

車両移動による運動視差を利用した半透明像注視判定

澤村 和樹^{*1} 浦宗 龍生^{*1} 池田 聖^{*1} 石塚 裕己^{*1} 大城 理^{*1}

Abstract — 電車やバスなどの車窓に情報提示が可能な透明ディスプレイでは、提示された半透明像（以下、仮想物体）とその背後にある実環境が重なって見える。視線計測器から得られる視線方向は、仮想物体と背後の実環境の二点で交差するため、利用者が両者のうちいずれを注視しているかを判定することが従来困難であった。本稿では、車両内の透明ディスプレイに対峙する利用者が注視する対象を判定する手法について提案する。提案法では、実環境の形状が与えられることを前提に、視線計測器から得られる眼球位置および視線方向から仮想物体と実環境のいずれかを注視しているか、もしくはそれ以外の状態（不定もしくは他の物体を注視）かを判定する。具体的には、視線と仮想物体および実環境との交点位置をそれぞれ算出した上で、両物体上の座標系において各交点に対して車両移動を考慮した固視検出法を適用する。各時刻の両座標系における固視の有無が注視の有無を表すとし、瞬きおよびサッケード等を考慮した上で注視対象を決定する。実験では、提案法の精度を確認するために、仮想物体と実環境の見かけの速度差（角速度差）が 10 から 40 deg/s および 10 から 60 deg/s となる仮想環境を構築し、HMD (Head-Mounted Display) 搭載の視線計測器により仮想環境を体験中の被験者 5 名の視線を取得し、提案法と従来の輻輳に基づく手法による注視判定を行った。実験の結果、輻輳に基づく手法では殆どの時刻で一方の対象に判定結果が偏ったのに対して、提案法では、全ての時刻で判定結果が得られた訳では無いが、判定結果が得られた場合には 80%以上の精度だった。

Keywords : 透明ディスプレイ, 3次元注視推定, 運動視差, 固視検出

1 はじめに

電車やバスでは近年、窓ガラスの一部に透明ディスプレイが試験的に設置されており、今後単純な案内や広告に加えて拡張現実感を利用した多様な情報が提示されることが予想される。マーケティング研究では、このような案内や広告に対して、どの程度の利用者がどの箇所に注目を示すのかを調べるために、頻繁に視線計測が用いられる [1]。しかし、視線計測は、基本的に視線方向を算出する技術であり、奥行きを含む注視箇所の 3次元位置を直接得る技術ではない。そのため、上述の透明ディスプレイに提示する半透明像（以下、仮想物体）と背後に見える実環境のように複数の対象が重なり同時に見える場合、単純な視線計測では注視する対象を区別できないという問題がある [2]。

従来、奥行き情報を含む 3次元的な注視位置を推定するため、輻輳開散運動 [3, 4] や前提動眼反射 [2] に基づく手法が提案されている。輻輳開散運動とは、異なる奥行きの対象に視線を移す際に両眼が反対方向に回転する眼球運動である。輻輳開散運動に基づく手法 [3, 4] は、利用者の意識的な動作を要さず注視位置の奥行き推定が可能であり、利用者から約 2 m 以下の距離にある対象については推定精度が高い反面、それよりも遠い距離の対象については推定精度が著しく低下する [2] 問題がある。これに対して、前庭動眼反射に基づく手法は、7 m 程度の距離の対象においても推

定精度が高い [2] という特徴がある。前庭動眼反射とは、頭部が回転する際に網膜像を安定させるために眼球が反対方向に回転し頭部の回転による影響を補償する反射運動である。しかしながら、この手法では、利用者が意識的に頭部を回転する必要がある。

