

Gaze Probing: イベント提示に基づく注視対象推定

米谷 竜[†] 川嶋宏彰^{††} 平山高嗣^{††} 松山隆司^{††}

[†] 京都大学大学院情報学研究科 ^{††} 〒606-8501 京都市左京区吉田本町

E-mail: [†]yonetani@vision.kuee.kyoto-u.ac.jp, ^{††}{kawashima, hirayama, tm}@i.kyoto-u.ac.jp

あらまし ディスプレイに提示されたコンテンツに対してユーザが注視している対象を推定する際、ユーザが自由な姿勢や立ち位置をとる場合には視線を十分な精度で計測できず、ディスプレイ上における注視座標とオブジェクト座標を直接比較する従来の手法はその精度に限界がある。本稿では、眼球動作に反映されやすい特徴的な動き（イベント）をオブジェクトの動きに組み込み、注視対象であるオブジェクトの動きと眼球動作の間に存在する時間的な同期構造を分析することにより、注視対象を推定する手法を提案する。実験の結果、注視対象推定に同期構造分析を用いることで、推定精度が大幅に向上することを確認した。

キーワード 注視対象推定, 同期構造, 動的コンテンツ, 情報提示

Gaze Probing: Event-Based Estimation of Gazed Object

Ryo YONETANI[†], Hiroaki KAWASHIMA^{††}, Takatsugu HIRAYAMA^{††}, and Takashi MATSUYAMA^{††}

[†] Graduate School of Informatics, Kyoto University Yoshida-Honmachi, Sakyo, Kyoto, 606-8501 Japan

E-mail: [†]yonetani@vision.kuee.kyoto-u.ac.jp, ^{††}{kawashima, hirayama, tm}@i.kyoto-u.ac.jp

Abstract In this paper, we propose a novel approach to estimate the gazed object by using the cue of timing structure between the movements of the objects and the user's eye movement. First, we define an *event* as a characteristic part of the object's movement. This event causes a readily-observable eye movement, therefore we can figure out the eye movement even if the user's facial direction and position are free. Then, we analyze the timing structure between the starting point of the events and that of the eye movement caused by the events to determine which object was gazed by the user. Our experimental results show the effectiveness of using the timing structure for gazed object estimation.

Key words gazed object estimation, timing structure, dynamic contents, information display

1. はじめに

街角に設置された情報提示ディスプレイにおいてさまざまなコンテンツを提示する際、訪れたユーザがコンテンツ中のどの画像や文章に興味を持っているかという非明示的な情報は、ユーザの興味に合わせてシステム側から提示内容を変化させる場合や、提示方法を評価する場合に有用である。この興味を推定するために、ユーザの注視対象、すなわち提示されたコンテンツ中のどのオブジェクトを注視しているかという情報はしばしば重要な役割を果たす。一般に注視対象は、視線計測により得られるディスプレイ上における注視座標と、既知であるオブジェクト座標とを直接比較することで推定されており、注視対象推定の精度を向上させるために視線計測の高精度化が進められてきた。その例として赤外光源などの非接触な機器を用いる手法 [1] [2] や、ビデオカメラのみを用いた手法 [3] [4] が挙げられる。ここで、実環境におい

てユーザがコンテンツを閲覧している状況を考えると、ユーザが自由な姿勢や立ち位置をとることが予想される。この場合、計測された視線データには比較的大きな誤差が発生してしまい、注視座標とオブジェクト座標の位置関係を直接用いる従来の注視対象推定の精度には限界があると考えられる。そこで本研究では、注視座標とオブジェクト座標の位置関係を直接用いず、コンテンツの提示に動的な変化を持たせることで、その変化と眼球動作との時間的な同期構造を獲得し、視線計測の精度の低い環境において注視対象推定の高精度化を図る。

オブジェクトが動いている状況での注視対象推定を行った先行研究として文献 [5] がある。この研究では、視線計測によって得られるディスプレイ上における注視座標の時系列データと、オブジェクト座標の時系列データとの差異に基づいて注視対象の推定を行っている。この手法の推定精度は視線データの計測精度に依存するが、ユーザの頭部を固定し視線計測にある程度の信頼性を与

えることで、98%以上の精度で注視対象推定を行うことに成功している。しかし提示オブジェクトが多数存在する場合、オブジェクト同士の動きの相関は大きくなる。そのためこの手法は、精度の高い視線計測を用いないと注視対象推定の精度が悪化するという問題点がある。

本研究では、提示内容が時間的に変化する動的コンテンツを扱う状況を想定し、コンテンツを構成するオブジェクトが出現、消滅、移動、提示内容の切り替えといった時間に伴った変化（以下、イベントと呼ぶ）をするような提示を行う。そして、視線計測の精度が低く、得られる視線データに大きな計測誤差が含まれる状況においても、検出可能な眼球動作を発生させるようなイベントをコンテンツに組み込む。ユーザがイベントを追視することで、視線データにはそのイベントに対応した眼球動作が現れると考えられる。そこで、オブジェクトにイベントが発生した時刻と、眼球動作においてイベントへの反応が検出された時刻の時間差を用いることで、高精度の注視対象推定を試みる。

ここで、実用的なインタフェースのデザインを考えると、イベントはユーザにとって違和感のないものであることが望ましい。このため、オブジェクトそれぞれの変化にコンテンツ全体としての統一感を持たせる方法が考えられる。本研究ではオブジェクト同士が非常に似通った変化、すなわち互いに空間的・時間的に大きな相関を持った変化をするという状況を考える。

