

顕著性と自己運動に基づく一人称視点における視覚的注意推定

山田健太郎[†] 菅野 裕介[†] 岡部 孝弘[†] 佐藤 洋一[†] 杉本 晃宏^{††}
開 一夫^{†††}

[†] 東京大学生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

^{††} 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

^{†††} 東京大学大学院総合文化研究科 〒153-8902 東京都目黒区駒場 3-8-1

E-mail: [†]{yamada,sugano,takahiro,ysato}@iis.u-tokyo.ac.jp, ^{††}sugimoto@nii.ac.jp,

^{†††}khiraki@idea.c.u-tokyo.ac.jp

あらまし 本稿では、視覚的顕著性と自己運動に基づく、一人称視点における視覚的注意の推定手法を提案する。提案手法は、頭部装着型カメラから撮影された一人称視点映像のみから、映像を撮影した本人の視覚的注意を推定することを目的とする。視覚的顕著性マップモデルは、一人称視点映像に限らず、一般の画像・映像のみから視覚的注意を推定する手法として近年注目を集めている。本研究では、一人称視点映像から自己運動を推定することにより、頭部回転と進行方向を利用した視覚的注意推定マップ生成し、これを視覚的顕著性マップと統合することで、より一人称視点に適した視覚的注意推定を実現する。提案手法の有効性は、複数の人物と環境の組み合わせで撮影した一人称視点映像において実際の視線と比較することで検証し、顕著性と自己運動の両方が視覚的注意推定に有効に寄与しており、両者を統合することでより高精度な視覚的注意推定が可能であることを示した。

キーワード 視覚的顕著性, 視覚的注意, 視線, 一人称視点映像, 自己運動推定

1. はじめに

人間の視覚的な注意は、我々の内部状態を推定するための重要な手がかりである。それゆえ、インタラクティブ広告や、Attentive User Interface といった、人間中心のインタラクティブシステムに効果的に用いられている。

人間中心のメディアの一つである一人称視点映像においても、視覚的注意を推定することは有用である。一人称視点映像とは、頭部に装着されたカメラ等の観察者と同じような視点（一人称視点）から撮影された映像のことである。一人称視点のカメラは日常的な自己の行動を観測するのに非常に適している。それゆえ、一人称視点映像を分析することで行動意図を理解し支援を行うこと等を目的とした研究が盛んに行われており、幅広いアプリケーションへの応用が期待されている。たとえば、モバイルアシスタント [1] や、福祉分野での患者のモニタリング [2] への応用が既に行われている。このような目的においても、一人称視点映像中のどこに注意が向けられているのか、という情報は有益である。

視覚的注意の範囲は、視線方向とは分離可能であることが知られている [3]。しかし、眼球運動による自由な視線移動が可能状態では、視線方向の周辺以外の領域に注意を払いづらくなることが被験者実験により明らかにされている [4], [5]。したがって、固視が強制されない日常的な環境においては、視線位置を視覚的注意が向けられている領域の中心として扱うことは妥当である。一人称視点の環境に限らず、これまで多くの視線計測技術が

開発されてきた [6]。しかしながら、高価な設備と煩雑なキャリブレーション等を必要とする手法が多く、拘束の少ない環境で視線を計測することは、現在もなお困難な課題である。

視線計測技術とは異なるアプローチで視覚的注意の推定を行う手法として、視覚的顕著性マップモデルが挙げられる。視覚的顕著性マップモデルは、人間の視覚処理を模した計算を行うことで、人間が注意を向けやすい画像中の領域を推定するための計算モデルである。顕著性マップはもともと人間の視覚処理・視覚的注意の機能を理解するための概念であり、心理学分野の視覚的注意に関する研究 [7] を踏まえ、Kock と Ullman によって提唱された [8]。その後、完全な計算モデルとしての視覚的顕著性マップモデルが Itti らにより提案され、人間の視覚ではなく、計算機による処理が可能となった [9]。それ以来、多数の視覚的顕著性マップモデルが提案されてきた [10] ~ [15]。たとえば、運動方向や輝度変化といった低レベルな動的特徴を導入することで、静止画像だけでなく動画にも適用可能なモデルが提案されている [10]。また、視線計測に基づいた研究により、顕著性マップは実際の人間の注意分布に適合していることが示されている [16] ~ [18]。このようなモデルの拡張、有用性の確認に基づき、近年、視覚的顕著性マップモデルはセグメンテーションや画像編集等、人間の知覚・注意を利用したアプリケーションに、広く応用されている。

本稿では、視覚的顕著性と自己運動に基づき、頭部装着型カメラから撮影された一人称視点映像のみから映像

を撮影した本人の視覚的注意を推定する手法を提案する。本研究では一人称視点特有の要素である自己運動として、頭部の回転速度と、映像中の進行方向位置であるオプティカルフローの拡大中心 (FOE; Focus of Expansion) を用いる。関連する研究としては、一人称視点のヴァーチャルリアリティ (VR) 環境において視点の回転速度と FOE を利用して視線の推定を行う手法が Hillair ら [19], [20] により提案されている。また、Fukuchi ら [21] は FOE が視覚的注意を誘導することを実験的に示し、FOE を組み込んだ顕著性マップモデルを提案している。本研究ではこれらの手法と異なり、実際の人物が装着したカメラから撮影された一人称視点映像を対象とした実証実験を行う。また、実際の一人称視点映像から頭部回転速度と進行方向を求めるために、提案手法では複数フレームの画像間の特徴点対応から自己運動パラメータの推定を行っている。さらに、頭部回転の方向と速度に応じて注意の中心が移動するマップと、映像中の FOE の位置を注意の中心とするマップの 2 種類の視覚的注意推定マップを生成し、顕著性マップと統合することで、より一人称視点映像に適した視覚的注意推定を実現する。実験では、複数の人物と環境の組み合わせで撮影した一人称視点映像に対し、提案手法により得られた視覚的注意推定マップと実際の視線の比較を行い、提案手法の有効性を示す。