ここで、動眼反射に基づく手法が輻輳に基づく手法と比較して、遠い距離における奥行き推定精度が高い要因を考えると、基線長の違いが挙げられる。眼球を三角測量におけるカメラに置き換えて考えると、輻輳に基づく手法では基線長が瞳孔間距離 (60-70mm) であるのに対して、動眼反射に基づく手法では基線長が眼球の移動距離すなわち頭幅程度 (150-160mm) に相当し長い。そのため、より遠い距離に配置された対象の推定精度を向上させたと考えられる。

本稿では、車両移動により生じる運動視差を用いる事で基線長の長さを確保し、輻輳に基づく手法よりも高精度に注視対象を判定する手法を提案する。提案法は、車両に設置された透明ディスプレイに利用者が対峙しており、その眼球位置や視線方向および実環境の 3次元形状が得られているものと仮定する。この仮定の下で、利用者が提示された仮想物体と背後に見える実環境のいずれを注視しているかを判定する。実験では、実際の車両や実環境を模した VR (Virtual Reality) 環境を構築し、提案法の精度を示し、従来の輻輳に基づく手法と比較する。VR 環境として、透明ディスプレイに張り付くステッカーのような平面的な像を仮想物

^{*1}大阪大学

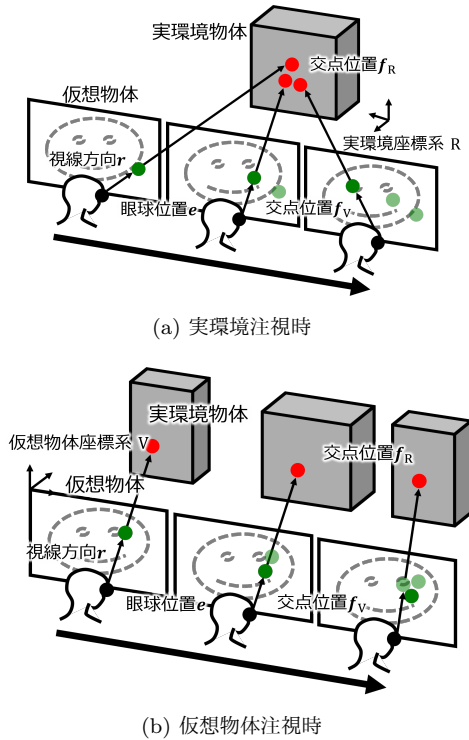


図1 運動視差に基づく注視判定の原理

体とする環境（以下、窓ステッカー環境）と実環境中に固定された立て看板状の仮想物体が配置された環境（以下、立て看板環境）を用いる。

2 運動視差に基づく注視判定法

2.1 提案法の前提と原理

図1は、窓ステッカー環境を例に挙げて提案法の原理を説明している。図1(a)では、利用者は実環境を注視しており、車両移動に伴い利用者が移動しても、視線と実環境との交点が一か所に密集しているが、仮想物体上ではばらつきが大きい。図1(b)では、利用者は仮想物体を注視しており仮想物体上の一か所に交点が密集しているが、実環境中では大きくばらついていて、交点の密集は、固視状態を表していると考えられるため、固視検出法により仮想物体が実環境のいずれを注視しているかを判定することができる。窓ステッカー環境と立て看板環境の違いは、それぞれ仮想物体が車両に固定されているか、実環境に固定されているかの違いであり、上記原理は変わらない。

提案法では利用者の視線情報（眼球位置 e 、視線方向 r ）、頭部の位置姿勢が得られるとし、仮想物体と実環境のいずれを注視しているかを判定する。利用者の視野をシーンカメラとして捉え、頭部の位置姿勢についてはシーンカメラの位置姿勢とみなす。また、車両の位置姿勢は得られるものとし、シーン（仮想物体、実環境）形状は既知とする。このような仮定は、自動走行車の車両や周辺地図を3次元的かつ実時間管理す

