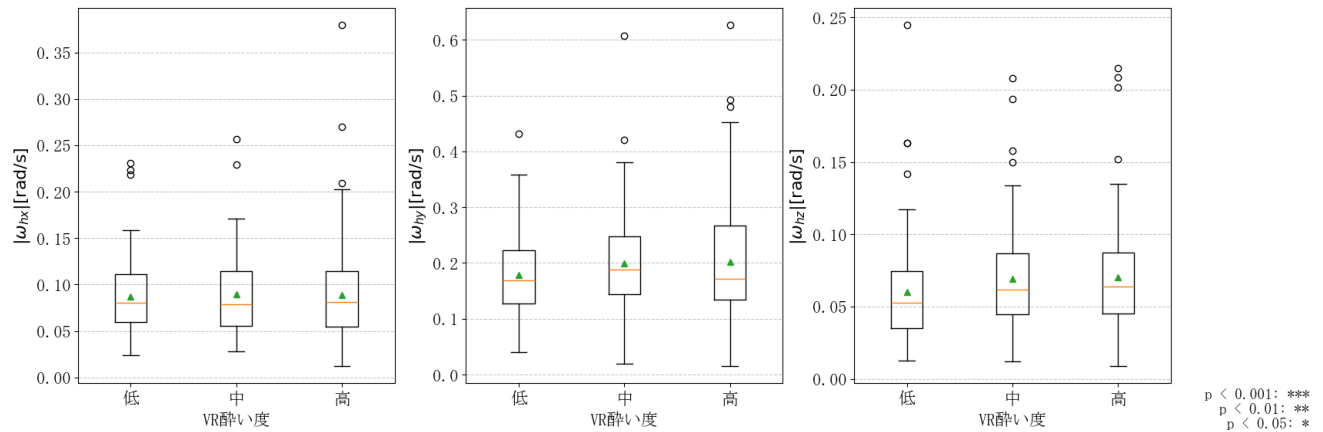
4.4.2 頭部姿勢の分析

頭部姿勢からはVR酔い度別の頭部速度の絶対値 ($|v_{hx}|, |v_{hy}|, |v_{hz}|$) と角速度の絶対値 ($|\omega_{hx}|, |\omega_{hy}|, |\omega_{hz}|$) の各軸について箱ひげ図をプロットしたものを、図4.5, 4.6に示す。VR酔い度が上昇するにつれて、 $|\omega_{hx}|$ 以外の $|v_{hx}|, |v_{hy}|, |v_{hz}|, |\omega_{hy}|, |\omega_{hz}|$ については平均値が上昇する傾向にあることが観察された。これはVR酔いによって、平衡感覚が乱れたことを示唆している。全ての特微量について、Friedman 検定を独立に行い、ボンフェロー二補正を適用した結果、 $|v_{hx}|, |v_{hz}|$ については統計的に有意差が認められた ($\chi^2(2) = 14.11, p_{corr} = 0.0052$, $\chi^2(2) = 11.15, p_{corr} = 0.023$) が、 $|v_{hy}|, |\omega_{hx}|, |\omega_{hy}|, |\omega_{hz}|$ については統計的に有意差が認められなかった ($|v_{hy}| : \chi^2(2) = 3.87, p_{corr} = 0.87$, $|\omega_{hx}| : \chi^2(2) = 2.79, p_{corr} = 1.49$, $|\omega_{hy}| : \chi^2(2) = 4.67, p_{corr} = 0.58$, $|\omega_{hz}| : \chi^2(2) = 6.09, p_{corr} = 0.29$)。有意差が確認された $|v_{hx}|, |v_{hz}|$ について、ボンフェロー二補正を適用したNemenyi 検定を行った結果、 $|v_{hx}|$ は低群と中群 ($p_{corr} = 0.0010$)、 $|v_{hz}|$ についても低群と中群 ($p_{corr} = 0.0084$) で有意差が認められた。

4.4.3 顕著性マップの分析

図4.7に、VR酔い度別の注視点周りにおける顕著性マップ画素値の最大値について箱ひげ図を示す。図より、VR酔い度が上昇するにつれて注視点周りの顕著性マップ画素値の最大値が減少する傾向にあり、VR酔いによって視覚的注意が分散していることを示唆している。元画像及び注視点を中心に1/4サイズで切り抜いた画像から顕著性マップを作成した場合で、Friedman 検定を独立に行い、ボンフェロー二補正を適用した結果、いずれも統計的に有意差が認められた ($\chi^2(2) = 14.60, p_{corr} = 0.0013$, $\chi^2(2) = 11.63, p_{corr} = 0.0060$)。また、ボンフェロー二補正を適用したNemenyi 検定による群間の比較を行った結果、元画像の場合は低群と高群 ($p_{corr} = 0.00040$)、1/4サイズの画像の場合では低群と高群 ($p_{corr} = 0.0021$) で有意差が認められた。

次に、各オプティカルフローのクラスタリング別で比較を行う。図4.8に、オプティカルフローのクラスで分けた際のVR酔い度別注視点周りの顕著性マップ最大値を一部抜粋して示す。また、図4.8に示しているもの以外の場合については、付録図A.3, A.4に示す。図4.8は、右側がk-meansを用いてオプティカルフローをクラスタリングした際のクラスタ中心を可視化させたものであり、左側がそのクラスに対応するVR酔い度別注視点周りの顕著性マップ画素値の最大値である。顕著性マップは元画像から生成したものを利用した。図4.8より、VR酔い度が上昇するにつれて注視点周りの顕著性マップ画素値の最大値が減

