

# モデルベース視線計測装置のインプリシットキャリブレーション手法の改良

廣江 葵<sup>1,2,a)</sup> 山本 倫也<sup>1,b)</sup> 長松 隆<sup>3,c)</sup>

**概要:** 近年、3D モデルベース手法の視線計測システムでは眼球の光軸をユーザキャリブレーションを行わずに推定可能である。3D モデルベースの視線計測システムでのユーザキャリブレーションは眼球の光軸と視軸（視線）の差であるカッパ角を推定しており、このカッパ角をインプリシット（暗黙的）に求めることができれば、ユーザに意図的注視を求めることなく視線計測を開始できるようになる。本論文では眼球の光軸周辺の顕著性マップを用いたインプリシットキャリブレーション手法を改良し、少ないフレーム数の顕著性マップを使ってキャリブレーションを達成する手法を提案する。光軸の運動から眼球運動の種類を識別しキャリブレーションに有効なフレームの顕著性マップだけをカッパ角の推定に用いることで、視線推定精度を維持したまま、インプリシットキャリブレーションにかかる計算コストを削減することができた。

## 1. はじめに

視線計測技術は、次世代のヒューマンインタフェースに応用可能な技術として注目されているが、従来の視線計測装置ではユーザが事前に画面に表示された点をいくつか意図的に注視する校正作業（キャリブレーション）が必要であった。そのため、不特定多数の視線を正確に計測することは困難であった。キャリブレーションを意図的な注視なしで、つまり、暗黙的（インプリシット）に実施できれば、例えば、通りすがりの人の視線を計測したり、心理実験においては自然な状態で視線計測できるようになる。

Shih ら [1]、Guestrin ら [2] のモデルベースの視線計測手法の研究により、角膜の球面モデルを用いることでキャリブレーション時に見る必要のある点の数は1点にまで減少させることができるようになっており、キャリブレーションをなくす、もしくは、簡略化・自動化する手法はいくつか提案されているが、決定的な手法は確立されていなかった [3]。そこで、廣江ら [4] は、ディスプレイに表示された画像の顕著性マップを使って、眼球の光軸を基準とした座標系で、眼球の視軸の位置を推定することにより、キャリブレーションを暗黙的に行える方法を提案した。

廣江らの手法においては、眼球を撮影した映像の全ての

フレームについて、対応するディスプレイ画像の顕著性マップの計算を行っていた。本研究は、多くの時間を要した顕著性マップの計算時間の削減を行うため、眼球運動を分析して、計算に用いるフレームを絞り込むことにより、計算不可の軽減を行った。

## 2. 関連研究

視線計測装置のキャリブレーションの簡略化は、活発に研究が行われている分野である。角膜を球面とモデル化するアプローチにより、幾何学的に眼球の光軸を求めることが可能となっている [1], [2]。この場合、眼球の光軸と視軸（視線と同等）の差であるカッパ角を求めることがユーザキャリブレーションに相当する。最少1点を注視することで、キャリブレーションを行うことが可能である。

幾何学的な両眼による制約を用いて、インプリシットキャリブレーションを行う研究も行われている。長松ら [5] は眼球の光軸とディスプレイとの交点を両眼について計算し、それらの中点を注視していると仮定してキャリブレーションを行う手法を提案した。この手法は、安定して注視点を推定することができるが、視線の推定精度に個人差が生じるので、確実に高精度で計測できるというわけではなかった。Model ら [6] は、計測した両眼の視軸がディスプレイ上で交差するという制約を利用して、両眼の光軸と視軸のずれを求めるキャリブレーション手法を提案した。しかしながらこの手法は光軸の推定精度に敏感であり、安定的な推定のためにピラミッド状の注視対象物を利用していた。以上のように現状では幾何学的な両眼の制約によるア

