自己運動と高次視覚的顕著性特徴に基づく一人称視点映像に対する視覚的注意予測技術の確立

研究代表者　滝　本　裕　則　岡山県立大学　情報工学部　助教
共同研究者　満　倉　靖　恵　慶應義塾大学　理工学部　准教授

1 はじめに

近年，カメラの小型化やウェアラブルコンピューティングの発達に伴い，Google glass 等のウェアラブルカメラが注目されている。このような頭部に装着されたカメラを用いて観察者と同じ視点から撮影された一人称視点映像は人間中心メディアとして注目されており，日常的な自己の行動を観測するのに適しているため，様々な分野への応用が期待されている。特に，一人称視点から撮影される動的なシーンを解析することにより，人の行動意図を理解し，様々な支援を行うことを目的とした研究が国内外を問わず活発に行われている[1]。

一方，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とって人の内部状態を推定することは重要な要素技術であり，その重要な手がかりとして注視情報や視覚的注意が注目されている。人の注視情報をリアルタイムで得るため，これまでに様々な視線計測技術が提案されてきたが，真に実用的な装着型の視線計測を実現するためには高価な設備と煩雑なキャリブレーションが必要であり，簡易な装置を用いて拘束の少ない環境で高精度に視線を計測することは今なお困難な課題である。そのため，視線計測とは異なるアプローチとして，一人称視点映像からの人の内部状態推定を目標とし，取得した映像から空間的な視覚的注意，いわゆる視覚的顕著性マップを推定する技術の実現が望まれている[2]。

ここで，視覚的注意について簡潔に述べる。人は，役割が異なる中心視と周辺視を用いることによって効率よく視覚情報を獲得している。中心視とは，視線を向けることで視野の中で解像度の高い中心窩と呼ばれる網膜中心部で見ることを意味する。一方，周辺視とは，中心視以外の網膜周辺部で見ることを意味する[3]。人の視野は約 200 度といわれているが，細部まで見ることのできる中心視は視野内で 1～2 度の範囲に過ぎない[4]。したがって，膨大な視覚情報の中から必要な情報の詳細を確認するためには，順々と中心視を移動させる必要がある。このとき，次に注視すべき対象箇所を検出する役割として周辺視が用いられる。そして，視線を移動させる際の高速な眼球運動はサッカードと呼ばれ，サッカード間の静止した時間を注視と呼ぶ。また，膨大な視覚情報全てを人の脳で処理することは困難であるため，網膜像の中から重要と思われる情報を選択的に収集し，処理するメカニズムが存在していると考えられている。このメカニズムのことを視覚的注意という[5]。この視覚的注意を視覚的メカニズムの初期段階で事前処理として用いることで，その後の認識や判断などのより高次な処理を簡素かつ高速に実現可能である。

視覚的顕著性マップモデルは，前述のような視覚的注意に関する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映像中の領域を推定するための計算モデルのことであり，様々な拡張や応用がなされている[6]-[13]。しかしながら，未だ一人称視点映像に特化した高精度な視覚的注意予測技術は確立されていない。ウェアラブルカメラを用いた人間中心のインタラクティブシステムを実現するためには，高次の視覚的顕著性特徴と自己運動に基づく一人称視点に対する視覚的注意推定技術の実現が急務である。

本研究では，特殊なデバイスに依存せず，実環境下にて得られる視覚情報から人の興味・意図を実時間で推定することによって人の活動を支援することを最終目的とし，高次の視覚的顕著性特徴と自己運動に基づき，一人称視点映像に対する視覚的注意推定技術の確立を図る。

2 自己運動と高次視覚的顕著性特徴に基づく一人称視点映像に対する視覚的注意予測

2-1 提案手法の概要

膨大な視覚情報全てを人の脳で処理することは困難であるため，網膜像の中から重要と思われる情報を選択的に収集し，処理するメカニズムが存在していると考えられている。このメカニズムのことを視覚的注意という[5]。視覚的注意はボトムアップ注意とトップダウン注意の 2 種類に分類され，相互に干渉しながら注意を制御している。ボトムアップ注意は色や傾きなどの外発的要因によって刺激される注意である。本研究

では，ボトムアップ注意に影響する視覚的特徴に基づく顕著性マップを実現するため，静的特徴と動的特徴それぞれに対して視覚的顕著性マップモデルを提案する。

次に，人にとって文字や人の顔は無意識下で注意を向ける重要かつ特殊なオブジェクトである。これは，人が日々の生活において経験的に習得する無意識的なトップダウンの注意機能であると考えられる。本研究では，一人称視点映像に対する視覚的注意マップの更なる高精度化を目的とし，映像からこれら無意識的なトップダウン注意に影響を及ぼすオブジェクトを自動で検出し，それら重要なオブジェクトの視覚的注意に対する寄与をモデル化する。

