

# Gaze Probing: イベント提示に基づく注視オブジェクト推定

米谷 竜<sup>\*1</sup> 川嶋 宏彰<sup>\*1</sup> 平山 高嗣<sup>\*1</sup> 松山 隆司<sup>\*1</sup>

Gaze Probing: Event-Based Estimation of Objects Being Focused On

Ryo Yonetani<sup>\*1</sup>, Hiroaki Kawashima<sup>\*1</sup>, Takatsugu Hirayama<sup>\*1</sup>, and Takashi Matsuyama<sup>\*1</sup>

**Abstract** – We propose a novel method to estimate the objects that a user is focusing on by using the synchronization between the movements of objects and a user's eyes as a cue. We first design an event as a characteristic motion pattern, and we then embed it within the movement of each object. Since the user's ocular reactions to these events are easily detected using a passive camera-based eye tracker, we can successfully estimate the object that the user is focusing on as the one whose movement is most synchronized with the user's eye reaction. Experimental results obtained from the application of this system to dynamic content demonstrate the effectiveness of the proposed method over existing methods.

**Keywords** : event-based gaze estimation, synchronization, dynamic content, eye movement

## 1. はじめに

街角や店内に設置されたディスプレイを用いてユーザに地域情報や商品といったさまざまな情報を提示する際には、(a) コンテンツ中のオブジェクト（画像や文章等）にスクロールなどの動きを取り入れる、(b) ユーザの興味に応じて提示コンテンツを切り替える、といった動的な提示方法を設計でき、ユーザが必要とする情報を効果的に提示・推薦することが実現可能となる。このとき (b) においてユーザの興味を知るうえで、ユーザの注視オブジェクト（提示された複数オブジェクトのうちユーザが実際に注視しているオブジェクト）が重要な手がかりとなる。本研究では、この注視オブジェクトを推定するために、(a) のようなオブジェクトの動きを伴うコンテンツ（以下、動的コンテンツと呼ぶ）を積極的に利用することを考える。

注視オブジェクトの推定には、視線計測により得られるディスプレイ上の注視座標と、既知であるオブジェクト座標とを直接比較する手法がしばしば用いられる。視線計測の精度はユーザ（被計測者）側の自由度とのトレードオフであり、ユーザの立ち位置や顔向きをある程度制限できる状況であれば、赤外光を照射するようなアクティブ計測を行うことで高精度な視線計測を行うことが可能である<sup>[1]~[4]</sup>。一方で、情報提示を行う実際の状況であれば、ユーザが自由な立ち位置や顔向きでコンテンツを閲覧することになるため、ビデオカメラのみを用いて視線計測を行う手法<sup>[5]~[8]</sup> が用いられることが多い。しかし、このような受動的な視線

計測手法では視線データに比較的大きな誤差が発生するため、注視オブジェクト推定の精度には限界がある。

注視座標の絶対位置を用いない注視オブジェクト推定としては、互いに大きく異なる動きを持つような複数オブジェクトを提示し、その際に得られるユーザの眼球運動と各オブジェクトとの相対的な動きの近さを評価する手法が挙げられる<sup>[9]</sup>。しかしながら、眼球運動の相対的なパターンは精度よく計測されることを前提としており、さらに、自然な（統一感のある）デザインにするという観点からオブジェクトの動きを互いに似通ったものにする場合や、オブジェクト数が増えた場合には、提示オブジェクト同士の動きの相関が大きくなり推定精度が低下する。その結果、眼球運動を安定に計測するためにやはりユーザの自由度を制限せざるを得ない。

そこで本論文では、計測誤差の影響を受けやすい眼球運動の変位パターンを直接評価するのではなく、眼球運動がいつ発生したかという時間的な情報を抽出し、コンテンツの動きと眼球運動との同期の程度を分析することで、ユーザの注視オブジェクトを高精度に推定する手法—Gaze Probing—を提案する。Gaze Probing では、まず、計測誤差が大きな視線データにおいても眼球運動の反応が観測できるような特徴的な短時間の動き（以下、イベントと呼ぶ）をデザインし、ユーザの注視オブジェクトを探るためのプローブとして、各オブジェクトの動きの一部に組み込んでおく。ユーザがこのような動的コンテンツを注視すると、注視オブジェクトのイベントに対応した反応が、ユーザ本人が気付かないうちに視線の計測データに現れる。このとき、イベントに反応した時刻を抽出することで、

\*1: 京都大学大学院情報学研究所

\*1: Graduate School of Informatics, Kyoto University

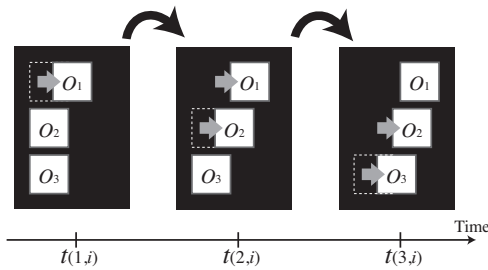


図1 イベントの組み込まれた動的コンテンツ。各オブジェクトにおいて一定の小さな時間差でイベントを順々に発生させる。

Fig.1 Events-embedded dynamic content. Each object orderly generates events in a constant and short time interval.

「ユーザがどの時刻のイベントに反応したか」という時間的な情報を評価する。これにより、視線計測の精度を十分に望めない(ユーザの自由度の高い)状況や、提示オブジェクトの動きが互いに大きな相関を持つ場合であっても、ユーザの注視オブジェクトを安定に推定することが可能となる。

次章では、Gaze Probing を用いる上でイベントが持つべき要件を整理し、適切なイベントのデザインを導く。これに基づき、眼球運動特性に基づく反応時刻検出方法、および注視オブジェクトの推定手法を示す。3章では、ディスプレイによる情報提示の具体的状況を設定し、注視オブジェクト推定精度に関する予備実験について述べる。4章では、いくつかの動的コンテンツデザイン事例を示し、これらを提示した際のユーザの注視オブジェクト推定精度を示す。5章では本手法の問題や解決法について検討し、6章でまとめる。

## 2. Gaze Probing: イベント提示に基づく注視オブジェクト推定

### 2.1 動的コンテンツにおけるオブジェクト

ユーザがディスプレイ上に提示される複数のオブジェクトを閲覧しながら、いずれかを選択しようと呼んでいる状況を考える。ここで、選択肢となる各オブジェクトは画像や文章などのメディアから構成されるものとし、各オブジェクトには動きを持たせることができるとする。このように動きを伴う複数オブジェクトからなる提示コンテンツを、本論文では動的コンテンツと呼ぶ。たとえば、図1は動きを伴う3つの画像オブジェクトからなる動的コンテンツである。注視オブジェクト推定とは、このような動的コンテンツを提示されたユーザが、いつどのオブジェクトを注視しているかといった注視状態を推定する問題である。

オブジェクト集合  $\{O_n | n = 1, \dots, N\}$  から構成される動的コンテンツを考える。ただし以下では、ある動的コンテンツを構成する各オブジェクトは同種のメディ

アとし、さらに、一時的なオーバーラップやディスプレイ上からの一時的な消滅は許すものの、各オブジェクトは互いに視覚的に分離・追跡可能であるとする。このとき、各オブジェクトに、ユーザの注視状態を探るためのイベントを組み込む。すなわち、ディスプレイ平面上での時刻  $t$  におけるオブジェクト  $O_n$  の重心を  $x_n(t) \in \mathbb{R}^2$  とし、その  $i$  番目のイベント発生時刻を  $t_{(n,i)}$  とすれば、 $t_{(n,i)}$  の近傍においてイベントの動き  $e_{(n,i)}(t)$  が現れるようにする。

$$x_n(t) = e_{(n,i)}(t) \quad (t_{(n,i)} - \delta_1 \leq t \leq t_{(n,i)} + \delta_2) \quad (1)$$