2. 提示イベントと眼球動作との同期構造分析に基づく注視対象推定

本章では、提示されたイベントと眼球動作から獲得される同期構造の分析手法および注視対象の推定手法について述べる。図1に提案手法の流れを示す。

2.1 動的コンテンツにおけるイベントのデザイン

動的コンテンツを構成する N 個のオブジェクトに対して、各オブジェクトの時刻 t におけるディスプレイ平面上での重心を以下の記号を用いて表す。

$$\mathbf{x}_n(t) \quad (n = 1, \dots, N, \mathbf{x}_n \in \mathbf{R}^2) \quad (1)$$

動的コンテンツのイベントを用いて、計測誤差が大きな視線データからでも検出可能な眼球動作を発生させるために、本研究ではイベントとしてオブジェクトの移動、特に並進移動の折り返しや静止状態からの動き出しなど、急激な速度の変化をともなう動きを用いる。これは、オブジェクトの移動において速度ベクトルが大きく変化する場合、その変化は計測誤差に影響を受けない眼球動作パターンとして観測できると考えられるためである。

ここで、動的コンテンツにおいて n 番目のオブジェクトに対して i 番目のイベントが発生する時刻を $t_{(n,i)}$ として表す。オブジェクト同士が非常に似通った変化をするという状況を考慮し、それぞれのオブジェクトにおけ

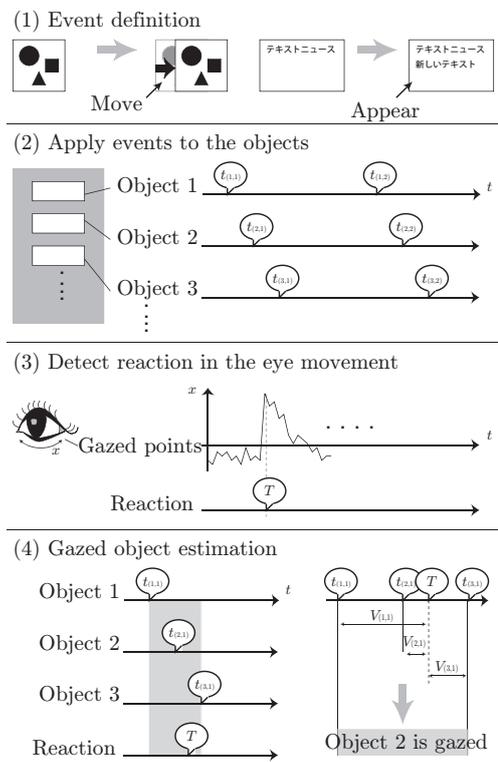


図1 提案手法の流れ

る i 番目のイベントが、一定の小さな時間差で順々に発生するように $t_{(n,i)}$ を設定する（図1(2)参照）。

2.2 眼球動作におけるイベント反応時刻の検出

ディスプレイ上におけるユーザの注視座標の計測系列を $\mathbf{x}_{\text{gaze}}(t)$ として定義する。 $\mathbf{x}_{\text{gaze}}(t)$ において、注視しているイベントに対応する眼球動作の反応が検出された時刻を T と表す。 T はイベントの発生時刻 $t_{(n,i)}$ 近傍におけるオブジェクトの動きと眼球動作との相関係数を用いて検出されるほか、眼球動作が2.1節で述べたような急激な速度変化にともなう動きや大きなサッカードである場合、眼球動作の加速度の大きい箇所を単純に探索することで検出できると考えられる。

なお、眼球動作の動特性として、視覚刺激に対するユーザの反応には潜時がともなう。一般に、滑らかな眼球動作の潜時は60~150msであり、サッカードの潜時は50~200msと言われている[6]。このように、 $t_{(n,i)}$ に対して T はある系統的な遅延をもって発生すると考えられる。これに加えて、イベントのデザインによってはユーザの眼球動作に以下のような認知特性の影響が生じることが予想される。

注視対象の動きに対する眼球動作の遅延 ユーザが注視対象であるオブジェクトのイベントによる変化を予測できない場合、 T が $t_{(n,i)}$ に対して遅れる。

注視対象の動きに対するユーザの予測 ユーザが注視対象であるオブジェクトのイベントによる変化を過度に予測する場合、 T が $t_{(n,i)}$ に対して先行する。

2.3 同期構造分析に基づく注視対象推定

各イベントの発生時刻 $t_{(n,i)}$ と眼球動作におけるイベント反応時刻 T との同期構造を分析する。そして、ユーザがどのイベントを注視していたかを推定することで、ユーザの注視対象であるオブジェクトを推定する。

まず、各イベントに対して、そのイベントをユーザが注視しているか否かを判別するための評価値 $V_{(n,i)}$ を導入する。ユーザがあるイベントを注視していれば、そのイベントの発生時刻と眼球動作におけるイベント反応時刻は同期すると期待できる。したがって、 $t_{(n,i)}$ と T 、2.2 節に述べた眼球動作の潜時や認知特性の影響 τ を用いて、評価値 $V_{(n,i)}$ を式 (2) に定義する。

$$V_{(n,i)} = (T + \tau) - t_{(n,i)} \quad (2)$$

ここで注視対象の推定時区間を、全オブジェクトの i 番目のイベントが含まれる時区間 $[t_{(\text{start},i)}, t_{(\text{end},i)}] = [\min_n t_{(n,i)}, \max_n t_{(n,i)}]$ として設定する。この推定時区間において、ユーザの注視対象は変化しないものとする。そして各オブジェクトの i 番目のイベントに対して得られる $V_{(n,i)}$ を比較することで、注視対象の推定を行う。例えば、各オブジェクトに p 番目のイベントが発生している推定時区間 $[t_{(\text{start},p)}, t_{(\text{end},p)}]$ において、 k 番目のオブジェクトが注視対象であれば、 $|V_{(k,p)}|$ は他のオブジェクトに比べて小さくなり、式 (3) が成立する。

$$k = \arg \min_n |V_{(n,p)}| \quad (3)$$

このように、眼球動作とイベントの同期構造を分析、比較することで、推定時区間において注視対象であるオブジェクトを推定することができる (図 1(4) 参照)。なお、各イベントの発生間隔を小さくすることで、推定時区間を十分に小さくすることが可能であり、推定の時間分解能を向上させることができる。