<sup>1</sup> 関西学院大学工学部

<sup>2</sup> 神戸大学大学院医学研究科

<sup>3</sup> 神戸大学大学院海事科学研究科

a) hiroe@ieee.org

b) michiya.yamamoto@kwansei.ac.jp

c) nagamatu@kobe-u.ac.jp

アプローチでは、精度よく、安定したキャリブレーションを行うのが困難であると思われる。

一方、画面に表示されている情報をもとにインプリシットにキャリブレーションを行う方法も提案されている。Chen ら [7] は顕著性マップとモデルベース手法を結びつけたインプリシットキャリブレーション手法を提案した。彼らは、眼球の光軸、視軸、光軸と視軸の成す角（カップ角）の確率的な関係を表すベイジアンネットワークを構築した。顕著性マップは注視点分布の事前確率として与えられ、カップ角の推定は、ディスプレイ画像と光軸が与えられたときに注視点分布の事後確率を計算する推論問題として定式化された。実施した実験は複数の静止画をそれぞれ 3~4 秒閲覧するものであった。Wang ら [8] は、Chen ら [7] の研究で利用した顕著性マップは浅いボトムアップの推定によるもので、人間の注意を引く意味情報を取り込むことが難しいと指摘し、Chen らの研究を発展させ、RCNN (regression based deep convolutional neural network) を用いて学習した fixation map を作成し、注視点の分布に着目してキャリブレーションを行う方法を提案した。この方法では fixation map と 3D モデルで推定した視線の分布が等しくなるようにパラメータを決定している。視線の分布を計算するため注視対象は静止画で、実験では各画像を 4 秒間注視していた。以上の手法は赤外線カメラを用いていたが、Sugano ら [9] は、可視光カメラからの眼球の画像と画面の顕著性マップのペアから、マッチング学習することで注視点を推定するアピランススペースの手法を提案した。動画を用いた評価実験を行ったが、約 3.5 度の精度であり、モデルベース手法の精度には及ばなかった。

廣江ら [4] は、Chen ら [7] の方法が、ベイジアンネットワークを用いた計算を行ったのに対して、よりシンプルな計算方法を提案した。また、Chen らは複数の静止画を対象とした検証を行ったのに対して、廣江らは動画を用いた検証を行った。Wang ら [8] の手法は注視点の分布に基づくため静止画にしか対応できなかったが、廣江らは静止画を一定時間閲覧する必要がなく、動画を含め自由に画面を閲覧している時でもキャリブレーションが可能な方法になっている。

### 3. 眼球の光軸周辺の顕著性マップを用いたインプリシットキャリブレーション手法 [4]

廣江らが提案した眼球の光軸周辺の顕著性マップを用いたインプリシットキャリブレーション手法は、1 点キャリブレーション手法が基礎となっている。1 点キャリブレーション手法は、眼球の光軸がキャリブレーション無しで推定可能であることと、場所が既知である点を注視することで眼球の視軸と光軸の差であるカップ角を求めることができることを組み合わせた手法である。ここで、眼球の光軸

は、眼球の角膜曲率中心と瞳孔中心を通る軸と定義し、視軸は注視点と眼球の中心窩を通る軸と定義している。1 点キャリブレーション手法で求められるカップ角は個人依存のパラメータであり、3D モデルベースの視線計測装置ではこのカップ角を求めることが、キャリブレーションに相当する。

廣江 [4] らは、キャリブレーションなしで求められる眼球の光軸周辺の顕著性マップを光軸基準の座標系に変換し、フレームごとに加算することで、そのピーク値の場所を、眼球の視軸の位置として推定している (図 1)。視軸が推定できた後、通常の 1 点キャリブレーションを行い、カップ角を求めている。1 点キャリブレーションでは、見ている点が正確にわかれば、必要な注視点は 1 点だけでよいが、一枚の顕著性マップでは正確に視軸が推定できないので、複数フレームにわたって顕著性マップを計算し、それらを足し合わせることで、より視軸らしい位置を求めている。これにより、ロバスト性と精度の向上を実現している。

顕著性マップの生成手法として、MIT/Tuebingen Saliency Benchmark[10] にコードが公開されている手法のうちスコアが上から 3 位 (2020 年 12 月 20 日現在) までの手法の UNISAL[11], MSI-Net[12], DeepGazeII[13] を用いている。

### 4. 光軸からの眼球運動推定によるデータの絞り込み

廣江らの手法では、光軸を推定できている限り全てのフレームの顕著性マップを足し合わせていた。そのため、キャリブレーションを行なっている間は常に顕著性マップの計算が発生し、リアルタイムでの推定の妨げとなっていた。顕著性マップの生成には、色相や明度などの複数のパラメータに着目して、画像を処理する必要があり、多くの計算コストを必要とする。この計算コストを削減することができれば、PC を操作している時に裏でインプリシットキャリブレーションが行えるようになる。

そこで、本論文では光軸の動きから眼球運動の種類を識別し、計算に使用するフレームを取捨選択することで計算に使用する顕著性マップの枚数を削減する手法を提案する。眼球運動はいくつかの種類に分類されるが、確実に何かのオブジェクトを注視しているであろう固視 (fixation) に着目し、固視を行っている時のフレームのみを用いることで、足し合わせに使用するマップを削減する (図 2)。具体的には推定された光軸の移動から眼球の回転角速度を計算し固視運動を識別する。本研究では  $30^\circ/sec$  以下を固視運動とし [14]、それ以上の速度で眼球が運動している時の顕著性マップを計算に使用しないようにした