ここで、 $e_{(n,i)}(t)$  をイベント発生時刻前から定義するのは、たとえば、動き出す直前にオブジェクトは静止している、といったことを規定するためである。一方、イベントが組み込まれている区間外では、 $O_n$  の動きがイベントより低い顕著性を持つように、適宜  $x_n(t)$  を補間する。

Gaze Probing では、オブジェクトに組み込まれたイベントへの反応が視線データに現れることを利用する。そのために、異なるオブジェクト  $O_m, O_n (m \neq n)$  にてイベント発生時刻を時間的にずらし、この時間間隔を  $|t_{(m,i)} - t_{(n,i)}| > \varepsilon$  として後述する眼球運動の潜伏時と区別できる程度に大きくする。これにより、ユーザの視線データに現れる反応時刻さえ検出できれば、反応時刻とイベント発生時刻の時間関係からどのオブジェクトを注視していたかを推定できるようになる。次節では、 $e_{(n,i)}(t)$  の適切なデザイン方法を検討する。

### 2.2 提示イベントのデザイン

提示されるイベントは、(1) 視線データに良く反映される動きであり、かつ (2) 反映された時刻を精度良く検出できる動きである必要がある。そこで、具体的に以下の条件を設けることにする。

#### (1) イベントの動きを反映させるための条件

(1a) 短時間の動きである ユーザは注視オブジェクトを頻繁に変更することもあるため、イベントが発生しているオブジェクトを短時間でも追視していれば、視線データにその動きが反映される状況が望ましい。よって、オブジェクトに対する視線停留時間に比べイベントの動きは短時間であるとする。

(1b) 追視される際の遅延が小さい イベントの発生時刻と眼球運動の反応時刻との時間関係を用いるため、イベントに対するユーザの反応には大きな遅れが生じないことが望ましい。そこで、イベントはユーザにとって予測しやすい単純な動きであり、かつ十分追視できる速度とする。

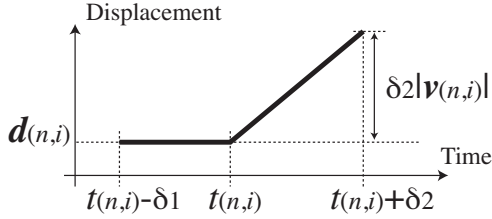


図2 イベント発生時刻近傍  $[t(n,i) - \delta_1, t(n,i) + \delta_2]$  におけるイベントの動き  $e_{(n,i)}(t)$  のデザイン．縦軸：運動方向での変位，横軸：時間．

Fig.2 Designed motion pattern around the starting time of an event.

## (2) 反応時刻を精度良く検出するための条件

(2a) 視線計測の誤差と区別できる オブジェクトの動きが小さすぎる場合，それに対応する眼球運動も小さくなり，視線データの計測誤差に埋もれてしまう．そのため，イベントは計測誤差と区別できる程度に大きな動きとする．

(2b) オブジェクト吟味時の眼球運動と区別できる ユーザの視線データには，オブジェクトの動きを追視する眼球運動だけでなく，オブジェクト内容を吟味する際に，自発的な眼球運動が生じる．そこで，イベントの動きの大きさはオブジェクトの面積に対して十分大きいものとする．

(2c) 注視切り替え時の眼球運動と区別できる 自発的な眼球運動は注視するオブジェクトを切り替える際にも生じる．そこで，イベントの運動方向は，複数オブジェクトが配置される方向とできるだけ直交するように定める．

これらの条件を満たすイベントには様々なものが考えられるが，ここではもっとも基本的なイベントのひとつとして，図2に示すような静止状態からの動き出しを用いることにする．つまり，各オブジェクトに組み込むイベントの動きとして以下の  $e_{(n,i)}(t)$  を用いる．

$$e_{(n,i)}(t) = \begin{cases} d_{(n,i)} & (t(n,i) - \delta_1 \leq t < t(n,i)) \\ (t - t(n,i))v_{(n,i)} + d_{(n,i)} & (t(n,i) \leq t \leq t(n,i) + \delta_2) \end{cases} \quad (2)$$

このイベントを組み込まれたオブジェクトは時区間  $[t(n,i) - \delta_1, t(n,i)]$  において座標  $d_{(n,i)}$  に静止しており，時刻  $t(n,i)$  から速度  $v_{(n,i)}$  で動き出す．

まず条件 (1a) より，上記の動きパターンが定義される時間  $\delta_1 + \delta_2$  は，オブジェクトに対する視線停留時間に比べ短くする．この視線停留時間は，用いるメディアの種類や内容の複雑さ，オブジェクトの面積や形状に依存するため，実際に用いるオブジェクトを考慮して定める．また，人間が滑らかな眼球運動で動く対象を追うことのできる最大の速さは  $40^\circ/\text{sec}$  であ

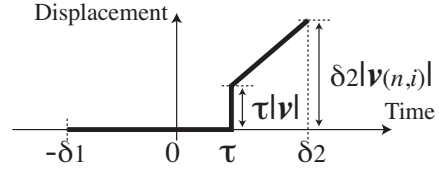


図3 潜時が反映された眼球運動テンプレート．縦軸：運動方向での変位，横軸：時間

Fig.3 Eye movement template for detecting reactions.

り [10]，条件 (1b) より， $|v_{(n,i)}|$  はこれよりも十分小さいとする．

次に，オブジェクトが  $[t(n,i), t(n,i) + \delta_2]$  において動く際の変位量  $\delta_2 |v_{(n,i)}|$  は，条件 (2a)，(2b) より，視線計測の誤差，およびオブジェクト内を吟味する自発的な眼球運動の変位量と比較して十分に大きいものとする．これらは実際に用いる視線計測システムおよびオブジェクトを考慮して定める．さらに条件 (2c) より，オブジェクトの運動方向は水平方向に限定し，複数オブジェクトは垂直方向に並べて配置する．

ここで，本研究では1章で述べたように，各オブジェクトの動きが非常に似通っており，時間変化の相関が大きな場合を想定する．また，イベントの動きはあくまで注視状態の探りを入れるためのプローブであるため，各オブジェクトの動きの顕著性は同程度であることが望ましい．そこで，全てのイベントについて  $v_{(n,i)} = v$  で同一とする．

### 2.3 視線データにおける反応時刻の検出

視線計測システムによって得られたユーザの視線データ（ディスプレイ上における注視座標系列） $X(t) (\in \mathbb{R}^2)$  から，イベントに対する反応時刻  $T$  を検出することを考える．理想的には，ユーザがオブジェクト  $O_n$  を追視する際は，イベントの動き  $e_{(n,i)}(t)$  が視線データ  $X(t)$  にそのまま反映され， $e_{(n,i)}(t)$  と  $X(t)$  とを照合することで  $T$  を検出できる．しかし，実際には眼球運動の生理学的特性が影響し，イベントの動きをそのまま反映した視線パターンとはならない．そこで，特に眼球運動の反応潜時を考慮し，以下の方法で  $T$  を検出する．

$e_{(n,i)}(t)$  を追視した際  $X(t)$  に現れるパターンとして，図3に示すようなテンプレート  $e'(t)$  を定義する． $e'(t)$  は  $t = 0$  をイベント発生時刻として，以下のように表される．

$$e'(t) = \begin{cases} \mathbf{0} & (-\delta_1 \leq t < \tau) \\ tv & (\tau \leq t \leq \delta_2) \end{cases} \quad (3)$$