3. 大画面情報端末における提案手法の適用

3.1 大画面情報端末を用いた提示・計測環境

提案手法を適用・評価する具体的な状況として、図 2 に示すような大画面情報端末を用いた情報提示システムを設定する。システムによる情報提示は、縦型 50 インチプラズマディスプレイ^(注1)によって行われる。このディスプレイに対してユーザが 1m 離れた位置に立ち、提示コンテンツを閲覧している状況を想定する。この位置からユーザは十分にコンテンツ全体を見ることができ、かつオブジェクトの動きに対して計測に十分な眼球動作が発生する。ディスプレイの下にはカメラ^(注2)を設置し、ユーザの撮影を行う。カメラはユーザの正面顔が常に映るよ

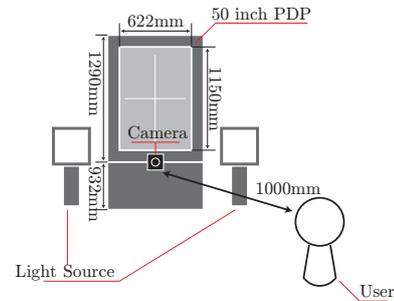


図 2 大画面情報端末を用いた環境

うに配置する。また、虹彩を検出するのに十分な光量を得るため、ディスプレイの両脇にライト^(注3)を配置する。ディスプレイ上にオブジェクトを提示する際には、以下の点に留意する。

オブジェクトの動く領域 オブジェクトの動く領域が小さすぎる場合、眼球動作が小さくなり計測誤差に完全に埋もれてしまう可能性がある。本研究における視線情報の計測は文献 [3] と同一の手法および環境で行っており、視線情報の計測誤差は平均 53.3mm である。オブジェクトはそれ以上の幅で動くことが好ましい。

オブジェクトの動く速さ 一般に、滑らかに動くオブジェクトに対して目の動きが追従できる速さは $2 \sim 40^\circ/\text{s}$ と言われている [6]。また、オブジェクトが写真画像など複雑な場合、その画像内容を吟味しながら追従できる速さはおよそ $2.8 \sim 4.2^\circ/\text{s}$ 程度と言われている [7]。

これらの知見を元にオブジェクトの動く領域および速さを設定する。

3.2 イベントのデザイン

本研究では、2.3 節で述べたように、推定時区間内で注視対象が変化しないという状況を想定し、注視対象推定に同期構造分析を用いることの有用性を検証する。動的コンテンツを構成するオブジェクトはディスプレイ上に縦に並べて配置し、水平方向の眼球動作が発生するようにイベントをデザインする。これは、目領域における虹彩の移動可能な範囲が垂直方向に比べて水平方向に大きいいため、イベントに対応する眼球動作の検出が容易になると考えられるためである。なお、一般に視線計測は垂直方向における精度が低く、縦に並べられたオブジェクトに対する注視対象の推定は困難であった。ここで、提示内容が画像か文章かによって、閲覧の際に発生する眼球動作が異なることが予想される。以下、注視座標の計測系列 $\mathbf{x}_{\text{gaze}}(t)$ の水平方向成分を $x_{\text{gaze}}(t)$ とする。

本研究では、オブジェクトの種類によって異なるイベントのデザインを考える。

a) オブジェクトが画像の場合

オブジェクトが画像の場合、提示内容を閲覧する際の

(注1)：富士通製 UBWALL

(注2)：Point Grey Research 社製 Grasshopper (UXGA (1600×1200pixels), 8bit グレー画像, 30fps, 1/1.8 インチ CCD)。レンズは FUJINON 社製 HF16HA-1B (f=16mm)

(注3)：写真電気工業社製 RIFA-F (50×50cm)

眼球動作はユーザに依存し、複雑になると考えられる。しかし、オブジェクトが時間変化にしたがって移動することで、オブジェクトに追従する眼球動作を発生させることができる。この場合、眼球動作のベクトル方向は移動のベクトル方向と一致すると考えられる。そこで、オブジェクトをある時区間において移動させることをイベントとする。具体的には、

- 静止中のオブジェクトが移動する。
- 移動中のオブジェクトが停止する。
- 移動中のオブジェクトが移動方向を変える。

といったイベントが考えられる。これはオブジェクトの移動における速度ベクトルの大きな変化は、眼球動作に大きく反映すると期待されるためである。

b) オブジェクトが文章の場合

オブジェクトが文章の場合、ユーザは停留とサッカー道を繰り返して数単語ずつ文章を読解していくことが知られている [7]。この状況においてオブジェクトに動きを持たせた場合、読解が著しく困難になることが予想される。そこで、イベントとして行単位での文章の出現を考える。これは、出現のタイミングを適切に設定することで、文章に新しい行が追加されるたびに行末から新しい行の文頭への大きなサッカー道が発生することが期待されるためである。

