

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

米谷 竜^{†1} 川嶋 宏 彰^{†1}
平山 高 嗣^{†1} 松山 隆 司^{†1}

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の有用性が示された。

Gazed Object Estimation Using the Timing Structure between Displayed Events and Eye Movements

RYO YONETANI,^{†1} HIROAKI KAWASHIMA,^{†1}
TAKATSUGU HIRAYAMA^{†1} and TAKASHI MATSUYAMA^{†1}

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 movements. 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 temporal relation 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.

1. はじめに

情報提示システムのディスプレイに提示されたコンテンツをユーザが閲覧しているという状況を考える際、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、システム側からの提示情報を動的に生成する場合や、提示された情報を評価する場合に大変有用である。そして、このユーザの興味を推定するためには、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がしばしば重要な役割を果たす。一般に注視対象は、視線情報として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情報推定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²⁾³⁾ や、ビデオカメラのみを用いた手法⁵⁾⁶⁾ 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。このような場合、推定された視線情報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、注視対象推定において注視座標とオブジェクト座標の位置関係を直接用いず、オブジェクトを動かすことで、その動きと眼球動作との時間的な同期構造を獲得し、視線推定の精度の低い環境において高精度の推定を行う手法を提案する。

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4)がある。この研究では、視線情報とし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。この手法における注視対象推定の精度は視線情報の推定精度に依存するが、ユーザの頭部を固定し視線情報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り、そのためこの手法は精度の高い視線情報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。

本研究ではユーザの閲覧対象として、コンテンツを構成するオブジェクトが出現、消滅、移動といった時間に伴った変化（以下、イベントと呼ぶ）をする動的コンテンツを用いる。

^{†1} 京都大学大学院情報学研究所
Graduate School of Informatics, Kyoto University

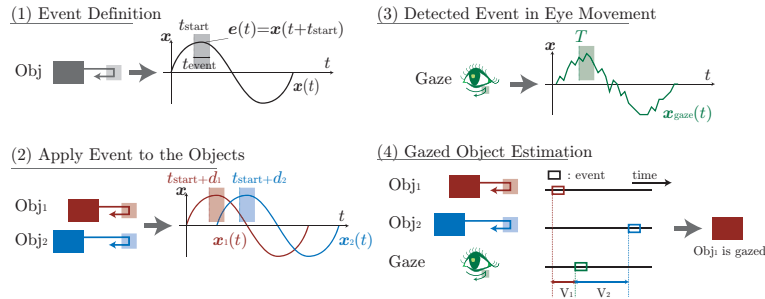


図 1 提案手法の流れ

Fig. 1 An overview of the proposed method

そしてオブジェクトに、計測誤差が大きな視線情報からでも検出可能な眼球動作を発生させるようなイベントを組み込む。そうすることで、ユーザが動的コンテンツを閲覧している状況において、オブジェクトにイベントが発生した時刻と、眼球動作からそのイベントへの反応が検出された時刻との時間差を用いた注視対象の推定を試みる。

また、実用的なインタフェースのデザインを考えると、動的コンテンツにおける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。このために、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、言い換えれば互いに空間的・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

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

とおく。さらにこれらのオブジェクト同士が非常に似通った変化として、小さな時間差で同一の変化をする状況を考える。すなわち、式 (1) はある連続関数 $\mathbf{x}(t)$ を用いて

$$\mathbf{x}_i(t) = \mathbf{x}(t + d_i) + \mathbf{b}_i \quad (\mathbf{x}, \mathbf{b}_i \in \mathbf{R}^2) \quad (2)$$

と表される。 d_i 、 \mathbf{b}_i はそれぞれ $\mathbf{x}_i(t)$ の $\mathbf{x}(t)$ に対する時間差、位置の差を表す。

ここで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情報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などの高加速度の動きを用いる。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として計測できると考えられるためである。ここで、イベントはオブジェクトの動きの一部であり、 $\mathbf{x}(t)$ に対してイベントの軌跡 $\mathbf{e}(t)$ は、

$$\mathbf{e}(t) = \mathbf{x}(t + t_{\text{start}}) \quad (0 \leq t \leq t_{\text{event}}) \quad (3)$$