一般に動くオブジェクトを追視する際，ユーザは潜時

$\tau$ の後、捕捉サッカド運動によってオブジェクトを中心窩に捉え、網膜像のぶれが少なくなるように追視を続ける<sup>[10]</sup>。これを考慮し、 $e'(t)$ では以下のような眼球運動の特性を仮定している。まず、ユーザの視線は時刻  $t = -\delta_1$  から  $t = \tau$  まで停留する。時刻  $\tau$  においてオブジェクトの変位量  $\tau|v|$  だけ捕捉サッカドを行う。サッカドは非常に高速な眼球運動であるため、ここでは非連続的な動きに近似する。その後少なくとも時刻  $\delta_2$  までオブジェクトの動きに追従するものとする。動く対象を滑らかに追視し始める際の反応潜時は  $0.06 \sim 0.15\text{sec}$  とされており<sup>[10]</sup>、潜時が大きい場合にも対応するために、本研究では  $\tau$  を  $0.15\text{sec}$  と設定した。

このテンプレート  $e'(t)$  を用い、 $X(t)$  から  $T$  を検出する。 $X(t)$  には突発性の誤差に加えて、ある一定量のバイアスやスケールの変化も起こりうることを想定し、ここでは  $e'(t)$  を用いて、 $X(t)$  における相関係数に基づくテンプレートマッチングを行う。オブジェクトの動きは水平方向に限定しているため、相関係数は水平方向成分のみを考慮すればよい。そこで、以降  $e'(t)$ 、 $X(t)$  の水平方向成分をそれぞれ  $e'(t)$ 、 $X(t)$  と表すものとし、反応時刻  $T$  を、相関係数  $c(t)$  が時間軸上で極大となり、かつその時の極大値が閾値  $c_{th}$  より大きくなる時刻として検出する。ここで  $c(t)$  は

$$c(t) = \frac{\int_{-\delta_1}^{\delta_2} (e'(t') - \bar{e}') (X(t+t') - \bar{X}(t)) dt'}{\sqrt{\int_{-\delta_1}^{\delta_2} (e'(t') - \bar{e}')^2 dt'} \sqrt{\int_{-\delta_1}^{\delta_2} (X(t+t') - \bar{X}(t))^2 dt'}} \quad (4)$$

であり、 $\bar{e}'$ 、 $\bar{X}(t)$  はそれぞれ区間  $-\delta_1 \leq t' \leq \delta_2$  における  $e'(t')$ 、 $X(t+t')$  の平均値である。閾値  $c_{th}$  は、イベントに無関連の反応を誤検出せず、かつ実際の視線データのばらつきを吸収するために、1 よりやや小さめに設定する必要があり、後述の実験では経験的に  $c_{th} = 0.9$  とした。

2.4 同期構造分析に基づく注視オブジェクト推定  
動的コンテンツには複数のイベントが組み込まれているため、前節の方法により、反応時刻  $T$  は複数検出される。検出された各反応時刻について、どのオブジェクトを注視していたかを、イベント発生時刻  $t_{(n,i)}$  との同期の程度に基づき、以下のように推定する。

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運動における反応時刻は同期すると期待できる。そこで、同期するほど値が小さくなる評価値として単純に以下のものを用いる。

$$V_{(n,i)} = |T - t_{(n,i)}| \quad (5)$$

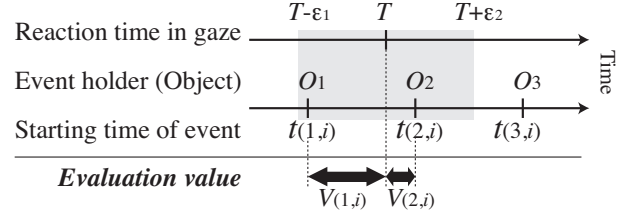


図4 同期構造分析に基づく注視オブジェクト推定。灰色で塗られた時区間  $[T - \epsilon_1, T + \epsilon_2]$  においてイベントを探索する。

Fig. 4 Estimation of an object being focused on. Events are searched in the gray area  $[T - \epsilon_1, T + \epsilon_2]$ .



図5 実験において、大画面情報端末に提示された動的コンテンツを閲覧する様子

Fig. 5 Experimental situation. Subject looks at a dynamic content displayed on a large screen.

次に、検出された時刻  $T$  に対して、 $T$  に発生した反応と関連するイベントが存在する時区間  $[T - \epsilon_1, T + \epsilon_2]$  を設定し、この時区間においてイベントを探索する(図4)。イベントが複数存在する場合、それらのイベントから得られる  $V_{(n,i)}$  を比較する。

$$(k, p) = \arg \min_{n,i} V_{(n,i)} \quad (6)$$

s.t.  $T - \epsilon_1 \leq t_{(n,i)} \leq T + \epsilon_2$

すなわち、 $[T - \epsilon_1, T + \epsilon_2]$  の範囲で最もよく同期したオブジェクト  $O_k$  を推定結果とする。この範囲は、オブジェクトの動きに対する眼球運動の遅延の最大値以上とし、かつ1個のオブジェクトにおけるイベントの発生間隔よりも小さく設定する。イベントが存在しない場合、時刻  $T$  に発生したユーザの反応はオブジェクトのイベントに関連しない自発的な眼球運動であると判定し、注視オブジェクトの推定は行わない。

### 3. 大画面情報端末における注視オブジェクト推定

提案手法を適用・評価する具体的な状況として、大画面情報端末に提示された複数の情報をユーザが閲覧、選択するという状況を設定する(図5)。ユーザは各オブジェクトを詳細に注視するため、オブジェクトの持つイベントが眼球運動に反映することが期待される。



### Gaze Probing: イベント提示に基づく注視オブジェクト推定

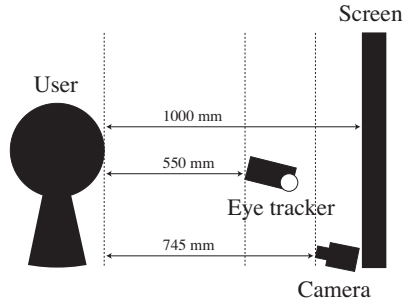


図6 大画面情報端末を用いた提示・計測環境  
Fig. 6 Displaying and sensing system using a large screen.

本章では、このような状況における提案手法の適用方法について述べるとともに、単純な矩形オブジェクトを用いて、注視オブジェクト推定精度の基礎的な評価を行う。

#### 3.1 大画面情報端末を用いた提示・計測環境

図6に本研究で用いる提示・計測環境を示す。情報提示は縦型50インチプラズマディスプレイ<sup>1</sup>を用いて行う。ディスプレイに対してユーザは1m離れた位置に立つ。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運動が発生する。ディスプレイの下にはカメラ<sup>2</sup>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うに配置する。カメラのパン・チルト・ズーム制御は行わないため、ユーザは立ち位置を大きく変化させることはできないが、カメラに両目が映る状態であれば姿勢、顔向きを自由に变化させることができる。なお、今回用いたカメラの感度はそれほど高くないため、虹彩を検出するのに十分な光量を得るために、ディスプレイの両脇にライト<sup>3</sup>を配置する。