図 4.5 VR 酔い度別の頭部の位置速度 ($|v_{hx}|$, $|v_{hy}|$, $|v_{hz}|$)図 4.6 VR 酔い度別の頭部の角速度 ($|\omega_{hx}|$, $|\omega_{hy}|$, $|\omega_{hz}|$)

少する傾向にあることが観察でき、特に非視野制限状態の静止以外のオプティカルフローの場合に、その傾向が強まることが観察できた。一方で視野制限状態では、VR 酔いによる減少傾向が見られず、顕著な領域に視線を合わせることが難しくないことを示唆している。また、オプティカルフローの32クラス全てにおいてFriedman検定を独立に行い、Benjamini-Hochberg法による補正をした結果、非視野制限状態の左方向、非視野制限状態の左上方向を表すクラスで有意差が認められた(非視野制限, 左: $\chi^2(2) = 15.61, p_{corr} = 0.0065$, 非視野制限, 左上: $\chi^2(2) = 17.61, p_{corr} = 0.0048$)。また、有意差が認められたクラスで、Nemenyi検定による群間の比較を行った結果、非視野制限状態の左方向を表すクラスでは、低群と中群 ($p_{corr} = 0.0064$)、低群と高群 ($p_{corr} = 0.0032$) で有意差が認められ、非視野制限状態の左上方向を表すクラスでは低群と高群 ($p_{corr} = 0.00048$) で有意差が認められた。

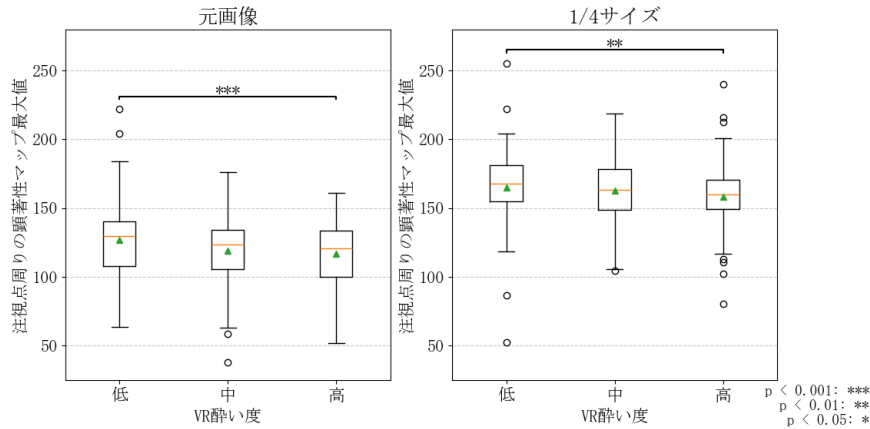


図 4.7 VR酔い度別の注視点周りにおける顕著性マップ最大値

4.5 実験結果

4.5.1 モジュール間の比較

表 4.3 に、動画データを処理する X3D、多変量時系列データを処理する Crossformer のそれぞれで、VR酔い推定を行った結果と、統合して VR酔い推定を行った結果について示す。Crossformer のみを用いた場合が最も RMSE が低く 0.94 となり、動画データを直接処理する X3D はむしろ推定精度を下げるという結果になった。

表 4.3 各モジュールの性能評価

モジュール	RMSE	MAE	PLCC
Crossformer	0.94	0.73	0.27
X3D	0.96	0.78	0.14
Crossformer+X3D	0.95	0.74	0.24

4.5.2 特徴量間の比較

表 4.4 に各特徴量を用いて VR酔いの度合いを推定した結果を示す。モジュールとしては、Crossformer のみを使用しており、表 4.4 の顕著性マップは注視点周りの顕著性マップ画素値の統計量、オプティカルフローは kmeans によるクラスタリング重心との類似度を表している。全ての特徴量を用いた場合に最も RMSE が低く 0.94 となった。また、視線情報、頭部姿勢、顕著性マップそれぞれを単体で用いた場合、顕著性マップが最も RMSE

