

# 修 士 論 文

## 自己運動と顕著性に基づく 一人称視点における視覚的注意予測

Predicting egocentric visual attention using low-level  
saliency map and egomotion estimation

2011 年 2 月 9 日提出

指導教員 佐藤 洋一 教授

東京大学大学院

情報理工学系研究科 電子情報学専攻

096427 山田 健太郎

# 内容梗概

画像・映像のみから視覚的注意を予測する手法として、視覚的顕著性マップモデルが近年注目を集めている。視覚的顕著性マップモデルは、コンピュータビジョン、画像処理、マルチメディア等の分野で応用されており、今後も幅広い分野で重要性が高まると期待される。しかし、頭部装着型カメラ等の一人称視点から撮影された映像に対しては、これまで視覚的顕著性マップモデルは用いられておらず、有用性も確認されていなかった。一人称視点映像は日常的な自己の行動を分析するのに非常に適しているため、行動意図の理解や、行動支援等を目的とした研究での利用が盛んである。こうした一人称視点映像を解析する上で、どこに注意が向けられていたのかという情報は、非常に重要なファクターとなる。そこで本研究では、一人称視点映像に対する視覚的顕著性マップモデルの有効性を実験的に検討し、さらに一人称視点映像に適応させるために視覚的顕著性マップモデルを拡張することを目的とする。

視覚的顕著性マップモデルを一人称視点映像に用いる場合の問題としては、自己運動の影響が考慮されていないことが予想された。特に、人間視覚の機能である、自己運動を補償するメカニズムが組み込まれていないことは重要な観点の欠如である。この予想を検証するため、視覚的顕著性マップモデルと一人称視点映像の視線データを比較する予備実験を行ったところ、やはり動的特徴によるマップが性能を低下させてしまうことが確認された。また同時に、静的な特徴を用いて計算されたマップは、一人称視点において視覚的注意推定を行う場合でも、有用であることも確認された。人間の視覚においては自己運動に伴う運動視覚刺激は抑制され注意を誘導しないが、動的特徴による顕著性マップはこの種の刺激に対して顕著度を付与してしまうことが、性能低下の原因であった。このように、顕著性マップにおいては、自己運動による視覚刺激が性能低下の原因となった。しかし、自己運動は視覚的注意と密接に関係しており、効果的に用いることで視覚的注意推定に役立てることも可能だと考えられる。特に、自己運動方向に対する知覚に、視覚が重要な役割を果たしていることは重要な知見である。そこで提案手法では、一人称視点映像か

---

ら推定した自己運動に基づき，自己運動方向マップを生成し，静的特徴による顕著性マップと統合することにより，一人称視点に適した視覚的注意推定を実現する．また，顕著性マップによる一人称視点での視覚的注意推定と比較し，提案手法は性能が向上しており，自己運動方向マップが注意予測に有用であることを実験により検証した．

# 目次

|       |                                     |    |
|-------|-------------------------------------|----|
| 第 1 章 | 序論                                  | 1  |
| 1.1   | 研究背景 . . . . .                      | 1  |
| 1.2   | 研究の目的とアプローチ . . . . .               | 2  |
| 1.3   | 本論文の構成 . . . . .                    | 3  |
| 第 2 章 | 関連研究                                | 5  |
| 2.1   | はじめに . . . . .                      | 5  |
| 2.2   | 視覚的注意と特徴統合理論 . . . . .              | 5  |
| 2.2.1 | 視覚的注意 . . . . .                     | 5  |
| 2.2.2 | 特徴統合理論 . . . . .                    | 6  |
| 2.3   | 視覚的顕著性マップモデル . . . . .              | 9  |
| 2.3.1 | 顕著性マップ . . . . .                    | 9  |
| 2.3.2 | 視覚的顕著性マップモデルの有用性評価 . . . . .        | 9  |
| 2.3.3 | 視覚的顕著性マップモデルの拡張 . . . . .           | 9  |
| 2.3.4 | 視覚的顕著性マップモデルを利用したアプリケーション . . . . . | 10 |
| 2.4   | 一人称視点映像と自己運動 . . . . .              | 14 |
| 2.4.1 | はじめに . . . . .                      | 14 |
| 2.4.2 | 一人称視点映像と視覚的顕著性マップモデル . . . . .      | 15 |
| 2.4.3 | 自己運動によるオプティカルフローと視覚的注意 . . . . .    | 16 |
| 2.5   | 本研究の位置づけ . . . . .                  | 16 |
| 第 3 章 | 視覚的顕著性マップモデルの一人称視点における課題            | 18 |
| 3.1   | はじめに . . . . .                      | 18 |
| 3.2   | 動画像の顕著性マップ生成の概要 . . . . .           | 19 |



|       |                                 |    |
|-------|---------------------------------|----|
| 3.3   | 被験者実験 . . . . .                 | 21 |
| 3.3.1 | 被験者実験の手順 . . . . .              | 21 |
| 3.3.2 | 被験者実験の結果 . . . . .              | 22 |
| 3.3.3 | 考察 . . . . .                    | 23 |
| 3.4   | 本章のまとめ . . . . .                | 24 |
| 第 4 章 | 自己運動と顕著性に基づく視覚的注意推定             | 29 |
| 4.1   | はじめに . . . . .                  | 29 |
| 4.2   | 自己運動と顕著性に基づく視覚的注意の推定 . . . . .  | 31 |
| 4.2.1 | 自己運動パラメータからの前方運動判定 . . . . .    | 31 |
| 4.2.2 | FOE の推定 . . . . .               | 32 |
| 4.2.3 | 複数フレームからの FOE マップ生成 . . . . .   | 34 |
| 4.2.4 | FOE マップと静的顕著性マップの統合による注意推定マップ生成 | 35 |
| 4.3   | 評価実験 . . . . .                  | 37 |
| 4.4   | 考察 . . . . .                    | 40 |
| 第 5 章 | 結論                              | 41 |
| 5.1   | 本論文のまとめ . . . . .               | 41 |
| 5.2   | 今後の課題と展望 . . . . .              | 42 |
| 付録 A  | 自己運動の推定                         | 44 |
| A.1   | エピポーラ幾何 . . . . .               | 44 |
| A.2   | 運動パラメータの推定 . . . . .            | 46 |
| 参考文献  |                                 | 48 |
| 発表文献  |                                 | 55 |