3.3 眼球動作の計測

注視座標の計測系列 $x_{\text{gaze}}(t)$ は、Active Appearance Model (AAM) [8] と 3次元眼球モデルを用いて計測できる [3]。まず、顔の特徴点を AAM を用いて抽出し、あらかじめ作成しておいた 3次元顔形状モデルと照合して顔姿勢の推定を行う。虹彩検出には単純な楕円探索を用い、それによって得られた虹彩中心と 3次元顔形状モデルが持つ眼球中心を結ぶ直線を視線と定義し、ディスプレイ平面と視線の交点から $x_{\text{gaze}}(t)$ を算出する。この $x_{\text{gaze}}(t)$ には高周波の計測誤差が含まれており、平滑化を行う必要がある。しかし、本手法では眼球動作の時間情報を重要視しており、平滑化を最低限にとどめる必要がある。文献 [6] より眼球が滑らかにオブジェクトに追従する際の最高速度は $40^\circ/\text{sec}$ 、周波数に換算すると 0.11Hz である。眼球はこの周波数以上では運動しないと仮定し、次章の評価実験においては $x_{\text{gaze}}(t)$ に対して、カットオフ周波数を 0.11Hz に設定したガウス関数の畳み込みによる平滑化を行う。

3.4 眼球動作におけるイベント反応時刻の検出

発生する眼球動作によって、イベント反応時刻 T は以下の方法で検出できる。

滑動性眼球運動 オブジェクトが動き、ユーザがそれを滑らかな眼球動作で追視する場合、 T は、推定時区間 $[t_{(\text{start},i)}, t_{(\text{end},i)}]$ における注視座標の計測系列 $x_{\text{gaze}}(t)$ とイベント発生時刻 $t_{(n,i)}$ 近傍のオブジェクトの動きとの

相関から検出できると考えられる。すなわち、

$$T = \arg \max_t (\text{corr} (x_{\text{gaze}}(t), e_{(n,i)}(t))) \quad (4)$$

$e_{(n,i)}(t)$ は $t_{(n,i)}$ 近傍におけるオブジェクトの動きである。

$$e_{(n,i)}(t) = x_n(t) \quad (t_{(n,i)} - \alpha < t < t_{(n,i)} + \alpha) \quad (5)$$

corr は正規化相互相関関数である。

$$\text{corr} (x_{\text{gaze}}(t), e_{(n,i)}(t)) = \frac{\int (x_{\text{gaze}}(m) - \bar{x}_{\text{gaze}}) (e_{(n,i)}(t+m) - \bar{e}_{(n,i)}) dm}{\sqrt{\int (x_{\text{gaze}}(m) - \bar{x}_{\text{gaze}})^2 dm} \sqrt{\int (e_{(n,i)}(t+m) - \bar{e}_{(n,i)})^2 dm}} \quad (6)$$

\bar{x}_{gaze} , $\bar{e}_{(n,i)}$ は $x_{\text{gaze}}(t)$, $e_{(n,i)}(t)$ の平均値である。

サッカー道 眼球動作として大きなサッカー道が予測される場合、 T は $[t_{(\text{start},i)}, t_{(\text{end},i)}]$ における $x_{\text{gaze}}(t)$ の 2階微分値の極値から検出することができる。

$$T = \arg \max_t \left| \frac{d^2}{dt^2} x_{\text{gaze}}(t) \right| \quad (t_{(\text{start},i)} < t < t_{(\text{end},i)}) \quad (7)$$

4. 評価実験

3.1 節で述べた環境において、提案手法の精度評価を目的とした実験を行った。4.1 節の実験では、まず単純な状況設定においてユーザの眼球動作や同期構造を分析することを目的として、単一色で塗られた 2つの小矩形図形に対する注視対象推定を行った。4.2 節、4.3 節の実験では、実環境のインタフェースへの適用を想定した際の提案手法の有用性を評価することを目的として、提示するオブジェクトとしてそれぞれ映画のシーン画像、テキストベースのニュースを用い、3つ以上のオブジェクトに対する注視対象推定を行った。

4.1 単純な小矩形図形に対する注視対象推定

4.1.1 実験方法

オブジェクトは 32.4mm 四方の灰色一色で塗られた小矩形図形 2つを用い、被験者がこれらのオブジェクトのうちどちらを注視しているかを推定した。生理学的には人間の中心視野は 2° 程度であり、被験者の立ち位置からは中心視野のみでこの図形全体を注視することができる。

視線計測の誤差を考慮し、それぞれのオブジェクトはディスプレイ中心から上下 54.0mm に配置した。この場合、得られる視線データの誤差の影響により、注視座標とオブジェクト位置を直接比較する従来手法では正しく注視対象推定が行えない可能性がある。

それぞれのオブジェクトには図 3, 4 に示すような水平方向に往復する動きを持たせた。これら 2種類の動きに対して、その動きの折り返しをイベントとして定義する。正弦波は折り返しから動き出す速さが遅く、周期的一次遅れ波は折り返しから動き出す速さが速いという特徴を持ち、それぞれユーザの認知特性に異なる影響を与えると予想される。オブジェクトの動きには何種類かの時間差を持たせ、先に動くオブジェクトを先行オブジェ

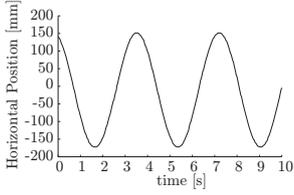


図3 正弦波.
往復幅 324.0mm
周期 3.67s

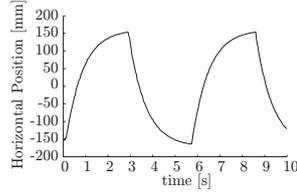


図4 周期的一次遅れ波.
往復幅 308.0mm
周期 5.70s

クト、後に動くオブジェクトを後行オブジェクトと呼ぶ。ここで、文献[3]より、本実験で用いる視線計測の精度は、ディスプレイ上の位置によって大きく異なる。提案手法の有用性を示す上で、上側オブジェクトに対する推定と下側オブジェクトに対する推定で精度に違いが現れないことを確かめる必要がある。よって、コンテンツの提示方法を以下の組み合わせから設定した。