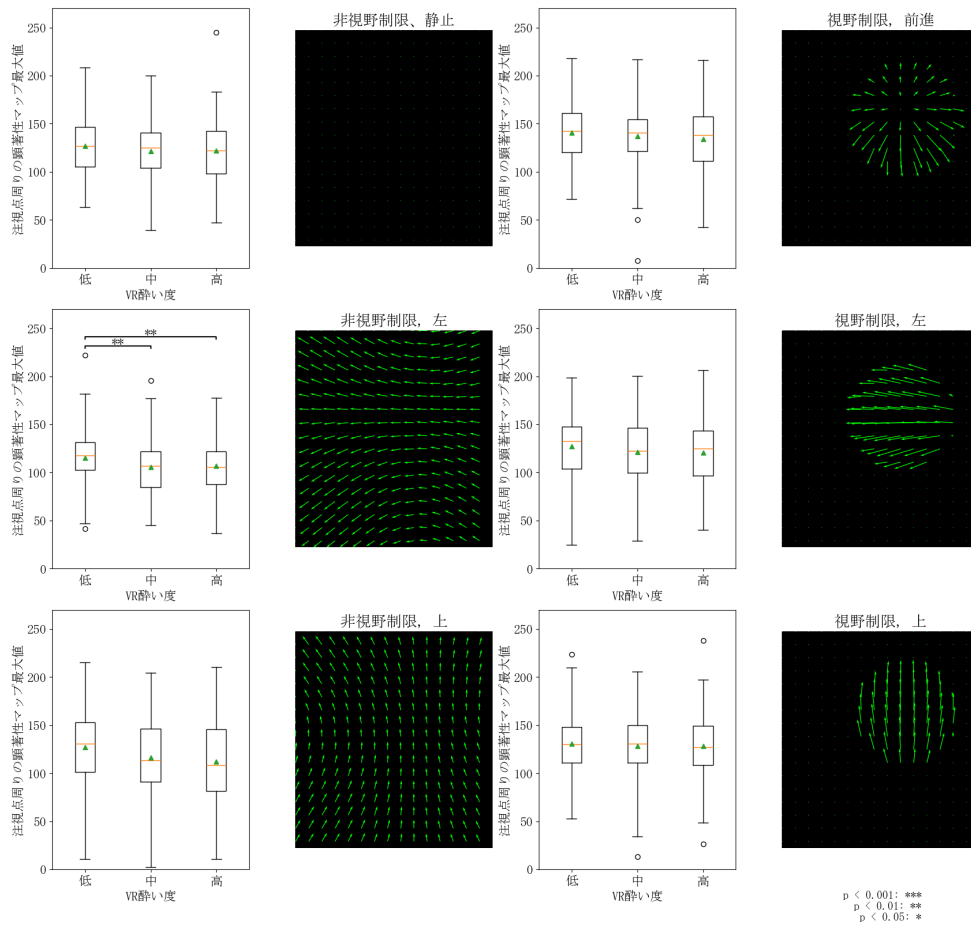


図 4.8 オプティカルフローによるクラスタリング別の注視点周りにおける顕著性マップの最大値

が低く 0.96 となった。しかし、視線情報、頭部姿勢、入力情報を用いた場合と、全ての特微量を用いた場合で比較すると RMSE は 0.01 しか減少せず、ほとんど変化しなかった。

4.5.3 先行研究との比較

表 4.4 に提案手法及び先行研究 [13, 34] のモデルを再実装し、VR.net に適用した結果を示す。先行研究 [34] では視線情報と頭部姿勢の他に生理学的データを用いていたが、VR.net には生理学的データが含まれていないため、視線情報と頭部姿勢のみを入力するようにした。また、先行研究 [13] では視線情報と頭部姿勢のみを用いており、この比較でも同様に視線情報と頭部姿勢のみの入力とした。表 4.4 に示すように、提案手法によって、すべての評価指標について改善が見られ、特に PLCC においては、既存手法と比べて大幅な向上

表 4.4 各特徴量ごとの推定結果

特徴量					評価指標		
視線情報	頭部姿勢	入力情報	顕著性マップ	オプティカルフロー	RMSE	MAE	PLCC
○					0.98	0.76	0.20
	○				0.98	0.75	0.19
○	○				0.95	0.75	0.21
○	○	○			0.95	0.74	0.25
			○		0.96	0.77	0.19
			○	○	0.95	0.77	0.21
○	○	○	○	○	0.94	0.73	0.27

が確認された。

表 4.5 先行研究との比較

モデル	RMSE	MAE	PLCC
Islam et al.[13]	0.97	0.77	0.12
Jeong et al.[34]	0.96	0.79	0.13
提案手法	0.94	0.73	0.27

4.6 考察

データ分析の結果から、注視点周りの顕著性マップ画素値の最大値について、VR酔いの症状が強くなるほど減少していくという結果が得られた。これはVR酔いの症状が強くなるほど、顕著な領域に視線が合わせられなくなっていることを示している。前章では、1/16サイズの画像の場合についてのみ有意差が認められたが、本章では元画像、1/4サイズの画像の場合について有意差が認められた。これについては、多様な被験者やVRゲームを含むVR.netによって、データの偏りが緩和されたためと考えられる。

また、オプティカルフローのクラスで分類した際の注視点周りの顕著性マップ画素値の最大値について、視野制限状態よりも非視野制限状態を表すクラスで、VR酔いの症状による減少傾向が強まることが観察された。この現象については、非視野制限状態では視野が広いこと、顕著な領域が分散しやすく、急に視野内に入ってきたオブジェクトに反応することが求められるため、視線を顕著な領域に合わせることが難しくなっていることが一因だと考えられる。また、非視野制限状態のオプティカルフローは、VR動画や慣性のような受動的な動きがない場合、頭部の動きによって生じるオプティカルフローが大部分を占

める。この場合、VR酔いによる頭のふらつきによって生じたオプティカルフローと、顕著な領域に視線を合わせようとする頭部の動きによって生じたオプティカルフローが混在している。特に前者の場合、視線が顕著な領域に合わないため、この割合が増えたことにより、注視点周りの顕著性マップ画素値の最大値の減少に影響を与えたと考えられる。

VR酔い推定の結果から、まず Crossformer と X3D のモジュール間の比較について、VR酔い推定には多変量時系列データと動画データを統合するよりも、多変量時系列データのみを処理の方が良いという結果が得られた。考えられる理由としては、動画データと視線の複雑な対応関係を学習するためには、より大規模なデータや高度なモデル設計が必要だった可能性や、ユーザがどのような刺激を受けているかという特徴量よりも、ユーザの頭部姿勢の変化や視線の挙動そのものがVR酔い度を直接反映していた可能性が挙げられる。またVRゲームでは、ユーザは酔いに応じて、刺激の量自体を制御することが出来てしまうという点も影響している可能性がある。

Crossformer のみを用いた場合の各特徴量毎の比較については、視線情報、頭部姿勢、入力情報を用いた場合と、全ての特徴量を用いた場合で比較すると RMSE は 0.01 しか減少せず、ほとんど変化しなかった。これは、注視点周りの顕著性マップ画素値の統計量という特徴量が、視線情報と頭部姿勢を合わせた特徴量と類似した傾向を持っている可能性を示唆している。オプティカルフローについても、VR動画のような受動的な刺激がない場合、ほとんどコントローラの入力や頭部の動きによって生じるオプティカルフローが大部分を占めるため、それらと類似した特徴量となり、効果がなかった可能性がある。先行研究で提案されたモデルとの比較では、PLCC について大幅な改善がみられた。これは Crossformer によって、視線情報や頭部姿勢などの特徴量間の依存関係や、長い時間の依存関係をうまく捉えることができ、推定精度の改善につながったと考えられる。