視線計測には、我々が従来提案した顔画像認識に基づく手法を用いる。詳細は文献<sup>[11]</sup>に委ねるが、撮影された映像から顔画像および各顔部位の特徴点を追跡し、3次元顔形状モデル当てはめと虹彩検出を行うことで、注視座標系列 $X(t)$ を計測する。得られた $X(t)$ は、あらかじめディスプレイ上のマーカを注視した際の注視座標と対応するマーカ座標の組を学習データとして、水平方向と垂直方向の線型回帰を独立に求めておき、これらを用いて補正するものとする。この計測システムにおける、ディスプレイ上での注視座標 $X(t)$

の計測誤差は水平方向に平均62mm(3.6°)、垂直方向に平均94mm(5.4°)である。

上記の視線計測システムに加え、注視オブジェクトの真値を求めるために、赤外光線を照射して計測を行うアイトラッカー<sup>4</sup>を用いる。この計測誤差は本環境において平均27mm(1.6°)であるが、赤外光源の照射範囲においてのみ視線が計測可能であり、ユーザの顔向きや位置・姿勢に関する許容範囲はカメラのみを用いる視線計測システムに比べて狭い。そのため、精度評価時には、ユーザの顔向き等をアイトラッカーでも計測可能な程度に制限するものとする。

#### 3.2 推定精度の評価方法および比較手法

視線データから検出されたイベントへの反応回数を $r_{\text{detected}}$ としたとき、推定結果と注視オブジェクトの真値を照合して正しく推定されている回数を $r_{\text{success}}$ と表すものとする。このとき推定精度 $R$ は、 $r_{\text{detected}}$ と $r_{\text{success}}$ を用いて以下のように求める。

$$R = r_{\text{success}}/r_{\text{detected}} \quad (7)$$

一方で、イベントの動きが視線データにどの程度反映されているかを検証するために、 $r_{\text{detected}}$ とユーザが実際に注視したイベントの個数 $r_{\text{all}}$ を用いて、以下のように反応の検出率 $R_d$ を定義する。

$$R_d = r_{\text{detected}}/r_{\text{all}} \quad (8)$$

なお本論文において $r_{\text{all}}$ は、アイトラッカーで求めた注視座標の真値からイベントへの反応回数を手作業でカウントすることで求める。

さらに、比較のため以下の手法でも推定を行う。

#### $M_{\text{pos}}$ : 位置に基づく推定手法

各フレームにおいて、注視座標 $X(t)$ と最もディスプレイ上で近い位置にあるオブジェクトを推定結果とする。すなわち、時刻 $t$ の注視オブジェクト $O_k$ を

$$k = \arg \min_n |x_n(t) - X(t)| \quad (9)$$

により推定する。なお、この手法に基づく推定精度 $R$ は視線計測された全フレーム数 $f_{\text{all}}$ と注視オブジェクトが正しく推定されたフレーム数 $f_{\text{success}}$ を用いて、以下のように求める。

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

#### $M_{\text{corr}}$ : 動きの類似性に基づく推定手法

反応時刻 $T$ の近傍時間区間 $[T - \epsilon_1, T + \epsilon_2]$ において、注視座標系列と最も類似性(相関係数)の高い動きを持つオブジェクトを推定結果とする。すなわち、オブ

1: FUJITSU 製 UBWALL . ディスプレイは高さ 1106mm , 幅 622mm , ディスプレイの中心位置は床から高さ 1462mm . コンテンツ提示にはディスプレイ上端から 562mm の領域を利用した .

2: Point Grey Research 社製 Grasshopper (1600×1200 pixel , 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) .

3: 写真電気工業社製 RIFA-F (50×50cm)

4: Tobii X120 Eye Tracker . サンプルレートは 60Hz , 頭部の移動可能な範囲は 40×22×30cm .

ジェクト座標  $x_n(t)$  および注視座標  $X(t)$  の水平方向成分  $x_n(t)$  および  $X(t)$  を用いて、反応時刻  $T$  における注視オブジェクト  $O_k$  を

$$k = \arg \max_n C_n \quad (11)$$

により推定する．ここで、相関係数  $C_n$  は

$$\frac{\int_{T-\epsilon_1}^{T+\epsilon_2} (x_n(t) - \bar{x}_n) (X(t) - \bar{X}) dt}{\sqrt{\int_{T-\epsilon_1}^{T+\epsilon_2} (x_n(t) - \bar{x}_n)^2 dt} \sqrt{\int_{T-\epsilon_1}^{T+\epsilon_2} (X(t) - \bar{X})^2 dt}} \quad (12)$$

であり、 $\bar{x}_n$ 、 $\bar{X}$  はそれぞれ  $[T - \epsilon_1, T + \epsilon_2]$  における  $x_n(t)$ 、 $X(t)$  の平均値である．推定は検出された反応時刻において行うので、推定精度は式 (7) で求められる．なお、この手法は動きの類似性から注視オブジェクトを推定するという点で、従来研究<sup>[9]</sup>と同様の考え方となっている．

M<sub>squ</sub> : 位置および動きの類似性に基づく推定手法

位置および動きの類似性の双方を考慮するために、各フレームの二乗誤差を時間方向に累積し、この累積二乗誤差が最も小さくなるオブジェクトを推定結果とする．すなわち、反応時刻  $T$  における注視オブジェクト  $O_k$  を

$$k = \arg \min_n \sum_{t=T-\epsilon_1}^{T+\epsilon_2} |x_n(t) - X(t)|^2 \quad (13)$$

により推定する．推定は検出された反応時刻において行うので、推定精度は式 (7) で求められる．

**3.3 予備実験: 小矩形オブジェクトでの精度評価**

3.1 節の状況設定において、提案手法の推定精度  $R$  を手法  $M_{pos}$ 、 $M_{corr}$  および  $M_{squ}$  と比較評価した．ただし、2.2 節の条件 (2b) と (2c) で述べたように、実際に用いるコンテンツのレイアウトやオブジェクト内容によっては自発的な眼球運動が生じ、推定精度に大きな影響を与える．そこで、オブジェクト内容の依存性を極力減らした状況での推定精度を評価するために、単純な小矩形図形を用いた．

オブジェクトは灰色で塗りつぶされた 2 個の小矩形図形 (20mm×20mm<sup>5)</sup> を用い、各オブジェクトは時間差を伴いながら 400mm の幅で水平方向に往復運動 (等速運動して両端では一時停止) を繰り返すものとした．本コンテンツでは往復における両端の動き出しをイベントとし、オブジェクト間のイベント発生時刻の時間差は、眼球運動の潜時  $\tau$  と区別できる間隔として  $t_{(2,i)} - t_{(1,i)} = 0.4\text{sec}$ 、同一オブジェクトにおけるイベント発生時刻の時間差  $t_{(n,i+1)} - t_{(n,i)}$  ( $n = 1, 2$ )

表 1 推定精度． $M_{pos}$ :位置に基づく推定手法、 $M_{corr}$ :動きの類似性に基づく推定手法、 $M_{squ}$ :位置および動きの類似性に基づく推定手法、 $M_{prop}$ :提案手法．

Table 1 Accuracies for estimating objects being focused on by the proposed method  $M_{prop}$  and the comparative methods  $M_{pos}$ ,  $M_{corr}$ , and  $M_{squ}$ .