のようにある時区間に対して定義される。 t_{start} はイベントの発生時刻、 t_{event} はイベント時区間の長さである。一方、式 (2) より $\mathbf{e}(t)$ が各オブジェクトにおいて $[t_{\text{start}} + d_i, t_{\text{start}} + d_i + t_{\text{event}}]$ の時区間に定義される。 $t_{\text{start}} + d_i$ によって定義される各時刻をそれぞれのオブジェクトにおけるイベント発生時刻と呼ぶ。

2.2 眼球動作からのイベントの検出

ユーザの眼球動作として得られるディスプレイ上の注視座標系列を $\mathbf{x}_{\text{gaze}}(t)$ として定義する。ユーザがイベントを追視することで、 $\mathbf{x}_{\text{gaze}}(t)$ にはイベントの軌跡 $\mathbf{e}(t)$ に似た特徴が現れると考えられる。そこで、 $\mathbf{x}_{\text{gaze}}(t)$ から注視対象のイベントに眼球動作が追従しはじめた時刻 T を検出する。以下 T を眼球動作におけるイベント発生時刻と呼ぶ。

2.3 同期構造分析に基づく注視対象推定

各オブジェクトにおけるイベント発生時刻 $t_{\text{start}} + d_i$ と眼球動作におけるイベント発生時刻 T との同期構造を分析することで、複数オブジェクトの中からユーザの注視対象であるオブジェクトを推定する。

まず、各オブジェクトに対して、そのオブジェクトがユーザの注視対象であるか否かを判別するための評価値 V_i を導入する。ユーザがあるオブジェクトの動きを追視していれば、そのオブジェクトにおけるイベント発生時刻と眼球動作におけるイベント発生時刻は同期すると期待できる。したがって、各オブジェクトにおけるイベント発生時刻 $t_{\text{start}} + d_i$ と眼球動作におけるイベント発生時刻 T との時間差を用いて、評価値 V_i を式 (4) と定義する。

$$V_i = T - (t_{\text{start}} + d_i) \quad (4)$$

ここで注視対象の推定時区間を、全オブジェクトのイベントが 1 つずつ含まれる時区

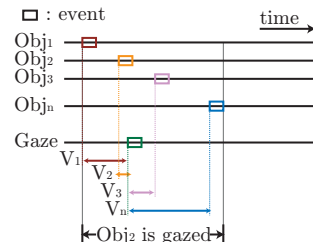


図2 複数オブジェクトに対する同期構造分析
Fig.2 Analysis of the timing structure

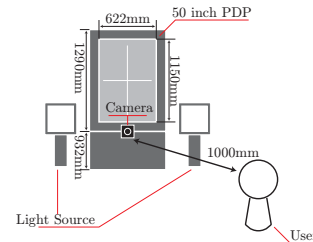


図3 大画面情報端末を用いた環境
Fig.3 Experimental situation

間 $[\min_i(t_{\text{start}} + d_i), \max_i(t_{\text{start}} + d_i + t_{\text{event}})]$ 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イベントに対して得られる V_i を比較することで、注視対象の推定を行う。今、 k 番目のオブジェクトが注視対象であれば、 V_k は他のオブジェクトに比べて小さくなる。すなわち、

$$k = \arg \min_i |V_i| \quad (5)$$

が成立する。眼球動作と各オブジェクトにおけるイベント発生時刻の同期構造を比較することで、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2)。

ただし、一般に滑らかな眼球動作の動特性として、対象の動きに眼球が追従するまでの潜伏時は 60~150ms とされており⁷⁾、 V_i を用いる上ではこの潜伏時や 2.4 節に述べる認知特性の影響を十分考慮する必要がある。なお、各オブジェクトに対して定義されている d_i 、 t_{event} を小さくすることができれば、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。

2.4 同期構造の分析における認知特性の考慮

オブジェクトの動き方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 ユーザが注視オブジェクトの動きを予測できない場合、眼球動作がオブジェクトの動きに対して遅れる。