以上の結果から、顕著性マップやオプティカルフローはデータ分析を通じて、意味のある指標となりうることが示唆された。一方で、VR酔い推定モデルへの寄与はほとんどみられず、これらの情報を十分に活用できていない可能性や、特徴量として冗長になってしまっている可能性が考えられる。また、本研究のモデルは個人差やタスク特性を考慮していないため、これらの影響による視線挙動や頭部動作のばらつきを捉えきれていないことが課題として挙げられる。たとえば、ユーザの普段の視線のパターンを取り入れることで、視覚刺激に対する反応という特徴量をよりの確にモデル化できる可能性がある。また、タスク特性を考慮した場合、顕著性マップの有効性がタスクに依存する形で変化する可能性もあり、今後のさらなる検討が必要である。

第5章 結論

5.1 結論

本研究では、動画データと視線との関係を VR 酔い推定に取り入れるため、顕著性マップの導入を検討した。まず、データ分析によって注視点周りの顕著性マップ画素値を用いて、VR 酔いの症状が強くなるほど、視覚的注意が分散することを定量的に示した。またオプティカルフローの分類によって、特に非視野制限状態でオプティカルフローが観測される場合に、視覚的注意が分散する傾向が見られた。

また、注視点周りの顕著性マップの統計量及びオプティカルフローのクラスを VR 酔い推定のための特徴量として導入した結果、VR.net データセットにおいて RMSE:0.94, MAE:0.73, PLCC:0.27 となり、わずかにしか推定精度が向上しなかった。提案した特徴量は頭部姿勢、視線情報、コントローラの入力情報を統合した特徴量と類似した傾向を持つため、大きな精度向上にはつながらなかった可能性がある。また、先行研究とのモデル比較では PLCC が大幅に向上したことから、Crossformer を用いた多変量時系列データの処理が VR 酔い推定に有効である可能性が示唆された。そして動画データ、顕著性マップ、オプティカルフローと視線情報から直接的に、3DCNN のベースのモデルである X3D によって特徴量の抽出を試みたが、現状の手法・データでは推定精度を向上させるには至らなかった。

5.2 今後の課題

本研究では、動画データから顕著性マップやオプティカルフローを生成することで、ユーザがどのような映像を見ているのかという情報を抽出し、それに対する視線や頭部の動きのパターンを見ることによって VR 酔い推定精度の改善を図った。データ分析を通して、顕著な領域への視線集中が弱まる傾向が統計的に確認できた。一方で、最終的な推定モデルに組み込んだ際の精度向上は限定的であったため、現状ではどのような刺激が与えられたかと、それに対する反応をうまく VR 酔い推定に活用しきれていない可能性がある。

そこで、今後はユーザの普段の視覚刺激に対する視線や頭部の動きのパターンや、酔いやすい、酔いにくい体質差などに起因するばらつきをモデルに組み込むといった課題が考えられる。例えば、ファインチューニングを活用して被験者ごとにモデルを最適化するアプローチや、年齢、VRゲーム経験、酔いやすさを問うアンケートなど、容易に取得できる情報をモデルの入力とすることでも、個人差に起因する誤差を低減できる可能性がある。

また、本研究ではオプティカルフローと k-means 法により、被験者が受けている視覚刺激を大まかにクラスタリングを行った。しかし、被験者自身の頭部の動きによって生じるオプティカルフローと、それ以外のオプティカルフローを厳密に分離することは出来ていない。VR 酔いを生じやすいような視覚刺激をより正確に特定するためには、仮想空間内におけるカメラの移動量と、ユーザの頭部姿勢の変化と比較することで、視覚刺激の原因を明確に切り分ける必要がある。こうした区別が可能になれば、更に詳細な分析や VR 酔い推定精度を向上させることができる可能性がある。

さらに、本研究で使用した VR.net データセットでは多数の被験者と VR ゲームが含まれているものの、1セッションにつき最大15分程度と限られている。したがって、より長時間にわたってプレイした際の酔い度の変化などについては考慮できていない。より現実的な利用状況をカバーするためにも、長時間や連続したプレイのデータの収集や分析が必要である。

謝辞

本研究を進めるにあたり，多大なるご指導を頂いた指導教員の小川剛史准教授に，心より感謝申し上げます．特に，毎週のミーティングを通して，一つ一つ丁寧に改善点などをご指摘いただくとともに，研究の方向性を示して下さいました．そのご指導がなければ，本研究をここまで進めることはできませんでした．

また，小川研究室のミーティングに参加してくださり，的確なアドバイスを惜しみなく与えてくださった中山雅哉准教授にも深く感謝いたします．その温かいご支援に心からお礼を申し上げます．

また，同じ研究室に所属する博士課程の朴燦鎬さんには，本研究の進展において多大なるご協力をいただきました．論文の添削においても，多くの時間を割いていただきました．そのご厚意に心より感謝申し上げます．

加えて，本研究を遂行するにあたり，日頃より温かく支えてくださった小川研究室のメンバーの皆様にも感謝申し上げます．毎週のミーティングの議論や交流を通じて，多くの知見を得ることができました．

最後に日々私を支えてくださった家族や友人に，深く感謝し，お礼を申し上げます．

発表文献

国内会議 (査読なし)

1. 岡本 一真, 朴 燦鎬, 小川剛史. “顕著性マップと視線方向の対応関係に着目した VR 酔い推定手法の検討”, サイバースペースと仮想都市研究会, (Jun. 2024).