2. 顕著性と自己運動に基づく視覚的注意推定

本研究の目的は一人称視点映像のみから映像を撮影した本人の視覚的注意を推定することである。顕著性マップは一人称視点映像に限らず一般の画像・映像のみから視覚的注意を推定することができるが、提案手法では一人称視点に特化した要素として、頭部回転と進行方向に基づく視覚的注意推定マップを生成し、顕著性マップと統合することで、より高精度な視覚的注意推定を実現する。提案手法のフローを図 1 に示す。以下、一人称視点映像からの顕著性マップの生成、一人称視点映像からの自己運動（頭部回転、進行方向）推定と頭部回転、進行方向に基づく視覚的注意推定マップの生成、各マップの統合の詳細について述べる。

2.1 一人称視点映像からの顕著性マップ生成

顕著性マップの生成には、Harel らによる GBVS (Graph-Based Visual Saliency) モデル [13] を用いる。映像に対して GBVS モデルを用いる場合、低レベルの画像特徴として、静的特徴である色、輝度、方向に加え、動的特徴である運動方向、輝度変化を用いることもできるが、一人称視点においては動的特徴が精度を低下させることが示されている [22] ため、本研究では静的特徴のみを用いる。視覚的顕著性マップモデルでは、これらの特徴量について周囲との差分が大きい領域は顕著性が高く、視覚的注意を誘導するという、そして、各特徴ごとに計算された顕著性マップを統合することで、より

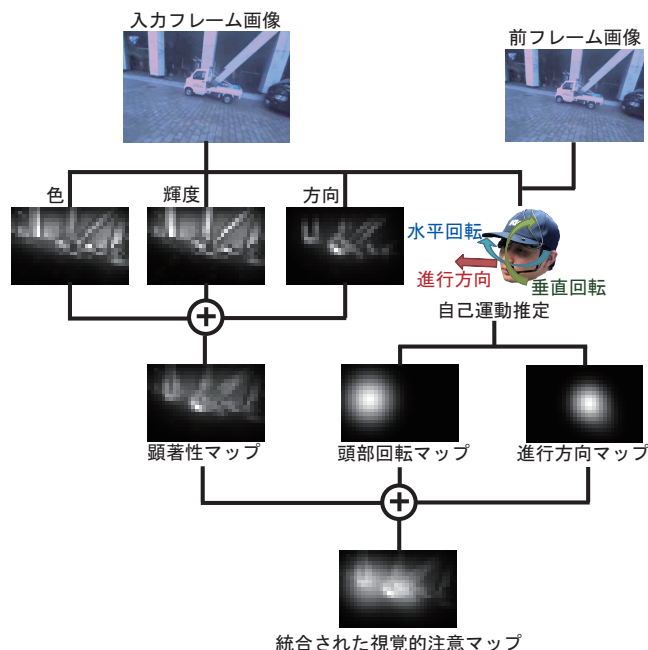


図 1 提案手法のフロー。

人間の視覚に適した顕著性マップが生成されることを基本的な概念としている。GBVS モデルでは特徴ごとの顕著性マップを、画素ノード間の重みを画像特徴の相違度として定義したマルコフ連鎖の定常分布解析により計算している。また、各特徴ごとの顕著性マップは等しい重みで足し合わされ、最終的な顕著性マップを得る。

2.2 自己運動に基づく視覚的注意推定マップ生成

自己運動に基づく視覚的注意推定マップの生成は、一人称視点映像からの自己運動推定、各軸周りの頭部回転の角速度の推定と視覚的注意推定マップ生成、映像中の進行方向位置推定と視覚的注意推定マップ生成、の 3 ステップで行われる。まず一人称視点映像の現在のフレームの画像と 1 つ前のフレームの画像から、エピポーラ幾何の関係を用いて 2 画像間の自己運動を推定し、回転行列を得る。次に、頭部回転に基づく視覚的注意推定マップは水平回転と垂直回転に基づいて生成するため、回転行列から各座標軸回りの回転角度の計算を行い、得られた頭部回転の角速度に基づき、一人称視点映像に対応した 2 次元マップ中のある座標を中心にガウス分布状に値を与えることで、頭部回転に基づく視覚的注意推定マップを生成する。また、進行方向に基づく視覚的注意推定マップはオプティカルフローの拡大中心 (FOE; Focus of Expansion, 動きの消失点, 動きの無限遠点) を複数フレーム間で求めた後、ガウシアンカーネルを用いた 2 次元カーネル密度推定により生成する。

映像から自己運動推定を行うステップでは、カメラの内部パラメータが必要となる。未校正の多数の画像間の対応点を用いることで、カメラの内部パラメータ（焦点距離、光軸点、幾何学的歪）を推定する理論も存在するが [23]、それは本研究の主題ではないため、本研究では

カメラの内部パラメータはキャリブレーションにより既知であり、一人称視点映像の幾何学的歪は補正済みであるとする。また、頭部の回転に基づく視覚的注意推定はカメラと頭部の位置関係に依存するため、本研究では、一人称視点カメラの水平軸と垂直軸およびの映像の中心が、人間の頭部の水平軸と垂直軸および視線を前方に向けた時の視野中心と一致するようにカメラを装着していると仮定する。以下、各ステップの詳細について述べる。