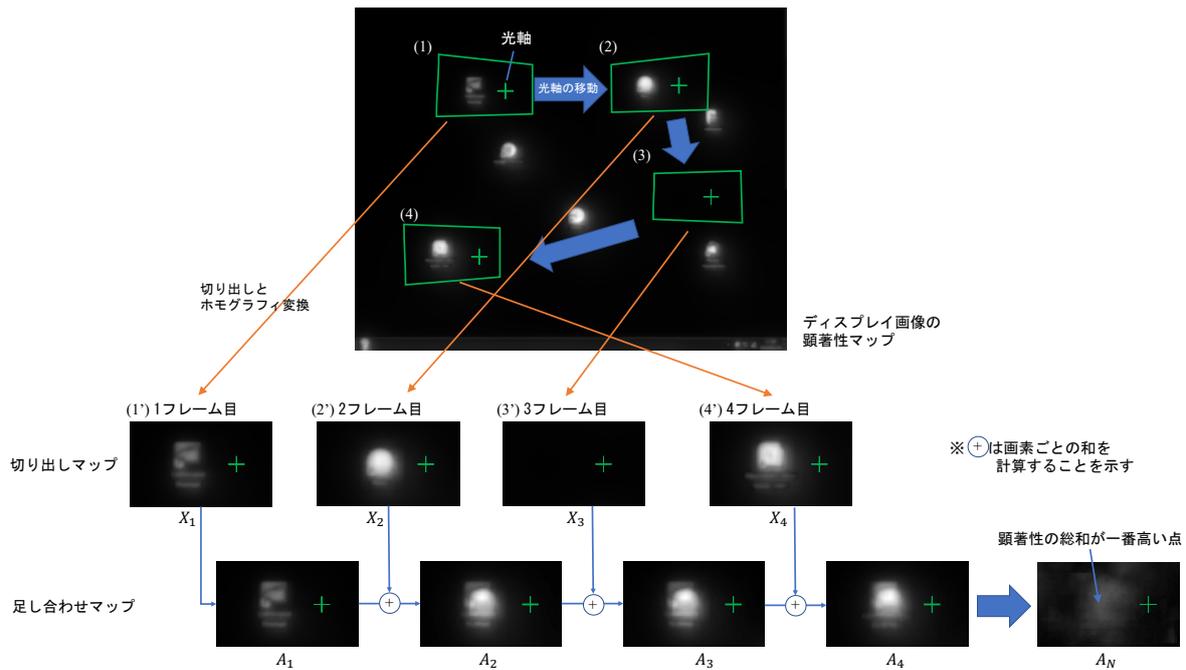


図 1 複数フレームを使った視軸の推定 出典:[4]

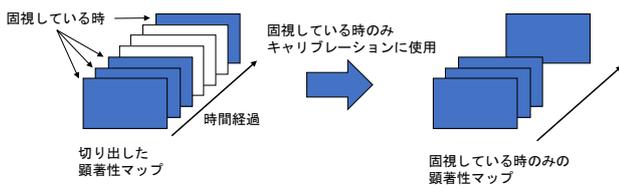


図 2 フレームの取捨選択

## 5. 評価

### 5.1 被験者データの取得

先行研究 [4] で取得した光軸データを使用した。先行研究におけるデータ取得実験の概略は以下の通りである。

被験者は、本実験への参加について承諾の得られた成人男性 7 名であった。被験者の年齢は 20 代 6 名、40 代 1 名であった。

実験システムには光学実験台上に 19 インチのディスプレイと可視光カットフィルタを取り付けた 2 台のカメラ (Baumer 社製, HXG20NIR), 顎台 (竹井機器工業製) を設置し、ディスプレイの両側に赤外線 LED を設置したシステムを用いた。眼球からディスプレイまでの距離は約 600mm であった。

実験に使用した 3 種類の動画は日常生活でよく見ることがあると想定される刺激 1. ニュース番組\*1 (198 秒), 刺激

2. スポーツ番組\*2 (150 秒), 刺激 3. 映画の予告\*3 (235 秒) の動画とした。なお実験に用いた動画は Youtube-8M Dataset[15] から選択した。