注視対象の動きに対するユーザの予測 ユーザが注視オブジェクトの動きを過度に予測する場合、眼球動作がオブジェクトの動きに対して先行する。

他のオブジェクトの動きに対する眼球動作のつられ 注視オブジェクトの近傍で他のオブジェクトが動いている場合、他のオブジェクトの動きにつられた眼球動作が発生する。

3. 大画面情報端末における提案手法の適用

本章では、提案手法を適用する具体的な状況設定の1つとして、大画面情報端末を用いた環境について述べる。また、その環境において最適な動的コンテンツのデザインおよび提案手法の適用方法について述べる。

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3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^{*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、提示コンテンツを注視している状況を想定する。この位置から提示コンテンツを閲覧する場合、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^{*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^{*3}を配置する。

3.2 動的コンテンツのデザイン

本研究では、注視対象推定に同期構造を利用することの有用性を検証するため、動的コンテンツを構成するオブジェクトの数を2つとし、そのどちらを注視しているかを推定するという単純化された状況を設定する。これらのオブジェクトを縦に並べて表示し、動きの方向は水平方向のみと限定する。これは、垂直方向の動きより水平方向の動きの方が眼球動作の検出が容易なためである。オブジェクトの動き $x_i(t)$ を決定する上で、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5)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。そこで、本環境ではそれぞれのオブジェクトをディスプレイ中心から上下54.0mmに配置し、106.6mm以上の幅で動かす。

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

*1 富士通製 UBWALL

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

*3 写真電気工業社製 RIFA-F (50×50cm)

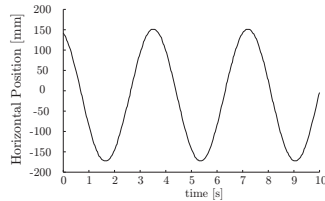


図 4 正弦波
Fig. 4 Sine wave

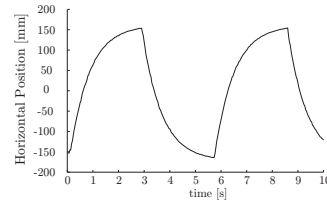


図 5 周期的一次遅れ波
Fig. 5 Periodic first-order lag wave

る速さは $2\sim 40^\circ/\text{s}$ であると言われている⁷⁾。また、オブジェクトが写真画像などである場合、その画像内容を吟味しながら追従できる速さはおよそ $2.8\sim 4.2^\circ/\text{s}$ 程度であると言われている⁸⁾。これらの知見を元にオブジェクトの動く速さを設定する。

オブジェクトの動き方 本研究では、オブジェクトを水平方向に往復させて動きに折り返しを持たせ、折り返しの近傍をイベントとして定義する。加えて、ユーザの認知特性を考慮し、動きを周期的にすることでユーザをその動きに慣れさせて、オブジェクトを追視しやすくすることを試みる。

以上の条件を満たす動きとして、 $x_i(t)$ として以下の2つの動きを採用する。なお、これらの動きはディスプレイの中心を座標原点として記述されるものとする。

正弦波

$$x_i(t) = a \sin\left(\frac{2\pi t}{T} + \theta_i\right) \quad (6)$$

振幅 $a = 162.0\text{mm}$ 、周期 $T = 3.67\text{s}$ とする (図 4 参照)。また、 θ_i はオブジェクト間の位相差として表される時間差である。

周期的一次遅れ波 (矩形波に対する一次遅れ応答)

$$x_i(t) = \begin{cases} K \left(1 - \exp\left(-\frac{t-T\cdot\theta_i/\pi}{\tau}\right)\right) - x_{\text{edge}} & \left(\frac{n}{2}T \leq t \leq \frac{n+1}{2}T\right) \\ K \cdot \exp\left(-\frac{t-T\cdot\theta_i/\pi}{\tau}\right) - x_{\text{edge}} & \left(\frac{n+1}{2}T \leq t \leq \left(\frac{n}{2} + 1\right)T\right) \end{cases} \quad (7)$$

振幅 $K = 308.0\text{mm}$ 、時定数 $\tau = 400\text{ms}$ 、周期 $T = 5.70\text{s}$ 、 $x_{\text{edge}} = 154.0\text{mm}$ とする (図 5 参照)。