# 目次

|     |   |    |
|-----|---|----|
| 2.1 | 視覚的探索 . . . . .   | 7  |
| 2.2 | 特徴統合理論の概念図 . . . . .  | 8  |
| 2.3 | 顔領域検出を用いた視覚的顕著性マップモデル [1] . . . . .   | 11 |
| 2.4 | Seam Carving による, Content-Aware なリサイズ [2] . . . . .                           | 12 |
| 2.5 | 顕著性マップを用いた自動トリミングの概要 [3] . . . . .  | 13 |
| 2.6 | サムネイル画像作成のためのオンライン顕著領域検出 [4] . . . . .  | 14 |
| 2.7 | 顕著性マップを利用したパノラマ画像からの高品質な構図の領域切り出し [5] . . . . .                               | 15 |
| 2.8 | オブ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) による進行方向知覚 . . . . . | 16 |
| 3.1 | 動画の顕著性マップ生成処理 . . . . .   | 19 |
| 3.2 | 被験者実験環境 . . . . .   | 25 |
| 3.3 | ROC 曲線 (視線推定性能曲線) . . . . .   | 26 |
| 3.4 | 運動物体に注意が向けられたシーン . . . . .  | 27 |
| 3.5 | 輝度変化に注意が向けられたシーン . . . . .  | 27 |
| 3.6 | 自己運動が動的特徴による顕著性マップに与える影響 . . . . .  | 28 |
| 4.1 | FOE とオブティカルフローの成分 . . . . .   | 30 |
| 4.2 | 自己運動と顕著性に基づく視覚的注意の推定 . . . . .  | 32 |
| 4.3 | オブティカルフローからの並進成分の分離 . . . . .   | 34 |
| 4.4 | FOE とエピポール . . . . .  | 35 |
| 4.5 | 複数フレームからの FOE マップ生成 . . . . .   | 36 |
| 4.6 | FOE マップと静的顕著性マップの統合による注意推定マップ生成例 . . . . .                                    | 36 |

|     |                                 |    |
|-----|---------------------------------|----|
| 4.7 | 提案手法の性能評価 (Itti らのモデル [6] を利用)  | 38 |
| 4.8 | 提案手法の性能評価 (Harel らのモデル [7] を利用) | 39 |
| A.1 | エピーポーラ幾何とエピーポーラ平面内の 3 つのベクトル    | 45 |

# 表目次

|     |   |    |
|-----|---|----|
| 3.1 | 特徴を統合した顕著性マップの性能評価 (AUC, 曲線下面積) . . . . .                         | 23 |
| 4.1 | Itti ら [6] の視覚的顕著性マップモデルとそれを用いた提案手法との性能比較 (AUC, 曲線下面積) . . . . .  | 37 |
| 4.2 | Harel ら [7] の視覚的顕著性マップモデルとそれを用いた提案手法との性能比較 (AUC, 曲線下面積) . . . . . | 37 |

# 第 1 章

## 序論

### 1.1 研究背景

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，一人称視点から撮影されるダイナミックなシーンを分析することで，行動意図を理解し，支援を行うこと等を目的とした研究も行われている．このような目的のためにも，一人称視点映像中のどこに注意が向けられているのか，という情報は非常に有益である．

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 [8]．しかし，眼球運動による自由な視線移動が可能な状態では，視線方向の周辺以外の領域に注意を払いつづけることが被験者実験により明らかにされている [9, 10]．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．それゆえ，多くの視線計測技術が，一人称視点の環境に限らず，様々なアプローチに基づいて開発されてきた [11]．しかしながら，多くの手法では，非常に高価な設備と，煩雑なキャリブレーション等が必要であり，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．画像だけを入力とし，画像中の重要な領域を推定するのに役立つため，近年注目を集めている．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 [12] を踏まえ，Kock と Ullman によって提唱された [13]．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 [14]．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 [1, 6, 7, 15–17]．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 [6]．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 [18–20]．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．

一人称視点においても，視覚的顕著性マップモデルを用いることで，一人称視点映像のみを入力とした視覚的注意の正確な推定が可能であれば，教育や医療，福祉における行動支援等，様々な分野での応用が期待できる．しかしながら，人間中心の視覚における視覚的注意のメカニズムは大幅に異なる．たとえば，自己運動によって生じた視覚刺激は，人間の視覚システムにおいては抑制されるが [21]，既存の視覚的顕著性マップモデルでは，自己運動によって生じる視覚的運動刺激といった重要な観点が考慮されていない．視覚的顕著性マップモデルでこの抑制メカニズムは明示的には組み込まれていないことは，一人称視点においては問題になると推察される．したがって，一人称視点映像から視覚的注意の推定を行う上で，既存の視覚的顕著性マップモデルには十分な改良の余地があることが予測される．

## 1.2 研究の目的とアプローチ

本論文では，一人称視点映像から，カメラを装着してその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．画像・映像から視覚的注意を推定する手法としては，前節のように視覚的顕著性マップモデルが広く応用されている．しかし，視覚的顕著性マップモデルは，これまで一人称視点映像における性能評価が行われていないため，一人称視点における視覚的注意の推定に有効であるかどうか明らかではない．顕著性マップの性

能評価を行った既存の研究では、既に撮影済みの画像や映像を提示した際の注意のみが検討されてきた。具体的には、ディスプレイ等を用いて被験者に提示された画像・映像から顕著性マップが計算され、ディスプレイ上の視線座標とマップの比較により、顕著性マップの性能が評価された。つまり、既存の研究において、視覚的顕著性マップモデルによる観察者の注意の推定や視線の計測は、提示された画像・映像に限られており、より自由な環境、すなわち、実空間の物体や周辺環境を観察するような、一人称視点における視覚的注意と顕著性マップの相関は確認されていない。また、上述のように、既存の視覚的顕著性マップモデルでは、自己運動によって生じる視覚的運動刺激が考慮されていないため、動的特徴により生成されるマップに問題が生じることが予想された。

我々はまず、一人称視点における視覚的顕著性マップモデルの性能を予備実験的に評価し、既存の視覚的顕著性マップモデルによる視覚的注意推定の問題点を明らかにした。一人称視点における既存の視覚的顕著性マップモデルの性能評価では、一人称視点映像に対する視線計測結果とその顕著性マップの比較を行った。被験者実験では、頭部に装着したカメラを用いて一人称視点映像を撮影し、同時に、頭部装着型の視線計測装置を用いて視線計測を行ったデータを取得した。また、複数の視覚的顕著性マップモデルと特徴の組み合わせについて定量的に性能の評価・比較を行い、一人称視点での視覚的注意の特性についての議論を行った。その結果、静的特徴と比較して動的特徴の性能が低く、自己運動の影響によって性能低下が生じることを確認した。しかし、同時に、静的な特徴を用いて計算されたマップは、一人称視点において視覚的注意推定を行う場合でも、有用であることもわかった。また、確かに自己運動による視覚刺激は顕著性マップの性能低下の原因となったが、自己運動は視覚的注意と密接に関係しており、効果的に用いることで視覚的注意推定に役立てることが可能である。そこで我々は、自己運動を考慮し、自己運動と視覚的注意の関係に基いた自己運動方向マップを生成し、静的な特徴による顕著性マップと統合することで、視覚的注意を推定する手法を提案する。また、評価実験により、自己運動に基づいて生成した自己運動方向マップが注意推定に有用であることを示す。本研究は、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を行った点と、一人称視点映像のみから視覚的注意推定を行うための手法を提案した点において、我々の知る限り初めての研究である。

### 1.3 本論文の構成

本論文の構成を以下に示す。

- 第 1 章 序論
- 第 2 章 関連研究
- 第 3 章 視覚的顕著性マップモデルの一人称視点における課題
- 第 4 章 自己運動と顕著性に基づく視覚的注意推定
- 第 5 章 結論