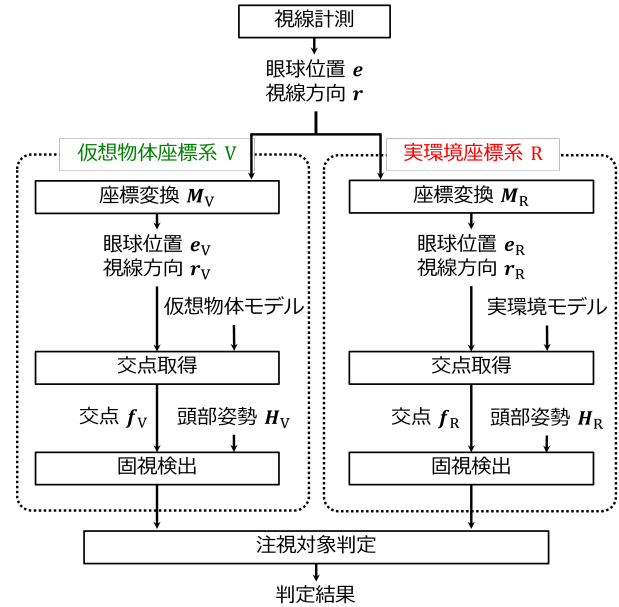


図2 注視判定の処理手順

るダイナミックマップ [5] の構想が現実化しつつある今日においては十分成り立つ。

本手法は、図2に示すように視線計測、座標変換、交点取得、固視検出、注視対象判定の五つの処理で構成されている。まず、視線計測により得られる視線情報を仮想物体上に固定された座標系（以下、仮想物体座標系 V ）と実環境上に固定された座標系（以下、実環境座標系 R ）の両方で扱うため座標変換する。次に、視線とシーンとの交点位置を両座標系において算出する。固視検出では、各座標における交点位置の時系列に対してそれぞれ固視を検出する。最後に、瞬きやサッケード中のデータ欠落を補間し、各時刻に対する注視対象の判定結果を算出する。

2.2 視線情報の座標変換と交点取得

前節の仮定の下では、視線計測器が頭部装着型であっても車両内の据え置き型であっても、視線情報や交点位置を容易に仮想物体座標系 V や実環境座標系 R に変換できる。この変換をそれぞれ、 M_V および M_R とする。頭部の位置姿勢についても、同様に座標系 V もしくは座標系 R からシーンカメラ座標系への変換 H_V もしくは H_R として記述できる。

本処理ではまず、各時刻で得られた視線情報 (e, r) を、変換 M_V および変換 M_R を用いて、仮想物体座標 (e_V, r_V) および実環境座標 (e_R, r_R) へと変換する。

次に、各座標系において視線とシーンとの交点位置 f_V, f_R を独立に算出する。交点の算出は、基本的には各形状モデルのポリゴン（平面）と視線との交点の算出と同じであり、実装上は一般的なゲームエンジンに標準搭載されている交点算出機能を流用する。

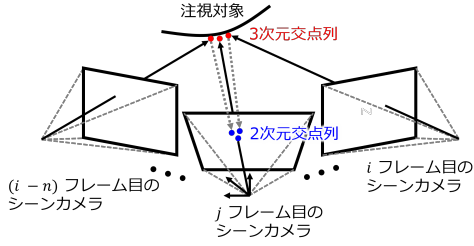


図3 固視検出区間におけるシーンカメラ

2.3 3次元固視検出

本処理では、前節で算出した各座標系における交点位置の時系列に対して、固視を検出する。固視検出は、交点位置の座標列に対して、各時刻が固視状態であるか否かをラベル付けする処理である。固視検出手法としては、固視中の交点の2次元的なばらつきが小さいことを利用するI-DT (Dispersion Threshold Identification) 法 [6] を3次元シーンに拡張して用いる。以下本節では、各座標系に対して同一の処理を行うため座標系を特定するための添え字 V, R や文言は省略する。