2.2.1 自己運動推定

まず、頭部回転の角速度と画像中の進行方向の推定に先立ち、一人称映像において、現在のフレームの画像と1フレーム前の画像間の相対的な自己運動推定を行い、回転行列 R と並進ベクトル t を求める。まず2つの画像から、KLT 特徴点追跡法 [24], [25] により求めた安定した特徴点对応のうち8点を用いて、8点アルゴリズム [26] により基礎行列を推定する。8点アルゴリズムを用いる際、RANSAC [27] により、運動物体や誤対応等のアウトライアを含まず、インライアが最大となるような8点を用いることでロバストな推定を行う。既知のカメラの内部パラメータ行列と、得られた基礎行列を用いて、回転行列 R と並進ベクトル t を求めることができる。回転行列 R と並進ベクトル t を求める際の詳細な計算については文献 [28], [29] 等を参照されたい。

2.2.2 頭部回転に基づく視覚的注意推定マップ生成

頭部回転に基づく視覚的注意推定マップは、頭部の水平回転速度と、垂直回転速度から計算するため、前ステップで得られた回転行列 R から、各軸回りの回転角度を求める必要がある。一人称映像において、水平右方向を x 軸、垂直下方向を y 軸、光軸の奥行き方向を z 軸、画像中心を原点とする。また、 x, y, z 各軸回りの回転角度をそれぞれ $\theta_x, \theta_y, \theta_z$ とすると、本研究ではカメラの水平・垂直軸と人間の頭部の水平・垂直軸を同一としているため、頭部の水平回転角度は θ_y 、垂直回転角度は θ_x となる。ここで、回転行列 R の各要素を以下のように置く。

$$R = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \quad (1)$$

また、回転の順番を x 軸、 y 軸、 z 軸の順とすると、 θ_x, θ_y は以下のように求められる。 [30]

(a) $\sin(\theta_y) \neq \pm 1$ のとき

$$\theta_x = \arctan\left(\frac{b_3}{c_3}\right) \quad (2)$$

$$\theta_y = \arcsin(-a_3) \quad (3)$$

(b) $\sin(\theta_y) = \pm 1$ のとき

$$\theta_x = \arctan\left(\pm \frac{b_1}{b_2}\right) \quad (4)$$

$$\theta_y = \pm \frac{\pi}{2} \quad (5)$$

ただし、複合同順。また (b) では値が一意に求まらないため、 $\theta_z = 0$ と仮定している。

一人称視点映像のフレームレートを $f[\text{fps}]$ とすると、水平回転の角速度 $\omega_y[\text{deg/s}]$ 、垂直回転の角速度 $\omega_x[\text{deg/s}]$ は、以下のように表すことができる。

$$\omega_x = 180f\theta_x/\pi \quad (6)$$

$$\omega_y = 180f\theta_y/\pi \quad (7)$$

Hillair ら [19] による、ヴァーチャルリアリティ環境における視点の回転と視線移動の関係によると、角速度が約 $100[\text{deg/s}]$ 以下では頭部回転の角速度と回転方向の視線の分布座標に正の相関があり、約 $100[\text{deg/s}]$ 以上ではほぼ一定の視線の座標分布となる。そこで、本研究では一人称映像に対応した2次元マップ上で、以下の座標 (x, y) を中心とした2次元ガウス分布状に値を与えることで、自己運動に基づく視覚的注意推定マップを生成した。

$$x = \begin{cases} \frac{\omega_y}{100} \cdot \frac{w}{k} & (|\omega_y| \leq 100) \\ \frac{w}{k} & (\omega_y > 100) \\ -\frac{w}{k} & (\omega_y < -100) \end{cases} \quad (8)$$

$$y = \begin{cases} -\frac{\omega_x}{100} \cdot \frac{h}{l} & (|\omega_x| \leq 100) \\ -\frac{h}{l} & (\omega_x > 100) \\ \frac{h}{l} & (\omega_x < -100) \end{cases} \quad (9)$$

ただし、 w, h はそれぞれ、視覚的注意推定マップの横幅、高さを表している。また、 k, l はカメラの画角に応じたパラメータで、本研究の実験環境では $k = l = 3$ とした。

2.2.3 進行方向に基づく視覚的注意推定マップ生成

視覚による自己運動知覚においては、前方への移動（前進）が最も重要であり、その際、進行方向を注視しやすいとされている [31]。人間は進行方向を知覚するために、オプティカルフローを重要な視覚の手がかりとしている。特に直進運動の場合、視覚刺激は放射状の拡大運動となり、拡大中心（FOE; Focus of Expansion, 動きの消失点、動きの無限遠点）を進行方向として知覚することが古くから知られている [32]。また、人間の視覚においてのみならず、コンピュータビジョンの分野においても、画像中の進行方向の位置として FOE が用いられている [33]。そこで、以下のようにして FOE を用いて進行方向に基づく視覚的注意推定マップを生成する。

他に運動物体のない環境におけるロボット等の、単純な直進運動として制御可能な移動物体から撮影された動画画像においては、オプティカルフローの交点を求めることで比較的容易に FOE を得ることができる。ただし、一人称視点映像におけるオプティカルフローは、通常、独立して運動する対象物や、自己運動による回転運動によるフローを反映したものとなり、これらを含む場合には