$M_{pos}[\%]$	$M_{corr}[\%]$	$M_{squ}[\%]$	$M_{prop}[\%]$
64.6	74.8	69.1	<b>85.3</b>

は 2sec (したがって往復運動の周期は 4sec) とした．また、このイベント発生間隔をもとに、式 (6) における  $\epsilon_1$  および  $\epsilon_2$  を  $\epsilon_1 = \epsilon_2 = 0.5$  に設定した．動き出しの速さ  $|v|$  は各イベント共通とし、2.2 節の条件 (1b) の議論により、ユーザは十分に追視することができる 466mm/sec (25.0°/sec) とした．2 個のオブジェクトは垂直方向に並べて配置し、垂直方向のオブジェクト間の距離は 150mm に設定した．

被験者数は 6 名 (眼鏡着用なし) とし、各被験者には、ディスプレイに提示されるオブジェクトのいずれかを注視する (ただし注視オブジェクトは途中切り替えてもよい) という 20sec のタスクを 2 回課した．そして、被験者が 2 個のオブジェクトのどちらを注視しているのかを前節の各手法にて推定し、推定精度を求めた．反応時刻検出用のテンプレート  $e'(t)$  では、 $\delta_1$  はオブジェクトの静止時間 1.14sec より小さく、かつ視線が停留していると判断できる時間として 0.6sec に設定した． $\delta_2$  はイベント近傍におけるオブジェクトの変位量  $\delta_2 |v|$  が計測誤差より大きくなるように 0.43sec に設定した (このとき  $\delta_2 |v| = 200\text{mm}$  となる) ．

推定結果

推定精度  $R$  の結果を表 1 に、得られた視線データとイベントに対する反応の一例を図 7 に示す．視線データから検出されたイベントへの反応回数 (式 (7) における  $r_{detected}$ ) は、全被験者のタスク合計時間 240sec の間で 109 回であり、 $R_d = 98.2\%$  であった．

比較手法  $M_{pos}$ 、 $M_{corr}$ 、 $M_{squ}$  において推定精度が低くなってしまうような視線データであっても、イベントの動きは視線データに良く反映されており、提案手法を用いることで高い推定精度が得られた．

考察

本実験で用いた動的コンテンツにおいて、オブジェクト間の距離は視線データ  $X(t)$  の計測精度に対して比較的近いことが、位置に基づく手法  $M_{pos}$  の推定精度低下を招いていると考えられる．また、各オブジェクトの動きの相関が大きいため、手法  $M_{corr}$ 、 $M_{squ}$  では、相関を調べる時区間において眼球運動がどちらのオブジェクトの動きにも類似してしまい、推定精度が低くなっている．提案手法は、このようにオブジェクト

5: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小矩形図形全体を注視することができる．

Gaze Probing: イベント提示に基づく注視オブジェクト推定

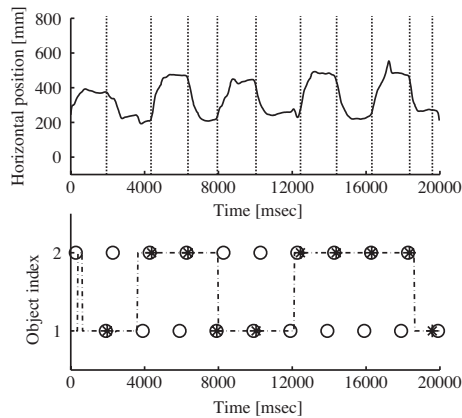


図7 上段: 計測された視線データ(実線), 検出された反応時刻(点線), 下段: オブジェクトに組み込まれたイベント(o), 視線データから検出された反応(\*), 注視オブジェクトの真値(1点鎖線)を示す.

Fig. 7 Above: gaze data (solid line) and reactions (dot line). Below: embedded events (o), reactions (\*) and the verified object (dashed line).

間の距離が近く, かつ類似した動きを持つ場合であっても, 高精度の推定が可能である. これは,  $X(t)$  の絶対位置や, 変位パターンそのものを用いず, 時間的な同期情報のみを考慮するためであると推測される.

4. 写真を用いた動的コンテンツ事例とその評価

動的コンテンツとして写真画像やテキストなど複雑なものが複数提示され, それらをユーザが自由に閲覧する状況では, 単純な矩形を追視する状況に比べて, 閲覧の際にオブジェクト内容の吟味や見比べが多く発生する. このような複雑な眼球運動を伴う場合でも, Gaze Probing が実用に耐えうる推定手法であるかどうかを評価するために, 本章では, オブジェクトとして50文字程度のキャプションの付いた写真画像<sup>6</sup> (150mm×150mm)を用い, 2種類の動的コンテンツを設定して注視オブジェクトの推定精度を評価した. 被験者は3.3節と同一の6名とし, 各被験者には提示されたオブジェクトのうち最も興味のあるものを選択するよう教示を与え, 60secのタスクを1回課した.

4.1 動的コンテンツのデザイン

本実験で用いた2種類の動的コンテンツを図8, 9に示す. いずれも4個のオブジェクトを10mmの間隔を空けながらディスプレイ上に垂直方向に配置した. 3.3節と同様にオブジェクトの動き出しをイベントとし, 自然さを与えるために, 静止時には滑らかに減速させる動きとした. オブジェクト間でのイベント発生時刻の最小時間差  $\min(t_{(m,i)} - t_{(n,i)})$  ( $m \neq n$ ) は0.4sec

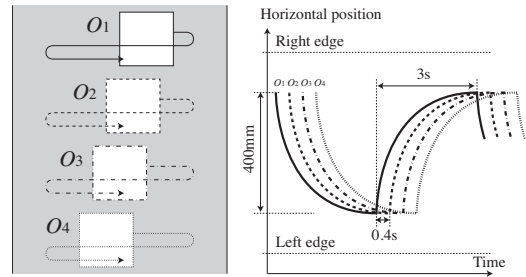


図8 [D-1] 画像の往復運動提示  
Fig. 8 [D-1] Swinging design.

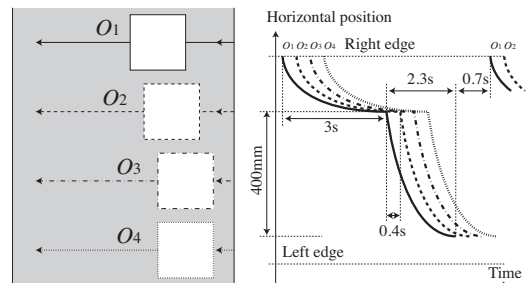


図9 [D-2] 画像のスクロール提示 (各オブジェクトは右端から出現, 左端で消滅, 画像を切り替えて再度右端から出現, を反復)  
Fig. 9 [D-2] Scrolling design.

とした. 式(6)における  $\varepsilon_1$  および  $\varepsilon_2$  は, 3.3節と同様に  $\varepsilon_1 = \varepsilon_2 = 0.5$  と設定した.

[D-1] 画像の往復運動提示 (図8)

3.3節で用いたデザインと同様に各オブジェクトは時間差を伴って400mmの幅で水平方向に往復運動する.  $|v| = 466\text{mm/sec}$  で動き始めたオブジェクトは, ディスプレイ端に近づくとき滑らかに減速し, 一時停止する. その後, 再び466mm/secで反対方向に並進を開始する. 往復における両端の動き出しをイベントとし, 同一オブジェクトにおけるイベント発生時刻の時間差  $t_{(n,i+1)} - t_{(n,i)}$  ( $n = 1, \dots, 4$ ) を3sec (往復運動の周期は6sec) とした.

[D-2] 画像のスクロール提示 (図9)

画像内容が更新されていく状況を想定し, 各オブジェクトは時間差を伴って右から左へスクロールする. ディスプレイ右端から出現したオブジェクトは150mm左へ並進して一時静止した後,  $|v| = 466\text{mm/sec}$  で動き始めてさらに400mm並進する. 左端に近づくとき滑らかに減速し, 左端で短時間静止した後, ディスプレイ上から消滅する. その後, 別の画像が右端から出現し, 同様の動きを繰り返す (各オブジェクトで3枚の画像が切り替わるとし, 合計12枚の画像を用いた). 右端での一時静止後の動き出しをイベントとし, 同一オブジェクトにおけるイベント発生時刻の時間差  $t_{(n,i+1)} - t_{(n,i)}$  ( $n = 1, \dots, 4$ ) を6sec とした.