一方，一人称視点映像に含まれる特有の情報としてカメラ装着者の自己運動が挙げられる。この自己運動には，装着者の身体の動きに加え頭部の動きが含まれている。ここで，頭部の動きは対象を視野の中心で捉えようとする際に発生することが多い。したがって，一人称視点映像から頭部の動き推定することにより，より高精度な視覚的注意予測が可能であると考えられる。そこで，自己運動と視覚的注意の関係に基づいた自己運動注意マップを提案する。

最後に，静的･動的な視覚的顕著性に影響を及ぼすボトムアップ特徴に基づく視覚的顕著性マップに加え，無意識的なトップダウン注意に基づく顕著性マップと自己運動注意マップを統合することにより，一人称視点に特化した視覚的注意推定モデルを実現する。

2-2 ボトムアップ注意を考慮した視覚的顕著性マップ

まず，静的特徴に基づく顕著性マップモデルについて述べる。はじめに，入力画像をRGB表色系からL*a*b*表色系に変換し，各成分のベース画像 L^*，a^*，b^*を作成する。次に，輝度成分L^*に対してガボール関数を適用することで方向成分Oを作成する。ガボール関数はガウス関数と正弦波の積で定義され，第一次視野にある単純型細胞の受容野の応答特性を近似することが可能である。単純型細胞は，特定の傾きを持つ情報に対して選択的に応答する方位選択性を持ち，視覚情報を方向線分に分解している。ガボール関数のパラメータを変えることで，さまざまなサイズ，方位，スケール，位相の空間構造を表すことができ，受容野構造を系統的に表すことが可能である。

次に，各成分のガウシアンピラミッドを作成する。ここで，ガウシアンピラミッドはガウシアンフィルタによる平滑化処理とダウンサンプリングを繰り返すことで作成される。それぞれのガウシアンピラミッドはスケール$\sigma \in [0,1,\ldots,8]$の9段階のスケール画像から成るように作成する。ここで，スケール$\sigma = 0$は原画像であり，$\sigma = 8$は原画像を1/256に縮尺した大きさとなる。

次に，各成分のスケール画像に対してCenter-surround difference処理を行いFeature mapを生成する。Center-surround difference処理はガウシアンピラミッドの画素の細かいスケール（center）と画素の粗いスケール（surround）間の差分を計算する処理であり，中心周辺型の受容野の働きをモデルにしている。centerに対応するスケールはガウシアンピラミッドの$c \in [2,3,4]$，surroundに対応するスケールは$s = c + \delta,\ \delta \in [3,4]$である。ここで，それぞれの成分のFeature map $FM_k(k = L^*, a^*, b^*, O)$は次のように定義される。

$$FM_k(c,s) = \left| k(c) \ominus k(s) \right| \tag{1}$$

なお，$\ominus$は異なるスケール間の対応する画素同士の差分を意味する。次に，異なるスケールのFeature mapをAcross-scale combination処理により結合してConspicuity mapを生成する。このときIttiらのモデル[6]では正規化を行っており，この処理に時間を要している。よって，提案手法では高速化のため正規化処理を行わない。各成分のConspicuity map $CM_k(k = L^*, a^*, b^*, O)$は次のように定義される。

$$CM_k = \sum_{c=2}^{4} \sum_{s=c+3}^{c+4} FM_k(c,s) \tag{2}$$

最後に，各成分の正規化したConspicuity mapを線形結合しSaliency mapを生成する。ボトムアップ注意に影響する静的特徴に基づくSaliency map SM_sは次のように定義される．

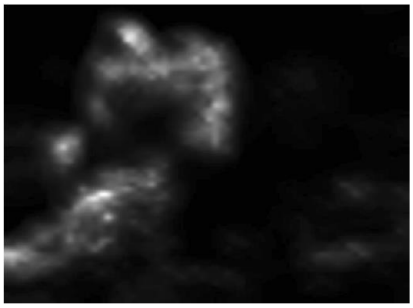
$$SM_s = \frac{1}{4} \left(\mathcal{N}(CM_{L*}) + \mathcal{N}(CM_{a*}) + \mathcal{N}(CM_{b*}) + \mathcal{N}(CM_O) \right) \tag{3}$$

ここで，$\mathcal{N}(\cdot)$は正規化処理を意味している[6]。例として，静的特徴に基づく顕著性マップモデルを図1に示す。図1において，(a)は入力フレームであり，(b)は得られる顕著性マップSM_sである。図1(b)においては，白っぽい画素ほど顕著性が高い，言い換えると人の注意を引き付けやすいことを意味している。