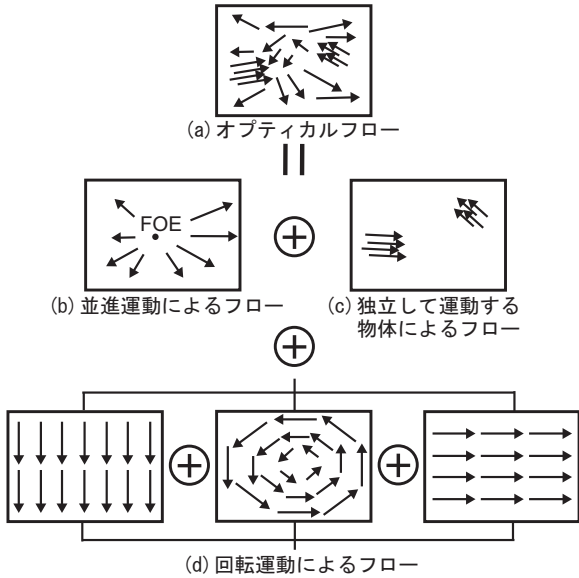


図2 FOE とオプティカルフローの成分．並進運動によるフロー (b) は放射状の拡大運動となり，その交点 (FOE; Focus of Expansion, 動きの消失点, 動きの無限遠点) として進行方向の位置が求められる．オプティカルフロー (a) は，独立して運動する対象物 (c) や自己運動による回転運動 (d) によるフローを反映したものとなり，これらを含む場合には進行方向とフローの交点は一致しない．(a) から，(c)・(d) の除去・分離を行うことで，並進運動のみによるフローを用いて FOE を求めることができる．

進行方向とフローの交点は一致しない (図2)．しかし，頭部回転を含む等の自己運動が複雑な場合でも，人間は眼球運動情報を利用し，進行方向をある程度正確に知覚することができ [34] ~ [38]，独立して運動する対象物があった場合にも進行方向は正しく知覚される [39]．コンピュータビジョンにおいても，運動物体によるフローの除去や，フローの回転成分と並進成分の分離をすることができれば，並進運動のみによるフローを用いて FOE を求め，進行方向として得ることができる．

オプティカルフローから FOE を推定するために，まず，運動物体や，誤対応によるフローを除去する．これには基礎行列 F を求めた際，RANSAC により，アウトライアとされたフローを運動物体と誤対応によるフローとして除去し，インライアとされたフローのみを用いることで実現する．次に，フローの回転成分と並進成分を分離する．フローの回転成分と並進成分の分離手法の一つとして，大西ら [40] により，ビデオカメラに取り付けたジャイロセンサから得られた回転行列を用いる手法が提案されている．本研究では映像のみを用いるが，既知の回転行列 R を利用することで同様の処理を以下のように実現する．カメラの光学中心 C に対応する画像を I ，カメラの光学中心 C' に対応する画像を I' とし，画像 I' を並進移動させずに R だけ回転させた画像 I'_R を考える．画像 I'_R の光学中心は C' であるから，画像 I'_R から



(a) 回転成分を含んだフロー (b) 並進成分のみのフロー
図3 オプティカルフローからの並進成分の分離．前フレームの画像を回転させ，特徴点座標を求めることで，回転成分と並進成分を含んだフロー (a) から，並進成分のみのフロー (b) を得ることができる．各画像において，白い四角は現在のフレームの画像における特徴点座標であり，黒い四角は，(a) では前フレームの画像における特徴点座標，(b) では前フレームを回転させた画像における特徴点座標である．赤い丸は，各フローまでのユークリッド距離の合計が最小となる点として推定されたフローの交点を示す．

画像 I への相対的な位置関係は並進ベクトル t のみとなる．FOE の計算に必要なのは並進運動のみによるフローであるから，画像 I'_R から画像 I へのフローを用いればよい．画像 I'_R における，投影点の正規化画像座標を x'_R ，一般的な画像座標を m'_R とし，既知のカメラの内部パラメータ行列を A とすると，次の関係式が得られる．

$$\tilde{m}' \sim Ax' \quad (10)$$

$$\tilde{m}'_R \sim Ax'_R \quad (11)$$

また，画像 I から，画像 I'_R への並進ベクトルは $t_0 = 0$ ，回転行列は R であるから，次の関係式が成り立つ．

$$\tilde{x}'_R \sim R\tilde{x}' + t_0 = R\tilde{x}' \quad (12)$$

式 10，式 11，式 12 より，次の式が成り立つ．

$$\tilde{m}'_R \sim ARA^{-1}\tilde{m}' \quad (13)$$

以上の手順で，計算された m'_R から m への移動が，点 P についての画像 I'_R から画像 I へのフローとなる．この処理を，用いるすべてのフローに対して行うことで，並進運動のみによるフローを抽出した．得られた並進運動のみによるフローは，理論的には 1 点 (FOE) で交わるが，3 本以上のフローがある場合，ノイズの影響ですべてのフローが 1 点で交わるとは限らない．そこで，各フローまでのユークリッド距離の合計が最小となる点を求めることで，ロバストに FOE を推定した．図3に，上述の手順で，実際の一人称視点映像のオプティカルフローから並進成分の分離を行った例を示す．

以上のように，2 フレーム間の自己運動に基づいて FOE を得たが，一人称視点映像においては歩行中の頭部運動や身体運動の影響により，2 フレーム間のみ関係では人間が知覚している進行方向を正確に得ることはできない．そこで，よりロバストに FOE マップを生成

するため、現在のフレームの画像と、現在のフレームから k フレーム前 ($k = 1, 2, \dots, n$, 本研究では $n = 15$) の画像を用いて自己運動パラメータと FOE の推定を行い、現在のフレームの画像中に検出されたすべての FOE を用いて FOE マップを生成する。また、FOE の周辺では注意が向きやすく、遠い場所では注意が向きにくいという性質があるため、FOE から離れるにしたがって小さくなるように顕著度を与えるように FOE マップを設計する。本研究では、複数の FOE から、ガウシアンカーネルを用いた 2 次元カーネル密度推定により、FOE マップを生成した。カーネル密度推定の平滑化バンド幅は、Silverman [41] の「経験則」(標準偏差と 1.34 倍で割った四分位偏差の小さい方を 0.9 倍したものを標本数の $1/5$ 乗で割る処理) により求めた。