6: 本実験では市販携帯電話の写真を用いた.

表2 推定精度．[D-1]: 画像の往復運動提示, [D-2]: 画像のスクロール提示． $M_{pos}$ : 位置に基づく推定手法,  $M_{corr}$ : 動きの類似性に基づく推定手法,  $M_{squ}$ : 位置および動きの類似性に基づく推定手法,  $M_{prop}$ : 提案手法．

Table 2 Accuracies for estimating objects being focused on by the proposed method  $M_{prop}$  and the comparative methods  $M_{pos}$ ,  $M_{corr}$ , and  $M_{squ}$ . [D-1]: swinging design, [D-2]: scrolling design.

	$M_{pos}$ [%]	$M_{corr}$ [%]	$M_{squ}$ [%]	$M_{prop}$ [%]
[D-1]	54.5	60.1	63.3	<b>68.4</b>
[D-2]	41.9	45.7	51.4	<b>76.8</b>

4.2 イベントおよびテンプレートのパラメタ設定  
視線データから反応時刻を検出するためのテンプレート  $e'(t)$  は、オブジェクトの動き出し直後が等速直線運動に近似できるとして、3.3 節と同様に式 (3) を用いた。ただし、本実験では被験者がオブジェクト同士の見比べやオブジェクト内容の吟味を行うことで、短時間で細かな眼球運動が発生することが予想される。そこで、2.2 節の条件 (2b) を考慮し、 $\delta_2$  はイベント近傍におけるオブジェクトの変位量  $\delta_2 |v|$  が計測誤差より大きくなり、かつ画像吟味時の細かな眼球運動が反応として検出されないように 0.64sec とした (このとき  $\delta_2 |v| = 300\text{mm}$  となる)。一方、 $\delta_1$  は視線が停留していると十分判定できる時間として、3.3 節と同様に 0.6sec に設定し、オブジェクトの静止時間は  $\delta_1$  より大きくなるように、[D-1] において 1.43sec, [D-2] において 1.63sec とした。

#### 4.3 推定結果と考察

##### 推定結果

推定精度  $R$  を表 2 に、得られた眼球運動とイベントに対する反応の一例を図 10 に示す。視線データから検出されたイベントへの反応回数 (式 (7) における  $r_{\text{detected}}$ ) は全被験者のタスク合計時間 360sec の間でデザイン [D-1] において 98 回 ( $R_d = 90.7\%$ )、[D-2] において 56 回 ( $R_d = 93.3\%$ ) であった。[D-1]、[D-2] のいずれにおいても、イベントの動きは視線データに良く反映されており、提案手法は他の比較手法より高い推定精度が得られた。しかしながら、図 10 の視線データからも分かるように、本実験は複数オブジェクトの見比べやオブジェクト内容の吟味による眼球運動が多く、これら自発的な眼球運動をイベント関連の反応として誤検出することがしばしば確認された。

##### 考察

本実験では被験者がオブジェクト内容の吟味や見比べといった自発的な眼球運動を頻繁に行っており、3.3 節の実験と比較して視線データが複雑になっている。しかし、イベントにおけるオブジェクトの動きを、自

発的な眼球運動とも区別できる程度の大きさに設計しておくことで、座標値を直接比較する手法  $M_{pos}$  や動きの相関を用いる手法  $M_{corr}$ ,  $M_{squ}$  に比べ、高い精度で推定することが可能であることが分かる。

デザイン [D-1] と [D-2] を比較すると、図 10 のように [D-2] は [D-1] に比べてオブジェクト内容を吟味することによる自発的な眼球運動が比較的少ないことが確認でき、推定精度も高くなっている。[D-1] においてオブジェクトが静止する時区間は [D-2] と比較して短時間ではあるが、静止する回数自体は [D-2] より多いため、その結果 [D-1] では詳細な内容吟味を行う眼球運動が全時間区間で偏りなく発生したと考えられる。[D-2] のようにオブジェクトの静止時間を長く設けることで、詳細な内容吟味をその時区間に集中させて、推定精度が向上する可能性があるが、一方で、推定間隔が疎になるというトレードオフがある。

## 5. 討議

### 5.1 視線計測の特性に着目した推定性能の評価

一般に、注視オブジェクト推定では視線計測システムによって得られた注視座標系列  $X(t)$  を何らかの形で評価するため、その精度は視線計測の誤差に影響される。手法  $M_{pos}$ ,  $M_{squ}$  では、式 (9), (13) に示すようにオブジェクト座標  $x_n(t)$  と  $X(t)$  に対して  $|x_n(t) - X(t)|$  という空間的な距離を評価するため、 $X(t)$  においてオブジェクトが並ぶ方向 (実験では垂直方向) にバイアス的に誤差が含まれると、推定精度は低下する。これに対して  $M_{corr}$  は、式 (12) に示すように  $X(t)$  に対して正規化を行うため、バイアス誤差の影響は受けない。しかしながら、相対的な眼球運動パターンが精度よく観測される必要があり、パターンの変位方向 (実験では水平方向) に突発性のランダム誤差が発生し、波形が大きく崩れてしまう場合は、正しく推定を行うことが難しくなる。これらと比較して提案手法は、式 (5), (6) に示すように推定の際に直接視線データを扱わないため、イベントに対する反応を精度良く検出することができれば、バイアス誤差やランダム誤差の影響を受けない推定を行うことが可能である。ただし、本研究では反応検出の際に、 $X(t)$  に対して相関係数 (式 (4)) に基づくテンプレートマッチングを行う。そのため、イベントパターンの変位方向におけるランダム誤差の大きさによっては正しい時刻に反応を検出できなくなる可能性がある。このように、提案手法を含めて本研究で用いた推定手法は、視線計測システムの特性によって性能が変化すると予想される。

そこで、3.3 節の実験に対して、真値として用いた視線データにガウシアンノイズを重畳し、その視線データ  $X'(t)$  によって提案手法の性能を評価すること



### Gaze Probing: イベント提示に基づく注視オブジェクト推定

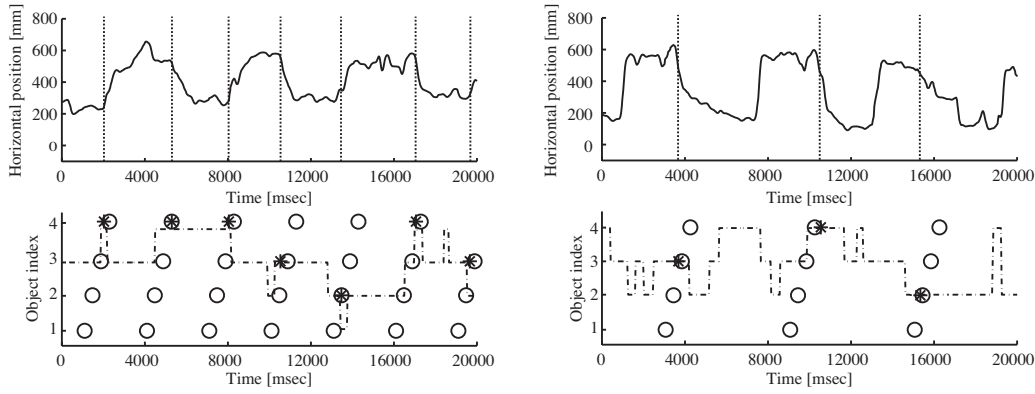


図 10 左: [D-1] 画像の往復運動提示 (図 8 参照), 右: [D-2] 画像のスクロール提示 (図 9 参照). 上段: 計測された視線データ (実線), 検出された反応時刻 (点線), 下段: オブジェクトに組み込まれたイベント (o), 視線データから検出された反応 (\*), 注視オブジェクトの真値 (一点鎖線) を示す.

