

映像の顕著性変動と視線運動の 時空間相関分析に基づいた集中状態推定

米 谷 竜^{†1} 川 嶋 宏 彰^{†1}
平 山 高 嗣^{†1,*1} 松 山 隆 司^{†1}

本稿は、米谷 竜, 川嶋 宏彰, 平山 高嗣, 松山 隆司: "映像の顕著性変動と視線運動の時空間相関分析に基づいた集中状態推定", 情報処理学会研究会資料, CVIM178-16, Sep. 2011, <http://ci.nii.ac.jp/naid/110008601902> の著者バージョンです. 図 2 を含め本研究で利用している映像はパナソニック株式会社の協力のもと提供されたものです.

Mental Focus Estimation Using the Spatio-temporal Correlation between Visual Saliency and Eye Movements

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

This article is the authors' version of Yonetani, Kawashima, Hirayama and Matsuyama: "Mental Focus Estimation Using the Spatio-temporal Correlation between Visual Saliency and Eye Movements", CVIM178-16, Sep. 2011, <http://ci.nii.ac.jp/naid/110008601902>. Videos used in this research including Figure 2 were provided by courtesy of Panasonic Corporation.

^{†1} 京都大学

Kyoto University

^{*1} 現在, 名古屋大学

Presently with Nagoya University

映像の顕著性変動と視線運動の 時空間相関分析に基づいた集中状態推定

米谷 竜^{†1} 川嶋 宏彰^{†1}
平山 高嗣^{†1,*1} 松山 隆司^{†1}

視線計測技術の一応用として、実環境計測された視線運動からの心的状態推定がある。本研究では人間の映像視聴状況を取りあげ、視聴映像に対する集中状態の推定手法を提案する。提案手法では、映像中に存在する顕著（視線が向きやすい）領域のダイナミクスを顕著性変動として抽出し、いくつかの変動パターンに分類する。そして、顕著性変動パターンとの相関性に基づいて視線運動を分類し、それぞれ集中状態を示す特徴の抽出・学習を行うことで、集中状態の推定を実現する。

Mental Focus Estimation Using the Spatio-temporal Correlation between Visual Saliency and Eye Movements

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

The spatio-temporal correlation analysis between visual saliency and eye movements is presented for estimation of mental focus toward videos. We extract spatio-temporal dynamics patterns of saliency areas from the videos, which we refer to as saliency-dynamics patterns, and evaluate eye movements based on their correlation with the saliency-dynamics patterns in view. Experimental results using TV commercials demonstrate the effectiveness of the proposed method for the mental-focus estimation.

^{†1} 京都大学

Kyoto University

^{*1} 現在、名古屋大学

Presently with Nagoya University

1. はじめに

「目は心の鏡」と言われるように、人間の目はしばしばその時の心的状態を映す。我々人間は、相手の目の動き（視線運動）によって、その相手が何を見ているかだけでなく、どういう状態にあるか（たとえば集中しているか）を知ることができる。このような視線運動と心的状態の関係は、古くより認知心理学、神経生理学分野などで研究されてきた¹⁾²⁾。しかしながら、そこで得られた知見は基本的に人工的環境におけるものであり、実世界における人間の振る舞いに関しては未だ明らかでないことが多い。一方、センシング技術の発達により、実環境における非拘束・非接触での視線計測が実現されている³⁾。我々はこのような技術を応用し、「目は心の鏡」のシステム化—視線運動の実環境計測に基づく心的状態推定の手法確立に取り組んでいる。

視線運動解析における困難さとして、人間がある心的状態にあるときにその視線運動が必ずしも同じ振る舞いとならないことが挙げられる。これは、人間の情報処理が制御処理（controlled processing）と自動処理（automatic processing）に分類される⁴⁾ ことから分かるように、視線運動には人間の心的状態に基づく運動（内因性視線運動; endogenous eye movements）だけでなく、外環境における視覚刺激に誘起される運動（外因性視線運動; exogenous eye movements）が存在し、観測される視線運動には両者が混在するためである。したがって心的状態を推定するためには、視線運動のみならず、視線運動が持つダイナミクス（視線ダイナミクス）とその視界に映る対象物が持つダイナミクス（対象ダイナミクス）間の関係性を考慮する必要がある。たとえばインタラクティブシステム構築の研究においては、提示コンテンツからユーザの視対象を検出し、視対象への注視時間⁵⁾ や複数対象の注視順序⁶⁾、対象の動的な切り替わりに対する反応潜時⁷⁾ などの特徴として抽出し、ユーザの心的状態推定を行っている。また運転環境を扱う研究では、視線運動と車外環境の相関性がしばしば解析対象となる⁸⁾⁹⁾。具体的には、optical flow や障害物、標識、歩行者認識技術を用いて車外環境から顕著性の高い物体を検出し、視線方向と物体位置との相関から運転手の注意状態を推定している。

以上の関連研究は、基本的に「人間が何を、どのように見ているか」を扱っており、視対象に向けられた人間の心的状態を、その視線運動に基づいて解析する研究である。言い換えれば、視線ダイナミクスと対象ダイナミクスの関係性を「視線を基準として」捉えた研究であると言える。一方で、一般に人間の視界には複数物体が存在し、それぞれが人間の視対象候補として何らかの視覚的顕著性を持つため、ある視対象に視線が向けられるという現象

は、視対象だけでなくその周囲の物体が持つ特性にも影響される。これは、心的状態を扱うにあたって「人間の視界中に何が存在し、どのように見られているか」という「対象を基準として」捉えた関係性も同様に重要であることを示唆している。

本研究では、このような対象を基準とした対象ダイナミクス—視線ダイナミクスの関係性に着目し、対象ダイナミクスに基づいた視線ダイナミクスの解析、心的状態の推定を行う。特に本稿では、ユーザがTV コマーシャルといった一般の映像を視聴する状況を設定し、対象ダイナミクスとして映像が持つダイナミクスを扱うとともに、視聴中の映像に対するユーザの集中状態を推定する。