第 1 章では、研究背景について述べ、本研究の目的を明らかにした。第 2 章では、本研究の関連研究として、まず視覚的顕著性マップモデルの背景にある視覚的注意について説明した後、視覚的顕著性マップモデルに関する研究と、一人称視点映像について説明する。また、先行研究の問題点を明らかにすることで、本研究の位置づけを明確にする。第 3 章では、本研究の予備実験として、視覚的顕著性マップモデルを一人称視点において用いた場合の性能評価を行い、視覚的顕著性マップモデルの問題点を明らかにする。さらに、一人称視点における視覚的注意推定における、低レベル特徴に基づく顕著性マップの有用性について検討する。第 4 章において、自己運動を考慮した視覚的注意推定手法を提案し、一人称視点映像に対して性能評価を行う。最後に第 5 章において本研究の成果と今後の課題・展望について述べる。



## 第 2 章

# 関連研究

### 2.1 はじめに

本章では，本研究の関連研究として，まず視覚的顕著性マップモデルの背景にある視覚的注意について説明した後，視覚的顕著性マップモデルに関する研究と，一人称視点映像について説明する．視覚的注意については，視覚的注意の役割とその種類について説明した後，視覚的注意と視線の関係に関する研究を紹介する．また，視覚的注意のメカニズムを理解する鍵となる理論として，Treisman らによって提案された特徴統合理論を紹介する．視覚的顕著性マップモデルに関しては，顕著性マップの概念や既存の複数の計算モデルの概要，そして，顕著性マップを利用したアプリケーションについて簡潔にまとめる．また，一人称視点映像に関しては，期待される応用領域や，現在の課題について説明する．

### 2.2 視覚的注意と特徴統合理論

#### 2.2.1 視覚的注意

認知心理学的な研究においては，人間を一種の情報処理機構と見なしている．人間の情報処理能力には限界があるため，大量の情報を同時に処理することは不可能である．そこで，情報の重要な部分を効率的に選択し，集中するために注意というメカニズムが必要とされる [22]．視覚システムに関しても同じことが言え，視覚情報処理における注意を視覚的注意と呼び，人間が目を通して受け取る大量の情報を扱うために必要となる．この視覚的注意には，ボトムアップなものと，トップダウンなものが存在するとされている．人間の視覚システムは初期処理において，視覚から取り入れた画像情報を，形，色，運動など

の単純な特徴に分解した後にパターン認識処理を行っていることが知られている [23] . この , パターン認識の前に , 形 , 色 , 運動などの単純な物理特徴に基づいて働く注意をボトムアップな注意と呼び , パターン認識の後に知識や予測に基づいて働く注意をトップダウンな注意と呼ぶ .

顕著性というキーワードは , ボトムアップな注意にも , トップダウンな注意においても用いられる . ボトムアップな注意に基づいた顕著性は視覚的顕著性マップモデルという 2 次元マップとして計算モデル化されており , 次節で詳しく述べる . それに対しトップダウンな注意 , 物体認識の後にその意味に基づいた注意が向けられるため , 顕著性は物体に与えられ , 顕著物体 (salient object) 等と呼ばれる . トップダウンな注意に基づいたアプローチで顕著物体領域を得るためには , 高度な物体認識を必要とし , 人間ではなくコンピュータによってそれを行う場合 , 非常に困難な問題である . それに対し , ボトムアップな注意は色 , 方向 , 輝度等の低レベルな特徴をもとに働くものであるため , コンピュータでも扱いやすく , コンピュータビジョン・画像処理で広く応用される要因となっている . また , ボトムアップな顕著性マップに , 低レベルなコンテキスト情報といったトップダウンな注意に基づいた要素を取り込んだモデルも , 近年特に用いられている [1, 24, 25] .

### 2.2.2 特徴統合理論

前節で述べたとおり , 人間の視覚系は , 取り入れた画像情報を形 , 色 , 運動などの単純な特徴に分解して処理していることが知られている . 一方で , 我々が知覚している物体は形や色などの複数の特徴を同時に併せ持っている . したがって , 視覚系は , 分解して別々に処理した単純な特徴をその後再び統合していると考えられる . Treisman らは , この特徴の統合が注意によってなされているという考えに基づき , 特徴統合理論 [12, 26, 27] を提唱した . 特徴統合理論は , ボトムアップな注意の働きを明確にモデル化したものである .

特徴統合理論は , 視覚的探索の実験結果に基づいて考えられたモデルである . 視覚的探索実験とは被験者に様々な妨害刺激 (distractor) の中から特定の目標刺激 (target) を探させ , それに要する時間を測定する実験である . 図 2.1 のような刺激を用いて実験を行った場合について説明する . 図 2.1 (a) の中から他と向きの異なる長方形を探した場合や , 図 2.1 (b) の中から他と色の異なる長方形を探し出す場合には , 全体の長方形の個数に関係なく一目でそれを発見することができる . それに対し図 2.1 (c) において , 横長赤色の長方形と縦長青色の長方形の妨害刺激の中から , 横長青色の長方形を探すのは妨害刺激の数に応じて困難になり , 探索反応時間は妨害刺激の数に比例して増加する . 図 2.1 の