2.3 マップの統合

上述のようにして得られた低レベルな視覚刺激に基づく顕著性マップと、自己運動に基づく視覚的注意推定マップを統合し、最終的な視覚的注意推定マップを生成する。本研究では、顕著性マップに統合する自己運動に基づく視覚的注意推定マップとして、頭部回転に基づくマップ、進行方向に基づくマップ、そしてその両方を利用した、3 種類の統合マップを提案する。統合の際には、顕著性マップと自己運動に基づく視覚的注意推定マップを各々同じ割合で足し合わせ、足し合わせ後の最大値が規定の最大値になるように単純なスケーリング調整を行い、最終的な視覚的注意推定マップを生成する。マップの統合による視覚的注意推定マップの生成例を図 4 に示す。

3. 評価実験

本節では、提案手法を評価するために行った被験者実験とその結果について述べる。提案手法による視覚的注意推定マップの生成モデルは、顕著性マップに加え、自己運動と関係した注意の要素を考慮しているため、低レベルの視覚刺激に基づいた概念である視覚的顕著性マップモデルの枠組みから外れているが、顕著性マップと同様の 2 次元マップの形で視覚的注意を表現している。そこで、顕著性マップの一般的な性能評価手法である、実際の視線位置との比較により計算される ROC (Receiver Operating Characteristic) 曲線を用いて提案手法の性能評価を行った。以下、実験の詳細について述べる。

3.1 実験設定

実験では、ナックイメーজテクノロジー製のモバイル型視線計測装置 EMR-9 [42] を用いて視線位置の計測と一人称視点映像の撮影を行った。図 5(a) に示すように、EMR-9 は 2 つのアイカメラと、2 つの赤外線光源を備え、眼球運動を 240[Hz] で計測することができる。また、それと同時に、頭部に装着された視野カメラにより、一

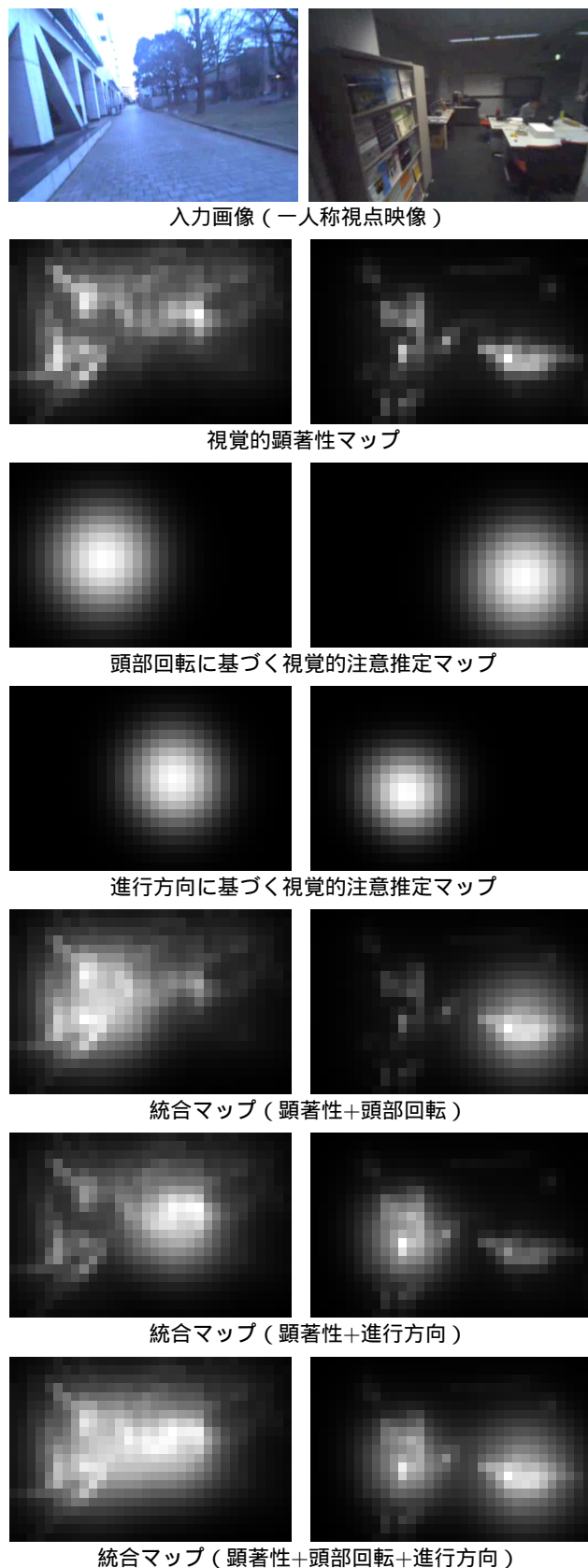


図 4 マップの統合による視覚的注意推定マップの生成例。各行は上から順に入力画像，顕著性マップ，頭部回転に基づく視覚的注意推定マップ，進行方向に基づく視覚的注意推定マップ，統合された視覚的注意推定マップ（顕著性+頭部回転，顕著性+進行方向，顕著性+頭部回転+進行方向，の 3 種類）である。

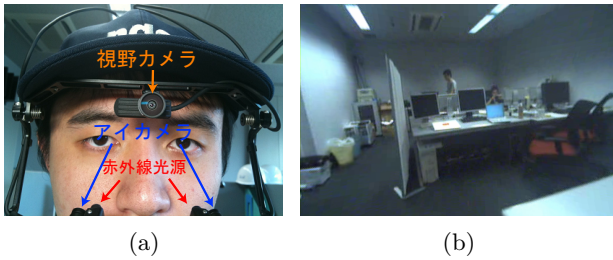


図 5 (a) EMR-9 [42], ナックイメージテクノロジー製のモバイル型視線計測装置. EMR-9 は 2 つのアイカメラと, 2 つの赤外線光源を備え, 眼球運動を 240[Hz] で計測することができる. また, それと同時に, 頭部に装着された視野カメラにより, 一人称視点映像を 30[fps] で撮影を行う. 視野カメラの水平画角は 121° であり, 撮影される映像の解像度は 640 × 480 である. (b) 実験中に視野カメラから撮影された映像のフレームの例.