2. 映像の顕著性変動と視線運動の時空間相関分析に基づいた集中状態推定

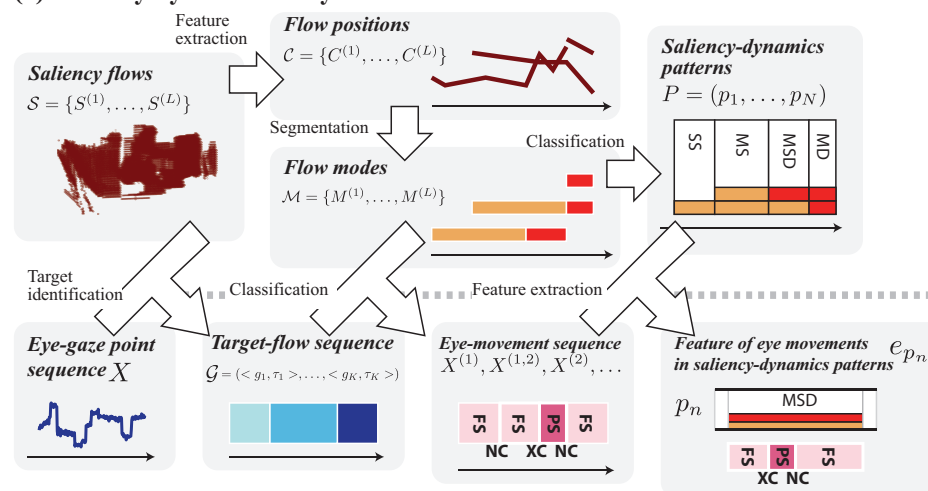
2.1 映像・視線ダイナミクスと集中状態

ディスプレイにTV コマーシャルといった一般の映像が提示され、ユーザがそれを視聴している状況を想定する。また、ディスプレイ下には視線計測装置が設置され、非拘束・非接触の状態ユーザの視線運動（ディスプレイ上の注視点系列）を精度良く計測できるものとする。このとき、ユーザが視聴映像に対してどれだけ注意を向けているかという注意の程度を集中の高さとして定義し、集中の高さは数段階に量子化できるものと仮定する。集中状態推定とは、映像および観測される視線運動に基づいて、その映像に対するユーザの集中の高さを推定する（たとえば、高集中・低集中を判別する）問題である。

1章で述べた通り、観測される視線運動には心的状態に基づく内因性運動と外環境の視覚刺激に誘起される外因性運動が混在する。このことに起因する視線運動の不確実性は、以下に示す2種類の問題として捉えることができる。

- (A) 同じ集中状態であっても、映像ダイナミクスによって視線ダイナミクスは異なる。
 - (B) 同じ映像ダイナミクスであっても、集中状態によって視線ダイナミクスは異なる。
- (A)を解決するにあたっては「人間の視界中に何が存在するか」に着目し、映像中において視線が向けられうる対象群、およびそれらのダイナミクスを抽出・解析する必要がある。1章で挙げた関連研究において、このような対象群はあらかじめ手動で、あるいは物体認識技術を用いてヒューリスティックに検出される。しかしながら一般の映像を扱う本研究において、上述の手法によって映像中から対象をすべて検出することは困難である。これに対して、画像の各画素が持つ視覚的注意の向きやすさ（顕著度）を計算するモデルとして、Saliency map¹⁰⁾が知られている。本研究ではこのSaliency mapを用いて、映像中に出現する対象に関する特定のヒューリスティクスを用いることなく、対象領域として視線の向きやすい顕

(a) Saliency dynamics analysis



(b) Eye movement analysis

図1 提案手法の概要

著領域を抽出する。そしてこれらの顕著領域の動きや画面中に存在する領域数の変化といったダイナミクス（顕著性変動）をいくつかのパターンに分類することを試みる。

一方(B)の問題により、このように映像から確定的に得られた顕著性変動パターンに対して、視線運動は一意に定まらない。しかしながらこの視線運動の違いこそが、集中状態を示す指標となると考えられる。そこで本研究では「これら顕著性変動パターンがどのように見られているか」に着目し、顕著性変動パターンとの相関性に基づいて発生しうる視線運動をいくつかのタイプに分類、そして視線運動のタイプごとに異なる基準で集中状態を示す特徴の抽出・統計的学習を行う。これにより、新たに観測された映像と視線運動の組に対して、ある時区間において顕著性変動パターンを同定することで、そのパターンに基づいて適切な基準で視線運動から集中状態を推定することが可能となる。

2.2 提案手法における集中状態推定の概要

図1に提案手法の概要を示す。図1(a)のように、提案手法ではまず Saliency flow と呼ばれる顕著領域の時空間ボリューム $S = \{S^{(1)}, \dots, S^{(L)}\}$ を抽出する。顕著性変動の解析では、これら Saliency flow が織りなす動きのパターンに着目する。具体的にはまず、 S の重心系列 $C = \{C^{(1)}, \dots, C^{(L)}\}$ を用い、Saliency flow における動きの状態（モード）を解

析することで、 $\mathcal{S}$ をモード系列 $\mathcal{M} = \{M^{(1)}, \dots, M^{(L)}\}$ に分節化する。そして $\mathcal{M}$ をもとに、ある時区間に存在する Saliency flow の数およびそれらのモードをいくつかのパターンに分類し、顕著性変動をパターン系列 $P = (p_1, \dots, p_N)$ として表現する (3 章)。

つづく視線運動解析 (図 1(b)) では、まず注視点系列 $\mathbf{X}$ を $\mathcal{S}$ と照合することにより、注視されている Saliency flow (以下では単に注視フローと呼ぶ) の ID と注視時区間からなる系列 $\mathcal{G} = (<g_1, \tau_1>, \dots, <g_K, \tau_K>)$ を得る。そして、各時区間 τ_k における $\mathbf{X}$ の部分系列を注視フローのモードとの相関に基づいていくつかの視線運動のタイプに分類し、視線運動系列 $X^{(1)}, X^{(1,2)}, X^{(2)}, \dots$ を獲得する。集中状態推定では、この視線運動タイプごとに異なる基準で特徴抽出を行う。ここで顕著性変動パターン $p_n \in P$ に着目すると、含まれる Saliency flow の数やそれらのモードに基づいて、それぞれに対してどのタイプの視線運動が観測されうるかを予測できる。提案手法ではこれをもとに、変動パターンそれぞれについて、観測されうる視線運動の特徴を組み合わせた特徴セットを集中状態の識別モデルとして学習する。そして、新たに観測された映像と視線運動の組に対して、映像から同定される顕著性変動パターンによって識別モデルを切り替えながら、集中状態を推定する (4 章)。