図3は、現時刻フレーム i から $n+1$ フレーム前までの間に得られた3次元交点列 $\{f_{i-n}, \dots, f_i\}$ に対応したシーンカメラを示す。フレーム数 $(n+1)$ は、固視検出の検出区間幅である。この幅は、検出に必要な最短時間 200 ms に相当するフレーム数に初期化される。3次元固視検出では、この検出区間内で最も平均的な視線方向に対応する j ($i-n \leq j \leq i$) フレーム目のシーンカメラ（以下、代表カメラ）の画像面上に3次元交点列 $\{f_{i-n}, \dots, f_i\}$ を投影し、従来の2次元シーン用の固視検出アルゴリズムであるI-DT法を適用する。 k フレーム目の視線方向を r_k としたとき、代表カメラのフレーム番号 j は次式により算出する。

$$j = \arg \min_{(i-n) \leq k \leq i} |r_k - \bar{r}| \quad (1)$$

ただし、 $\bar{r} = \sum_{k=(i-n)}^i r_k / (n+1)$ である。

次に、 $(n+1)$ フレーム分の交点を代表カメラに投影して、2次元の交点座標列 $\{g_{i-n}, \dots, g_i\}$ を得る。次式のとおり、これらの交点列の重心に対する最大偏差 s が閾値 T 以下であるとき、フレーム i の視線が固視状態であるとラベル付けされる。

$$s = \max_{(i-n) \leq k \leq i} |g_k - \bar{g}| \quad (2)$$

ただし、 $\bar{g} = \sum_{k=(i-n)}^i g_k / (n+1)$ 、閾値 T は、見かけの角度に換算して 1 deg に相当する値に設定される。上記条件を満たさない場合は固視以外の状態（以下、非固視と呼ぶ）となる。I-DT法では、 i フレームが固視の場合は、フレーム番号 i と検出区間幅 $(n+1)$ がインクリメントされ、非固視の場合は i が $(i+n+1)$ に更新され、検出区間幅 $(n+1)$ は、再度 200 ms に

初期化される。以上の処理が繰り返され全フレームに対して固視／非固視のラベルが振られる。

2.4 注視対象判定

既に述べたとおり、提案法は、ある対象を注視している状態であることを固視から判断する手法である。非固視の時間については、瞬きやサッケードが生じていると考えられる。この非固視中の区間は、同一対象上の別の箇所に視線を移動させるサッケードが生じている場合や、異なる対象に視線を移動（仮想物体から実環境、実環境から仮想物体など）させる場合に分けられる。前者の場合においては、注視している対象に変化はないため、注視対象判定においては、この区間を固視状態とみなす。具体的には、非固視時間の前後で同じ座標系において固視が検出されており、非固視の時間が 200 ms 未満の場合は、その時間を固視としてラベルを付け直して固視データの欠損を補間する。それ以外の場合は、非固視のまま維持する。

次に、各座標系に対して得られた固視／非固視ラベルの時系列データから、表1に示すルールで各時刻の注視対象を判定する。本実験においては、車両の速度が十分小さいとき、仮想物体と実環境の両座標系において固視が検出される（表中の「不定」）。提案法は仮想物体と実環境の見かけの速度差を利用するため、この場合は、注視する対象を定められない。一方、車両の速度が十分大きく仮想物体と実環境の奥行きが異なるとき、原理的にいずれかの座標系にしか固視は検出されない。仮想物体を注視していれば、仮想物体座標系においてのみ固視が検出され、実環境座標系で検出されない（表中の「仮想物体」）。反対に、実環境を注視するとき、実環境座標系において固視が検出されるが、仮想物体座標系においては検出されない（表中の「実環境」）。何も注視していない、もしくは他の移動物体を注視している場合は、仮想物体と実環境の両座標系において固視が検出されない（表中の「その他」）。提案法では、以上の四つの状態に判定する。

3 実験

提案法では、原理的に仮想物体と実環境の見かけの速度の差（角速度で表される）を手掛かりにして、注視対象が判定される。そこで実験では、見かけの速度の差と判定精度の関係を調べ、従来法として輻輳に基づく手法と比較する。