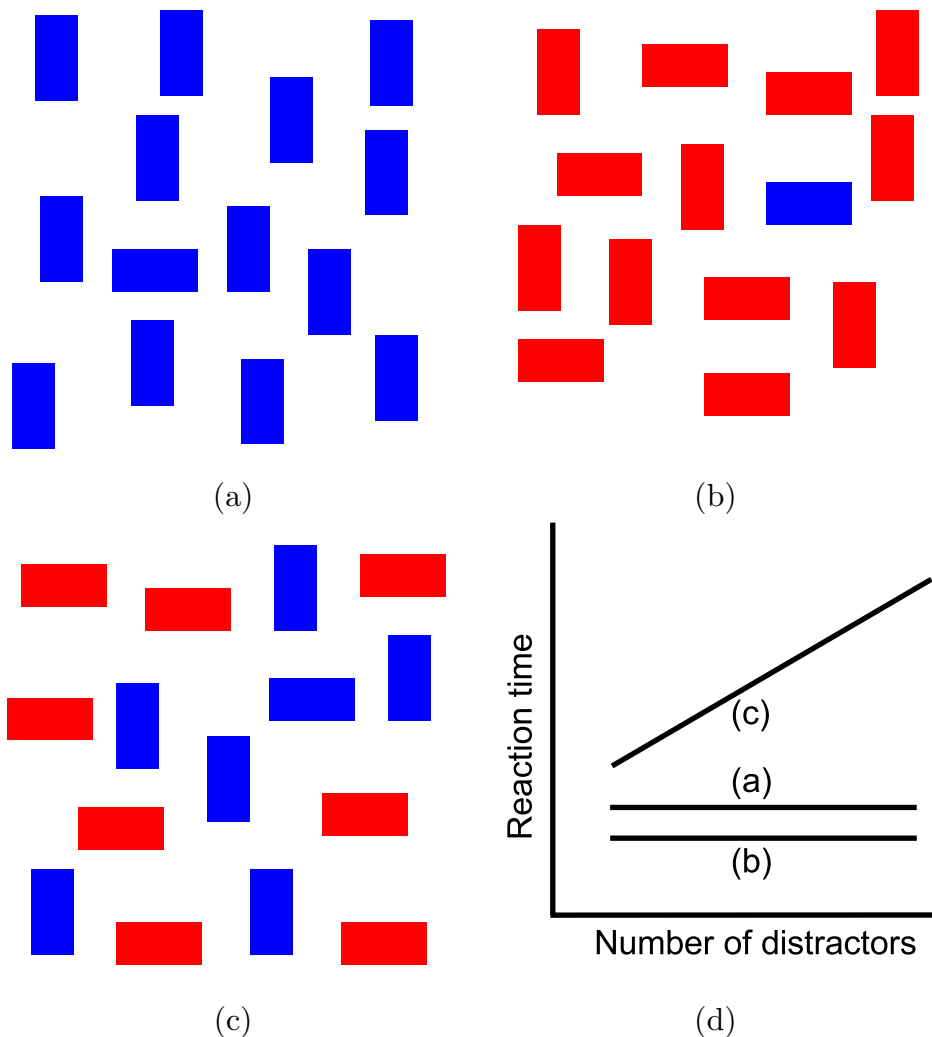


図 2.1: 視覚的探索実験に用いられる視覚刺激の例。(a)の中から他と向きの異なる長方形を探した場合や、(b)の中から他と色の異なる長方形を探し出す場合には、全体の長方形の個数に関係なく一目でそれを発見することができる。それに対し(c)において、横長赤色の長方形と縦長青色の長方形の妨害刺激の中から、横長青色の長方形を探すのは妨害刺激の数に応じて困難になり、探索反応時間は妨害刺激の数に比例して増加する。(a),(b),(c)の刺激を用いて視覚的探索実験を行った場合に得られる結果の概念図は(d)のようになる。

(a),(b),(c)の刺激を用いて視覚的探索実験を行った場合に得られる結果の概念図は図 2.1 (d) のようになる。このような視覚的探索実験の結果から Treisman らは、色、形、明る

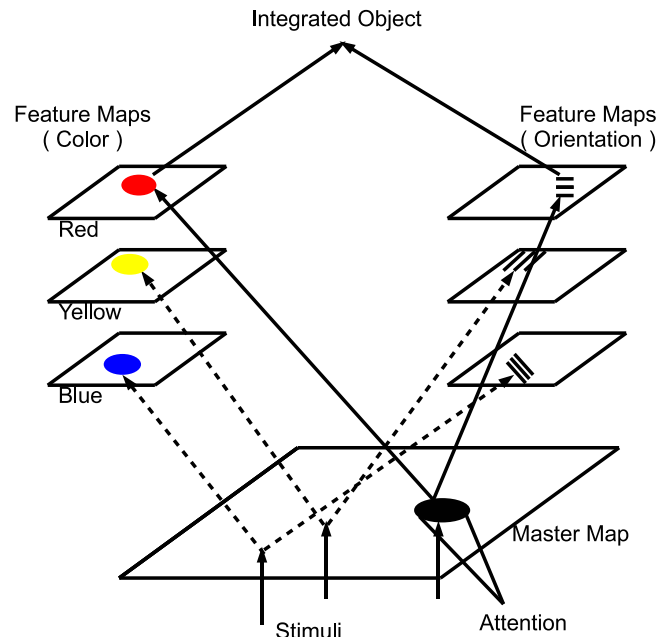


図 2.2: 特徴統合理論の概念図．視覚系は色や傾きなど，画像の基本的な特徴をそれぞれ空間的に並列に抽出し，それらの特徴ごとにその値を記録した特徴マップ群を形成する．ここでは，例として色と方向の特徴マップ群を示す．

さなどは空間的に並列に処理できるが，それらを組み合わせたものは一度に一箇所ではしか処理できない逐次的なものであると考えた．彼らは，並列的に処理できる特徴をより基本的な特徴と考え，単純特徴と呼び，それらの組み合わせを結合特徴と呼んだ．そして，結合特徴の処理が逐次的になるのは，単純特徴同士を結合する際にその物体に注意を集中する必要があるためと考えた．このような考えをもとに彼らが提案した注意モデルが特徴統合理論である．この理論では，視覚系は色や方向など，画像の基本的な特徴をそれぞれ空間的に並列に抽出し，それらの特徴ごとにその値を記録した特徴マップ群を形成する（図 2.2）．これらの特徴マップは，位置の情報のみを持つ位置マップと空間的な対応関係を保って繋がっている．観察者は画像上の特定の位置に注意を向けることで，位置マップに連結した各特徴マップ上の対応する特徴値を結合する．特徴を正確に結合するためには，注意を狭くする必要があるため，特徴の結合処理は逐次的になる．

## 2.3 視覚的顕著性マップモデル

### 2.3.1 顕著性マップ

視覚的顕著性マップモデルの分野では,Itti ,Koch らにより最初に提案された Itti-Koch モデルをベースとしたものが支配的である。まず,Koch ら [13] は,特徴統合理論 [12] をもとに,単純な特徴が注意を誘導する性質をそれぞれ独立に持っており,それを統合することで,ボトムアップな注意を誘導する値(顕著度)が得られると考え,それを符号化して 2 次元マップ上に配置する視覚的顕著性マップモデルの概念を提案した。これを,完全な計算モデルとして実装したのが Itti ら [14] である。実際の画像において視覚的注意を引く領域を計算することができるため,視覚心理の分野のみならず,コンピュータビジョン・画像処理の分野においても基礎的なモデルとして広く用いられている。

### 2.3.2 視覚的顕著性マップモデルの有用性評価

Itti らの研究においても被験者実験により,定性的な評価がなされているが,注意と密接な関係のある注視を計測した研究においても,自然情景で人が注視する位置の有用な指標となると示された [20]。

また,動的情景においても,離心率が大きくなるにつれコントラスト感度が減少する網膜をシミュレートするなどの現実的特徴を加えれば,偶然以上の予測を行うことができると報告されている [19]。

しかし,顕著性マップは注視の確率分布を予測しているものの,注視の順序を正確には予測できていないという実験結果も得られている [18]。

また,Elazary ら [28] は,顕著性マップ上の非常に顕著度が高い領域は,手作業によるラベル付けにより,トップダウンに注意を引きつけるとラベル付けされたオブジェクトの領域であることが多いという結果をもとに,トップダウンな注意を引きつける領域についても,ボトムアップな注意から非常に大きな影響を受けていると主張している。

### 2.3.3 視覚的顕著性マップモデルの拡張

Itti-Koch の視覚的顕著性マップモデルを拡張・改良したものには,主に,ボトムアップな注意の特徴マップを拡張したもの,トップダウンな注意のアプローチを加えたものな

どがある。

Itti らによるオリジナルの視覚的顕著性マップモデルでは特徴マップは色，輝度，方向の 3 種類のみが用いられていたが，他にもいくつかのボトムアップな注意に基づく特徴マップが提案されている．応用面での有用性が高い特徴マップとしては，映像に対応するための，輝度変化特徴，運動方向特徴といった動的特徴マップ [6] が挙げられる．

また，異なるアプローチにより，顕著性マップの高精度化を目指したモデルも提案されている．たとえば，近年グラフベースのアプローチによるモデルがいくつか提案されている [7, 16, 17]．入力画像のグラフはノード間の相違度関数と距離関数によって定義され，顕著度はグラフの定常状態を解析することで計算される．このようなグラフベースのアプローチをとった研究では，用いられる画像特徴の性能の向上よりも，単純な画像特徴を用いて，より高精度な顕著度を計算することに重点が置かれていた．