3. 顕著性変動の解析

3.1 Saliency flow の抽出

映像には、人間の視線を引きつける顕著領域が存在する。本研究では Saliency map¹⁰⁾ と呼ばれる計算モデルを用いて、映像中からこの顕著領域を抽出する。Saliency map は視覚的注意の計算モデルであり、入力映像から明度、色差、エッジ方向といった基礎的画像特徴量のコントラストを複数スケールで抽出し、単一の顕著度画像 (Saliency map) に統合する。顕著領域の抽出にあたってはまず、入力画像 i_t より Saliency map $\hat{i}_t$ を計算し、各ピクセル $\mathbf{c} \in \mathbb{N}^2$ について顕著度 $\hat{i}_t(\mathbf{c})$ を得る^{*1}。さらに、各 $\hat{i}_t$ に対して顕著度 π_s での閾値処理を行い、閾値以上の顕著度を持つピクセルを 8 近傍でラベリングすることによって、顕著領域集合 $S_t = \{s_{(t,1)}, \dots, s_{(t,N_t)}\}$ を抽出する (図 2)。

2 章で述べたように、本研究では顕著領域の時空間ボリュームを Saliency flow として抽出する。ここでは Saliency flow を、時空間的に連続した複数顕著領域が単連結した (領域の分岐や併合を含まない) 時空間ボリュームとして定義する。映像にはしばしば複数の Saliency flow が含まれることを考慮し、ここでは各フローに対して ID を与えるものとする。

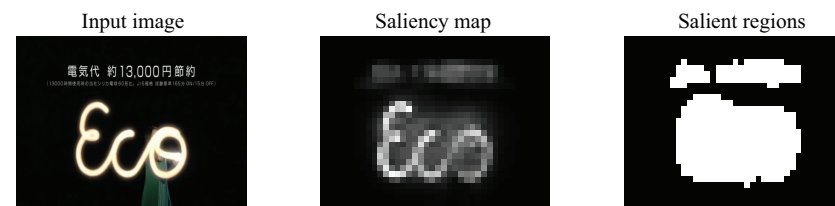


図 2 顕著領域の抽出。領域抽出にあたっての閾値は $\pi_s = 0.2$ とした。

る。具体的にはまず、各顕著領域 $s_{(t,n)}$ に対して、以下の手続きにより ID $l_{(t,n)}$ を与える。

(1) 1 時刻前の顕著領域集合 S_{t-1} のうち、 $s_{(t,n)}$ と同一領域を共有する部分集合 $\hat{S}_{t-1}$ を考える。 $\hat{S}_{t-1}$ の要素数に基づき、以下の (1a)~(1c) のうちいずれかを行う。

(1a)要素数が 0 の場合、 $s_{(t,n)}$ は Saliency flow の始点であり、 $l_{(t,n)}$ には新たな ID 番号を割り当てる。

(1b)要素数が 1 の場合、 $s_{(t,n)}$ は $s_{(t-1,m)} \in \hat{S}_{t-1}$ と同一の Saliency flow を構成する領域であり、 $l_{(t-1,m)}$ を継承する。

(1c)要素数が 2 以上の場合、 $s_{(t,n)}$ は複数フローが併合する領域である。ここでは $\hat{S}_{t-1}$ のうち重心が $s_{(t,n)}$ の重心に最も近い領域 $s_{(t-1,\tilde{m})}$ の ID $l_{(t-1,\tilde{m})}$ を継承する。

(2) 与えられた $l_{(t,n)}$ について、 S_t の他領域に与えられた ID との重複を調べる。もし $l_{(t,n)}$ と同一の ID を持つ領域が既に存在する場合、 $s_{(t,n)}$ は時刻 t において分岐を発生するフローの一端であるため、ここでは $l_{(t,n)}$ に対して新しい ID 番号を再度割り当てる。

このようにして、全ての $s_{(t,n)}$ に対して ID $l_{(t,n)} \in \{1, \dots, L\}$ (L は最大 ID 番号) が与えられる。ID l を持つ Saliency flow $S^{(l)}$ は、同じく ID l を持つ顕著領域の集合 $S^{(l)} \triangleq \{s_{(t,n)} \mid l_{(t,n)} = l\}$ によって構成される。また以下では、 $S^{(l)}$ が定義される時区間を $[b_l, e_l]$ と表すものとする。

3.2 顕著性変動のモデル化と解析

Saliency flow 集合 $\mathcal{S} = \{S^{(1)}, \dots, S^{(L)}\}$ が含む顕著性変動をモデル化する。 $\mathcal{S}$ の各要素は、その重心位置や形状、顕著度の時間変化パターンを内包する。その中でも特に領域の動き (ここでは重心位置の変化) は、顕著度計算のための特徴としてもしばしば着目される重要な特徴である¹²⁾。したがって Saliency flow の動きは自身の誘目度合いに影響し、またそれゆえそのフローを注視する視線ダイナミクスにも影響を与えることが期待される。