オブジェクトの動き方 正弦波、周期的一次遅れ波

オブジェクト間の位相差 $\frac{1}{16}\pi$, $\frac{1}{8}\pi$, $\frac{1}{4}\pi$ rad

オブジェクトの順序 上側が先行, 下側が先行

以上 $2 \times 3 \times 2 = 12$ 通りの組み合わせについて、2名の被験者（以下A, Bとする）に上下それぞれのオブジェクトを20秒間注視し続けてもらった（すなわち1被験者あたり24試行）。そして被験者がオブジェクトを注視する様子をカメラで撮影し、得られた顔画像系列に対して3.3節で述べた方法で注視座標系列を計測した。

眼球動作におけるイベント反応時刻 T は、注視対象のオブジェクトが常に動いていることから、オブジェクトの動きと眼球動作との相関から検出した。式(5)において $\alpha = 500\text{ms}$ に設定した。

同期構造分析を行う上では2.3節で述べたように、眼球動作の潜時や被験者の認知特性の影響 τ を考慮しなければならない。本実験では往復するオブジェクトに対して被験者は滑らかな眼球動作で追従を行うと予想される。この場合、2.2節より、眼球動作の潜時は60~150msであると考えられる。また、単純な小矩形図形の往復運動に対する認知特性の影響は小さいと考えられる。そこで、本実験では潜時が最も大きくなってしまった場合を補償するために $\tau = 150\text{ms}$ に設定して推定を行った。

推定精度として得られる認識率 R は、1試行に対して正弦波から9個、周期的一次遅れ波から6個のイベントを検出し、以下のように求めた。

$$R = e_{\text{success}}/e_{\text{all}} \quad (8)$$

e_{success} は認識に成功したイベントの数であり、 e_{all} は1試行におけるイベントの総数である。

4.1.2 比較対象とする推定手法

提案手法との比較のため以下の手法で推定を行った。

オブジェクト座標と注視座標の位置関係に基づいた推定
各オブジェクトの座標 $\mathbf{x}_n(t)$ と注視座標 $\mathbf{x}_{\text{gaze}}(t)$ とのユー

クリッド距離を求め、小さい値をとる方のオブジェクトを注視対象とする。すなわち、 $\mathbf{x}_n(t) = (x_n(t), y_n(t))$, $\mathbf{x}_{\text{gaze}}(t) = (x_{\text{gaze}}(t), y_{\text{gaze}}(t))$ としたとき、時刻 t において注視対象である k 番目のオブジェクトは

$$k(t) = \arg \min_n \sqrt{(x_n(t) - x_{\text{gaze}}(t))^2 + (y_n(t) - y_{\text{gaze}}(t))^2} \quad (9)$$

と求められる。1試行に対して推定精度として得られる認識率 R は以下の式で与えられる。

$$R = f_{\text{success}}/f_{\text{all}} \quad (10)$$

ここで、 f_{success} は認識に成功したフレームの数、 f_{all} は1試行における全フレーム数であり、 $f_{\text{all}} = 600$ である。

オブジェクトの動きと眼球動作の相関係数に基づいた推定

オブジェクトの動きに眼球動作が追従するという点に着目し、推定時間区間 $[t_{(\text{start},i)}, t_{(\text{end},i)}]$ における各オブジェクト座標と注視座標の時系列の相関係数を求め、大きい方のオブジェクトを注視対象とする。ここで、オブジェクトの動きが水平方向のみであるという点を考慮し、推定には水平方向成分の相関係数を扱う。すなわち、注視対象である k 番目のオブジェクトは以下のように求まる。

$$\begin{aligned} k &= \arg \max_i \left(\frac{A}{B \cdot C} \right) \\ A &= \int_{t_{(\text{start},i)}}^{t_{(\text{end},i)}} (x_n(m) - \bar{x}_i)(x_{\text{gaze}}(m) - \bar{x}_{\text{gaze}}) dm \\ B &= \sqrt{\int_{t_{(\text{start},i)}}^{t_{(\text{end},i)}} (x_n(m) - \bar{x}_i)^2 dm} \\ C &= \sqrt{\int_{t_{(\text{start},i)}}^{t_{(\text{end},i)}} (x_{\text{gaze}}(m) - \bar{x}_{\text{gaze}})^2 dm} \end{aligned} \quad (11)$$

\bar{x}_i , \bar{x}_{gaze} はそれぞれ $[t_{(\text{start},i)}, t_{(\text{end},i)}]$ における $x_n(t)$, $x_{\text{gaze}}(t)$ の平均値である。この手法においても眼球動作に τ のバイアスをかけることで推定精度が向上すると考えられる。なお、認識率 R は式(8)を用いて求められる。

4.1.3 推定結果と考察

a) 注視対象推定の結果

表1に被験者A, Bの推定結果の平均を示す。表中の(a)はオブジェクト座標と注視座標の位置関係に基づいた推定手法、(b)はオブジェクトの動きと眼球動作の相関係数に基づいた推定手法、Proposedは提案手法を示す。

垂直方向の視線データには大きく計測誤差が含まれるため、手法(a)に注目すると、全体的に上側のオブジェクトを注視した場合と下側のオブジェクトを注視した場合で認識率が大きく異なっている。これに対し手法(b)および提案手法は認識率がオブジェクトの位置に依存していない。

b) 考察

図5に評価値 $V_{(n,i)}$ の頻度分布を示す。図5(i)より正弦波における $V_{(n,i)}$ の最頻値は -60ms であり、設定した $\tau = 150\text{ms}$ から差し引くと、 90ms の眼球潜時が最も多く発生していたことになる。この結果より、被験者がオブジェクトの動きに対して滑らかな眼球動作で追従できていたと考えられる。一方(ii)より周期的一次遅れ波

表 1 推定結果. (a) はオブジェクト座標と注視座標の位置関係に基づいた推定手法, (b) はオブジェクトの動きと眼球動作の相関係数に基づいた推定手法, (c) は提案手法を示す.