参考文献

- [1] Jaekyung Kim, Woojae Kim, Sewoong Ahn, Jinwoo Kim, and Sanghoon Lee. Virtual reality sickness predictor: Analysis of visual-vestibular conflict and vr contents. In *2018 tenth international conference on quality of multimedia experience (QoMEX)*, pp. 1–6. IEEE, 2018.
- [2] Jiawei Yang, Guangtao Zhai, and Huiyu Duan. Predicting the visual saliency of the people with vims. In *2019 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1–4. IEEE, 2019.
- [3] Richard Droste, Jianbo Jiao, and J Alison Noble. Unified image and video saliency modeling. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pp. 419–435. Springer, 2020.
- [4] Triangle Factory. HyperDash, February 2021. https://store.steampowered.com/app/1386890/Hyper_Dash/?l=japanese.
- [5] Motorsport Games. KartKraft, November 2018. <https://store.steampowered.com/app/406350/KartKraft/?l=japanese>.
- [6] JT Reason. Motion sickness, 1975.
- [7] Sue VG Cobb, Sarah Nichols, Amanda Ramsey, and John R Wilson. Virtual reality-induced symptoms and effects (vrise). *Presence: Teleoperators & Virtual Environments*, Vol. 8, No. 2, pp. 169–186, 1999.
- [8] EunHee Chang, InJae Hwang, Hyeonjin Jeon, Yeseul Chun, Hyun Taek Kim, and Changhoon Park. Effects of rest frames on cybersickness and oscillatory brain activity. In *2013 International Winter Workshop on Brain-Computer Interface (BCI)*, pp. 62–64, 2013.

- [9] J.J.-W. Lin, H.B.L. Duh, D.E. Parker, H. Abi-Rached, and T.A. Furness. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings IEEE Virtual Reality 2002*, pp. 164–171, 2002.
- [10] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, Vol. 3, No. 3, pp. 203–220, 1993.
- [11] Yi-Tien Lin, Yu-Yi Chien, Hsiao-Han Wang, Fang-Cheng Lin, and Yi-Pai Huang. 65-3: The quantization of cybersickness level using eeg and ecg for virtual reality head-mounted display. In *SID Symposium Digest of Technical Papers*, Vol. 49, pp. 862–865. Wiley Online Library, 2018.
- [12] Eugene Nalivaiko, Simon L Davis, Karen L Blackmore, Andrew Vakulin, and Keith V Nesbitt. Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiology & behavior*, Vol. 151, pp. 583–590, 2015.
- [13] Rifatul Islam, Kevin Desai, and John Quarles. Cybersickness prediction from integrated hmd’ s sensors: A multimodal deep fusion approach using eye-tracking and head-tracking data. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 31–40, 2021.
- [14] Dayoung Jeong, Seungwon Paik, YoungTae Noh, and Kyungsik Han. Mac: multi-modal, attention-based cybersickness prediction modeling in virtual reality. *Virtual Reality*, pp. 1–16, 2023.
- [15] Elliott Wen, Chitrlekha Gupta, Prasanth Sasikumar, Mark Billingham, James Wilmott, Emily Skow, Arindam Dey, and Suranga Nanayakkara. Vr. net: A real-world dataset for virtual reality motion sickness research. *IEEE Transactions on Visualization and Computer Graphics*, 2024.
- [16] Christoph Feichtenhofer. X3D: expanding architectures for efficient video recognition. *CoRR*, Vol. abs/2004.04730, , 2020.

- [17] James T Reason. Motion sickness adaptation: a neural mismatch model. *Journal of the royal society of medicine*, Vol. 71, No. 11, pp. 819–829, 1978.
- [18] Willem Bles, Jelte E Bos, Bernd De Graaf, Eric Groen, and Alexander H Wertheim. Motion sickness: only one provocative conflict? *Brain research bulletin*, Vol. 47, No. 5, pp. 481–487, 1998.
- [19] Stephen Palmisano, Robert S Allison, Mark M Schira, and Robert J Barry. Future challenges for vection research: definitions, functional significance, measures, and neural bases. *Frontiers in psychology*, Vol. 6, p. 193, 2015.
- [20] Behrang Keshavarz, Bernhard E Riecke, Lawrence J Hettinger, and Jennifer L Campos. Vection and visually induced motion sickness: how are they related? *Frontiers in psychology*, Vol. 6, p. 472, 2015.
- [21] Richard HY So, WT Lo, and Andy TK Ho. Effects of navigation speed on motion sickness caused by an immersive virtual environment. *Human factors*, Vol. 43, No. 3, pp. 452–461, 2001.
- [22] Lorenzo Terenzi and Peter Zaal. Rotational and translational velocity and acceleration thresholds for the onset of cybersickness in virtual reality. In *AIAA Scitech 2020 forum*, p. 0171, 2020.
- [23] Jiwon Kim and Taezoon Park. The onset threshold of cybersickness in constant and accelerating optical flow. *Applied Sciences*, Vol. 10, No. 21, p. 7808, 2020.
- [24] Simon Davis, Keith Nesbitt, Eugene Nalivaiko, et al. Comparing the onset of cybersickness using the oculus rift and two virtual roller coasters. In *Proceedings of the 11th Australasian conference on interactive entertainment (IE 2015)*, Vol. 27, p. 30. Australian Computing Society Sydney, Australia, 2015.
- [25] Matti Pouke, Arttu Tiirio, Steven M. LaValle, and Timo Ojala. Effects of visual realism and moving detail on cybersickness. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 665–666, 2018.
- [26] 藤木卓, 市村幸子, 寺嶋浩介, 小清水貴子. Vr コンテンツの精度が現実感と酔いに与える影響. *日本教育工学会論文誌*, Vol. 36, No. Suppl., pp. 73–76, 2012.