そこでここでは、前節で得られた Saliency flow から動き特徴を抽出し、動きの状態 (以

*1 Saliency map の計算には、Saliency Toolbox¹¹⁾ を用いた。

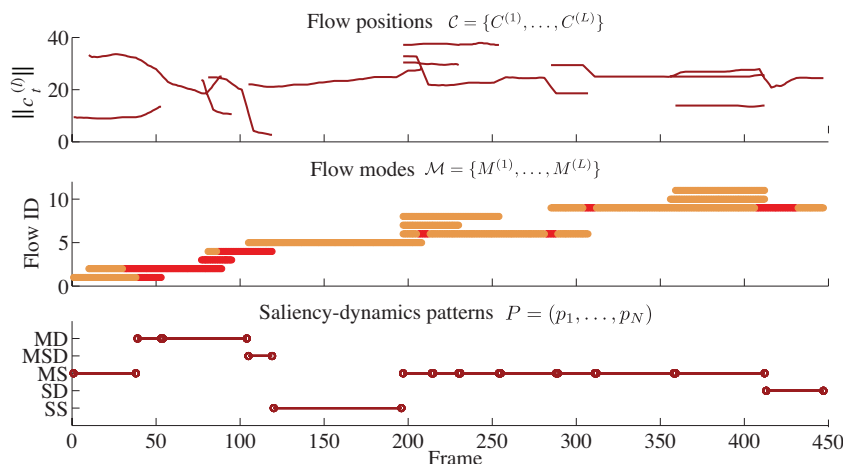


図3 顕著性変動の解析。上段: 各 Saliency flow の重心位置の変化 (縦軸は画像左上を原点にとった重心位置のノルム)、中段: 各 Saliency flow のモード (赤: 動的状態, 橙: 静的状態)、下段: 顕著性変動パターン (SS: Single Static, SD: Single Dynamic, MS: Multi Static, MSD: Multi Static/Dynamic, MD: Multi Dynamic)。各パラメータは $\pi_c = 2^\circ/\text{s}$, $\omega_d = 2^\circ$, $w_s = 0.1 \text{ sec}$, $\omega_p = 0.5 \text{ sec}$ に設定した。

下ではモードと呼ぶ) の系列によって $S^{(l)}$ のダイナミクスを表現するとともに, S が含む顕著性変動を, ある時区間において共起するモードの組み合わせによって表現する (図3)。なお以下では, $S^{(l)}$ から得られる重心系列を $C^{(l)}$ と表す (図3 上段)。

Saliency flow の分節化

本研究ではモードとして, 単純に m_d : Saliency flow が動きを持つ状態, m_s : Saliency flow が静止する状態の2種類を考える。ここではモード $m \in \{m_d, m_s\}$ が定義されたある時区間 τ を $\langle m, \tau \rangle$ と表現するものとし, Saliency flow $S^{(l)}$ の重心系列 $C^{(l)}$ をモード系列 $M^{(l)} = (\langle m_1^{(l)}, \tau_1^{(l)} \rangle, \dots, \langle m_{N_l}^{(l)}, \tau_{N_l}^{(l)} \rangle)$ ($\sum_n \tau_n^{(l)} = e_l - b_l + 1$) に分節化する (図3 中段)。まず, $C^{(l)}$ の速度 (時間方向の一階微分値) ノルムに対して閾値 π_c を与え, 閾値以上の区間に m_d , 閾値未満の区間に m_s を割り当てることで, 大まかな分節化位置を決定する。次に m_s が与えられた時区間 $\tau_n^{(l)}$ のうち, 区間長が閾値 w_s より小さなものを隣接する m_d 区間と併合する。また, m_d が与えられた時区間 $\tau_n^{(l)}$ のうち, $\tau_n^{(l)}$ 中の動きの振幅が一定値 w_d より小さなものを隣接する m_s 区間と併合する。そして, 併合後のモード—時区間系列を, 改めて $M^{(l)} = (\langle m_1^{(l)}, \tau_1^{(l)} \rangle, \dots, \langle m_{N_l}^{(l)}, \tau_{N_l}^{(l)} \rangle)$ として定義する。

顕著性変動のパターン分類

顕著性変動は, Saliency flow $S = \{S^{(1)}, \dots, S^{(L)}\}$ が織りなす動きのパターンである。いま $S^{(l)}$ の動き情報として, モード系列 $M^{(l)}$ が得られている。ここでは $M = \{M^{(1)}, \dots, M^{(L)}\}$ において同時区間に共起するモードの組み合わせに着目し, 顕著性変動をいくつかのパターンに分類する。特に本研究では, 簡単化された以下の5パターンを想定する。

Single Static (SS) 単一の静的状態フローが存在する。

Single Dynamics (SD) 単一の動的状態フローが存在する。

Multi Static (MS) 複数の静的状態フローが存在する。

Multi Static/Dynamics (MSD) 静的・動的状態の混在した複数のフローが存在する。

Multi Dynamic (MD) 複数の動的状態フローが存在する。

これらのパターンを用いて, M をパターン系列として表現することを考える。まず, あるフレームに存在する Saliency flow の数およびそれらのモードに基づいて, M を分節化する。そして, 分節化された各時区間について, その区間長が閾値 w_p より小さなものを隣接区間に併合する。このようにして, 顕著性変動パターンからなる系列 $P = (p_1, \dots, p_N)$ ($p_n \in \{\text{SS, SD, MS, MSD, MD}\}$) が得られる (図3 下段)。

4. 視線運動の解析と集中状態の推定

4.1 顕著性変動パターンと視線運動

本章では, 前章で得られた顕著性変動パターンおよび個々の Saliency flow のダイナミクスに基づいて, 視線運動を分類, そして集中状態の推定を行う。Saliency flow に向けられた視線運動は, そのフローが持つダイナミクスによっていくつかに分類される。たとえば, 動いている対象を吟味する場合には滑動性眼球運動 (smooth eye pursuit) が発生する一方で, 静止している対象を注視する場合には固視運動 (fixation) が発生する。また1章で述べたように, 視線運動には内因性運動と外因性運動の2種類が存在し, 特に注視フローの切り替えの際に発生する衝動性眼球運動 (saccade) は, 視聴者自ら注視フローを切り替える内因性運動と, フローの出現や静的状態から動的状態へのモード遷移 (これらをまとめて, 以下ではイベントと呼ぶ) に誘起される外因性運動に分類できる。

そこで本研究では, 映像視聴中の視線運動として **静止対象の吟味 (Fixation scan; FS)**, **動的对象の吟味 (Pursuit scan; PS)**, **意図的对象切り替え (Endogenous target change; NC)**, **外因的对象切り替え (Exogenous target change; XC)** を想定する。以下ではまず, 注視フローが持つモードやイベントとの相関性に基づき, 視線データ (ディスプレイ上の注

表 1 顕著性変動パターンと観測される視線運動タイプ (1,2 列目) および特徴セット (3 列目) .