一人称視点映像を 30[fps] で撮影する. 視野カメラの水平画角は 121° であり, 撮影される映像の解像度は 640 × 480 である. ただし, 映像の上下に同期信号が表示されるため, それを除いて実際に用いた映像の解像度は 640 × 426 である.

被験者実験は, 室内で着席した状態 (頭部の運動は可能), 室内で自由移動, 屋外で自由移動の 3 パターンの実験設定で行った. 5 人の被験者が実験に参加し, 各被験者につき, 約 1 分間の一人称視点映像を 3 パターン撮影し, 合計 15 セットのデータセットを作成した. 図 5(b) に, 実験中に視野カメラから撮影された映像のフレームの例を示す. 瞬目等によるエラーと高速な視線移動中の視線位置を取り除いた上で, 性能評価への寄与を同一にするため, 各一人称視点映像につき 8,000 点の視線位置を実験に用いた.

3.2 実験結果

提案手法により生成された視覚的注意推定マップが, 一人称視点映像における視覚的注意をどの程度有効に表現できているかを評価するため, 視覚的注意推定マップと, 視線計測結果の関係を分析した. 図 6 に示す曲線は, 被験者実験で取得した各データセットを用い, 閾値を最小から最大まで変化させることで算出された ROC(Receiver Operating Characteristic) 曲線である. 図 6 では, 提案手法による結果と顕著性マップのみを用いた結果を比較している. 縦軸は検出率であり, 各マップにおいて閾値よりも高い値を持つ領域に含まれた実際の視線位置の割合に対応する. 横軸は誤検出率であり, 各マップ中で閾値よりも高い値を持ち, 視線が存在しない領域の割合を示している. すなわち, より左上に近い曲線は, より高精度に視覚的注意を推定できているといえる. この結果から, 一般的な画像・映像に対する視覚的注意をモデル化した顕著性マップに比べ, 一人称視点映像に特化した提案手法の方が, より高精度に一人称視点における視覚的注意を推定できていることがわかる.

表 1 提案手法の性能評価 (AUC, 曲線下面積). 全データセットを用いた結果における各マップの視覚的注意の推定性能を表している. 提案手法による結果 (顕著性+頭部回転, 顕著性+進行方向, 顕著性+頭部回転+進行方向, の 3 種類) と顕著性マップのみを用いた結果 (顕著性のみ), 頭部回転に基づく視覚的注意推定マップのみを用いた結果 (頭部回転のみ), マップの中央 (中心座標) を中心としてガウス分布状に値を与えたマップを用いた結果 (中央のみ / 顕著性+中央) について比較している.

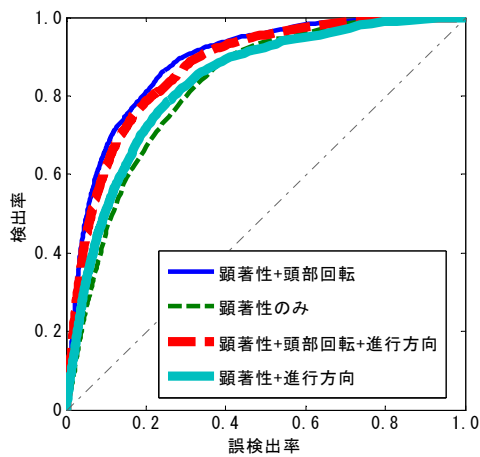
| マップ生成の手法 | AUC |
|----------------------|--------------|
| 提案手法 (顕著性+頭部回転) | 0.900 |
| 提案手法 (顕著性+進行方向) | 0.841 |
| 提案手法 (顕著性+頭部回転+進行方向) | 0.893 |
| 顕著性のみ | 0.809 |
| 頭部回転のみ | 0.892 |
| 中央のみ | 0.884 |
| 顕著性+中央 | 0.890 |

表 1 に示す AUC (Area under the curve, ROC 曲線の曲線下面積) は全データセットを用いた結果における各マップの視覚的注意の推定性能を表している. 表 1 では, 3 種類の提案手法による結果 (顕著性+頭部回転, 顕著性+進行方向, 顕著性+頭部回転+進行方向) と顕著性マップのみを用いた結果 (顕著性のみ), 頭部回転に基づく視覚的注意推定マップのみを用いた結果 (頭部回転のみ) に加え, マップの中央 (中心座標) を中心としてガウス分布状に値を与えたマップを用いた結果 (中央のみ/顕著性+中央) についても比較している. この結果から, 進行方向も視覚的注意推定に有用ではあるが, 顕著性と頭部回転を用いた提案手法が最も高精度に視覚的注意を推定できることがわかる. また, 単純に一人称視点映像の中央 (顔の前方) にガウス分布状の値を与えるよりも, 頭部の回転速度に応じてガウス分布の中心を移動させた提案手法の方が有用であるといえる.

