

## 注視オブジェクト推定のための動的コンテンツデザインとその評価

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

### 1 イベント提示に基づく注視オブジェクト推定

ディスプレイに提示されたコンテンツに対する閲覧者の興味・関心を探るうえで、閲覧者が注視しているオブジェクト（画像や文章等）は重要な手がかりとなる。本稿では、オブジェクトに移動・停止といった動きを組み込んだコンテンツ（以下では動的コンテンツと呼ぶ）を積極的に利用することで、注視オブジェクト推定を行う枠組みを検討する。

注視オブジェクトは、視線計測から得られるディスプレイ上の注視座標とオブジェクト座標を直接比較することでしばしば推定される。しかしながら、視線計測の精度と閲覧者側の自由度はトレードオフの関係にあるため、閲覧者が自由な顔向きを取るなど視線計測に大きな誤差が出る状況では、従来の注視オブジェクト推定の精度には限界がある。

そこで本研究では、動的コンテンツを構成する各オブジェクトに特徴的な動き（イベント）を組み込む。閲覧者がイベントを追視することで、視線データにはイベントに対応した眼球運動パターンが反応として現れる。この反応の発生時刻を検出し、どの時刻のイベントに反応したかという時間的情報から注視オブジェクトを推定する。このとき、イベントを適切にデザインすることで、反応は計測誤差の大きな視線データからも検出可能であり、視線計測精度が低い環境においても高精度の推定を行うことができる。

### 2 提示イベントのデザイン

イベントは計測誤差の大きな視線データからでも正しい時刻に反応が検出できるようにデザインする必要があり、以下にイベントが満たすべき要件を示す。

(1) 視線データによく反映されること イベントは閲覧者にとって予測しやすい単純な動きであり、かつ十分に追視できる速度とする。そして、注視オブジェクトを頻繁に変更する状況にも対応するために、イベントが定義される時区間を1個のオブジェクトに対する視線停留時間に比べて短くする。

(2) 反映された箇所を精度良く検出できること イベントの動きの大きさは視線計測の誤差と区別できる程度に大きな動きとする。加えて、オブジェクトの内容吟味による眼球運動とも区別できるように、イベント

はオブジェクトの面積に対して十分に大きな動きとする。さらに、注視オブジェクト変更による眼球運動と区別できるように、イベント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方向の動きとする。

本研究ではこれらの要件を満たす基本的なイベントとして、静止状態からの動き出しを用いる（図1(a)）。イベント発生時刻近傍において、オブジェクトを以下の式に従って動かす。

$$e(t) = \begin{cases} \mathbf{0} & (-\delta_1 \leq t < 0) \\ tv & (0 \leq t \leq \delta_2) \end{cases} \quad (1)$$

ここでは、イベントの発生時刻を時刻0に定義している。このイベントを組み込まれたオブジェクトは時区間 $[-\delta_1, 0]$ において静止しており、時刻0から速度 $v$ で動き出す。まず(1)より、 $|v|$ は閲覧者が十分追視できる程度の大きさとし、イベントが定義される時区間 $\delta_1 + \delta_2$ をオブジェクトに対する視線停留時間に比べて短くする。さらに(2)より、動きの大きさ $\delta_2 |v|$ は計測誤差、オブジェクトの面積と区別できる程度に大きくする。動きの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。

このイベントを、動的コンテンツの構成オブジェクト $\{O_n | n = 1, \dots, N\}$ の動きに以下のように組み込む。まず、 $O_n$ の重心を $x_n \in \mathbb{R}^2$ とし、 $O_n$ における $i$ 番目のイベント発生時刻を $t_{(n,i)}$ とする。このとき、 $t_{(n,i)}$ の近傍 $[t_{(n,i)} - \delta_1, t_{(n,i)} + \delta_2]$ における $O_n$ の動きが

$$x_n(t) = e(t - t_{(n,i)}) + x_n(t_{(n,i)}) \quad (2)$$

によって表されるものとする。さらに、異なるオブジェクト $O_m, O_n (m \neq n)$ についてイベントの発生時刻をずらし、 $|t_{(m,i)} - t_{(n,i)}| > \varepsilon$ とする。イベントの発生間隔 $\varepsilon$ を後述する眼球運動の潜時と区別できる程度に大きくしておくことで、イベントに対する反応時刻さえ検出できれば、どのオブジェクトを注視していたかを推定できる。なお、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。

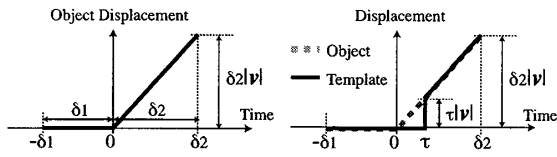
### 3 注視オブジェクト推定手法

まずはじめに、2でデザインしたイベントの動き $e(t)$ を組み込んだ動的コンテンツを閲覧した際、ディスプレイ上における閲覧者の注視座標系列が、視線計測システムによって視線データ $X(t)$ として得られたとする。このとき、 $e(t)$ を追視した際 $X(t)$ に現れる眼球運動パターンのテンプレート $e'(t)$ を定義することで、イ

Dynamic Contents Design for the Estimation of the Gazed Object and Its Evaluation