Fig. 10 Left: [D-1] swinging design, right: [D-2] scrolling design. Above: gaze data (solid line) and reactions (dot line), below: embedded events (o), reactions (\*) and the verified object (dashed line).

を考える. 実験で用いた視線データ  $X(t)$  の誤差における平均 (垂直方向 94mm) および標準偏差 (水平方向 45mm) を参考に,  $X'(t)$  のバイアス誤差を表すパラメータとして誤差の平均  $\mu$  を 70~110mm, ランダム誤差を表すパラメータとして誤差の標準偏差  $\sigma$  を 25~65mm に設定し, 各手法によって推定を行った.

$X'(t)$  に対して各手法で注視オブジェクト推定を行った結果を図 11, 表 3 に示す. 図 11 上段より, 提案手法は  $M_{\text{pos}}$  や  $M_{\text{squ}}$  と異なりバイアス誤差の影響は受けない. また図 11 下段より,  $\sigma$  が  $X(t)$  における誤差の標準偏差と同程度であれば, 提案手法は  $M_{\text{corr}}$  より安定して推定を行うことができるが,  $\sigma$  が大きくなるにしたがって精度は低下する. 一方で表 3 より, 反応の検出率  $R_d$  は  $\sigma$  の変化にともなって低下するが, いずれの推定においても 92~98% であり, 推定精度の低下と比較して変化の幅が小さい. したがって, 2.2 節で述べた条件 (1a), (1b) は達成されているため反応の検出には成功しているが, ランダム誤差の影響によって条件 (2a) が達成できなくなり, その結果正しい時刻での検出に失敗していると考えられる.

このように, 提案手法において推定精度を向上させるうえでは正しい時刻に反応を検出する必要があり, 適切なイベント設計を行うほかに, 反応検出手法を工夫することが重要となる. 2.3 節で述べたように, 本研究ではテンプレートマッチングによって反応時刻の検出を行っており,  $X(t)$  の局所的な情報のみを用いているためランダム誤差の影響を受けやすい. これに対して,  $X(t)$  のより大局的な情報を利用する例として,  $X(t)$  に対して多重解像度解析を行う手法が挙げられる. すなわち  $X(t)$  について, 低解像度における大まかな時間変化に対する反応検出と, 高解像度にお

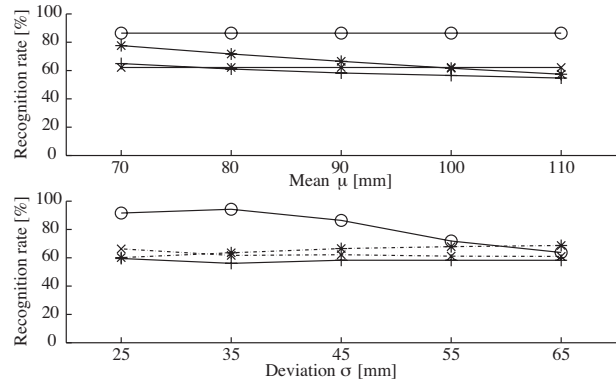


図 11 上段: 誤差平均の変化に対する精度評価の一例 ( $\sigma=45\text{mm}$ ), 下段: 誤差の標準偏差の変化に対する精度評価の一例 ( $\mu=90\text{mm}$ ).  $M_{\text{pos}}(*)$ ,  $M_{\text{corr}}(x)$ ,  $M_{\text{squ}}(+)$ ,  $M_{\text{prop}}(o)$  を示す.

Fig. 11 Above: estimation accuracies at five different  $\mu$  ( $\sigma=90\text{mm}$ ), below: estimation accuracies at five different  $\sigma$  ( $\mu=90\text{mm}$ ).  $M_{\text{pos}}(*)$ ,  $M_{\text{corr}}(x)$ ,  $M_{\text{squ}}(+)$ ,  $M_{\text{prop}}(o)$ .

表 3 誤差の標準偏差の変化に対するイベント検出率 ( $R_d$ ) 評価の一例 ( $\mu=90\text{mm}$ )

Table 3 Detection rate of reactions to events at five different  $\sigma$  ( $\mu=90\text{mm}$ )

$\sigma$ [mm]	25	35	45	55	65
$R_d$ [%]	98.2	94.6	92.3	92.3	93.7

ける細かな時間変化に対する反応検出を併せて行う. これにより, ランダム誤差の影響を抑えつつ, 正しい時刻での反応検出が可能になると考えられる.

### 5.2 間欠的な推定結果の補間法

本手法では, イベントが発生し, かつそれが視線データへの反応として検出された時刻でのみ推定結果

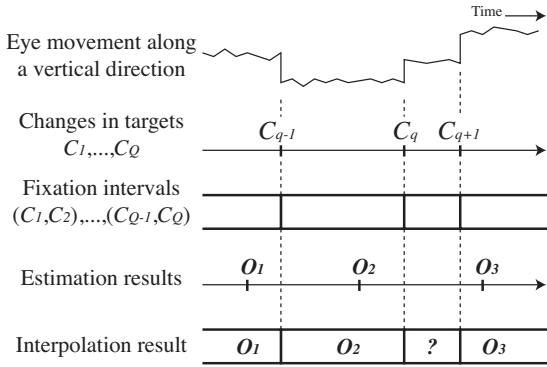


図 12  $X(t)$  における垂直方向の加速度を用いた注視オブジェクト推定の補間  
 Fig.12 Interpolation of estimation results using observed accelerations of eye movement along a vertical direction.

が得られる。提示オブジェクト数が増加する場合でも、各オブジェクトに適切な時間間隔でイベントを組み込むことで、同様に推定を行うことが可能である。一方で本手法は、手法  $M_{pos}$  のように各フレームでの注視オブジェクトを推定する手法に比べると、間欠的な推定となる。目的によっては、より密な推定が必要になり得るため、ここでは推定結果の補間法を検討する。

視線データから注視オブジェクトの切り替え時刻を検出することで、推定結果を伸展することを考える。注視オブジェクトの切り替えの際には、一般にはサッカーのような高速の眼球運動が発生する。2.2 節で述べた条件 (2c) より、この眼球運動は垂直方向の動きであり、イベントへの反応と区別できる。そのため注視オブジェクトの切り替え時刻は、注視座標の絶対位置を用いなくとも、視線計測の垂直方向の加速度などから比較的容易に検出できると考えられる。注視オブジェクトの切り替えが時刻  $\{C_1, \dots, C_Q\}$  に検出されるとすると、反応時刻  $T$  以外の時区間における注視オブジェクトは、限定的ではあるが以下のように補間可能となる(図 12)。まず、 $\{C_1, \dots, C_Q\}$  に対して、互いに隣接する切り替え反応によって定義される時区間  $[C_1, C_2), \dots, [C_{Q-1}, C_Q]$  をそれぞれ 1 個の注視区間として定義する。各注視区間においてユーザはある 1 個のオブジェクトを注視している(あるいは、どれも注視していない)と考えることができる。そこで定義された注視区間に対して、区間内の時刻  $T$  においてイベントへの反応が検出される(すなわち、ある  $[C_{q-1}, C_q)$  に対して  $C_{q-1} \leq T < C_q$  が成立する)場合、その区間全体に推定結果を伸展して適用する。