また，Itti-Koch の視覚的顕著性マップモデルは，ボトムアップな注意をモデル化していたが，最近の応用例ではこれにトップダウンな注意の要素を加えたモデルを用いたものが多い．Cerf ら [1] は，人間にとって人の顔は非常に顕著度が高いということから，画像中の顔領域を検出して，それを顕著性マップに足し合わせたモデルを提案している (図 2.3)．

また，上述のように単純な特徴マップの足し合わせを行うのではなく，それらの特徴を機械学習に用いて顕著性マップを生成するモデルも提案されている．Judd ら [25] は約 1000 枚の画像・15 人の被験者に対して視線計測を行い，アイトラッキングデータのデータベースを構築した．彼らは，その大量のアイトラッキングデータから，実際に注視されやすかった点と，注視されにくかった点をそれぞれサンプリングし，各点における各種特徴量を用いて SVM による学習を行った．これにより，実際の注視を反映した，より精度の高い顕著性マップを生成するモデルを提案している．彼らは機械学習のための特徴量として，輝度，色，方向の低レベルな特徴に加えて，地平線検出器，顔検出器，人物検出器によって得られる，中レベル，高レベルな特徴も用いている．

#### 2.3.4 視覚的顕著性マップモデルを利用したアプリケーション

視覚的顕著性マップモデルは，人間の注意のメカニズムを計算モデルとして実装している．そのため応用技術としては，人間の代わりにコンピュータに自動的に選択を行わせる，人間の興味を引くものを提示する，といった技術に用いられる傾向がある．本節では，具体的な応用例を紹介する．

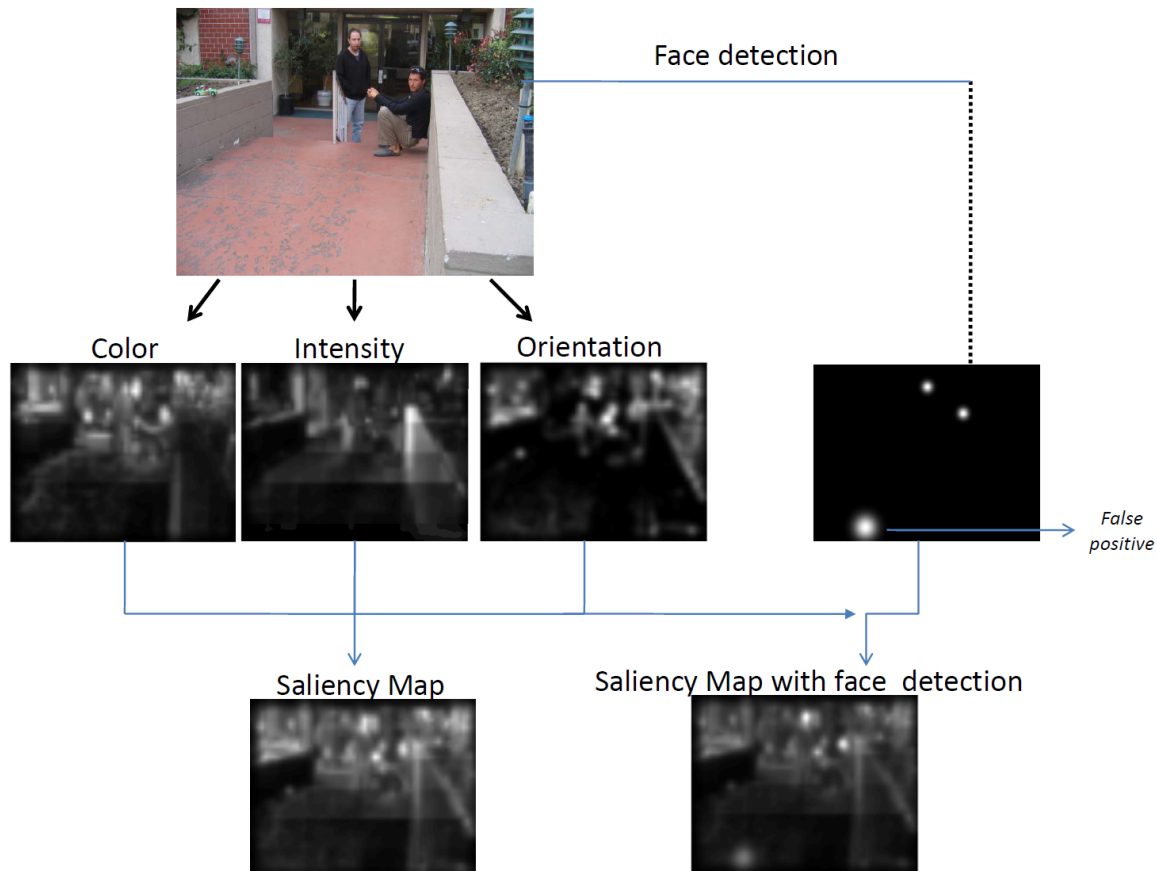


図 2.3: 顔領域検出を用いた視覚的顕著性マップモデル [1]。既存のボトムアップ注意に基づく顕著性マップに、顔領域という、トップダウンな注意による顕著な特徴を組み合わせている。

画像中の前景領域と背景領域を分割するセグメンテーションは、古くからコンピュータビジョン分野での課題として研究されてきた。この問題に対し、視覚的顕著性マップモデルを用いた研究がいくつかなされている [29–31]。人間が注意を向けやすい領域は前景領域である可能性が高く、注意が向きにくい領域は背景領域である可能性が高いといった考えを用いて、視覚的顕著性マップモデルと、グラフカットを組み合わせたり [29, 31]、GAC を組み合わせる [30] といった研究が行われている。このうち、グラフカットは高精度なセグメンテーションができるために近年よく用いられる技術であるが、典型的な前景/背景領域を与える必要がある。インタラクティブグラフカットなどでは人手によりそれが与えられていたが、顕著性マップの利用でそれを自動化している。人間の代わりにコ





図 2.4: Seam Carving による, Content-Aware なアスペクト比のリサイズ [2]. 左の元画像が前景領域の大きさを変えずに, 背景領域だけを伸縮させることにより, 幅を広げられている.

ンピュータが選択するという意味で, これは典型的な視覚的顕著性マップモデルの応用例といえる. また, 顕著性マップをパターン認識に応用した研究もある. Wang ら [32] は, 視覚的属性や, 物体のクラス分けのために, 顕著度を含めて機械学習を行っている.

画像処理分野での応用例とは, まず, Content-Aware な画像・映像編集技術があげられる. これは, 例えば, 画像をリサイズする際に, 顕著度の高い領域を変化させず, 顕著度の低い領域を拡大・縮小することにより, 人間が見て不自然でない編集を行うことができる, という技術である. Avidan ら [2] によって提案された Seam Carving が有名である (図 2.4). この Seam Carving を改良したものや, 映像に拡張したものなどが最近特に多く提案されている [4, 33–36]. また, 画像圧縮 [37], 映像圧縮 [38, 39] にも応用されている. 顕著度の高い領域は高レートで, 低い領域は低レートに圧縮することで, 重要な領域を綺麗に残しつつ, 圧縮を行うといった形で用いられている.