Saliency-dynamics pattern p	Observable eye movements	feature set e_p
Single Static (SS)	FS	$\{e_{fs1}, e_{fs2}\}$
Single Dynamic (SD)	PS	$\{e_{ps1}, e_{ps2}\}$
Multi Static (MS)	FS, NC, XC	$\{e_{fs1}, e_{fs2}, e_{nc}, e_{xc}\}$
Multi Static/Dynamic (MSD)	FS, PS, NC, XC	$\{e_{fs1}, e_{fs2}, e_{ps1}, e_{ps2}, e_{nc}, e_{xc}\}$
Multi Dynamic (MD)	PS, NC, XC	$\{e_{ps1}, e_{ps2}, e_{nc}, e_{xc}\}$

視点系列) に上述の視線運動タイプを同定するとともに, 集中状態を示す特徴を抽出する.

一方でこれらの視線運動特徴の振る舞いは, 注視フローが存在する顕著性変動パターンによっても異なることが予想される. 顕著性を扱ううえで, 対象の動きはその対象の顕著度を強める特徴としてしばしば用いられる¹²⁾. したがって, 静的状態のフローと動的状態のフローが混在する顕著性変動パターン (Multi Static/Dynamic) では, 動的フローへの視線運動が優位となり, 静的フローへの視線運動が抑制される可能性がある. また注視フローの切り替え運動は, 複数フローが存在する顕著性変動パターン (Multi Static, Multi Static/Dynamic, Multi Dynamic) においてのみ発生する.

そこでここでは各顕著性変動パターンに対して, 表 1 のように発生しうる視線運動タイプを整理し, そこから得られる視線運動特徴を組み合わせた特徴セットを構築する. たとえば Multi Static に対する視線運動としては FS, NC, XC が考えられるため, FS, NC, XC から得られる特徴を組み合わせたものを特徴セットとする. このような特徴セットを, 各顕著性変動パターンについて集中状態の識別モデルとして学習しておく. これにより, いったん顕著性変動パターンがある時区間で同定されることで, そのパターンに応じて適切な識別モデルを適用して集中状態を推定することが可能になる.

4.2 視線運動タイプの同定

視線運動タイプの同定にあたって, はじめに視線データを注視フローのモードに基づいて分節化する. 時刻 t におけるディスプレイ上の注視点を $x_t \in \mathbb{N}^2$ と表す. いま, 視線計測装置によって, ディスプレイ上の注視点系列 $\mathbf{X} = (x_1, x_2, \dots)$ が視線データとして得られているものとする. まず, $\mathbf{X}$ 上に存在する Saliency flow の ID を調べることで, $\mathbf{X}$ より注視フロー系列 $\mathcal{G} = \langle g_1, \tau_1 \rangle, \dots, \langle g_K, \tau_K \rangle, g_k \in \{1, \dots, L, \zeta, \eta\}$ を得る. ただし, 注視フロー ID が g_k である時区間 τ_k を $\langle g_k, \tau_k \rangle$ と表し, ζ, η はそれぞれ, 注視点上に Saliency flow が存在しない状態, 瞬きなどにより注視点が観測できない状態を指す. 次に,

時区間 $\langle g_k, \tau_k \rangle$ における注視フロー $g_k (\notin \zeta, \eta)$ のモードに基づき, これをモード系列 $\langle m_{k1}, \tau'_{k1} \rangle, \dots, \langle m_{kM}, \tau'_{kM} \rangle$ へとさらに分節化する. 最後に, 各要素にインデックスを与えなおすことで, 全体のモード系列 $\mathcal{G}' = \langle m_1, \tau'_1 \rangle, \dots, \langle m_{K'}, \tau'_{K'} \rangle$ を得る. なおここで, 時区間 τ'_k の始点, 終点を $[b_k, e_k]$ と表すものとする.

つづいて, 注視点系列 $\mathbf{X}$ のうち, $[b_k, e_k]$ に定義される部分系列を $X^{(k)} \triangleq (x_t \mid t \in [b_k, e_k])$, $[e_{k-1}, b_k]$ に定義される部分系列を $X^{(k-1,k)} \triangleq (x_t \mid t \in [e_{k-1}, b_k])$ と表すものとし, これらの分節化された部分系列に対して各視線運動タイプを同定する. まず FS, PS について, これらは定常的な視線運動であり, $X^{(k)}$ に対してそのいずれかが同定される. 具体的には時区間 $[b_k, e_k]$ におけるモード m_k に基づき, $m_k = m_s$ のときには FS, $m_k = m_d$ のとき PS が $X^{(k)}$ に対して同定される. 一方で NC, XC は過渡的な視線運動であり, $X^{(k-1,k)}$ に対してそのいずれかが同定される. 前節で述べたように, 両視線運動はそれぞれ内因性, 外因性の運動であるため, 注視フローにおけるイベント (フローの出現や静的状態から動的状態へのモード遷移) との同期の程度によって両者を判別する. 具体的には, $X^{(k-1,k)}$ 中において新たなフローに注視が移動した時刻 b_k と, その注視フローにおいてイベントが発生した時刻のうち b_k から最も近いもの $T (\leq b_k)$ を比較する. ここでは, 外因性 saccade の潜時 (latency) ϵ に着目し (文献 13) より ϵ は 0.2sec 程度), $b_k - T < \epsilon$ のとき $X^{(k-1,k)}$ には XC を, それ以外の場合 NC を同定する.

以上のようにして, $\mathbf{X}$ より視線運動タイプの同定された注視点系列 $(X^{(1)}, X^{(1,2)}, X^{(2)}, \dots, X^{(K-1,K)}, X^{(K)})$, $X^{(k)} \in \{\text{FS, PS}\}, X^{(k-1,k)} \in \{\text{NC, XC}\}$ が得られる (図 4) .

4.3 視線運動からの特徴抽出と集中状態推定

集中状態が視線ダイナミクスに影響を与える可能性を考慮し, 以下の特徴抽出を考える.