実験では事前にカメラキャリブレーションを行った後、被験者に実験装置の前に座るように指示し、ディスプレイ上に表示される動画 3 種類を見るように教示した。その後、画面の中心点を注視させて 1 点キャリブレーション用のデータを取得し、次にキャリブレーション後の精度を評価するために 9 点を注視させデータを取得した。実験中は被験者がディスプレイを見ている時の眼球画像とディスプレイに表示されている画像を取得した。

本論文では、先行研究で取得した被験者データの中から、20 代の成人男性 (ソフトコンタクトレンズを着用) 1 名について、以下のように評価を行った。

### 5.2 提案手法の評価結果

表 1 は、足し合わせに用いる顕著性マップのフレームを光軸の回転速度  $30^\circ/sec$  の閾値で絞り込んだ時のフレーム数をそれぞれの刺激動画に対して示したものである。絞り込みによって平均約 12.3% のフレームが削減された。表 2 は表 1 のように削減されたフレーム数の顕著性マップで足し合わせを行うことで求められたカッパ角の値を使って 9 点注視時の視線推定精度を示したものである。 $30^\circ/sec$  でフレーム数を削減しても視線推定精度は同程度であった。

\*1 <https://www.youtube.com/watch?v=f23fNM8JIUI> (2021 年 4 月 28 日)

\*2 <https://www.youtube.com/watch?v=KK9AH3G7gBc> (2021 年 4 月 28 日)

\*3 <https://www.youtube.com/watch?v=ZcFzeAwextg> (2021 年 4 月 28 日)

表 1 30°/sec の閾値で絞り込んだ時のフレーム数の変化

	フレーム数 [frames]		
	絞り込み無し	30°/sec の閾値	削減率
刺激 1	2346	2034	-13. 3%
刺激 2	1737	1488	-14. 3%
刺激 3	3189	2895	-9. 2 %

### 5.3 考察

表 1 に示すように光軸の回転速度を 30°/sec の閾値でフィルタリングすることで足し合わせに用いるフレーム数を削減することができた。この時、刺激によって削減率が異なっているのは、被験者が各刺激の動画を閲覧している際に見ているオブジェクトの動きや、注視していたオブジェクトが映っている時間が異なっただけが理由として考えられる。しかしながら、視線の推定精度はフレーム数の削減率に依存していないことから、本手法が有効であることが示唆される。図 3 は閾値を変更した時の各刺激のフレーム数がどれだけ削減されるかを示したものである。表 3 で示すように閾値を下げて精度が大きく下がらないことがわかった。

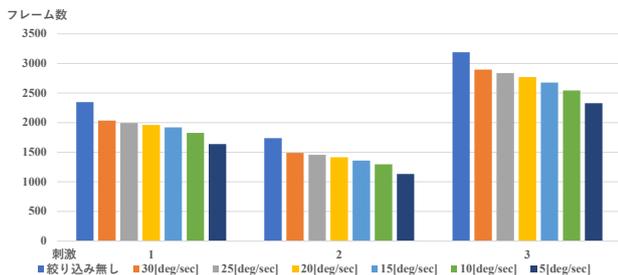


図 3 速度の閾値を変更した時のフレーム数

## 6. 結論

本論文では廣江らが提案したインプリシットキャリブレーション手法において、カップ角の推定に用いる顕著性マップを光軸の回転速度 30°/sec の閾値でフィルタリングすることで、視線推定精度を下げることなく、計算量の削減を図る手法を提案した。今後は被験者の人数を増やし、閾値の値と精度の関係をより詳細に明らかにする予定である。また、本研究では、光軸の回転速度のみから注視しているかどうかを推定し、キャリブレーションに使用するフレームを選択したが、画面上のオブジェクトをじっくり見ている可能性が高い追従眼球運動にも着目して改良を行っていく予定である。