一方，人は視界においてユニークな動きをする領域を注視する性質がある。この動的視覚特徴に対する注

(a) 入力フレーム　　(b) 顕著性マップSM_s

図 1　静的特徴に基づく顕著性マップの出力例



(a) 入力フレーム 1 (t=0)　　(b) 入力フレーム 2 (t=1)　　(c) 顕著性マップSM_d

図 2　動的特徴に基づく顕著性マップの出力例

意をモデル化するため，Seo らの動的特徴顕著性マップモデル[13]の考え方を利用する。Seo らの計算モデルでは，映像中の輝度変化や色変化や雑音によらず，安定して顕著領域を検出することが可能である。Seo らの動的特徴顕著性マップSM_dの導出過程を以下に示す。

Step 1：静止画に対してはローカルステアリングカーネル(LSK) を用い局所領域のエッジ特徴を抽出する。一方，動画に対しては，時空間ローカルステアリングカーネル(ST-LSK) を用い時間と空間領域に対して局所領域のエッジ特徴を抽出する。

Step 2：ノンパラメトリカルカーネル密度推定によって注目局所領域とその周りの局所領域との類似度を計算し，各局所領域の顕著度を決定する。

Step 2 では，LSK もしくは ST-LSK によって抽出された特徴ベクトルからノンパラメトリックカーネル密度推定によって顕著領域を推定する。注目する特徴ベクトルとその周りの特徴ベクトルから，前述の Center-surround difference 検出と同様の処理を行う。動的特徴に基づく顕著性マップモデルの例を図 2 に示す。図 2 において，(a) と(b) は連続する入力フレームであり，(c) は両フレームから得られる動的特徴顕著性マップSM_dである。

しかし、Seo らによって提案されたモデルは固定カメラより撮影された映像に対する動的視覚的顕著性マップモデルであるため，一人称視点映像にそのまま適用した場合，本来は制止している物体であっても装着者（カメラ）自身の運動によって顕著領域として過検出されることが課題であった。そこで本研究では，カメラの自己運動を推定し，閾値処理によって一定以上の運動が発生している場合を検出し動的顕著性マップモデルの出力を抑制する。なお，自己運動推定の詳細は 2.5 節にて述べる。

2-3 トップダウン注意を考慮した視覚的顕著性マップ

ここでは，事前知識に特化したトップダウン顕著性マップについて述べる。人は普段から顔や文字に無意

識に注意を向ける傾向がある。よって，一人称視点映像に対する視覚的注意マップの更なる高精度化を目的とし，顔や文字といった無意識的なトップダウンの注意機能に影響を及ぼす対象を映像から自動で検出し，視覚的注意予測モデルに組み込むことを考える。

顔検出については Haar-like 特徴に基づくカスケード型分類器[14]を用いて行った。Haar-like 特徴量とは，局所特徴量のうち輝度に着目したものの一つである。局所特徴量とは，抽出対象物体の局所的な領域に着目した特徴量のことであり，環境変化や物体の形状変化にロバストな特徴である。Haar-like 特徴量によって顔領域の輝度による部分的な明暗のパターンと顔でない領域の明暗パターンを大量に記憶し，識別器によって探索範囲内の物体が顔か否かを判断する。

次に、検出された顔領域に顕著度を与えることを考える。ここで、人は相手の顔の大きさから対象との距離を無意識的に得ており、より距離が近い対象に注目しやすいと考えられる。よって、フレームTから検出されたN個の顔領域F_i $(i = 1,2, \dots, N)$について、顔領域の各画素(x, y)の顕著性マップ$SM_F(x, y)$は、顔領域F_i の面積S_i に応じて以下のように定義される。

$$SM_F(x,y) = \begin{cases} \alpha_i & (x,y) \in F_i \\ 0 & otherwise \end{cases} \tag{4}$$

ここで，α_iは以下のように定義される。

$$\alpha_i = \frac{S_i}{\max(S_i)} \tag{5}$$

これは、顔領域F_i の面積が大きいほど高い顕著度が与えられることを意味している。ここで，人の顔への注視は，輪郭部よりも中央部に向きやすいと考えられる。したがって，中央部になるほど顕著度が高くなるよう，顕著性マップSM_Fに対して Gaussian カーネルを畳み込む。Gaussian カーネルの分散σは文献[15]を参考にし，半値幅が中心窩の範囲とおおよそ一致するようにして求めた。