静止対象の吟味 (FS) 静止対象に対する吟味は, その対象がある程度の大きさを持つ場合, ある点の注視 (fixation) と注視点移動 (saccade) より構成される. そこでまず, FS の同定された注視点系列 $X^{(k)} \triangleq (x_t \mid t \in [b_k, e_k])$ について, その速度成分 $\dot{X}^{(k)}$ に対して値 π_v による閾値処理を行うことで, saccade を含む (閾値以上の速度での運動を含む) 時区間集合 $\{O_1^{(k)}, \dots, O_{N_v}^{(k)}\}$ を得る. これを用い, saccade の移動長平均およびその頻度を特徴 e_{fs1}, e_{fs2} として以下のように抽出する.

$$e_{fs1} = \frac{1}{N_v} \sum_{v=1}^{N_v} \frac{1}{\sqrt{a_v}} \left(\max_{(i,j) \in O_v^{(k)}} (||x_i - x_j||) \right), \quad e_{fs2} = \frac{N_v}{e_k - b_k + 1} \quad (1)$$

ただし saccade の移動長は注視フロー領域の面積に依存すると考えられるため, e_{fs1} はその注視フロー面積 a_v を用いて正規化されるものとする. また π_v については fixation

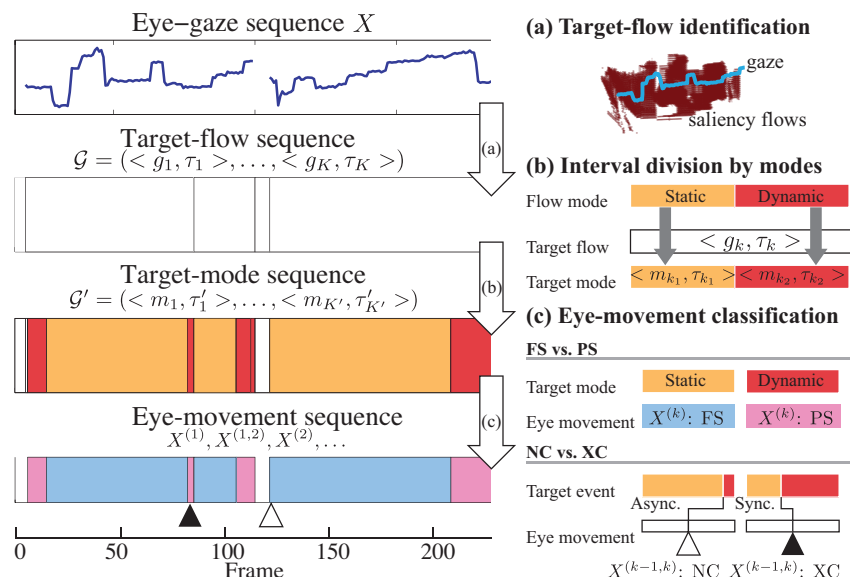


図4 視線運動タイプの同定。(a) 注視対象の同定, (b) モードに基づく分節化, (c) 視線運動タイプの同定。

を誤って検出しないよう、5章では実験的に $8^\circ/\text{sec}$ に設定した。

動的对象の吟味 (PS) 動的对象の吟味は、対象の追跡 (smooth eye pursuit) および注視点移動 (saccade) より構成される。以下では $X^{(k)}$ の速度成分 (時間方向への1階微分) を $\dot{X}^{(k)} \triangleq (\dot{x}_t \mid t \in [b_k, e_k])$ と表す。まず pursuit を表す特徴として、 $\dot{x}_t$ における対象運動方向成分 $\dot{x}_t \cos \theta_t$ (時刻 t における対象の運動速度を $\dot{c}_t$ としたとき、 $\cos \theta_t = \frac{\dot{x}_t \cdot \dot{c}_t}{\|\dot{x}_t\| \|\dot{c}_t\|}$) を抽出し、それと $\dot{c}_t$ との比を時間平均して e_{ps1} とする。

$$e_{\text{ps1}} = \frac{1}{e_k - b_k} \sum_{t \in [b_k, e_k]} \frac{\|\dot{x}_t\| \cos \theta_t}{\|\dot{c}_t\|} \quad (2)$$

そして saccade を表す特徴として、 $\dot{x}_t$ のうち対象運動方向との直交成分を抽出し、時間平均したものを e_{ps2} とする。

$$e_{\text{ps2}} = \frac{1}{e_k - b_k} \sum_{t \in [b_k, e_k]} \|\dot{x}_t\| \sin \theta_t \quad (3)$$

意図的对象切り替え (NC) NC は saccade のみにより構成されているため、FS と同様

にその発生頻度に着目する。この発生頻度は画面中のフロー数に依存すると考えられる

ため、ここでは顕著性変動パターン中の NC 発生頻度を、そのパターンを構成するフロー数により正規化したものを特徴 e_{nc} とする。すなわち、 F 個のフローが存在するパターン (区間長 L) 中で K 回の NC が観測されたとき、 $e_{\text{nc}} = K/(L \cdot F)$ とする。

外因的对象切り替え (XC) 前節で述べたように、XC と NC は直近のイベントとの時間差によって特徴づけられる。そこでここでは、顕著性変動パターン中において K 回の XC が観測される場合、それらの平均潜時 (前節における $b_k - T$ の平均値 $\sum_{k=1}^K (b_k - T)/K$) を特徴 e_{xc} として定める。

以上の特徴を、顕著性変動パターン p に対して観測されうる視線運動に基づいて組み合わせることで、特徴セット e_p を得る。たとえば p が Multi Static の場合、FS, NC, XC より抽出される特徴を組み合わせることで、 $e_p = \{e_{\text{fs1}}, e_{\text{fs2}}, e_{\text{nc}}, e_{\text{xc}}\}$ となる (その他の特徴セットについては表1参照)。なおここで、FS, PS は顕著性変動パターンが定義される区間中に複数存在する可能性がある。このとき e_p の当該要素には、その区間における特徴量の平均値を与えるものとする。

集中状態の推定にあたって、複数の集中状態が $R = \{R_1, \dots, R_N\}$ と表されるものとする。このとき、各顕著性変動パターン p に対して識別モデル $P(e_p \mid R)$ を学習しておく。そして、同一のパターンによって定義された区間において新たに獲得された特徴セット e_p^* に対して、以下を解くことでその区間における集中状態 $\hat{R}$ を推定する。

$$\hat{R} = \arg \max_R P(R \mid e_p = e_p^*) \propto \arg \max_R P(e_p = e_p^* \mid R) P(R) \quad (4)$$