表1 固視検出結果からの注視判定

		仮想物体座標系	
		固視	非固視
実環境座標系	固視	不定	実環境
	非固視	仮想物体	その他



図4 実環境として用いたモデル

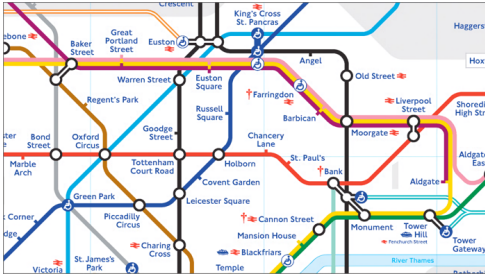


図5 透明ディスプレイ上の提示像

3.1 実験環境

本実験は、シーン中の物体の形状を容易に取得できる VR 環境で行った。VR 環境内で直線移動する車両から移動方向と直角な方向に見ると仮想物体と実環境が重なるように仮想の透明ディスプレイを配置した。本節では、実験に使用した実験装置、VR 環境、タスクを示す。

実験装置 視線計測器と VR 環境を提示する HMD として、HTC Vive Pro Eye を用いた。頭部姿勢は HMD 付属の Lighthouse トラッキングシステムにより、視線情報は HMD 内蔵の視線計測器より取得し、いずれも標本化周波数 90 Hz で計測した。HMD の解像度は片目あたり 1440×1600 画素、視野角 110 deg、垂直同期周波数 90 Hz であり、VR 環境は Unity を用いて構築した。今回使用した PC (CPU : Intel(R) Core(TM) i7-10700F, GPU : NVIDIA GeForce RTX 2060 SUPER) では常時 90 fps 以上で両眼立体視像を提示されていた。

VR 環境 実環境として図 4 に示すような建造物を Japanese Street¹から取得し City Generator²により車両の軌道と平行に透明ディスプレイから約 15 m の距離に並べた。透明ディスプレイは、車両の床から 0.5 m の高さに水平視野角約 100 deg、垂直視野角約 25 deg の大きさで、利用者の前方約 2 m に配置した。

窓ステッカー環境では、仮想物体としての路線図(図 5)を車両の窓ガラス上に固定した。仮想物体の水平視野角は約 50 deg であった。この環境では、仮想物体



図6 立て看板のテクスチャ

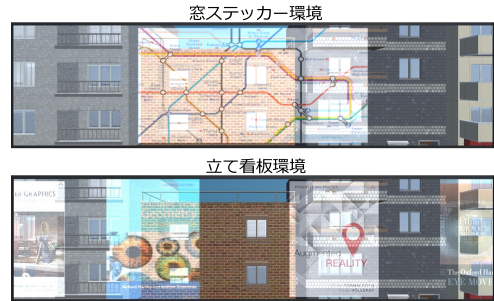


図7 被験者視野

と実環境の見かけの速度差が 10, 20, ..., 60 deg/s となるように、車両速度を約 9.42, 18.9 (=9.42×2), ..., 56.6 (=9.42×6) km/h とした。車両に固定される仮想物体の見かけの速度は常に 0 deg/s となる。

立て看板環境では、仮想物体として図 6 に示す本の表紙をテクスチャにもつ長方形の看板を車外に車両の軌道と平行に配置した。立て看板環境では、仮想物体と実環境との見かけの速度差が約 10, 20, 30, 40 deg/s となるように、車両速度を約 9.4 km/h に固定したまま仮想物体までの最短距離を約 7.2, 4.5, 3.1, 2.2 m に変化させた。このとき、仮想物体は垂直視野角が常に約 25 deg となるように距離に応じて大きさを変えた。車両速度は一定なため、実環境の見かけの速度は常に 10 deg/s となる。各環境における被験者の視野例を図 7 に示す。