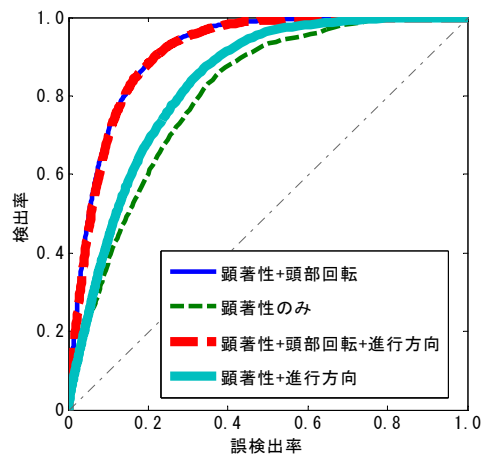
4. 結論

本稿では, 視覚的顕著性と自己運動に基づき, 一人称視点における視覚的注意を高精度に推定する手法を提案した. これは頭部装着型カメラから撮影された一人称視点映像のみから, 映像を撮影した本人の視覚的注意を推定する初めての研究であるといえる. 実験により, 顕著性と自己運動の両方が視覚的注意推定に有効に寄与しており, 両者を統合することでより高精度に一人称視点における視覚的注意を推定できることを確認した. 特に, 顕著性と頭部回転を用いた視覚的注意推定マップが最も高精度に視覚的注意を推定できることがわかった.

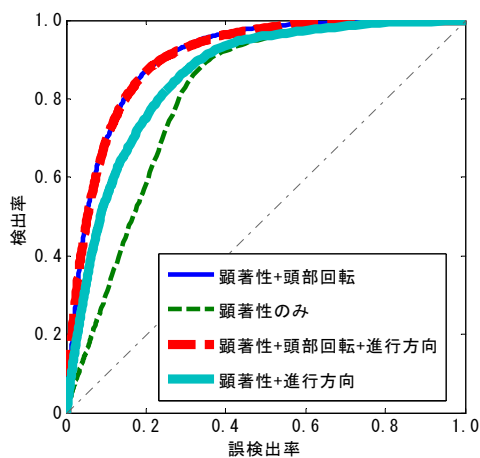
提案手法により, 頭部に装着したカメラ以外の特殊なデバイスを用いずに, 一人称視点映像における視覚的注意を推定することが可能になる. これにより, 安価な装



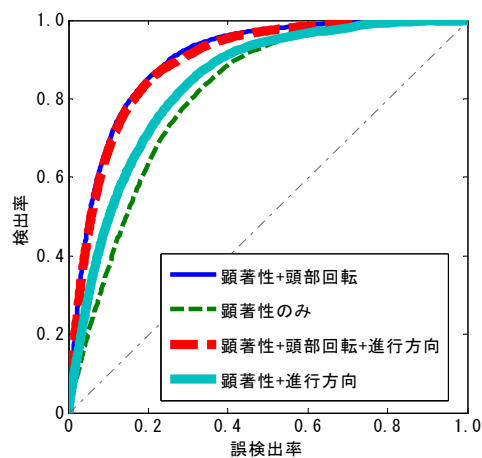
(a) 室内で着席状態



(b) 室内で自由移動



(c) 屋外で自由移動



(d) 全てのデータセット

図 6 各データセットの ROC 曲線 ((a) 室内で着席状態, (b) 室内で自由移動, (c) 屋外で自由移動, (d) 全てのデータセット). ROC 曲線は, 被験者実験で取得した各データセットを用い, 閾値を最小から最大まで変化させることで算出した. 縦軸は検出率であり, 各マップにおいて閾値よりも高い値を持つ領域に含まれた実際の視線位置の割合に対応する. 横軸は誤検出率であり, 各マップ中で閾値よりも高い値を持ち, 視線が存在しない領域の割合を示している.

着型視線推定装置の開発や, 一人称視点映像中の視覚的注意領域に着目したライフログシステムなど, 様々な用途が期待できる.

本研究においては, 一人称視点映像は頭部装着型カメラから撮影されるものとしたが, 頭部以外の身体(肩等)に装着されたカメラから一人称視点映像が撮影される場合もある. その場合, 本研究で用いた視点の回転に基づいた視覚的注意推定は困難と考えられるが, 進行方向に基づいた視覚的注意推定は, 映像中に進行方向が含まれている限り有用であると考えられる. そこで, 頭部以外に装着された一人称視点映像における提案手法の性能評価とその対応のための拡張が今後の課題として挙げられる.

謝辞 本研究の一部は JST CREST「共生社会に向けた人間調和型情報技術の構築」領域採択課題「日常生活空間における人の注視の推定と誘導による情報支援基盤

の実現」により実施した.

文 献

- [1] T. Starner, "Wearable agents," IEEE Pervasive Computing, vol.1, no.2, pp.90-92, 2002.
- [2] R. Megret, D. Szolgay, J. Benois-Pineau, P. Joly, J. Pinquier, J.F. Dartigues, and C. Helmer, "Wearable video monitoring of people with age dementia: Video indexing at the service of healthcare," International Workshop on Content-Based Multimedia Indexing, 2008. (CBMI 2008.)IEEE, pp.101-108 2008.
- [3] M.I. Posner, "Orienting of attention," The Quarterly journal of experimental psychology, vol.32, no.1, pp.3-25, 1980.
- [4] B. Khurana and E. Kowler, "Shared attentional control of smooth eye movement and perception," Vision Research, vol.27, no.9, pp.1603-1618, 1987.
- [5] J.E. Hoffman and B. Subramaniam, "The role of visual attention in saccadic eye movements," Perception and Psychophysics, vol.57, no.6, pp.787-795, 1995.