†Ryo YONETANI †Hiroaki KAWASHIMA †Takatsugu HIRAYAMA †Takashi MATSUYAMA

†Graduate School of Informatics, Kyoto University



(a) イベント発生時刻近傍でのオブジェクトの動き (b) 反応検出のための眼球運動テンプレート

図 1: イベントデザインと眼球運動テンプレート

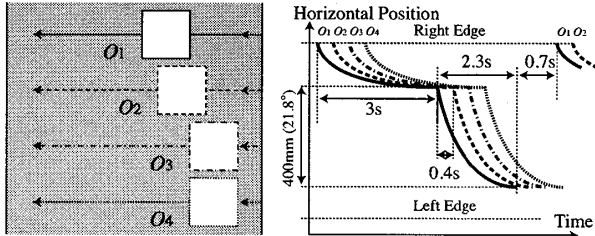


図 2: オブジェクトが並進移動する動的コンテンツ

イベントに対する反応時刻を検出する。一般に動くオブジェクトの追視をする際、閲覧者は潜時  $\tau$  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 [1]。これを考慮して、以下の式に示す眼球運動パターン  $e'(t)$  を用いて、 $X(t)$  において相関係数に基づくテンプレートマッチングを行い、反応時刻  $T$  を得る (図 1(b))。

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

次に、反応時刻  $T$  と各イベントの発生時刻  $t_{(n,i)}$  との同期構造を分析し、反応に関連するイベントを決定することで注視オブジェクトを推定する。まず、時刻  $T$  に検出された反応の近傍においてイベントを探索する。そして、イベントが複数存在する場合には、それぞれのイベント発生時刻  $t_{(n,i)}$  と時刻  $T$  の時間差を比較することで、時刻  $T$  における注視オブジェクトの推定を行う。たとえば、オブジェクト  $O_k$  の  $p$  番目のイベントを注視して時刻  $T$  に反応が発生した場合、以下の式が成立することが期待できる。

$$(k, p) = \arg \min_{n,i} |T - t_{(n,i)}| \quad (4)$$

## 4 評価実験

### 4.1 評価実験の概要

提示内容を順々に更新して閲覧者に商品紹介をするような動的コンテンツをデザインし、提案手法により注視オブジェクト推定を行った。

ディスプレイ (高さ 1106mm, 幅 622mm) に対してオブジェクトには 50 文字程度のキャプションの付いた市販携帯電話の写真 (150mm×150mm) を用いた。6 名の被験者は、提示された 4 種類の商品の中で最も購入したい 1 個を選択した。図 2 のように、オブジェクトはディスプレイ右端から出現した後に短時間静止し、左方向へ並進移動する。ディスプレイ左端に近づくと

滑らかに減速し、左端で短時間静止した後に消滅する。その後同一商品の別の写真が右端から出現し、同様の動きを繰り返す (各オブジェクトで 3 枚の写真が切り替わるとし、合計 12 枚の写真を用いた)。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は 6s, オブジェクト間のイベント発生時刻の時間差は 0.4s とした。視線計測は [2] に基づいており、単眼カメラの映像から抽出した被験者の顔特徴点に対し、あらかじめステレオ計測で構築しておいた 3 次元顔形状モデルを当てはめることで、3 次元空間上での眼球中心と虹彩中心を推定し、それらの延長線からディスプレイ上の注視座標系列  $X(t)$  を求めた。被験者はディスプレイから 1m 離れた位置に立ち、映像に両目が映る状態であれば自由に姿勢、顔向きを変えることができ、 $X(t)$  の計測誤差は水平方向に平均 3.6°, 垂直方向に平均 5.4° であった。

### 4.2 実験結果と考察

注視オブジェクトの真値を、赤外光線を照射して視線計測を行うイトラッカー (60Hz) を用いて求め、推定精度を算出した。注視座標とオブジェクト座標の位置関係に基づく比較手法の推定精度 (推定成功フレーム数 / 全フレーム数) は 41.9%, 提案手法の推定精度 (推定成功回数 / 反応検出数) は 76.8% であった。今回設定したオブジェクト間の距離は、視線データ  $X(t)$  の計測誤差より小さく、比較手法では推定がしばしば困難となるが、一方で提案手法は  $X(t)$  の絶対位置を考慮しないため、計測誤差の影響を受けない。さらに、オブジェクトの内容吟味や見比べによる眼球運動と反応が区別できるようにイベントをデザインしており、提案手法によって高精度で推定を行うことができる。提案手法を用いることで、オブジェクト数が増加しても同様に推定は可能であるが、一方で推定結果は間欠的なものであり、反応時刻  $T$  以外の区間における注視オブジェクトは得られた推定結果を補間して求める必要がある。これに関して、時刻  $T$  の近傍で注視オブジェクトの変更が起こる時点までは、推定結果を伸展できる。

なお、本研究で用いた動的コンテンツは注視オブジェクト推定を目的としたデザインであり、オブジェクトが動くということが閲覧者に対して不自然さを与える可能性がある。これに関して、詳細な印象評価を今後の課題とする。

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] 日本視覚学会 (編): “視覚情報処理ハンドブック”, 朝倉書店 (2000).
- [2] 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).