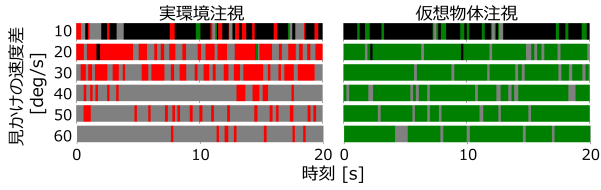
タスク 被験者として、実験の目的に同意が得られた 20 代の男性 5 名を対象とした。被験者には、座位で頭部移動は自由とし、各速度に対して 45 s 間並進移動する車両内からのシーンを提示し、3 名には仮想物体、実環境の順に、2 名には実環境、仮想物体の順に注視させた。前後端 1 s および切り替え時のピープ音中 3 s を破棄し、実環境、仮想物体それぞれ 20 s ずつの視線データを取得した。各速度での映像提示の間に、被験者は十分な休息を得られるようにした。

3.2 視線解析法

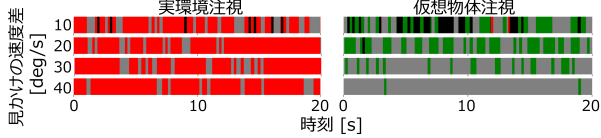
提案法では、両眼の中央に仮想的な眼 (cyclopean eye) があるものとして両眼の平均の視線情報を扱う。平均の視線情報は、両眼の眼球位置と視線方向をそれぞれ平均することで算出する。視線計測器の較正については、予め被験者とは異なる 1 名により Vive Pro Eye の標準搭載の較正プログラムを実行後に各被験者に対してマーカを用いて再度較正した。マーカを用い

¹<https://assetstore.unity.com/packages/3d/environments/urban/japanese-street-170162>

²<https://assetstore.unity.com/packages/templates/systems/city-generator-24993>

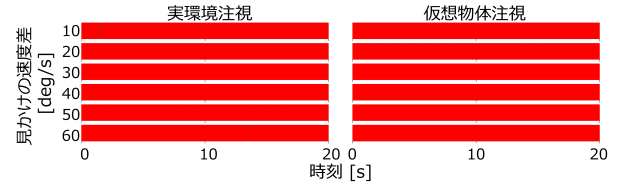


(a) 窓ステッカー環境

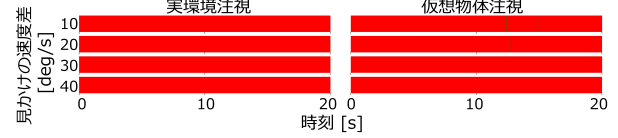


(b) 立て看板環境

図 8 提案法での判定結果例 (1 名分)

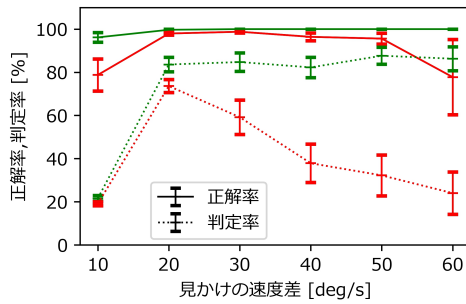


(a) 窓ステッカー環境

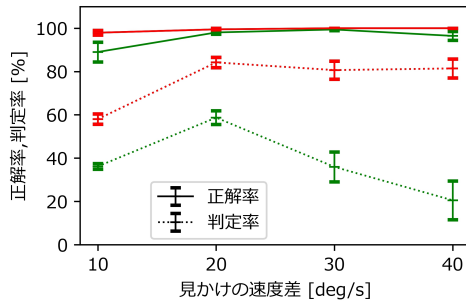


(b) 立て看板環境

図 10 従来法の判定結果例 (1 名分)