一方，シーンからの文字の検出方法として，テキストのサイズや方向，フォントなどの違いに頑健な連結成分ベース[16]の手法を用いた。これは Extremal Regions(ERs)と呼ばれる局所領域を検出し，分類器によって文字であるかを識別するアルゴリズムである。文字の候補となる領域に対し，構造木を作成し閾値処理を行うことで ERs が得られる。次に，ERs を 2 段階の識別器にかけることにより文字領域かどうか識別を行う。第 1 段階より面積や周囲長といった基本的特徴，第 2 段階より包含比や境界の変曲点の数といった高レベルな特徴で識別を行う。このようなレベルの異なった識別器を使用することにより精度の高い識別が可能である。これを入力チャンネルごとに行った後，グループ化手法[17]を用いることで単語やテキスト行の領域を検出する。

次に，検出された文字領域に対して顕著度を与えることを考える。フレームTから矩形領域として検出されたM個の文字領域C_i $(i = 1,2, \dots, M)$について、文字領域の各画素(x, y)の顕著度$SM_C(x, y)$は、文字領域C_i の長辺の長さL_i に応じて以下のように定義される。

$$SM_C(x,y) = \begin{cases} \beta_i & (x,y) \in C_i \\ 0 & otherwise \end{cases} \tag{6}$$

ここで，β_iは以下のように定義される。

$$\beta_i = \frac{L_i}{\max(L_i)} \tag{7}$$

これは、文字領域C_i が大きいほど高い顕著度が与えられることを意味している。なお，人の文字への注視は，外郭部よりも中央部に向きやすいと考えられる。したがって，顕著性マップSM_Cに対して上記 Gaussian カーネルと同じ分散を持つ Gaussian カーネルを畳み込む。

2-4 自己運動の推定

既存の視覚的顕著性マップモデルは，既に撮影済みの映像等を観察者に提示した場合の視覚的顕著度を計算するために設計されており，自己運動によって生じた視覚刺激の取り扱いは考慮されていない。そこで本研究では，一人称視点映像から自己運動の一つである頭部運動を推定し，顕著性と頭部運動に基づいた視覚的注意推定モデルを提案する。本節では，一人称視点映像からの頭部運動推定について述べる。

初めに、一人称視点映像の隣接する異なる 2 フレームに対してそれぞれ特徴点を検出する。我々は，特徴

点検出と特徴点を表す局所特徴量記述のために Accelerated KAZE 法[18]を用いた。Accelerated KAZE は KAZE[19]をベースとし，非線形かつ非等方のフィルタリングを行うことで特徴点を求める手法である。また，Accelerated KAZE は，Modified local difference binary を用いてスケール変化と回転変化に頑健な特徴量をバイナリコードにより記述する。なお、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，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。

次に，検出された特徴点に対して注目する隣接フレーム間で対応点探索を行うことによりオプティカルフローを求める。例として，Accelerated KAZE によって求めたオプティカルフローの例を図 3 に示す。

カメラからの特徴点の 3 次元空間上の点に向かう光線のベクトルを$\boldsymbol{r}$とする。前後フレームの画像の光線ベクトルをそれぞれ$\boldsymbol{r_i} = [x_i, y_i, z_i]^T, \boldsymbol{r'_i} = [x'_i, y'_i, z'_i]^T$とし，式(8)で表現されるフレーム間の位置と姿勢情報からなる基本行列 $\boldsymbol{E}$ を求める。

$$\boldsymbol{r'_i}^T \boldsymbol{E} \boldsymbol{r_i} = 0 \tag{8}$$

基本行列 $\boldsymbol{E}$ は 8 点以上の光線ベクトル対（オプティカルフロー）に対応する連立方程式を解くことで求めることが可能である。得られた基本行列 $\boldsymbol{E}$ よりカメラの回転行列 $\boldsymbol{R}$ と並進移動ベクトル $\boldsymbol{t}$ が得られる。なお，AKAZE 特徴点追跡によって得られた対応点は必ずしも全てが正しく対応しているとは限らない。誤対応点を用いて基本行列 $\boldsymbol{E}$ を求めた場合，得られるカメラの位置・姿勢の推定精度が低下する。そこで、外れ値除去として RANSAC[20]を用いた。

2-5 自己運動注意マップの生成と視覚的顕著性マップへの統合

本節では，複数提案した各顕著性マップの統合に加え，自己運動注意マップの生成と顕著性マップとの統合について述べる。まず，提案した複数の顕著性マップを 1 つの顕著性マップに統合する方法について述べる。

前節までに提案した 4 種類の視覚的顕著性マップモデルについて，L*a*b*色特徴を考慮した静的特徴顕著性マップをSM_s，動的特徴顕著性マップをSM_d，顔特徴顕著性マップをSM_f，文字特徴顕著性マップをSM_cとしたとき，統合顕著性マップSMは次の通り定義される。

(a) 入力フレーム 1 (t=0)

(b) 入力フレーム 2 (t=1)