これら2つの動きは、時間に伴って速さが増える。正弦波は折り返しから動き出す速さが遅く、その一方で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。また、イベント時区間

の長さ t_{event} を 1.00s とし、これらの2つの動きにおける折り返しに相当する部分、すなわち1周期中の極大点および極小点の前後 0.50s の時区間にイベントを定義する。

3.3 眼球動作の計測

注視座標系列 $x_{\text{gaze}}(t)$ は、Active Appearance Model (AAM)¹⁾ と3次元眼球モデルを用いて検出できる⁵⁾。まず、顔の特徴点をAAMを用いて抽出し、予め作成しておいた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注視座標系列 $x_{\text{gaze}}(t)$ を検出する。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7)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発生時刻の検出

眼球動作におけるイベント発生時刻 T は、注視座標系列 $x_{\text{gaze}}(t)$ とイベントの軌跡 $e(t)$ との相関を用いて検出する。

$$T = \arg \max_t (\text{corr}(x_{\text{gaze}}(t), e(t))) \quad (8)$$

ここで corr は正規化相互相関関数である。

$$\text{corr}(x_{\text{gaze}}(t), e(t)) = \frac{\int (x_{\text{gaze}}(m) - \bar{x}_{\text{gaze}}) (e(t+m) - \bar{e}) dm}{\sqrt{\int (x_{\text{gaze}}(m) - \bar{x}_{\text{gaze}})^2 dm} \sqrt{\int (e(t+m) - \bar{e})^2 dm}} \quad (9)$$

\bar{x}_{gaze} 、 \bar{e} はそれぞれ $x_{\text{gaze}}(t)$ 、 $e(t)$ の推定時区間における平均値である。

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。注視対象は、単一色で塗られた小矩形図形 (4.1節) および写真画像 (4.2節) を用いた。両実験において、コンテンツの背景色は黒とした。

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

本実験では、オブジェクトとして 32.4mm 四方の灰色一色で塗られた小矩形図形を用い

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。動的コンテンツを構成する 2 つのオブジェクトの動きには時間差を持たせ、先に動くオブジェクトを**先行オブジェクト**、後に動くオブジェクトを**後行オブジェクト**と呼ぶ。これら 2 つのオブジェクトに対して提案手法による注視対象推定を行う。ただし、2.4 節で述べたように、注視対象であるオブジェクトのイベント発生時刻と眼球動作におけるイベント発生時刻の間には、被験者の認知特性に起因する時間的ずれが発生する可能性があり、先行オブジェクトと後行オブジェクトで推定精度に違いが現れることが予想される。また文献 5) より、本実験で用いる視線情報の計測精度は、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、実験に用いる動的コンテンツのデザイン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 $\frac{1}{16}\pi, \frac{1}{8}\pi, \frac{1}{4}\pi$

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検出した。なお、オブジェクトを周期的な動きにしたことで、各試行に対してオブジェクトの動きから複数回のイベントを検出できる。本実験では正弦波から 9 個、周期的一次遅れ波から 6 個のイベントを検出し、1 試行に対して推定精度として得られる認識率 R を以下のように求めた。

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

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総数（正弦波は 9 回、周期的一次遅れ波は 6 回）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 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、各フレームにおいて注視座標の水平成分 $x_{\text{gaze}}(t)$ とオブジェクトの座標の水平成分 $x_i(t)$ との差分をそれぞれのオブジェクトに対して求め、差分が小さい方のオブジェクトを注視対象とする。すなわち、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以下のように求まる。

$$k(t) = \arg \min_i |x_i(t) - x_{\text{gaze}}(t)| \quad (11)$$

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。

$$R = f_{\text{success}}/f_{\text{all}} \quad (12)$$

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、全オブジェクトのイベントが含まれる区間 $[t_b, t_e] = [\min_i(t_{\text{start}} + d_i), \max_i(t_{\text{start}} + d_i + t_{\text{event}})]$ 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。すなわち、2.3 節で定義した提案手法の推定区間 $[t_b, t_e]$ において、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。

$$k = \arg \max_i \left(\frac{\int_{t_b}^{t_e} (x_i(m) - \bar{x}_i)(x_{\text{gaze}}(m) - \bar{x}_{\text{gaze}}) dm}{\sqrt{\int_{t_b}^{t_e} (x_i(m) - \bar{x}_i)^2 dm} \sqrt{\int_{t_b}^{t_e} (x_{\text{gaze}}(m) - \bar{x}_{\text{gaze}})^2 dm}} \right) \quad (13)$$