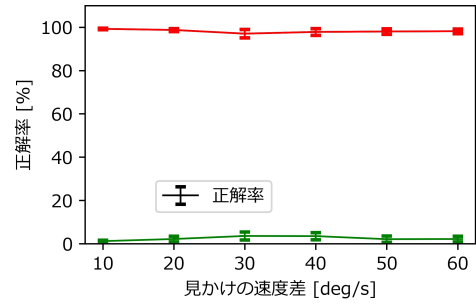


(a) 窓ステッカー環境

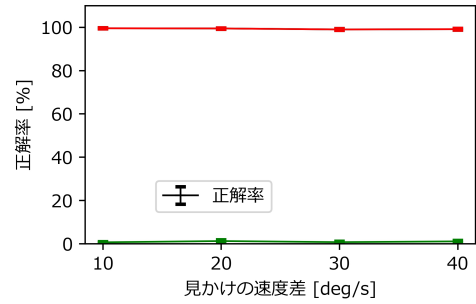


(b) 立て看板環境

図 9 提案法における見かけの速度差と精度



(a) 窓ステッカー環境



(b) 立て看板環境

図 11 従来法における見かけの速度差と精度

た校正では、奥行き 1 m に配置した 16 個のマーカを用いてアフィン変換パラメータを推定した上で、異なる 9 個のマーカにより評価し、注視点の推定誤差が平均 1.5 deg 未満であることを確認して用いた。マーカを用いた校正は、提案法においては平均視線方向に、従来法においては片目ずつ同様の校正を行った。

比較対象とする輻輳に基づく手法では、両眼の視線情報を独立に扱う。両眼の視線と仮想物体および実環境と交点を算出した上で、cyclopean eye のシーンカメラ画像上での両眼の交点間の水平距離が短い方を注視中の対象として判定する。

また、提案法では、注視対象が仮想物体か実環境の二択として判定されるのではなく、「不定」や「その他」という形で正しく推定できない事自体が判定結果

として得られる。この事を考慮して、本実験では正解率と判定率という二つの指標で評価する。正解率とは、注視対象が実環境、または仮想物体と判定したとき、被験者が判定した対象を注視していた割合を表し、判定率とは、全フレームのうち、実環境、または仮想物体と判定されたフレームの割合を表す。例えば、実際の注視対象が仮想物体の場合、各フレームの判定結果のうち注視対象が仮想物体であると判定されたフレーム数を n_V 、実環境であると判定されたフレーム数を n_R 、20 s 間のフレーム数を n とするとき、正解率は $n_V/(n_V+n_R)$ 、判定率は $(n_V+n_R)/n$ として表される。

3.3 実験結果

図 8 に、それぞれ窓ステッカー環境および立て看板環境の結果例として 1 名の被験者に対する時系列の注

視対象判定結果を示す。図中の色は表1の判定に対応しており、緑色は仮想物体注視、赤色は実環境注視、黒色は不定、灰色はその他を表す。図9に、それぞれ窓ステッカー環境および立て看板環境における、見かけの速度差と全被験者の正解率（実線）と判定率（破線）を示す。エラーバーは標準誤差を示す。図中の色は緑色は仮想物体、赤色は実環境に対応する。

一方、輻輳に基づく手法では、図10に示す例のように半透明像の奥行きに依存せず、ほとんどの時刻で実環境を注視していると判定された。図11の5名の被験者の平均においても、見かけの速度差に依存せず偏った判定結果になった。図10, 11の色は、図8, 9と同様である。

3.4 考察

提案法における、正解率はいずれの速度差においても高く、両環境のいずれの注視対象においても10 deg/sと窓ステッカー環境における60 deg/sの場合を除いて90%以上の高い精度を維持している。窓ステッカー環境においても立て看板環境においても、共通して速度差が10 deg/sのとき正解率は80%以上と高いものの、20 deg/sの場合と比べて判定率は大きく低下する場合があった。この結果は見かけの速度差が小さいとき、固視検出結果が混同したことで、両座標系上で固視が生じ多くのフレームが不定となったと考えられる。実際に、図9(a)および(b)では、どちらも10 deg/sの場合にいずれの注視対象においても注視対象が不定（黒い部分）である箇所が多いことから確認できる。この結果は、提案法が見かけの速度差を利用していることと矛盾なく説明できる。見かけの速度差が不十分な場合には原理的に注視対象を判定できなくなる。