(c) Accelerated KAZE による特徴点探索の結果

図 3　Accelerated KAZE による対応点探索例

$$SM = \max\left\{\frac{1}{\sum w}(w_1 SM_s + w_2 SM_d), SM_f, SM_c\right\} \quad (9)$$

なお，$w_i(i = 1\sim2)$はボトムアップ顕著性マップにかかる係数を表しており，その値は0~1の範囲をとるものとする。なお, 一人称視点映像には被験者の頭部回転や前進後退による自己運動によって背景移動が生じる。このため，動的特徴顕著性マップSM_dでは，本来は制止している物体であってもカメラの運動によって顕著領域として過検出される場合がある。我々は，一人称視点映像によって起こる背景の移動に対して顕著度を割り当てないために, 大きな自己運動が発生した際の動的特徴顕著性マップの統合を行わないことを考える。2.4 節で述べた方法により得られる回転行列 $\boldsymbol{R}$ と並進移動ベクトル $\boldsymbol{t}$ より，運動量がある一定の閾値Thを超えた場合は動的特徴顕著性マップをSM_dに対する係数w_2を 0 とする。

次に，推定された自己運動によって統合顕著性マップSMを制御する方法について述べる。人間の視界は水平方向約 200°に開けているが，その細部まで見ることができるのは中心視野とよばれる約 2°に限られている。周辺視野の基本的な特性についてはこれまで多くの研究が行われており, 周辺視の視力は離心角度 10 度で中心視の約 20%にまで低下することが知られている[21]。この中心視野と周辺視野の機能の違いは視細胞の分布によるものである。そこで, 相対視力特性をモデル化するため 2 次元ガウス分布を用いる。図 4 では，被験者が画像の中央を注視しているものとし，その際の 2 次元相対視力分布DVA表しており，白っぽい画素ほど相対視力が高いことを意味している。なお，相対視力特性を参考にし，ガウス分布の分散σについては離心角度 10 度の領域にて中心の約 20%となるように求めた。

一方，一人称視点映像から推定した頭部運動に基づく自己運動注意マップについて述べる。まず，得られた回転行列 $\boldsymbol{R}$ より各軸方向の角速度を求める。人の頭部の水平方向の角速度ω_x[deg./s]と垂直方向の角速度ω_y[deg./s]は以下のように定義される。

$$\omega_x = f_s \theta_x \frac{180}{\pi} \quad (10)$$

$$\omega_y = f_s \theta_y \frac{180}{\pi} \quad (11)$$

なお，f_sは映像のフレームレート[fps]である。また，θ_xとθ_yはそれぞれ一人称視点映像での水平・垂直方向周りの回転角度[rad.]である。得られた各軸方向の角速度に基づいて，2 次元ガウス分布の中心位置を変更することにより，頭部姿勢変動に応じた相対視力分布である自己運動注意マップDVAを生成する。

最後に，統合顕著性マップSMと自己運動注意マップDVAの積を求めることにより，最終的な一人称視点映像に特化した視覚的顕著性マップSM'を生成する。頭部姿勢変動に応じた自己運動注意マップDVAとの積を求めることで，相対視力が低い領域の顕著度は低く抑えることが可能であり，自己運動の影響による視覚的注意の変動をモデル化することが可能であると考えられる。

3 評価実験

提案した顕著性マップの有効性を検証するため，頭部に装着したカメラから取得した動画に対して提案手法を適用した。我々は，撮影装置として Pupil labs 社製 Mobile Eye Tracking Headset[22]を用いた。本装置は人の視線を計測するためのものであり，その外観を図 5 に示す。本装置は右目上部に World カメラが 1 つと，

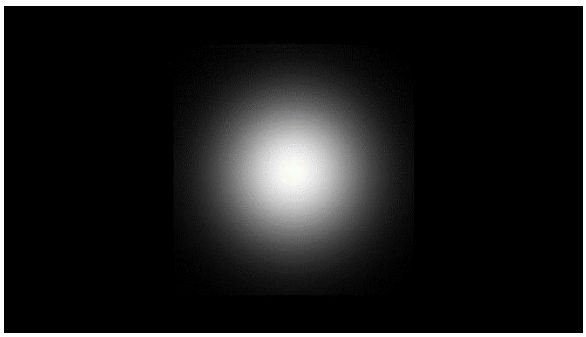


図 4　相対視力分布DVAの例

図 5　Mobile Eye Tracking Headset の外観

両目を撮影するための Eye カメラが 2 つの合計 3 つのカメラを備えている。World カメラの仕様は 1920x1080 画素，30fps，視野角は約 90 度である。一方 Eye カメラについては，赤外カメラと赤外 LED をセットで備えており，その仕様は 640x480 画素，120fps である。なお，視線は暗瞳孔検出に基づいて検出される。