Order	Phase [rad]	Sine			Periodic first-order lag		
		(a)[%]	(b)[%]	Proposed[%]	(a)[%]	(b)[%]	Proposed[%]
Top/Preceding	$\pi/4$	39.4	100.0	100.0	67.8	100.0	100.0
	$\pi/8$	56.3	100.0	94.4	75.4	100.0	100.0
	$\pi/16$	44.7	100.0	52.8	56.8	100.0	81.3
Top/Following	$\pi/4$	55.6	100.0	100.0	55.0	100.0	100.0
	$\pi/8$	82.3	97.2	97.2	57.5	100.0	100.0
	$\pi/16$	57.5	31.3	86.1	65.7	100.0	100.0
Bottom/Preceding	$\pi/4$	38.6	100.0	100.0	16.7	100.0	100.0
	$\pi/8$	19.4	100.0	100.0	10.1	100.0	100.0
	$\pi/16$	15.0	100.0	100.0	11.6	88.8	61.3
Bottom/Following	$\pi/4$	34.6	100.0	100.0	19.0	100.0	100.0
	$\pi/8$	42.5	97.2	100.0	6.9	100.0	100.0
	$\pi/16$	37.9	27.8	80.6	10.5	93.8	100.0
Average		43.6	87.8	92.6	37.7	98.6	95.2

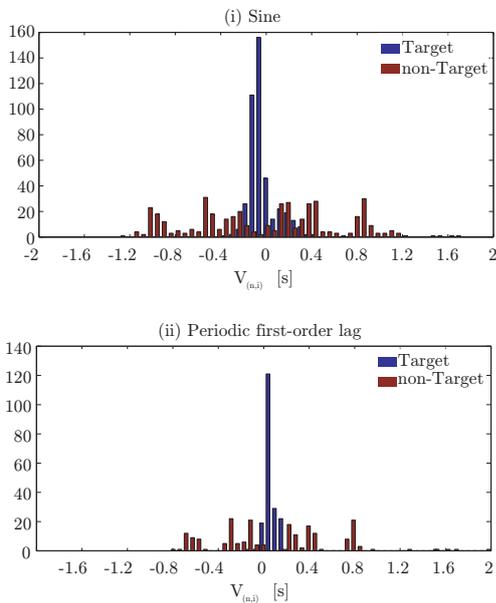


図 5 評価値 $V_{(n,i)}$ のヒストグラム ((i) 正弦波 (ii) 周期的な一次遅れ波)

における $V_{(n,i)}$ の最頻値は 50ms であり, 200ms の眼球動作の潜時が最も多く発生していたことになる. この結果は滑らかな眼球動作の潜時から外れている. これに関して, 周期的な一次遅れ波は動きの折り返しタイミングが予測しにくく, それに加えて折り返し始めの初速が速すぎるため, オブジェクトを追視することが難しいという意見が被験者 A, B 双方から得られた. また, 得られた潜時は 2.2 節で述べたサッカードの潜時 50~200ms と合致することから, 被験者はオブジェクトの動きの折り返しをサッカードで追視していたと考えられる.

提案手法を用いることで, 大きな計測誤差を含む垂直方向の視線データを全く使わずに, 垂直方向に並べられた複数のオブジェクトに対する注視対象を推定すること

を可能にしている. なお, 手法 (b) においても提案手法と同程度の推定精度を得ることができている. これは, 本実験は往復運動をするオブジェクトに対してユーザが常に 1 つのオブジェクトを注視し続けるという単純な状況であり, 相関係数の計算 (式 (11)) で, 比較的長い時間区間で積分をとることができたためであると考えられる. しかし, コンテンツの閲覧中にユーザの注視対象が変化する場合, 眼球動作がより複雑になり, 相関係数を用いた推定は困難になることが予想される. これに対して提案手法での推定では, 眼球動作におけるイベント反応時刻 T の検出を, 比較的短い時間区間で積分計算 (式 (6)) や加速度の極値 (式 (7)) から求めることができる. したがって, 1 つのオブジェクトに複数個のイベントを組み込み推定時間区間の数を増やすといった拡張が可能であり, 推定時間区間で注視対象が変化しないという制約があっても, 推定の時間分解能を向上させて注視対象の変化に対応できると考えられる.

4.2 写真画像に対する注視対象推定

4.2.1 実験方法

本実験では, オブジェクトを写真画像とし, 3.1 節で述べた環境において, 映画のいくつかのシーンを紹介するコンテンツを閲覧する状況を想定し, コンテンツおよびイベントのデザインを行った (図 6 参照).