ただしここでは、簡単のため $P(R)$ には無情報事前分布として定数を与えるものとする。

5. 実験

5.1 実験の概要

提案手法の性能を、評価実験を通して確認する。映像コンテンツには12種類のTVコマーシャルを用いた。TVコマーシャルは本来視聴者の注意を引き付けるデザインが製作者によってなされており、映像中より明快な顕著領域が容易に抽出できることが期待される。

本評価実験では比較的簡単化された問題として、高集中と低集中の2状態の判別を試みる。10名の被験者 (いずれも20~30代の男女) は、以下の2条件で映像視聴を行った。

条件1 被験者は映像を視聴し、その後映像に関する簡単なアンケートに回答する。

条件2 被験者は映像を視聴しつつ、別のタスク (ここでは暗算) を遂行する。

これらの2条件はそれぞれ映像視聴に対する高集中、低集中状態に対応する。実験は、映像セットA (12種類の映像コンテンツのうち6種) —条件1, 映像セットB (残り6種) —

条件 2, 映像セット B—条件 1, 映像セット A—条件 2 の順に行った。なお, 映像セットの組み合わせは, 被験者によって変えている。被験者はディスプレイ*1前に着席し (ディスプレイと被験者間の距離はおよそ 1m), 映像を視聴した。ディスプレイ下には視線計測装置*2を設置し, 被験者の視線を計測した。その際被験者には, 条件に関わらず可能な限りディスプレイ方向に視線を向けるよう教示した。

5.2 集中状態の推定

獲得されたデータ (10 名 $\times$ 12 映像 $\times$ 2 条件 = 240 試行, 各集中状態について 120 試行) に対して, ここでは Leave one out 法を用いて推定精度を評価する。図 5~9 に, 獲得データより構成される相対頻度分布を示す (図ではデータに対して線形判別分析を適用し, 判別関数によって次元に投影したものを示している)。評価にあたってはデータより 1 試行を取り除き, 残りのデータを用いて各顕著性変動パターンに対する視線特徴セットの条件付確率密度分布を学習する。この際, 訓練データにより構成される相対頻度分布に対して加算スムージングを行うことによる補間を行う。そして, 取り除いたデータをテストデータとして用い, 集中状態を正しく判別できるかを評価する。これをテストデータを取り替えて全データについて行い, 各顕著性変動パターンに対する平均推定精度を算出した。

表 2 に推定結果を示す。本研究ではベースライン手法として, 以下の 2 種類を用いた。

Duration 顕著性変動パターンが定義される時区間において, 顕著領域を注視する時間の割合を特徴として用いる。

PERCLOS 顕著性変動パターンが定義される時区間において, 眼が閉じている時間の割合を特徴として用いる*3。裏を返せば, 顕著領域および非顕著領域を注視する時間の割合を扱っている。

表 2 から分かるように, 提案手法を用いることで, いずれの顕著性変動パターンにおいてもベースライン手法と比較して推定精度が改善されている。このことより, 提案手法において顕著性変動パターンごとに視線解析手法を切り替えることの有用性が確認できる。また提案手法では, FS や PS といった定常的視線運動の評価にあたって内部の saccade やフロー運動速度との同期の程度といったダイナミクスに着目していることから, ベースライン手法で扱われている「何を見ているか」という情報だけでなく, 「どのように見ているか」という情報が重要であることが示唆される。

*1 MITSUBISHI Diamondcrysta RDT262WH, 25.5 inch, W550 mm/H344 mm.

*2 Tobii X60 Eye Tracker. 頭部の移動可能な範囲は 400 $\times$ 220 $\times$ 300mm.