まず，カメラを装着した被験者に対して動かないように指示し，野外にて撮影した連続するフレームの一部を図 6(a)と(b)に示す。また，図 6(a)と(b)に対して提案手法によって求めた各顕著性マップを図 6(c)～(h)に示す。なお，パラメータ$w_i(i = 1\sim2)$については，視覚的注意はどの刺激も均等に働いていると仮定し全て1としている。図 6(d)では，被験者の微妙な動きにより車だけではなく，本来静止しているはずの物体も顕著領域として現れている。また，図 6(f)の文字顕著性マップでは，「岡山県立大学」という 6 文字からなる 1 つの単語が 2 つの単語（領域）として分割して検出されたため，それぞれの領域に異なる顕著度が与えられていることがわかる。一般的に顔に比べて文字列はその構造が複雑であるため，文字列検出の精度向上が今後の課題である。また，文字検出後に隣接する領域は統合するなどの対策が必要である。なお，この例では被験者に動かないよう指示しているため、2 次元相対視力分布DVAは図 4 と同じである。

次に，被験者が室内で自由に移動した場合についての結果を示す。図 7(a)と(b)に撮影した連続するフレームの一部を示す。また，図 7(a)と(b)に対して提案手法によって求めた各顕著性マップを図 7(c)～(h)に示す。なお，パラメータw_iについては，動画よりある一定以上の頭部運動を検出したため，動的顕著性マップの重みを 0 としている。また，フレーム中に文字列は複数存在するものの，その文字のサイズが非常に小さいため，今回用いた文字検出法では正確に文字を検出することが困難であった。この例では，被験者は実際に首を左方向に回転させており，図 7(a)と(b)から検出した頭部運動によって制御された自己運動注意マップDVAからもその動作を確認可能である。なお，この例では文字と顔は検出されておらず，動的顕著性マップの重みw_2には 0 に設定されているため，統合顕著性マップSMは静的顕著性マップSM_dと同じになる。

最後に，図 7 の例に対して視線計測装置を用いて検出した被験者の注視の移動を可視化したものを図 8 に示す。図において，注視点を緑の丸で，注視の移動の軌跡を赤の線でそれぞれ示している。図 7(h)と図 8 を見比べると，顕著性マップSM'にて高い顕著度を持つ領域と実際の注視点の位置に若干のずれはあるものの，視線が移動しようとしているおおよその方向を顕著性マップが推定できていることを確認できた。しかしながら，今後，2 次元相対視力分布DVA，さらには，検出した自己運動によるDVAの制御を改善する必要がある。

ところで，人は頭部運動だけではなく，眼球運動も利用して視線方向を変更している。他の注視点へと視線方向を変える際,眼球と頭部は同じ速さで遷移先の注視方向へ直線的に動くといった単純な動きではなく，複雑に連動していることが示唆されている。具体的には，注視点を変更しようとする際，まず眼球が急速に注視の遷移方向に動き出す。その後，頭部は少し送れて眼球の動きを追うように，かつ，眼球よりも遅い角速度で同じ方向に動く。また，頭部が注視の遷移方向に動くにつれて眼球はその頭部運動とは逆方向の運動をする，いわゆる前庭動眼反射と呼ばれる神経制御が働くとされている[23]。本研究で提案した視覚的顕著性マップモデルでは眼球と頭部の複雑な動的性質は考慮できていない。今後，このような前庭動眼反射に関する挙動を考慮したモデルの提案が急務である。

4 おわりに

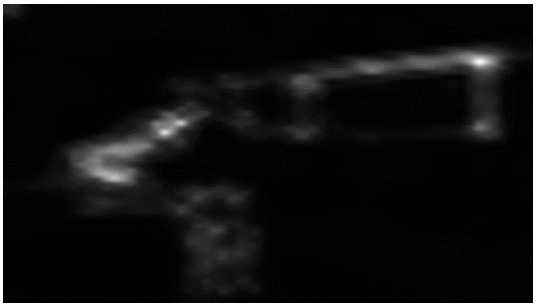
本研究では，一人称視点映像から人の興味・意図を実時間で推定することによって人の活動を支援することを目的とし，高次の視覚的顕著性特徴と自己運動を映像から抽出し，得られた複数の視覚的顕著性マップと自己運動注意マップを統合することにより，一人称視点映像に対する視覚的注意推定技術を提案した。

今後の課題として,眼球と頭部の複雑な強調動作の考慮,身体の移動方向の検出と顕著性マップへの適用,より複数のトップダウン注意に影響を及ぼす特徴のモデル化などが挙げられる。また，評価実験に関して，視線計測装置を用いて視線情報が紐付けされた大規模な映像データベースを構築し，それらを用いて有効性を検証することが挙げられる。