マルチメディアの分野においては, 画像・映像の知覚的品質に関して, 顕著性マップを



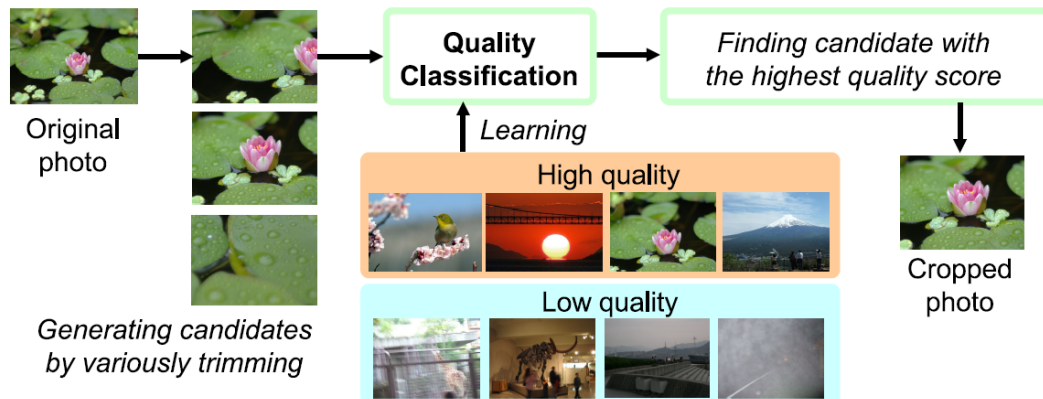


図 2.5: 顕著性マップを用いた自動トリミングの概要 [3]。品質分類器を作成する際に、顕著性マップに対して k-Means クラスタリングを行い、画像中の主要な領域を獲得し、それを学習に用いている。

用いた研究がいくつかなされている。Sun ら [40] は、ボトムアップとトップダウンを組み合わせた顕著性マップの分布をもとに、画像の知覚的品質のスコア付けを行っている。また、You [41] らは、映像の知覚的品質の評価に顕著性マップを用いている。更に、画像中の知覚的品質の高い領域をトリミングすると言った研究も行われている。Nishiyama ら [3] は、Web 上から得られる、多数のユーザーによって評価がなされた大量の画像を学習することで、品質分類器を作成し、それを用いて、画像中の高品質な領域を自動的に切り出す手法を提案している（図 2.5）。品質分類器を作成する際に、顕著性マップに対して k-Means クラスタリングを行い、画像中の主要な領域を獲得し、それを学習に用いている。また、Marchesotti ら [4] は、サムネイル画像として適した領域の切りだしに顕著性マップを利用している。彼らは、全体的な見た目を共有している画像群は、似たような Saliency を共有しているという考えに基づいた、オンライン処理による顕著性マップの作成についても提案している。k-NN 法を用いて検索された類似画像を用いて顕著度の高い領域と低い領域の分類器を作り、それを用いて入力画像の顕著性マップを作成し、サムネイル画像を切りだしている（図 2.6）。さらに、Chang [5] らは、パノラマ画像から、良い構図の領域を切り出す研究を行っている。彼らは、プロの写真家による写真のデータベースを用意し、それぞれ、GIST descriptor [42] と呼ばれるコンテキスト特徴と、顕著性マップを用いて構図の比較を行うことで、パノラマ写真の中からプロの写真家のどれか 1 枚の画像と類似した構図の領域を切りだし、提示している（図 2.7）。

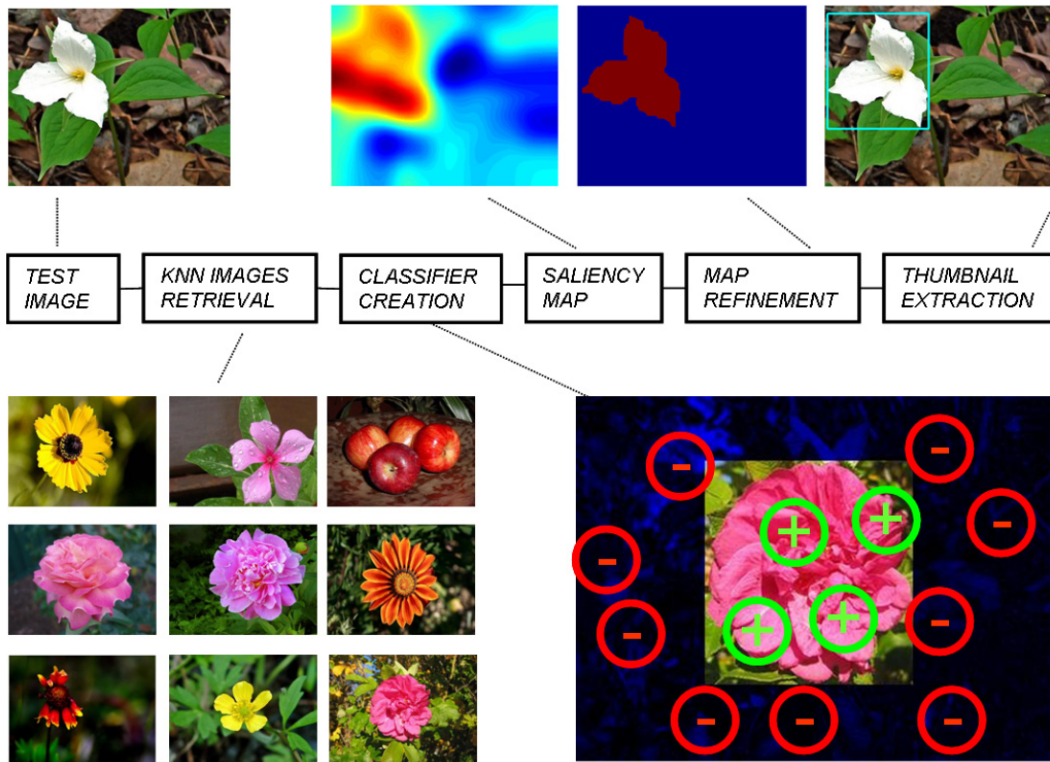


図 2.6: サムネイル画像作成のためのオンライン顕著領域検出 [4] . k-NN 法により検索された類似画像を用いて顕著度の高い領域と低い領域の分類器を作り，顕著性マップを生成する．顕著度の高い領域を元に矩形をサムネイルとして切り出す．

## 2.4 一人称視点映像と自己運動

### 2.4.1 はじめに

一人称視点映像とは，カメラセンサを人の頭部等に装着することで得られる，人の視界と同じような視点の映像のことである．近年，カメラの小型化やモバイルコンピューティングの発達に伴い，一人称視点の自己動作解析技術がコンピュータビジョンの分野で注目されている．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ため，一人称視点から撮影されるダイナミックなシーンを解析することで，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であり，幅広いアプリケーションへの応用が期待されている．既にモバイルアシスタント [43] や，福祉分野での患者のモニタリ



図 2.7: 顕著性マップを利用したパノラマ画像からの高品質な構図の領域切り出し [5] . パノラマ画像 (上段) から, コンテキスト特徴と顕著性マップの構図を用いてプロの写真家の画像 (下段画像の各セットの右側) と類似した構図の領域 (下段画像の各セットの左側) を切り出している .