$\bar{x}_i, \bar{x}_{\text{gaze}}$ はそれぞれの $x_i(t), x_{\text{gaze}}(t)$ の推定区間 $[t_b, t_e]$ における平均値である。なお、認識率 R は式 (10) を用いて求められる。

4.1.3 推定結果と考察

推定結果

注視対象推定の結果として、被験者 A, B の推定結果の平均をとったものを表 1 に示す。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。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の方が下側のオブジェクトを注視した場合より認識率が高い。これに対して、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が、全体的に後行オブジェクトを注視した場合の方が先行オブジェクトを注視した場合より認識率が高い。また、注視対象が先行オブジェクトの場合に着目すると、周期的一次遅れ波に対する推定は正弦波に対する推定に比べて認識率が低い。

認知特性の検討

被験者がオブジェクトを注視する際の認知特性を検討するために、眼球動作におけるイベント発生時刻 T と各オブジェクトにおけるイベント発生時刻 $t_{\text{start}} + d_i$ の間の時間差

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.

Table 1 Estimation results. (a) Method based on the position relation between objects and gazed points. (b) Method based on the correlation between the movements of objects and eye movement. (c) Proposed method.

Order	Phase	Sine			Periodic first-order lag		
		(a) [%]	(b) [%]	(c) [%]	(a) [%]	(b) [%]	(c) [%]
Top/Preceding	$\pi/4$	77.9	100.0	97.2	70.4	100.0	100.0
	$\pi/8$	82.8	100.0	77.8	59.0	76.3	6.3
	$\pi/16$	47.6	66.7	19.4	49.5	0.0	0.0
Top/Following	$\pi/4$	79.0	100.0	100.0	57.3	100.0	100.0
	$\pi/8$	85.6	100.0	100.0	46.1	100.0	100.0
	$\pi/16$	56.5	100.0	100.0	43.8	93.8	100.0
Bottom/Preceding	$\pi/4$	16.8	100.0	100.0	27.0	100.0	100.0
	$\pi/8$	33.3	100.0	94.4	34.9	12.5	0.0
	$\pi/16$	48.7	75.0	44.4	42.0	0.0	0.0
Bottom/Following	$\pi/4$	5.7	100.0	100.0	27.7	100.0	100.0
	$\pi/8$	12.8	100.0	100.0	30.1	100.0	100.0
	$\pi/16$	20.8	100.0	100.0	46.8	100.0	100.0

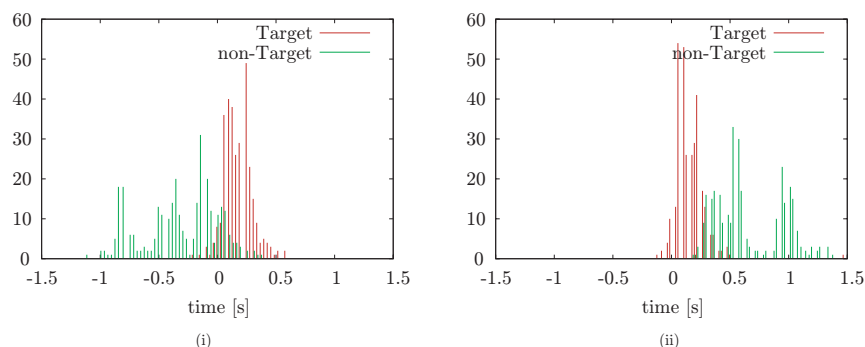


図 6 評価値 V_i のヒストグラム ((i) 先行オブジェクトが注視対象 (ii) 後行オブジェクトが注視対象)
Fig. 6 V_i Histograms: (i) Preceding object is gazed. (ii) Following object is gazed.