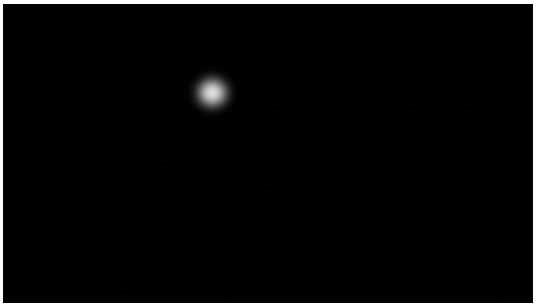
(a) 入力フレーム 1

(b) 入力フレーム 2

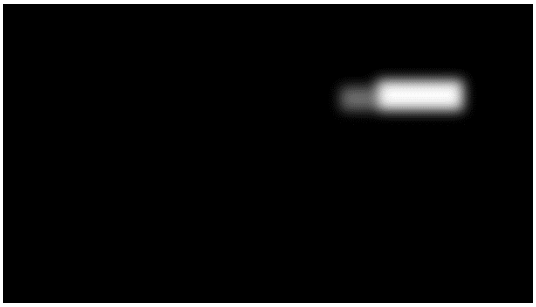
(c) 静的顕著性マップSM_s



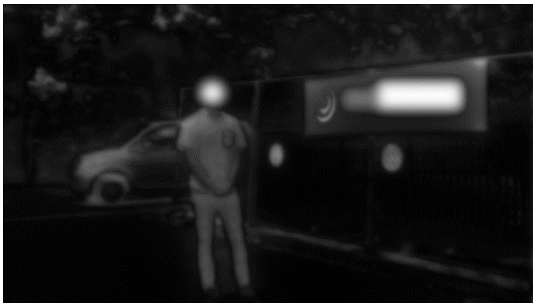
(d) 動的顕著性マップSM_d



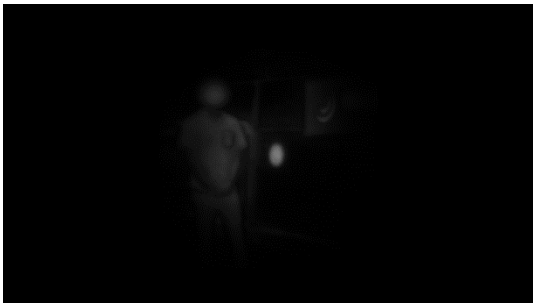
(e) 顔顕著性マップSM_f



(f) 文字顕著性マップSM_c

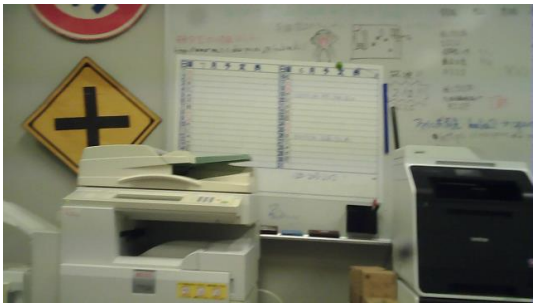


(g) 統合顕著性マップSM



(h) 一人称視点に特化した顕著性マップSM'

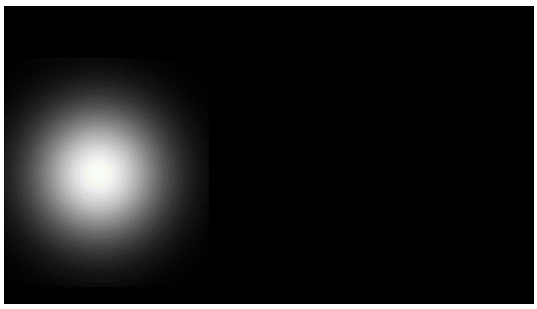
図 6　提案手法による視覚的顕著性推定の結果例 1



(a) 入力フレーム 1

(b) 入力フレーム 2

(c) 静的顕著性マップSM_s



(d) 相対視力分布DVA

(e) 一人称視点に特化した顕著性マップSM'

図 7　提案手法による視覚的顕著性推定の結果例 2

図 8　図 7 の例に対する注視点とその移動

【参考文献】

[1] S. Mann, K. Kitani, Y. J. Lee, M. S. Ryoo, and A. Fathi: ``An Introduction to the 3rd Workshop on Egocentric (First-Person) Vision'', IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014.

[2] A.Kimura, R.Yonetani, and T.Hirayama: ``Computational models of human visual attention and their implementations: A survey'', IEICE Transactions on Information and Systems, Vol. 96, No. 3, pp. 562‐578, 2013.