このコンテンツは図 6 の Display Area に示すオブジェクト提示領域 (287.0mm×159.0mm) を 6 つ持ち, それぞれ異なる映画の画像を提示している. 各提示領域間の距離は 5.3mm である. 各オブジェクトは 4 枚の画像からなり, 提示領域に画像を順番に提示する. 新たな画像は 2400ms ごとに提示領域の左側 235mm の位置に出現する. その際, 提示領域の画像は薄暗く暗転する. 出現した新たな画像は $11^\circ/s$ の速さで提示領域まで移動する. この新たな画像の出現を本コンテンツのイベントとして

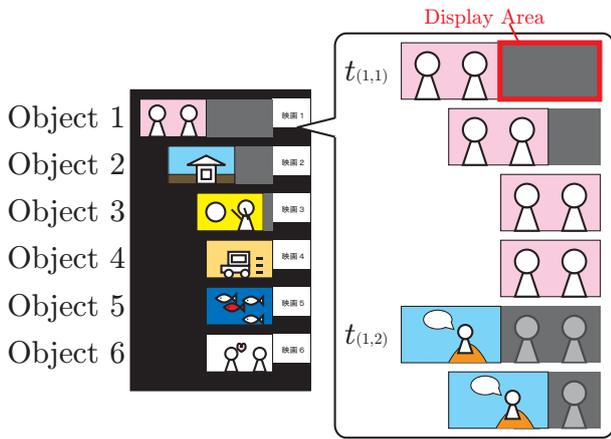


図 6 映画のいくつかのシーンを紹介するコンテンツ

定義する. 3.1 節より, この速さで動く画像に対して, 被験者は画像内容を詳細に吟味することはできず, イベント発生時刻周辺で複雑な眼球動作が起こることを避けられると考えられる. 各オブジェクトのイベントの発生タイミングの時間差は 400ms に設定した. 以上の状況において, 被験者 A は各オブジェクトの提示内容をそれぞれ 20 秒間閲覧した.

本実験におけるイベントによって, 被験者の眼球動作には提示中の画像から新たに提示される画像に対してサッカードが発生することが予想される. そこで, 眼球動作におけるイベント反応時刻 T を推定時区間 $[t_{(start,i)}, t_{(end,i)}]$ における注視座標の計測系列 $x_{gaze}(t)$ の 2 階微分値の極値から求めた. またこの場合, 2.2 節より, 潜時は 90~200ms であると考えられる. 本実験は提示内容が複雑なため, 認知特性の影響を受けると考えられるが, 少なくとも眼球動作の潜時が最も大きくなってしまった場合を補償するために $\tau = 200ms$ に設定して推定を行った.

4.2.2 推定結果と考察

表 2 に推定結果を示す. また, 図 7 に得られた眼球動作とイベントに対して検出された反応の例を示す. 図 6 より, 被験者はオブジェクトが提示領域で静止している際に画像内容の吟味を行い, 提示領域の左側に新たな画像が提示されるとその動きに反応, 追視していることが分かる. オブジェクトの提示内容が複雑になることで眼球動作がより複雑になっており, 相関係数を用いた比較手法 (b) の推定精度は大きく低下している. これに対して提案手法は, 図 7 のように新たな画像の出現に反応することで発生するサッカードをイベントとして検出することで, 精度の高い推定を行うことができている.

4.3 テキストに対する注視対象推定

4.3.1 実験方法

本実験では, テキストベースのニュースコンテンツを閲覧する状況を想定し, コンテンツおよびイベントのデザインを行った (図 8 参照).

このコンテンツは図 8 の Display Area に示すオブジェ

表 2 推定結果. (a) はオブジェクト座標と注視座標の位置関係に基づいた推定手法, (b) はオブジェクトの動きと眼球動作の相関係数に基づいた推定手法, (c) は提案手法を示す.

Object	(a)[%]	(b)[%]	(c)[%]
1	0.0	0.0	33.3
2	65.0	0.0	62.5
3	0.0	75.0	80.0
4	0.0	75.0	100.0
5	58.2	0.0	100.0
6	0.0	25.0	80.0
Average	16.4	25.0	76.0

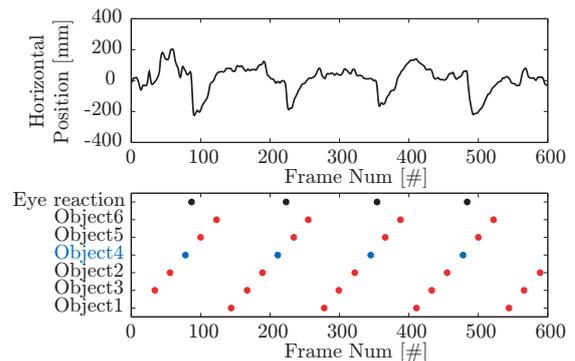


図 7 眼球動作と検出された反応の例 (注視対象: Object 4)

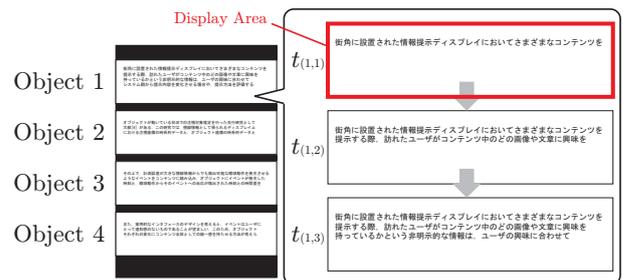


図 8 テキストベースのニュースコンテンツ

クト提示領域 (768.0mm×191.0mm) を 4 つ持ち, それぞれ複数行からなるテキストベースのニュースを提示している. 各提示領域間の距離は 15.9mm であり, テキスト 1 行の高さに等しい. 各オブジェクトは 5 行のテキストからなり, 提示領域に 1 行ずつテキストを追加して提示する. 1 行あたりの長さは最大 538.5mm (日本語 32 文字) である. 新たな行は 1800ms ごとに追加して提示される. この新たな行の出現を本コンテンツのイベントとして定義する. 各オブジェクトのイベントの発生タイミングの時間差は 400ms に設定した. 以上の状況において, 被験者 A は各オブジェクトの提示内容をそれぞれ 20 秒間閲覧した.

本実験におけるイベントによって, 被験者の眼球動作には提示中の行の文末から新しい行の文頭へのサッカードが発生することが予想される. そこで, T の検出方法や τ の値は 4.2 節の実験と同様とした. なお, 本実験で

表 3 推定結果. (a) はオブジェクト座標と注視座標の位置関係に基づいた推定手法, (c) は提案手法を示す.