$V_i = T - (t_{\text{start}} + d_i)$ を, 注視対象が先行オブジェクトの場合と後行オブジェクトの場合それぞれについて図 6 にヒストグラムで示した. 注視対象 (図中の Target) の場合, V_i は 0.00~0.50s の周辺に集中しているが, 非注視対象 (図中の non-Target) の場合 V_i がばらばらしている. このことから, 注視対象の動きに対して眼球動作は時間差を持って同期してい

ることが確かめられる. 図中 (ii) における注視対象の V_i の最頻値は 0.06s であり, 2.3 節に述べた 60ms~150ms という眼球動作の潜時と合致していることから, 後行オブジェクトの動きに対して眼球動作は滑らかに追従できていることが分かる. その一方で, 図中 (i) における注視対象の V_i の最頻値は 0.25s であり, 被験者が先行オブジェクトの動きを予測できないことで, その動きに対して眼球動作は滑らかに追従ができず, 遅延が発生していたことが分かる. これに関して, 周期的一次遅れ波は動きの停止から動き出しのタイミングが予測しにくく, それに加えて動き出しの初速が速すぎるため, 先行オブジェクトを注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.

図 6 で観測される他の現象として, 図中 (i) では注視対象である先行オブジェクトにおけるイベント開始時刻に対して, 眼球動作におけるイベント開始時刻が先行する状況が見られる. このことから, 注視対象の動きが正弦波であったり遅かったりした場合, 被験者はしばしばその動きを予測していたと考えられる. 図中 (ii) で注視対象である後行オブジェクトに対しても同様の状況が見られる. これは被験者が注視対象の動きを予測していたか, 注視対象でない先行オブジェクトの動きにつられたものと考えられる. これらの状況は全体の 5%程度であり, 推定精度に影響がないと考えてよい.

推定精度向上の試み

眼球動作に前項で述べたような遅延が存在してしまう場合, 眼球動作におけるイベント開始時刻 T を早めるバイアスを加えることによって, 推定精度の向上を図ることができると考えられる. 図 6 における V_i の最頻値の差から, 先行オブジェクトを注視する際に 0.19s 程度の遅延が最も発生していたと推測することができる. そこで, T をその遅延分早めることによって, 先行オブジェクトに対する推定精度が向上すると予想される. ここで, 図 6(ii) において非注視対象の V_i の多くは 0.19s より十分大きく, 後行オブジェクトに対する推定において T を 0.19s 早めても推定精度にさほど影響しないと予想される. 一方, 相関係数に基づく推定手法 (b) においても, 式 (13) における注視座標系列 $x_{\text{gaze}}(t)$ を 0.19s 早めることにより推定精度が向上すると予想される. T にバイアスをかけた場合の推定結果を表 2 に示す. このバイアスによって, 後行オブジェクトを注視した場合の認識率がそれほど下降しないのに対し, 先行オブジェクトを注視した場合の認識率は, バイアスをかけない場合に比べて大きく上昇していることが分かる.

また, 手法 (b) においても提案手法と同程度の推定精度が得られている. これは, 本実験で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実験を行ったため, 相関係数の計算 (式 (13)) で, 比較的長い時区

表 2 眼球動作におけるイベント開始時間 T に対するバイアスによる認識率変化。(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。

Table 2 An improvement of the recognition rate by biasing starting point of the event in eye movement. (b) Method based on the correlation between the movements of objects and eye movement. (c) Proposed method.

Order	Phase	(b) [%]		(c) [%]	
		no bias	0.19s	no bias	0.19s
Preceding	$\pi/4$	100.0	100.0	99.3	100.0
	$\pi/8$	72.2	100.0	44.6	98.6
	$\pi/16$	35.4	97.2	16.0	82.3
	average	69.2	99.1	53.3	93.6
Following	$\pi/4$	100.0	100.0	100.0	100.0
	$\pi/8$	100.0	98.6	100.0	98.6
	$\pi/16$	98.4	63.2	100.0	79.5
	average	99.5	87.3	100.0	92.7
average		84.3	93.2	76.6	93.2

間で積分をとることができたことが精度向上につながったと考えられる。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ることで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検出を、比較的短い時区間で積分計算のみで行うことができる(式(9))。したがって、1つのオブジェクトに複数のイベントを組み込んで推定時区間の数を増やすといった拡張も合わせて行うことで、注視対象の変化に対応できると考えられる。