謝辞 本研究は JSPS 科研費 JP20H04229, JP22K21317 の助成を受けたものです。

## 参考文献

- [1] Shih, S.-W. and Liu, J.: A novel approach to 3-D gaze tracking using stereo cameras, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 34, No. 1, pp. 234–245 (online), DOI: 10.1109/TSMCB.2003.811128 (2004).
- [2] Guestrin, E. and Eizenman, M.: General theory of remote gaze estimation using the pupil center and corneal reflections, *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 6, pp. 1124–1133 (online), DOI: 10.1109/TBME.2005.863952 (2006).
- [3] 長松 隆, 菅野裕介, 竹村憲太郎: キャリブレーションフリー視線計測手法の研究動向, *ヒューマンインタフェース学会論文誌*, Vol. 23, No. 1, pp. 73–88 (オンライン), DOI: 10.11184/his.23.1.73 (2021).
- [4] 廣江 葵, 山本倫也, 長松 隆: 眼球の光軸周辺の顕著性マップを用いた視線計測装置のインプリシットキャリブレーション手法, *ヒューマンインタフェース学会論文誌*, Vol. 23, No. 4, pp. 431–442 (オンライン), DOI: 10.11184/his.23.4.431 (2021).
- [5] Nagamatsu, T., Kamahara, J. and Tanaka, N.: Calibration-Free Gaze Tracking Using a Binocular 3D Eye Model, New York, NY, USA, Association for Computing Machinery (2009).
- [6] Model, D. and Eizenman, M.: User-Calibration-Free Remote Gaze Estimation System, *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, New York, NY, USA, Association for Computing Machinery, p. 29–36 (online), DOI: 10.1145/1743666.1743672 (2010).
- [7] Chen, J. and Ji, Q.: Probabilistic gaze estimation without active personal calibration, *CVPR 2011*, pp. 609–616 (online), DOI: 10.1109/CVPR.2011.5995675 (2011).
- [8] Wang, K., Wang, S. and Ji, Q.: Deep Eye Fixation Map Learning for Calibration-free Eye Gaze Tracking, *Proceedings of the 2016 ACM Symposium on Eye Tracking Research & Applications*, ACM, pp. 47–55 (online), DOI: 10.1145/2857491.2857515 (2016).
- [9] Sugano, Y., Matsushita, Y. and Sato, Y.: Appearance-Based Gaze Estimation Using Visual Saliency, *IEEE Transactions Pattern Anal. Mach. Intell.*, Vol. 35, No. 2, pp. 329–341 (online), DOI: 10.1109/TPAMI.2012.101 (2013).
- [10] Kümmerer, M., Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A. and Torralba, A.: MIT/Tübingen Saliency Benchmark, <https://saliency.tuebingen.ai/>.
- [11] Droste, R., Jiao, J. and Noble, J. A.: Unified Image and Video Saliency Modeling, *Proceedings of the 16th European Conference on Computer Vision (ECCV)*, (online), DOI: 10.1007/978-3-030-58558-7\_25 (2020).
- [12] Kroner, A., Senden, M., Driessens, K. and Goebel, R.: Contextual encoder-decoder network for visual saliency prediction, *Neural Networks*, Vol. 129, pp. 261–270 (online), DOI: 10.1016/j.neunet.2020.05.004 (2020).
- [13] Kümmerer, M., Wallis, T. S. A., Gatys, L. A. and Bethge, M.: Understanding Low- and High-Level Contributions to Fixation Prediction, *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4799–4808 (online), DOI: 10.1109/ICCV.2017.513 (2017).
- [14] Olsen, A. and Matos, R.: Identifying Parameter Values for an I-VT Fixation Filter Suitable for Handling Data Sampled with Various Sampling Frequencies, *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, New York, NY, USA, Associa-

表 2 30[°/sec] の閾値で絞り込んだ時の精度

		各手法での平均精度 [°]			
	絞り込み	UNISAL	DeepGazeII	MSI-Net	平均
刺激 1	全データ	1. 96	1. 96	2. 04	1. 99
	30[°/sec]	1. 92	2. 01	1. 85	1. 93
刺激 2	全データ	2. 21	2. 25	1. 95	2. 13
	30[°/sec]	2. 19	2. 18	2. 18	2. 18
刺激 3	全データ	1. 96	1. 82	2. 08	1. 95
	30[°/sec]	1. 93	1. 95	1. 95	1. 94
1 点キャリブレーション					1. 83

表 3 各刺激で閾値を変更した時の視線推定精度

各刺激での平均精度 [°]				
閾値 [°/sec]	刺激 1	刺激 2	刺激 3	平均
5	1. 95	2. 13	2. 06	2. 05
10	1. 97	2. 10	2. 00	2. 02
15	1. 92	2. 08	1. 99	2. 00
20	1. 91	2. 12	2. 10	2. 04
25	1. 91	2. 20	2. 06	2. 06
30	1. 93	2. 18	1. 94	2. 02
全データ	1. 99	2. 13	1. 95	2. 02

tion for Computing Machinery, p. 317–320 (online), DOI: 10.1145/2168556.2168625 (2012).

- [15] Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B. and Vijayanarasimhan, S.: YouTube-8M: A Large-Scale Video Classification Benchmark, *arXiv:1609.08675 [cs.CV]*, p. <http://arxiv.org/abs/1609.08675> (online), DOI: 10.48550/arXiv.1609.08675 (2016).