ング [44] への応用が行われている .

#### 2.4.2 一人称視点映像と視覚的顕著性マップモデル

一人称視点映像を解析する上で, 映像のどこに注意が向けられていたのかというのは, 非常に重要なファクターである . 一人称視点から行動認識を行う目的で, 一人称視点カメラだけでなく, 視線計測のためのカメラを装備し, 計測された視線情報を利用して, 一人称視点カメラが注視点をトラッキングする, といったデバイスも開発されている [45] . しかし, 特殊なデバイスを用いずに, 一人称視点映像のみから, そのカメラの装着者の視覚的注意を推定することができれば, 非常に有用であろう . 画像や映像から視覚的注意を推定する手法としては, 前節のように, 視覚的顕著性マップモデルが広く応用されているが, 一人称視点映像に対する顕著性モデルの有用性については, 確認されていない . 一人称視点の視覚における視覚的注意のメカニズムは, これまで視覚的顕著性マップモデルが用いられた環境のものと大幅に異なる . たとえば, 自己運動によって生じた視覚刺激は, 人間の視覚システムにおいては抑制されるが [21], 既存の視覚的顕著性マップモデルでは, 自己運動によって生じる視覚的運動刺激といった重要な観点が考慮されていない .

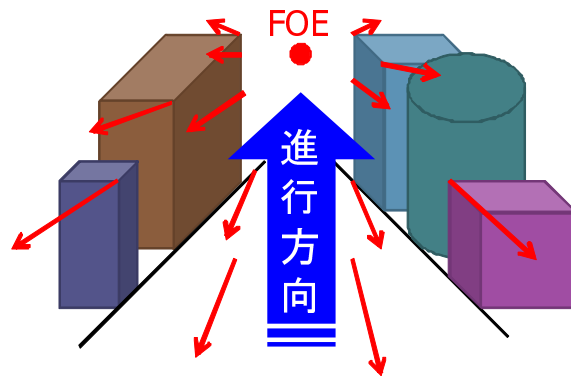


図 2.8: オプティカルフローの拡大中心 (FOE; Focus of Expansion, 動きの消失点, 動きの無限遠点) による進行方向知覚。

### 2.4.3 自己運動によるオプティカルフローと視覚的注意

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 [46]。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、視覚刺激は図 2.8 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 [47]。さらに、最近の研究では、視覚刺激としての FOE は、非常に強力に視覚的注意を誘導することが実験的に確かめられている [48, 49]。また、ヴァーチャルリアリティ環境において、FOE を用いて注意を予測する研究も行われている [50]。

## 2.5 本研究の位置づけ

本研究の目的は 2 つである。1 つは一人称視点における視覚的顕著性マップモデルの有用性を実験的に評価し、その特性を検証することである。そしてもう 1 つは、自己運動に基づく注意予測マップを提案し、顕著性マップと統合することで一人称視点における視覚的注意の予測を高精度化することである。第 2.3.2 項で紹介したように、ディスプレイ上で画像・映像を提示した環境においては視覚的顕著性マップモデルの有用性が確認されているが、一人称視点映像における視覚的顕著性マップモデルの有効性については、これま

であまり研究されてこなかった．また，第 2.4.2 項で述べたとおり，一人称視点映像において撮影された映像のみから視覚的注意の予測が可能であれば，様々な応用が期待できる．そこで我々は，一人称視点映像に対して視覚的顕著性マップモデルを用いた視覚的注意の予測について，その有効性や特性を実験的に検証した．ディスプレイ上での視線計測結果と顕著性マップの比較により視覚的顕著性マップモデルの有効性評価を行った既存の研究と同様に，本研究では一人称視点映像における視線計測結果と，顕著性マップを比較する被験者実験を行った．特に，一人称視点映像で生じる自己運動の影響を検討するため，静的特徴と動的特徴の違いに着目して比較・議論を行った．仮想空間における一人称視点ともいえる，ヴァーチャルリアリティ環境においては，顕著性マップと視線の関係についての検討がなされている [51–53]．ただし，これらの研究が扱う対象は仮想空間であって，我々が用いる現実空間の一人称映像映像とは異なる．さらに，これらは特定のタスクを与えた際の顕著性マップと視線の関係を検討しているものであり，静的特徴のみを用いているため，自己運動の影響を検討したものではない．我々の知る限り本研究は，一人称視点映像に対する視覚的顕著性マップモデルの性能評価を実験的に行った初めての研究である．

また，第 2.3.3 項，第 2.3.4 項で述べたとおり，視覚的顕著性マップモデルは様々な拡張や応用が提案されている．しかし，視覚的顕著性マップモデルを一人称視点映像のために拡張した研究や，一人称視点映像に対して応用した研究はほとんど行われていない．そこで，本研究では視覚的顕著性マップモデルを一人称視点において用いるだけでなく，自己運動に基づいた FOE による注意予測マップを提案し，顕著性マップと統合することで，その精度の向上を目指す．FOE を用いて注意を予測する研究も，ヴァーチャルリアリティ環境においては既に行われている [50] が，本研究では現実世界の一人称視点映像を対象とすることがまず異なる．またこの先行研究 [50] では，オプティカルフローの大きさが小さいほど FOE に近いと計算されているが，一人称視点映像においては複雑な頭部回転運動や身体運動が影響し，このような単純な方法では正確な FOE を求めることはできない．本研究ではフレーム間の特徴点对応により，自己運動パラメータを推定し，オプティカルフローの並進成分と回転成分を分離することで並進成分による FOE をより正確に推定している．さらに，生成した FOE マップを，低レベル特徴から計算された顕著性マップと統合している点が最大の相違点である．



## 第 3 章

# 視覚的顕著性マップモデルの一人称視点における課題

### 3.1 はじめに

前章で確認したように，視覚的顕著性マップモデルは，画像・映像から視覚的注意を推定するために用いられている．しかし，一人称視点にでの視覚的注意推定においては，視覚的顕著性マップモデルの有効性は確認されていない．また，既存の視覚的顕著性マップモデルでは，一人称視点から撮影された映像に生じる自己運動による影響が考慮されていない．そこで本章では，一人称視点における視覚的顕著性マップモデルを用いた視覚的注意推定の性能評価をするための実験を行い，その妥当性と既存のモデルの問題点を検討する．

評価実験では，一人称視点において，低レベルな刺激の特徴に基づいた，既存の視覚的顕著性マップモデルを適用することの有効性を，2 種類の代表的なモデルを用いて評価した．1 つは center-surround メカニズムに基づいた Itti らのモデル [6] であり，もう 1 つは Harel らのグラフベースモデル [7] である．これらのモデルの詳細については，次節で説明する．頭部装着型カメラにより撮影された一人称視点映像から，この 2 種類のモデルを用いて顕著性マップを計算し，同時に計測された視線データとの比較を行う．特に，自己運動の影響を検討するため，各特徴，及び，静的マップと動的マップについて，顕著性マップによる視線推定の性能を表す ROC 曲線による解析を行い，それをもとに考察する．