一方、判定率については、見かけの速度差が30 deg/s以上で大きくなるにつれて低下した。窓ステッカー環境では実環境の判定率が低下し、立て看板環境では、仮想物体の判定率が低下した。速度差が大きくなるにつれて判定率が低下したのは、仮想物体と実環境の一方を注視するときに、他方を見かけの動きにより視覚運動性眼振が発生するなどして、注視対象上の固視検出が困難になっている可能性が考えられる。視覚運動性眼振などの影響を考慮して固視検出することにより判定率の低下を軽減できる可能性がある。一方の対象において判定率が低下した原因は、見かけの速度の増加に伴い固視の正確さが低下し、交点のばらつきが大きくなった可能性がある。

従来の輻輳に基づく手法では、いずれの環境、注視対象においても殆どのフレームにおいて実環境を注視しているという判定結果になり、正確に判定が得られたとは言えない。文献[2]は、輻輳に基づいた手法においては、奥行き0.8, 1.5, 7.0 mのときの平均誤差

が約0.15, 0.48, 4.86 mと報告している。三角測量による誤差は奥行きの二乗に比例することから、奥行き15 mの場合の誤差は誤差が約22 mであると予測できる。本実験のシーンのように遠距離の対象を注視するとき、輻輳の精度は極めて悪いことから、正解率は低くなったと考えられる。

4 おわりに

本稿では、電車やバスなどに設置された透明ディスプレイにて、半透明な仮想物体像と仮想物体越しに見える実環境のいずれを注視しているか判定する手法を提案した。提案法は、車両移動により運動視差が生じ、その結果車両内外に固定された座標系によって視線と対象との時系列交点位置データが停留することを利用した。実験では、HMD装着の下VR環境で行い、提案法が見かけの速度差に依存して使用できる範囲の限界を示した。これらの結果から、想定する環境において提案法によって注視対象が判定可能な範囲を示した。今後は、十分な規模の被験者実験による統計解析や仮想物体のテクスチャや動物体の仮想物体に対応した判定法を開発する予定である。

謝辞 本研究の一部は、JSPS 科研費(21K03964)の支援による。実験は、大阪大学大学院基礎工学研究科人を対象とした研究に関する倫理委員会の承認(R2-9-1)を得て実施した。

参考文献

- [1] 里村. 視線計測による消費者行動の理解. オペレーションズ・リサーチ, Vol. 62, No. 12, pp. 775–761, 2017.
- [2] D. Mardanbegi, T. Langlotz, and H. Gellersen. Resolving target ambiguity in 3d gaze interaction through vor depth estimation. In *Proc. CHI Conf. on Human Factors in Computing Systems (CHI2019)*, No. 612, 2019.
- [3] F. Alt, S. Schneegass, J. Auda, R. Rzayev, and N. Broy. Using eye-tracking to support interaction with layered 3d interfaces on stereoscopic displays. In *Proc. 19th Int. Conf. on Intelligent User Interfaces (IUI2014)*, p. 267–272, 2014.
- [4] E. G. Mlot, H. Bahmani, S. Wahl, and E. Kasneci. 3d gaze estimation using eye vergence. In *Proc. Int. Conf. on Health Informatics*, Vol. 6, pp. 125–131, 2016.
- [5] 高田, 佐藤. ダイナミックマップ. システム/制御/情報, Vol. 60, No. 11, pp. 457–462, 2016.
- [6] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proc. Symp. on Eye Tracking Research & Applications (ETRA2000)*, p. 71–78, 2000.

© 2022 by the Virtual Reality Society of Japan (VRSJ)