Object	(a)[%]	(c)[%]
1	0.0	66.7
2	23.0	100.0
3	52.9	100.0
4	0.0	100.0
Average	19.0	91.7

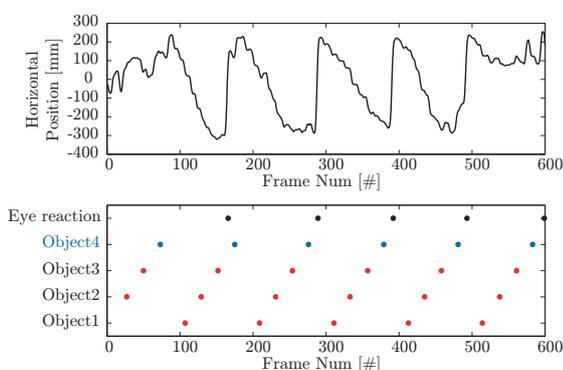


図 9 眼球動作と検出された反応の例 (注視対象: Object 4)

はオブジェクトに対して時間変化にともなう動きを与えていないため, オブジェクトの動きと眼球動作の相関係数に基づいた比較手法 (b) での推定は行えなかった.

4.3.2 推定結果と考察

表 3 に推定結果を示す. また, 図 9 に得られた眼球動作とイベントに対して検出された反応の例を示す.

3.2 節で述べたように, 読者は基本的には注視の停留とサッカドを繰り返すことで文章を数文字ずつ読解することが知られている. イベントによって発生した文末から文頭へのサッカドは読解におけるサッカドに比べて十分大きく, 図 9 のようにイベントに対する眼球動作の反応として検出することができる.

本実験ではこのように高い推定精度を得ることができたが, 新たな行の出現がユーザの読解の速さに対して遅すぎる場合, ユーザが新たな行の出現を予測して, 文頭位置をあらかじめ注視する状況も考えられる. 逆に出現のタイミングが早すぎる場合, ユーザは出現に対して読解が追いつかず, 前の行の読解が終わってからイベント発生時刻より遅れて新たな行の文頭を注視するため, 正しく推定ができない状況も考えられる. これに関して, ユーザの読解状況を逐次観察し, 1 行を読みきる時間をフィードバックとして得ることで, 適切な出現タイミングを設定することができると思われる.

5. おわりに

本研究では, 動的コンテンツを用いてオブジェクトの動きとユーザの眼球動作との時間的な同期構造を分析することで, 実環境においてユーザの姿勢及び立ち位置の自由さを許容する精度の高い注視対象推定法を提案した.

単純な往復運動をする小矩形図形と, 実環境におけるインタフェースを想定した映画情報の紹介コンテンツ, テキストベースのニュースコンテンツに対する注視対象推定の実験から, ユーザが 1 つのオブジェクトを注視し続ける状況において, 注視対象推定に同期構造分析を用いることの有用性が示された.

推定において被験者の眼球動作の潜時や認知特性の影響 τ は非常に重要な要素である. τ はイベントのデザインに加えて提示コンテンツ, ユーザの特性によって様々に変化することが想定される. これに関して, 実際の状況では事前にキャリブレーションフェーズを設けてユーザの眼球動作の特性を観察することで, 発生する眼球動作やその潜時を決定できると考えられる. また, 同時にユーザに提示コンテンツの動きに慣れさせることで, 動きを予測できないことによる反応の遅延などの認知特性の影響を低減できると考えられる.

実環境において提案手法を用いる上で, 注視対象の変化にどの程度対応できるかは重要な課題である. 本研究では, オブジェクトに組み込むイベントの数を増やす方法を挙げたが, その際ユーザの閲覧を阻害しないように動的コンテンツのデザインを工夫する必要がある. また, ユーザが自由な立ち位置や顔向きをとる場合についても本研究では十分に扱うことができなかった. ユーザの注視対象が変化する場合とあわせて, これらの場合にどの程度の精度で推定ができるかという点について, 今後検討していきたい.

謝 辞

本研究の一部は, 科学研究費補助金 18049046 の補助を受けて行った.

文 献

- [1] C. Hennessey, B. Noureddin and P. Lawrence: "A single camera eye-gaze tracking system with free head motion", *Advances in eye tracking technology*, pp. 87–94 (2006).
- [2] C. H. Morimoto and M. R. M. Mimica: "Eye gaze tracking techniques for interactive applications", *Computer Vision Image Understanding*, **98**, 1, pp. 4–24 (2005).
- [3] 佐竹, 小林, 平山, 川嶋, 松山: "高解像度撮影における実時間視線推定の高精度化", *電子情報通信学会技術報告*, **107**, 491, pp. 137–142 (2008).
- [4] 山添, 内海, 米澤, 安部: "3次元眼球モデルを利用した単眼カメラによる遠隔視線推定", *画像の認識・理解シンポジウム (MIRU2008)*, pp. 1650–1655 (2008).
- [5] 岡本, 藤江, 小林: "視線方向とオブジェクトの動きの関係に基づいた注目オブジェクト推定", *画像の認識・理解シンポジウム (MIRU2006)*, pp. 31–36 (2006).
- [6] 日本視覚学会 (編): "視覚情報処理ハンドブック", 朝倉書店 (2000).
- [7] 荻阪, 中溝, 古賀: "眼球運動の実験心理学", 名古屋大学出版会 (1993).
- [8] T. F. Cootes and C. J. Taylor: "Statistical models of appearance for medical image analysis and computer vision", *Proc. SPIE Medical Imaging*, **4322**, pp. 236–248 (2001).