- [6] D.W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol.32, no.3, pp.478–500, 2010.
- [7] A.M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol.12, no.1, pp.97–136, 1980.
- [8] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry.," *Human neurobiology*, vol.4, no.4, pp.219–227, 1985.
- [9] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol.20, no.11, pp.1254–1259, 1998.
- [10] L. Itti, N. Dhavale, F. Pighin, et al., "Realistic avatar eye and head animation using a neurobiological model of visual attention," *SPIE 48th Annual International Symposium on Optical Science and Technology*, vol.5200, pp.64–78, 2003.
- [11] T. Avraham and M. Lindenbaum, "Esaliency (extended saliency): Meaningful attention using stochastic image modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol.32, no.4, pp.693–708, 2010.
- [12] M. Cerf, W. Harel, J. and Einhäuser, and C. Koch, "Predicting human gaze using low-level saliency combined with face detection," *Advances in Neural Information Processing Systems (NIPS)*, vol.20, pp.241–248, 2007.
- [13] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems (NIPS)*, vol.19, pp.545–552, 2006.
- [14] L.F. Costa, "Visual saliency and attention as random walks on complex networks," *ArXiv Physics e-prints*, arXiv:physics/0603025, pp.1–6, 2006.
- [15] W. Wang, Y. Wang, Q. Huang, and W. Gao, "Measuring visual saliency by site entropy rate," *Computer Vision and Pattern Recognition (CVPR)IEEE*, pp.2368–2375 2010.
- [16] T. Foulsham and G. Underwood, "What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition," *Journal of Vision*, vol.8, no.2:6, pp.1–17, 2008.
- [17] L. Itti, "Quantitative modelling of perceptual salience at human eye position," *Visual Cognition*, vol.14, no.4, pp.959–984, 2006.
- [18] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision Research*, vol.42, no.1, pp.107–123, 2002.
- [19] S. Hillaire, A. Lécuyer, G. Breton, and T.R. Corte, "Gaze behavior and visual attention model when turning in virtual environments," *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pp.43–50, VRST '09, ACM, New York, NY, USA, 2009.
- [20] S. Hillaire, A. Lécuyer, T. Regia-Corte, R. Cozot, J. Royan, and G. Breton, "A real-time visual attention model for predicting gaze point during first-person exploration of virtual environments," *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, pp.191–198, VRST '10, ACM, New York, NY, USA, 2010.
- [21] M. Fukuchi, N. Tsuchiya, and C. Koch, "The focus of expansion in optical flow fields acts as a strong cue for visual attention," *Journal of Vision*, vol.9, no.8, p.137a, 2009.
- [22] 山田健太郎, 菅野裕介, 岡部孝弘, 佐藤洋一, 杉本晃宏, 開 一夫, "一人称視点における視覚的顕著性マップモデルの性能評価," *電子情報通信学会技術研究報告. HIP, ヒューマン情報処理*, vol.110, no.422, pp.81–86, 2011-02-22 .
- [23] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, 2nd ed., Cambridge Univ Pr, 2003.
- [24] C. Tomasi and T. Kanade, "Detection and tracking of point features," *Carnegie Mellon University Technical Report CMU-CS-91-132*, pp.1–22, 1991.
- [25] J. Shi and C. Tomasi, "Good features to track," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)IEEE*, pp.593–600 1994.
- [26] R.I. Hartley, "In defense of the eight-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol.19, no.6, pp.580–593, 1997.
- [27] M.A. Fischler and R.C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol.24, no.6, pp.381–395, 1981.
- [28] 金谷健一, *画像理解 3次元認識の数理*, 森北出版, 1990 .
- [29] K. Kanatani, *Statistical optimization for geometric computation: theory and practice*, Elsevier, Amsterdam, 1996.
- [30] 島田静雄, *CAD・CGのための基礎数学 (インターネット時代の数学シリーズ (7))*, 共立出版, July 2000 .
- [31] 日本視覚学会, *視覚情報処理ハンドブック*, 朝倉書店, 2000 .
- [32] J.J. Gibson and L. Carmichael, *The perception of the visual world*, Houghton Mifflin Boston, 1950.
- [33] D.H. Ballard and C.M. Brown, *Computer Vision*, Prentice Hall, May 1982.
- [34] W.H. Warren, Jr and D.J. Hannon, "Eye movements and optical flow," *Journal of the Optical Society of America A*, vol.7, no.1, pp.160–169, 1990.
- [35] C.S. Royden, M.S. Banks, and J.A. Crowell, "The perception of heading during eye movements," *Nature*, vol.360, no.6404, pp.583–585, 1992.
- [36] C.S. Royden, J.A. Crowell, and M.S. Banks, "Estimating heading during eye movements," *Vision research*, vol.34, no.23, pp.3197–3214, 1994.
- [37] E.C. Hildreth and C.S. Royden, "Computing observer motion from optical flow," *High-level motion processing: Computational, neurobiological, and psychophysical perspectives*, pp.269–293, 1998.
- [38] W. Warren, "The state of flow," *High-Level Motion Processing: Computational, Neurobiological and Psychophysical Perspectives*, pp.315–358, 1998.
- [39] C.S. Royden and E.C. Hildreth, "Human heading judgments in the presence of moving objects.," *Perception & psychophysics*, vol.58, no.6, p.836, 1996.
- [40] 向井利春, 大西 昇, "2次元動画画像中の特徴点の追跡方法," *特開 2000 - 339473 号*, pp.1–10, 2000 .
- [41] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, London : Chapman and Hall, 1986.
- [42] "nac image technology inc. emr-9". <http://www.nac inc.com/products/Eye-Tracking-Products/EMR-9/>