3.3 節の動的コンテンツについて、 $X(t)$  の垂直方向の加速度から注視オブジェクトの変更を検出した。このとき、補間により推定可能な時区間は全体の 94.7% であり、その時区間におけるフレーム単位での推定精度

を算出すると 84.5% となった。3.3 節の考察で述べたように、本研究で用いている視線計測は垂直方向の計測誤差がオブジェクト間の距離と同程度の大きさであり、注視オブジェクト変更の検出が安定しない。しかし、この実験においてユーザは常にいずれかのオブジェクトを注視しており、かつ注視オブジェクトの頻繁な切り替えがなかったため、推定可能な時区間は十分大きく、その時区間におけるフレーム単位での推定精度も  $M_{pos}$  と比較して高くなった。4. 章の評価実験では、[D-1] について、補間により推定可能な時区間は全体の 74.0%、その時区間における推定精度は 63.4%、[D-2] について、補間により推定可能な時区間は全体の 45.4%、その時区間における推定精度は 61.6% となった。この実験では反応時刻の近傍においても注視オブジェクトの変更が多く発生しており、推定可能な時区間が短くなっている。推定可能な時区間を増加させるとともに推定精度の向上を図るには、ひとつにはイベントの発生間隔をより密にすることが考えられるが、これは次節で述べるように、コンテンツの自然さを考慮して行う必要がある。

### 5.3 動的コンテンツのデザイン的側面

通常の動的コンテンツでは「(A) いかに関心を引くか」に基づいて各オブジェクトの動きがデザインされるのに対し、本手法では「(B) いかに関心の注視状態を探るか」を目的としたデザインになっており、特に本論文で述べたようなデザインでは、いずれのオブジェクトも動くということが不自然さを与える可能性がある。詳細な印象評価は今後の課題であるが、被験者に簡易なアンケート調査を行ったところ、[D-1] の画像の往復運動提示は [D-2] の画像のスクロール提示よりも不自然さやストレスを感じるという結果となった。これは、[D-2] における提示内容を切り替えるためのスクロールと異なり、[D-1] では動くことの必然性を感じないためであるとも考えられ、背景やオブジェクト内容、その動き方に関して総合的にデザインする枠組みを、今後検討する必要がある。

ただし、(A) と (B) の目的は必ずしも背反しない。すなわち本手法では、ユーザの注目を引くためにオブジェクトを動かした際に、実際にそのオブジェクトが注目されたか否かを判定することも可能であり、実際の状況ではこのような応用も有効であろう。

## 6. むすび

本論文では、動的コンテンツにおける各提示オブジェクトの動きと眼球運動とがどの程度同期するかを評価することで、注視オブジェクト推定を行う手法—Gaze Probing—を提案した。視線データによく反映され、かつ反映された時刻を精度よく検出できるように

イベントの動きをデザインし、動的コンテンツに組み込むことにより、カメラベースの視線計測のように、比較的大きな計測誤差が生じる状況においても、高精度の注視オブジェクト推定を行うことが可能であることが確認された。今後は、5章で述べた課題や、注視オブジェクト推定の精度向上に加え、興味推定や情報推薦を適切なタイミングで効果的に行うための動的コンテンツデザインを検討する。

## 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。

## 参考文献

- [1] J. Chen, Y. Tong, W. Gray and Q. Ji: "A robust 3d eye gaze tracking system using noise reduction", Proc. of the symposium on Eye tracking research & applications, pp. 189–196 (2008).
- [2] C. Hennessey, B. Noureddin and P. Lawrence: "A single camera eye-gaze tracking system with free head motion", Proc. of the symposium on Eye tracking research & applications, pp. 87–94 (2006).
- [3] C. H. Morimoto and M. R. M. Mimica: "Eye gaze tracking techniques for interactive applications", Computer Vision and Image Understanding, **98**, 1, pp. 4–24 (2005).
- [4] Z. Zhu and Q. Ji: "Eye gaze tracking under natural head movements", Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, **1**, pp. 918–923 (2005).
- [5] H. Yamazoe, A. Utsumi, T. Yonezawa and S. Abe: "Remote and head-motion-free gaze tracking for real environments with automated head-eye model calibrations", Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshop, **0**, pp. 1–6 (2008).
- [6] T. Ishikawa, S. Baker, I. Matthews and T. Kanade: "Passive driver gaze tracking with active appearance models", Technical report, Robotics Institute (2004).
- [7] J.-G. Wang, E. Sung and R. Venkateswarlu: "Estimating the eye gaze from one eye", Computer Vision and Image Understanding, **98**, 1, pp. 83–103 (2005).
- [8] D. Beymer and M. Flickner: "Eye gaze tracking using an active stereo head", Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, **2**, 2, pp. 451–458 (2003).
- [9] 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).
- [10] 日本視覚学会(編): "視覚情報処理ハンドブック", 朝倉書店 (2000).
- [11] 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).

(2002年1月1日受付)

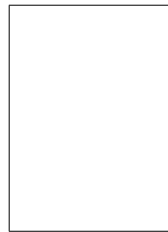
## 著者紹介

米谷 竜



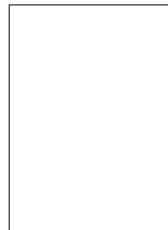
2009年京都大学工学部電気電子工学科卒業。現在、同大学院情報学研究所修士課程在学中。ヒューマンコンピュータインタラクションの研究に興味を持つ。

川嶋 宏彰 (正会員)



2001年京大大学院情報学修士課程了。2007年より同大学院講師。2010年よりジョージア工科大学客員研究員(JSPS海外特別研究員)。博士(情報学)。時系列パターン認識、ハイブリッドシステム、実世界インタラクションの研究に従事。2004年FIT論文賞, 2005年船井ベストペーパー賞, 2007年FITヤングリサーチャー賞。電子情報通信学会, 情報処理学会, IEEE 会員。

平山 高嗣 (正会員)



2000年金沢大学工学部電気情報工学科卒業。2005年大阪大学大学院基礎工学研究科博士課程修了。2005年より京都大学大学院情報学研究所特任助教。博士(工学)。顔画像認識, ヒューマンコンピュータインタラクションの研究に従事。2009年第9回ヒューマンインタフェース学会論文賞受賞。電子情報通信学会, 情報処理学会会員。

松山 隆司 (正会員)



1976年京大大学院修士課程修了。京大助手, 東北大助教授, 岡山大教授を経て1995年より京大大学院電子通信工学専攻教授。現在同大学院情報学研究所知能情報学専攻教授。2002年学術情報メディアセンター長, 京都大学評議員, 2005年情報環境機構長。2008年副理事。工博。画像理解, 分散協調視覚, 3次元ビデオの研究に従事。最近は「人間と共生する情報システム」「エネルギーの情報化」の実現に興味を持っている。1980年情報処理学会創立20周年記念論文賞, 1990年人工知能学会論文賞, 1993年情報処理学会論文賞, 1994年電子情報通信学会論文賞, 1995年第5回国際コンピュータビジョン会議 Marr Prize, 1999年電子情報通信学会論文賞, 2000年画像センシングシンポジウム優秀論文賞。2004年, 2005年FIT優秀論文賞, 2009年ヒューマンインタフェース学会論文賞, 文部科学大臣表彰科学技術賞(研究部門)。国際パターン認識連合, 情報処理学会, 電子情報通信学会フェロー。日本学術会議連携会員。