- [27] Juno Kim, Wilson Luu, and Stephen Palmisano. Multisensory integration and the experience of scene instability, presence and cybersickness in virtual environments. *Computers in Human Behavior*, Vol. 113, p. 106484, 2020.
- [28] Xiao Dong and Thomas A Stoffregen. Postural activity and motion sickness among drivers and passengers in a console video game. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 54, pp. 1340–1344. SAGE Publications Sage CA: Los Angeles, CA, 2010.
- [29] Yi-Chou Chen, Xiao Dong, Jens Hagstrom, and Thomas A Stoffregen. Control of a virtual ambulation influences body movement and motion sickness. In *BIO Web of Conferences*, Vol. 1, p. 00016. EDP Sciences, 2011.
- [30] Weina Jin, Jianyu Fan, Diane Gromala, and Philippe Pasquier. Automatic prediction of cybersickness for virtual reality games. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pp. 1–9. IEEE, 2018.
- [31] Sangmin Lee, Seongyeop Kim, Hak Gu Kim, and Yong Man Ro. Assessing individual vr sickness through deep feature fusion of vr video and physiological response. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 32, No. 5, pp. 2895–2907, 2021.
- [32] Shogo Shimada, Peerawat Pannattee, Yasushi Ikei, Nobuyuki Nishiuchi, and Vibol Yem. High-frequency cybersickness prediction using deep learning techniques with eye-related indices. *IEEE Access*, 2023.
- [33] Junhong Zhao, Kien TP Tran, Andrew Chalmers, Weng Khuan Hoh, Richard Yao, Arindam Dey, James Wilmott, James Lin, Mark Billingham, Robert W Lindeman, et al. Deep learning-based simulator sickness estimation from 3d motion. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 39–48. IEEE, 2023.
- [34] Dayoung Jeong and Kyungsik Han. Precyse: Predicting cybersickness using transformer for multimodal time-series sensor data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 8, No. 2, pp. 1–24, 2024.

- [35] Shogo Shimada, Peerawat Pannattee, Yasushi Ikei, Nobuyuki Nishiuchi, and Vibol Yem. High-frequency cybersickness prediction using deep learning techniques with eye-related indices. *IEEE Access*, Vol. 11, pp. 95825–95839, 2023.
- [36] Ripan Kumar Kundu, Rifatul Islam, John Quarles, and Khaza Anuarul Hoque. Litevr: Interpretable and lightweight cybersickness detection using explainable ai. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 609–619, 2023.
- [37] Wei Quan, Linxuan Li, Cheng Han, Yuechen Zhang, Yongqing Cai, Zhen Liang, Mingming Zheng, and Rui Wang. Objective evaluation of vr sickness and analysis of its relationship with vr presence. In *Advanced Intelligent Computing Technology and Applications: 20th International Conference, ICIC 2024, Tianjin, China, August 5 – 8, 2024, Proceedings, Part X*, p. 416 – 427, Berlin, Heidelberg, 2024. Springer-Verlag.
- [38] Po-Chen Kuo, Li-Chung Chuang, Dong-Yi Lin, and Ming-Sui Lee. Vr sickness assessment with perception prior and hybrid temporal features. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 5558–5564, 2021.
- [39] Tae Min Lee, Jong-Chul Yoon, and In-Kwon Lee. Motion sickness prediction in stereoscopic videos using 3d convolutional neural networks. *IEEE transactions on visualization and computer graphics*, Vol. 25, No. 5, pp. 1919–1927, 2019.
- [40] Minghan Du, Hui Cui, Yuan Wang, and Henry Been-Lirn Duh. Learning from deep stereoscopic attention for simulator sickness prediction. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 29, No. 2, pp. 1415–1423, 2021.
- [41] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, pp. 1254–1259, 1998.
- [42] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

- [43] Ming Jiang, Shengsheng Huang, Juanyong Duan, and Qi Zhao. Salicon: Saliency in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1072–1080, 2015.
- [44] Wenguan Wang, Jianbing Shen, Fang Guo, Ming-Ming Cheng, and Ali Borji. Revisiting video saliency: A large-scale benchmark and a new model. In *Proceedings of the IEEE Conference on computer vision and pattern recognition*, pp. 4894–4903, 2018.
- [45] Katherine J. Mimnaugh, Evan G. Center, Markku Suomalainen, Israel Becerra, Eliezer Lozano, Rafael Murrieta-Cid, Timo Ojala, Steven M. LaValle, and Kara D. Federmeier. Virtual reality sickness reduces attention during immersive experiences. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 29, No. 11, p. 4394 – 4404, November 2023.
- [46] Phil Lopes, Nana Tian, and Ronan Boulic. Eye thought you were sick! exploring eye behaviors for cybersickness detection in vr. In *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pp. 1–10, 2020.
- [47] Jeffrey S Perry and Wilson S Geisler. Gaze-contingent real-time simulation of arbitrary visual fields. In *Human vision and electronic imaging VII*, Vol. 4662, pp. 57–69. SPIE, 2002.
- [48] Behrang Keshavarz and Heiko Hecht. Validating an efficient method to quantify motion sickness. *Human factors*, Vol. 53, No. 4, pp. 415–426, 2011.
- [49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, Vol. abs/1512.03385, , 2015.
- [50] Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*, 2023.
- [51] Meta. Meta Quest Pro. https://www.meta.com/jp/quest/quest-pro/?srsltid=AfmB0oodW62pNfVssXvpdy8qsSbC_WIAcDBt6tfeXC2K4wA02jr1FLT.