眼球動作の遅延は動的コンテンツのデザインや提示内容、ユーザの個人差に依存するものであると考えられる。しかし、提示された動的コンテンツに対してオブジェクトの動きと眼球動作から得られる V_i の傾向を調べることによって、本実験で行ったように眼球動作の遅延がある状況においても推定精度の向上を図ることができる。さらには、オブジェクトの動きをユーザにとって予測しやすいようデザインすることで眼球動作の遅延を軽減でき、高精度の注視対象推定が可能であると考えられる。

4.2 写真画像に対する注視対象推定

4.2.1 実験方法

前節で得られた知見を元にオブジェクトを写真画像として評価実験を行った。被験者 A は前節と同様の環境において先行オブジェクトおよび後行オブジェクトをそれぞれ 20 秒間注視し続けた。オブジェクトとして 108.0mm 四方の洋服の写真画像を用いた。このオブジェクトの大きさは被験者の視点から見て 6.16° の大きさを持っている。生理学的にはユーザの

表 3 推定結果。 T は周期, v はオブジェクトの最大の速さ, t_{event} はイベント時区間, (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。

Table 3 Estimation results. T :Period. v :max speed. t_{event} :event segment. (a) Method based on the position relation between objects and gazed points. (b) Method based on the correlation between the movements of objects and eye movement. (c) Proposed method.

T [s]	v [$^\circ$ /s]	t_{event} [s]	(a) [%]	(b) [%]	(c) [%]
3.67	6.72	1.00	55.7	94.4	100.0
4.07	6.06	1.13	57.3	92.9	92.9
4.94	4.99	1.33	59.9	100.0	83.3
6.07	4.07	1.67	67.3	70.0	100.0
7.47	3.31	2.07	63.4	66.7	100.0
Average			48.0	84.8	95.2

中心視野は 2° 程度であり、ユーザの立ち位置からは、視線を動かさなければこの画像の内容を完全に吟味することはできない。オブジェクトの動きは、前節で推定精度の良かった正弦波(式(6))を用いた。振幅 a および周期 T は前節と同様の設定とし、位相 θ_i の差は $\frac{1}{8}\pi$ とした。ここで、3.2 節で述べたように、オブジェクトが写真画像である場合、その画像内容を吟味しながら動きを追視することのできる速さは $2.8\sim 4.2^\circ/\text{s}$ 程度であることを考慮して、オブジェクトの持つ速さを設定する必要がある。注視対象が前節と同様に $T = 3.67\text{s}$ で正弦波の動きをする場合、ディスプレイ中心付近での動きが最も速く $6.72^\circ/\text{s}$ である。この速さでは被験者が画像内容を吟味しながらオブジェクトの動きを追うことが難しいと考えられる。そこで、式(6)における周期 T を $T = 3.67, 4.07, 4.94, 6.07, 7.47\text{s}$ に設定し、どの T の場合もイベントの軌跡が同じ形になるように、イベント時区間 t_{event} を設定した。 $T \geq 6.07\text{s}$ において、オブジェクトの最大の速さ v は $v \leq 4.07^\circ/\text{s}$ と十分遅く、ユーザはその画像内容を吟味しながら動きを追視することができると考えられる。

4.2.2 推定結果と考察

推定結果

表 3 に比較手法 (a), (b), 提案手法の推定結果を示す。また前節の考察から、眼球動作の遅延を考慮した推定精度の向上を図ったが、注視対象に対する V_i の最頻値は 0.1s であったため遅延はほとんど発生しなかったと判断し、 T にはバイアスをかけずに推定を行った。

眼球動作に見られる特徴

本実験では、提案手法は比較手法 (a) や (b) に比べて高精度の推定を行うことができた。これに関して相関係数を用いた手法 (b) は、画像内容の吟味により眼球動作がより複雑にな