*3 なおここでは, 視線計測装置によって視線が獲得できない区間を, 眼が閉じている区間として扱った。

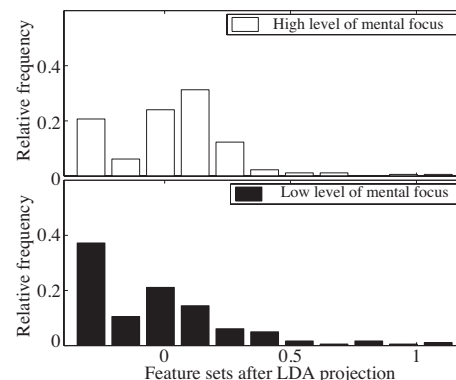


図 5 獲得された相対頻度分布 (SS)。

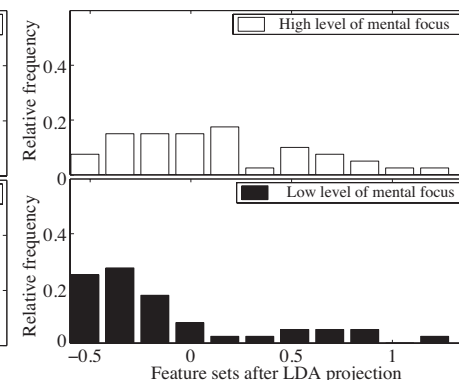


図 6 獲得された相対頻度分布 (SD)。

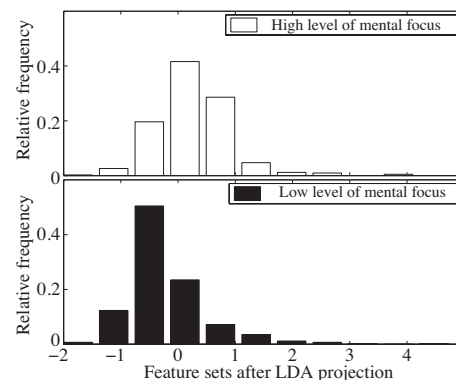


図 7 獲得された相対頻度分布 (MS)。

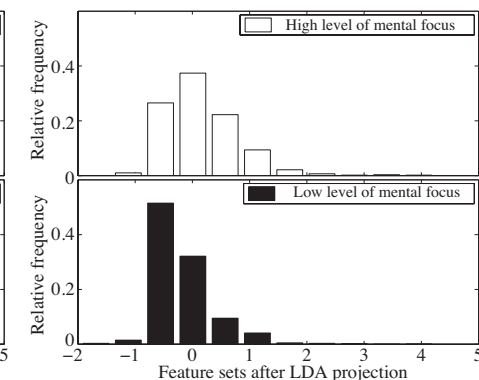


図 8 獲得された相対頻度分布 (MSD)。

提案手法では, 顕著領域に対して向けられた視線運動についてのみ分類, 特徴抽出を行っている。その一方で, 顕著領域への注視時間のみを特徴として扱う Duration より顕著領域および非顕著領域への注視時間を扱う PERCLOS の方が高精度であることから, 非顕著領域に対する視線運動も集中状態を推定するうえで重要であることが示唆される。本研究では顕著, 非顕著領域の判別にあたって固定値による単純な閾値処理を行っており, 実際に視線が向きやすい箇所が非顕著領域に判別されている可能性がある。今後は顕著度のヒストグラムを考慮し適応的に閾値処理を行う, 顔領域などヒューリスティックに視線が向きやすい傾

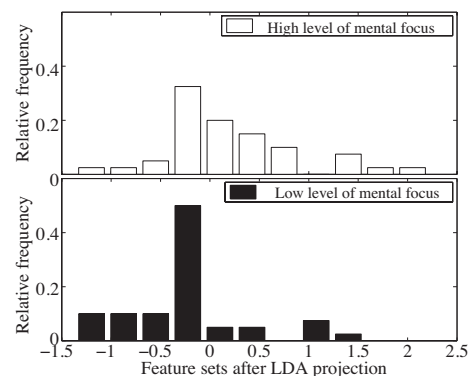


図9 獲得された相対頻度分布 (MD)。

表2 高集中・低集中状態の推定結果。SS, SD, MS, MSD, MD は変動パターンに対する推定精度, Average はその平均を示す。Duration および PERCLOS はベースライン手法の推定精度を示す。

Duration	PERCLOS	
53.8 %	65.0%	
SS	SD	MS
66.9 %	75.0 %	78.8 %
MSD	MD	Average
81.5 %	76.3 %	78.2 %

域の抽出を組み込む¹⁴⁾ など, より詳細な顕著領域抽出を検討する。

6. 結 論

本研究では, 視線運動の実環境計測に基づく人間の心的状態推定の一つとして, 映像視聴時における視聴者の集中状態推定手法を提案した。提案手法では, 映像が持つ顕著性を基準とした顕著性変動—視線ダイナミクスの相互関係性に着目し, 顕著性変動のパターンに基づいて視線運動の分類, 集中状態の判別を行っており, TV コマーシャルを用いた評価実験を経て, その有用性を確認した。しかし一方で, 平坦な自然風景といった顕著性が表れにくい映像や雑踏を映すような顕著性が非常に多く現れる映像などは, 特定の領域に視線が集まりにくくなることが予想される。このような状況では顕著性を基準とするのではなく, 視線運動を基準とし, 「人間が視線を向けている対象がどのようなダイナミクスを持っているか」を詳細に評価する必要があると考えられる。今後は5章で述べた課題とともに, 映像ダイナミクスと視線ダイナミクス間の関係性のより詳細な解析に取り組む。

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) Just, M. and Carpenter, P.: Eye fixations and cognitive processes., *Cognitive Psychology*, Vol.480, pp.441–480 (1976).

2) Emery, N.: The eyes have it: the neuroethology, function and evolution of social gaze, *Neuroscience & Biobehavioral Reviews*, Vol.24, No.6, pp.581–604 (2000).
3) Morimoto, C.H. and Mimica, M. R.M.: Eye gaze tracking techniques for interactive applications, *Computer Vision and Image Understanding*, Vol.98, No.1, pp. 4–24 (2005).
4) Atkinson, R. and Shiffrin, R.: Human memory: A proposed system and its control processes, *The psychology of learning and motivation: Advances in research and theory*, Vol.2, pp.89–195 (1968).
5) Qvarfordt, P. and Zhai, S.: Conversing with the user based on eye-gaze patterns, *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, New York, NY, USA, ACM, pp.221–230 (2005).
6) Nakano, Y. and Ishii, R.: Estimating User's Engagement from Eye-gaze Behaviors in Human-Agent Conversations, *Proceedings of the International Conference on Intelligent User Interfaces (IUI2010)* (2010).
7) Hirayama, T., Dodane, J.-B., Kawashima, H. and Matsuyama, T.: Estimates of User Interest Using Timing Structures between Proactive Content-display Updates and Eye Movements, *IEICE Transactions on Information and Systems*, Vol.E-93D, No.6, pp.1470–1478 (2010).
8) Doshi, A. and Trivedi, M.: Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions, *IEEE Intelligent Vehicles Symposium*, Vol.10, No.3, pp.887–892 (2009).
9) Fletcher, L. and Zelinsky, A.: Driver Inattention Detection based on Eye Gaze-Road Event Correlation, *International Journal of Robotics Research*, Vol.28, No.6, pp.774–801 (2009).
10) Itti, L., Koch, C. and Niebur, E.: A Model of Saliency-based Visual Attention for Rapid Scene Analysis, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.20, No.11, pp.1254–1259 (1998).
11) Walther, D. and Koch, C.: Modeling attention to salient proto-objects, *Neural networks : the official journal of the International Neural Network Society*, Vol.19, No.9, pp.1395–407 (2006).
12) Ma, Y.-F. and Zhang, H.-J.: A model of motion attention for video skimming, *Proceedings of 2002 International Conference on Image Processing*, Vol.1, pp.I–129–I–132 (2002).
13) Saslow, M.G.: Latency for Saccadic Eye Movement, *Journal of the optical society of America*, Vol.57, No.8, pp.1030–1033 (1967).
14) Cerf, M., Harel, J., Einhauser, W. and Koch, C.: Predicting human gaze using low-level saliency combined with face detection, *Advances in neural information processing systems*, Vol.20, pp.241–248 (2008).