- [52] Tobias Senst, Volker Eiselein, and Thomas Sikora. Robust local optical flow for feature tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 9, pp. 1377–1387, 2012.
- [53] J.S. Pointer and R.F. Hess. The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research*, Vol. 29, No. 9, pp. 1133–1151, 1989.
- [54] Haoqi Fan, Tullie Murrell, Heng Wang, Kalyan Vasudev Alwala, Yanghao Li, Yilei Li, Bo Xiong, Nikhila Ravi, Meng Li, Haichuan Yang, Jitendra Malik, Ross Girshick, Matt Feiszli, Aaron Adcock, Wan-Yen Lo, and Christoph Feichtenhofer. PyTorchVideo: A deep learning library for video understanding. In *Proceedings of the 29th ACM International Conference on Multimedia*, 2021. <https://pytorchvideo.org/>.

付 録 A

オブティカルフローを非視野制限状態と視野制限状態に分けた後, kmeans によってオブティカルフローを分類した結果を図 A.1,A.2 に示す. そして, オブティカルフローの分類に基づき, 各 VR 酔い度毎に注視点周りの顕著性マップ画素値の最大値について箱ひげ図をプロットした結果を図 A.3A.4 に示す.

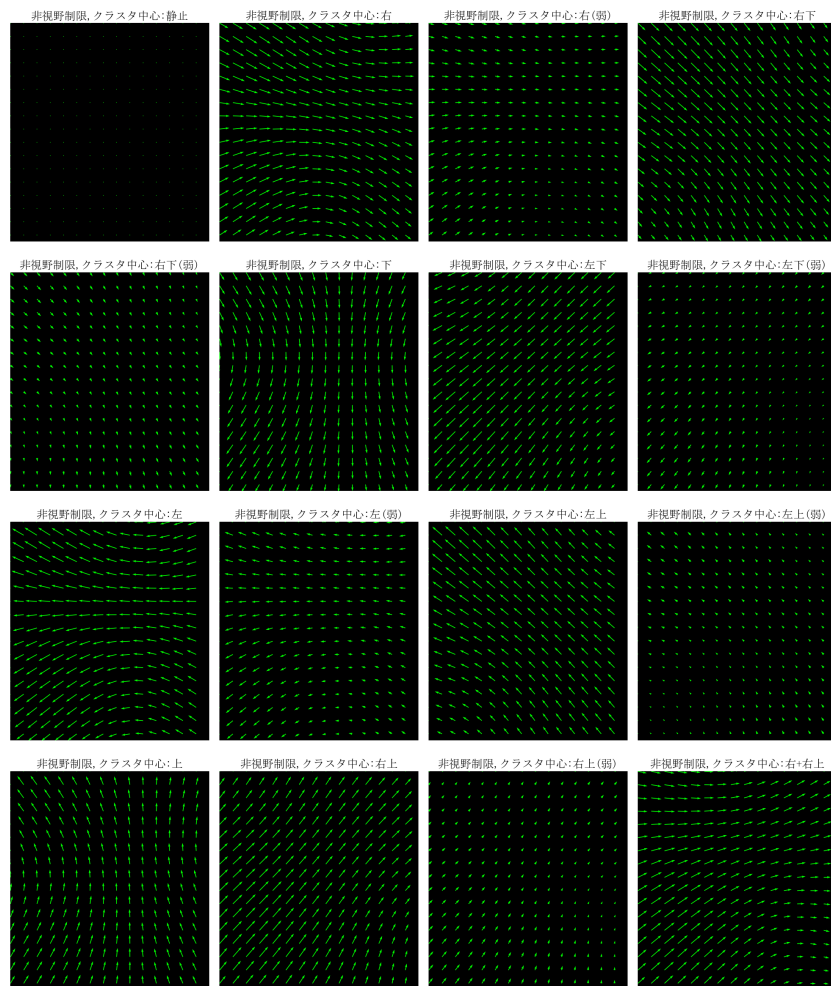


図 A.1 k-means による非視野制限状態でのオブティカルフローの分類結果

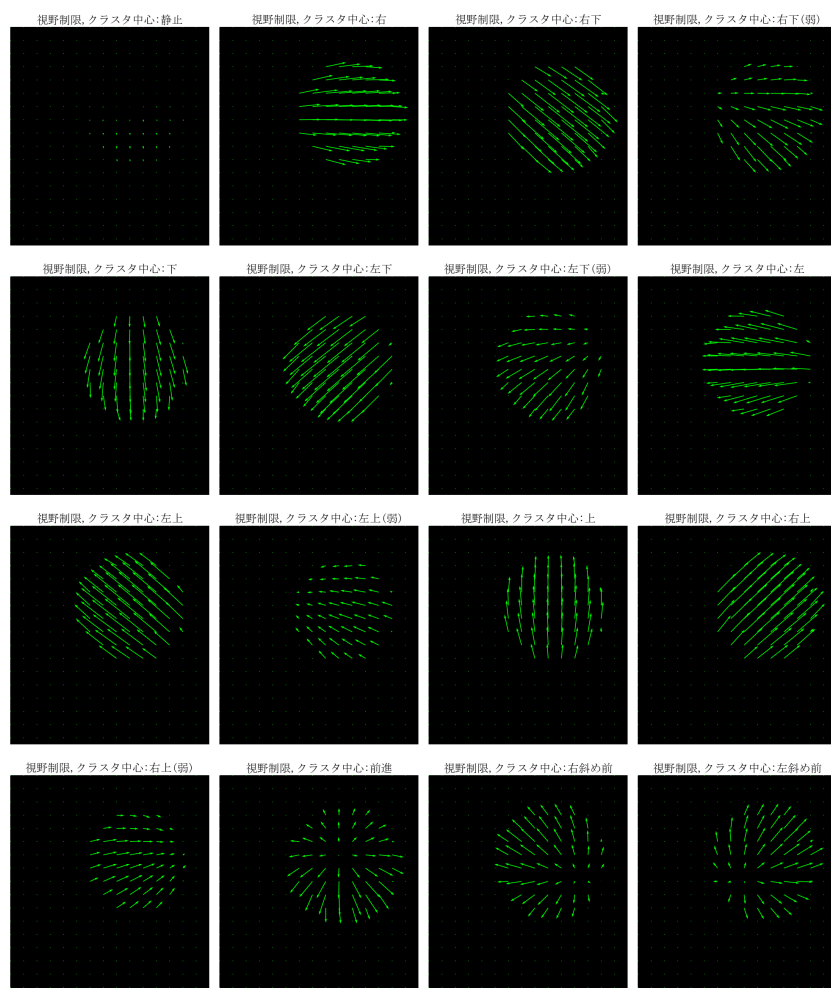


図 A.2 k-means による視野制限状態でのオプティカルフローの分類結果

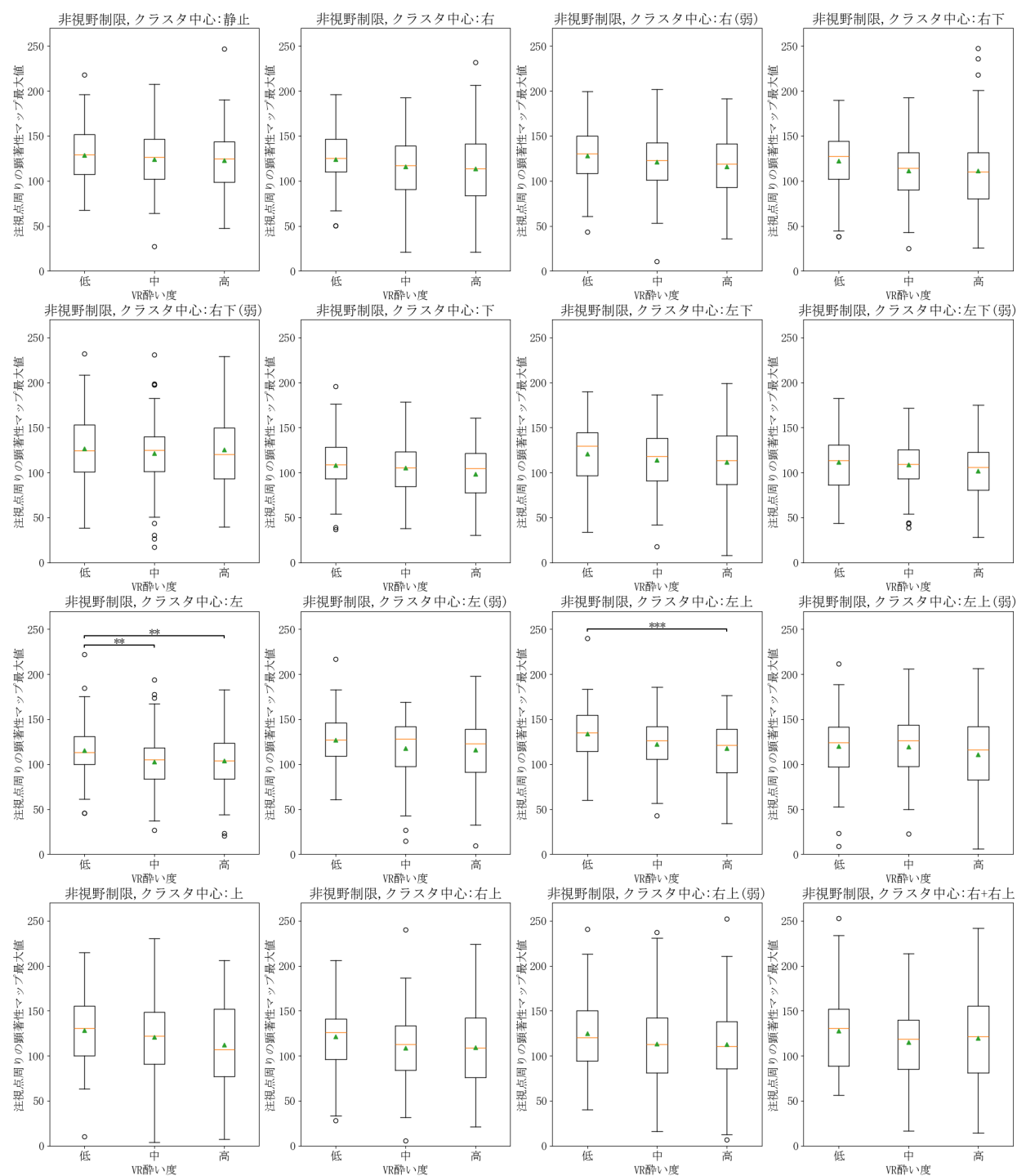


図 A.3 非視野制限状態のオプティカルフローの分類結果に対応する注視点周り (< 2.5°) の顕著性マップ画素値の最大値



図 A.4 視野制限状態のオプティカルフローの分類結果に対応する注視点周り (< 2.5°) の顕著性マップ画素値の最大値