[3] J. Shen, E. M. Reingold, and M. Pomplun: ``Distractor ratio influences patterns of eye movements during visual search'', Perception, Vol. 29, pp. 241‐250, 2000.

[4] S. E. Palmer: ``Vision science: Photons to phenomenology'', MIT Press, 1999

[5] 原口 健, 岡崎 克典: ``視覚探索における誘目性の定量化", 日本視覚学会誌 VISION, Vol. 23, No. 1, pp. 1‐18, 2011.

[6] L. Itti, C.Koch, and E.Niebur: ``A model of saliency-based visual attention for rapid scene analysis'', IEEE Trans. on PAMI, Vol. 20, No. 11, pp. 1254‐1259, 1998.

[7] A. Hagiwara, A. Sugimoto, and K. Kawamoto: ``Saliency-based image editing for guiding visual attention'', Proc. of 6th International Workshop on Pervasive Eye Tracking and Mobile Eye-based Interaction 2011, pp. 43‐48, 2011.

[8] T. Shi and A.Sugimoto: ``Video saliency modulation in the HSI color space for drawing gaze'', Proc. of the The Pacific-Rim Symposium on Image and Video Technology 2013, pp. 206‐219, 2013.

[9] T. Kokui, H.Takimoto, H.Yamauchi, M.Kishihara, and K.Okubo: ``Image modification based on bottom-up saliency map for directing user's gaze'', Proc. of 2014 RISP International workshop on NCSP'14, pp. 637‐640, 2014.

[10] H. Takimoto, S. Hitomi, H. Yamauchi, M. Kishihara, and K, Okubo: ``Image Modification Based on Spatial Frequency Components for Visual Attention Retargeting'', IEICE Transactions on Information and Systems, Vol. E100-D, No. 6, pp.1339-1349, 2017.

[11] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk: ``Frequencytuned salient region detection'', Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1597‐1604, 2009.

[12] Y. Chuang, L. Chen, G. Chen, and J. Woodward: ``Isophote Based Center-Surround Contrast Computation for Image Saliency Detection'', IEICE Transactions on Information and Systems, Vol. E97-D, No. 1, pp. 160‐163, 2014.

[13] H. J. Seo and P. Milanfar: ``Static and space-time visual saliency detection by self-resemblance'', Journal of Vision, Vol. 9, No. 12, pp. 1‐27, 2009.

[14] R. Lienhart and J. Maydt: ``An Extended Set of Haar-like Features for Rapid Object Detection", Proceedings of the 2002 IEEE International Conference on Image Processing, Vol. 1, pp. 900‐903, 2002.

[15] S. Marat, A. Rahman, D. Pellerin, N. Guyader, and D. Houzet: ``Improving visual saliency by adding 'face feature map'and 'center bias' Cognitive Computation", Vol. 5, No. 1, pp. 63‐75, 2013.

[16] L. Neumann and J. Matas: "Real-Time Scene Text Localization and Recognition'', Proceedings of the 25th IEEE Conference on Computer Virsion and Pattern Recognition, pp. 16‐21, 2012.

[17] L. Gomez and D. Karatzas: ``Multi-script Text Extraction from Natural Scenes'', Document Analysis and Recognition (ICDAR), pp. 467‐471, 2013.

[18] P. F. Alcantarilla, J. Nuevo, and A. Bartoli: ``Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces", In British Machine Vision Conference, 2013.

[19] P. F. Alcantarilla, A. Bartoli, and A. J. Davison: ``KAZE features'', In Eur. Conf. on Computer Vision (ECCV), pp. 214‐227, 2012.

[20] 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.

[21] 福島 邦彦: ``視覚の生理とバイオニクス", 電子通信学会, 1976.

[22] https://pupil-labs.com/

[23] 沖中 大和, 満上 育久, 八木 康史: ``人の眼球と頭部の協調運動を考慮した視線推定", 研究報告コンピュータビジョンとイメージメディア（CVIM）, Vol. 2016-CVIM-202, No. 18, pp. 1‐8, 2016.

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
自己運動を考慮した一人称視点映像のための視覚的顕著性マップ	2016 年度 瀬戸内合同信号処理研究会	2016 年 9 月
目立ち度の見える化技術と福祉・デザイン支援への応用	第 21 回岡山リサーチパーク研究・展示発表会	2017 年 1 月
目立ち度の見える化技術と福祉・デザイン支援への応用	平成 28 年度岡山県立研究機関協議会交流発表会	2017 年 2 月
Guiding visual attention based on visual saliency map with projector-camera system	The 19th International Conference on Human-Computer Interaction (HCII2017)	2017 年 7 月 (Accepted)