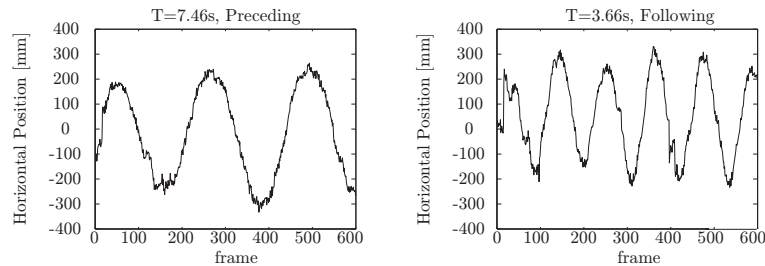


図 7 眼球動作の一例
Fig. 7 Examples of the eye movements

ることで推定が困難になったと考えられる。ここで、図 7 に検出された眼球動作の一例を示す。この図において動きの折り返し周辺は他の箇所と比べて波形が比較的大きく崩れている。このことから被験者は、イベントが定義されている折り返し周辺、すなわち注視対象の動きが十分に遅くなる箇所とその画像内容を吟味しており、ディスプレイ中心付近では主に注視対象への追従のみを行っていたと考えられる。

画像内容を吟味することで眼球動作が複雑化する場合、被験者が画像を吟味する前後、すなわち追従の動きが現れている区間も利用してイベントを検出することで、比較手法 (b) および提案手法の推定精度を向上させることが可能である。このようにイベント時区間を伸長することで、注視対象推定の時間解像度が悪化することが予想されるが、4.1.3 で述べたように、本提案手法ではイベント時区間という短い区間のみでの相関 (積分) 計算で推定が可能であり、その影響は比較的小さいと考えられる。ただし、よりロバストな推定が可能なイベントのデザインや推定手法については今後検討する必要がある。

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およ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。単純な小矩形図形と写真画像とのそれぞれ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という限定的な状況ではあるが、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。

提案手法を既存の情報提示インタフェースに適用する上では、閲覧中に注視対象が頻繁に変化する場合への対応はもちろんのこと、コンテンツのメディア (画像、テキスト、映像)

の特性を十分考慮してイベントをデザインすることが重要である。オブジェクトが画像の場合、イベントとしてオブジェクトの移動を用いて、移動中に速度ベクトルを大きく変化させることで、計測誤差に埋もれない大きな眼球動作を発生させることができる。本研究では並進移動の折り返しをイベントとして用いたが、その他にも静止しているオブジェクトを動かす、動いているオブジェクトを静止させるなどのデザインが考えられる。ただし、ユーザが画像内容を十分に吟味できるように、動きの速さは十分に考慮する必要がある。一方オブジェクトがテキストの場合、その内容を読解しやすいように、イベントとしてオブジェクトの移動を用いるのではなく、行 (もしくは数単語) 単位でテキストを出現させることをイベントとすることが考えられる。またオブジェクトが映像の場合、映像内の動きの切り替わりをイベントとして、眼球動作との同期構造を詳細に分析することが考えられる。これら様々なイベントを用いた場合の有効性評価については今後の課題としたい。

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。

参考文献

- 1) Cootes, T.F. and Taylor, C.J.: Statistical models of appearance for medical image analysis and computer vision, *Proc. SPIE Medical Imaging*, Vol.4322, pp.236-248 (2001).
- 2) Hennessey, C., Noureddin, B. and Lawrence, P.: A single camera eye-gaze tracking system with free head motion, *Advances in eye tracking technology*, pp.87-94 (2006).
- 3) Morimoto, C.H. and Mimica, M. R.M.: Eye gaze tracking techniques for interactive applications, *Computer Vision Image Understanding*, Vol.98, No.1, pp.4-24 (2005).
- 4) 岡本宇正, 藤江真也, 小林哲則: 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).
- 5) 佐竹純二, 小林亮博, 平山高嗣, 川嶋宏彰, 松山隆司: 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, Vol.107, No.491, pp.137-142 (2008).
- 6) 山添大丈, 内海 章, 米澤朋子, 安部伸治: 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).
- 7) 日本視覚学会 (編): 視覚情報処理ハンドブック, 朝倉書店 (2000).
- 8) 菅阪良二, 中溝幸夫, 古賀一男 (編): 眼球運動の実験心理学, 名古屋大学出版会 (1993).