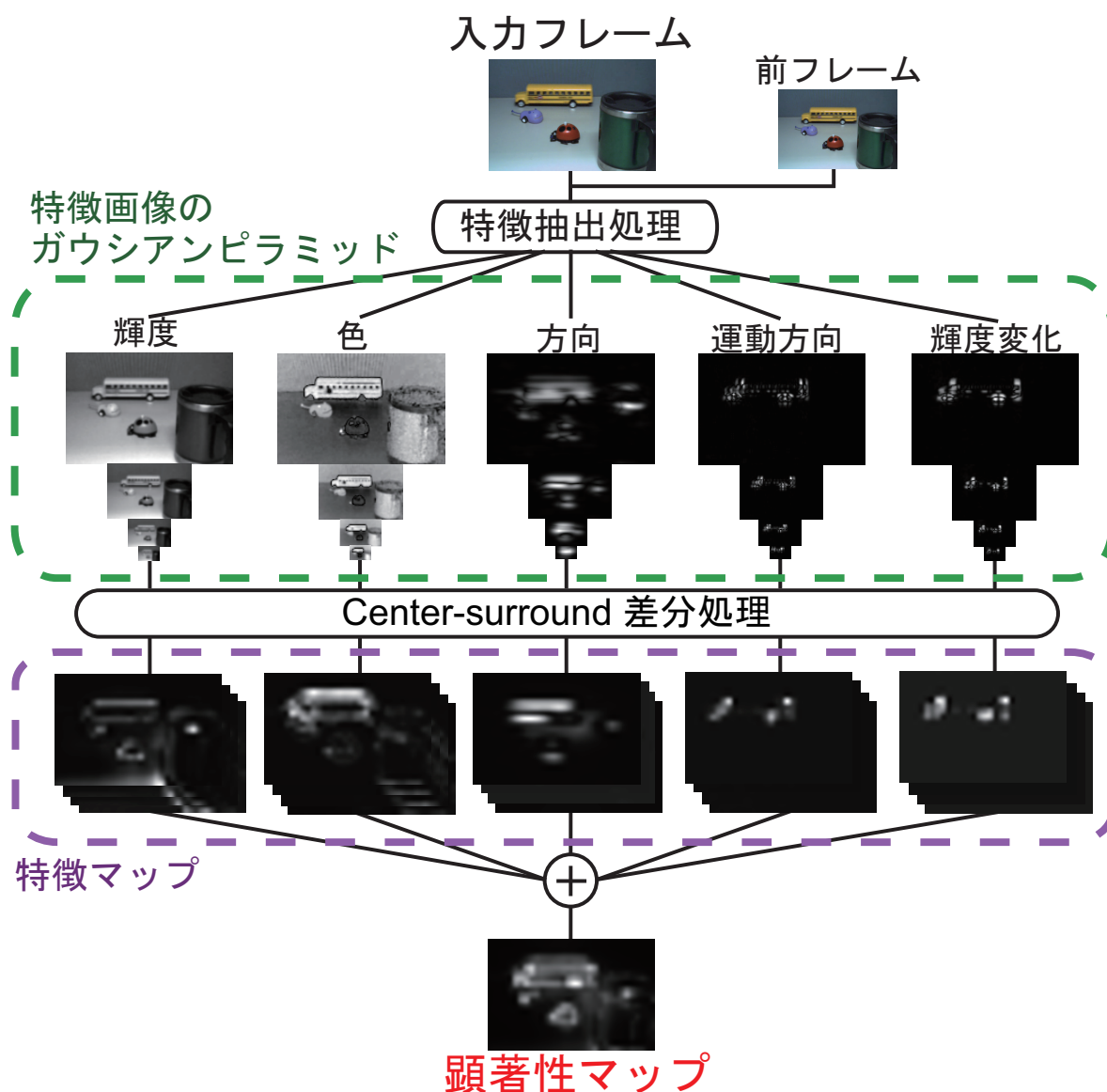


図 3.1: 動画像の顕著性マップ生成処理

### 3.2 動画像の顕著性マップ生成の概要

本節では、動画像の顕著性マップを生成する代表的なモデルとして、center-surround メカニズムに基づいた Itti らのモデル [6] と、Harel らのグラフベースモデル [7] の処理を説明する。まず Itti らのモデルの計算処理について説明した後に、それを改良した Harel

らのモデルの処理について説明する．

主に 3 つのステージで構成される動画像を入力とした顕著性マップの計算処理を図 3.1 に示す．第 1 ステージでは，特徴抽出処理により，入力フレームから特徴画像のガウシアンピラミッドを生成する．第 2 ステージでは，center-surround メカニズムの処理により，特徴画像から，特徴マップを生成する．第 3 ステージでは，全ての特徴マップを正規化し，統合することで最終的な単一の顕著性マップを生成する．

第 1 ステージでは，単純な線形フィルタにより，入力画像を 5 種類の視覚的特徴画像に分解する．動画像からの顕著性マップ生成においては，輝度，色，方向が静的特徴として，運動方向，輝度変化が動的特徴として，一般に用いられる．輝度特徴画像は，入力画像の RGB (赤，緑，青) チャンネルの平均として得られる．Itti らのモデルでは，赤と緑，青と黄色，という補色の関係にある 2 つの色チャンネルのセットから生成された，2 つの異なる画像が色特徴画像として用いられている．一方，本研究ではこの RGB 色空間による色特徴画像の代わりに，Derrington-Krauskopf-Lennie (DKL) 色空間 [54] を用いて生成された色特徴画像を用いた．DKL 色空間は，網膜と視床における生理学的な色処理の 3 つのチャンネルの軸によって表現された色空間で，人間の脳内の色処理を扱うのに適している．方向特徴画像は，4 方向 ( $0^\circ, 45^\circ, 90^\circ, 135^\circ$ ) のガボールフィルタを用いて，輝度画像から計算される．

運動方向と輝度変化の特徴画像を得るためには，2 つの入力フレームが必要となる．運動方向特徴画像は，現在のフレームと，前フレームの間で各 4 方向の方向特徴画像を空間的にシフトした差分画像から生成される．輝度変化特徴画像は，現在のフレームと，前フレームの間の輝度特徴画像の差分画像として計算される．結果として，輝度が 1 種類，色が 2 種類，方向が 4 種類，運動方向が 4 種類，輝度変化が 1 種類で，合計 12 種類の特徴画像が生成される．続いて，各特徴画像ごとに，9 つのスケール (スケール 0 = 1:1 ~ スケール 8 = 1:256) ガウシアンピラミッド [55] を生成する．

第 2 ステージでは，この特徴画像のガウシアンピラミッドから，center-surround メカニズムを用いて特徴マップが計算される．まず，2 種類の異なるスケールサイズのガウシアンピラミッドを組み合わせ，6 つのセットを作成する．続いて，各セットにおけるスケール間の補間と各ピクセルごとの減算により得られた，スケール間の差分画像を用いて，各特徴画像毎に 6 種類の特徴マップが計算される．

第 3 ステージでは，輝度が 6 種類，色が 12 種類，方向が 24 種類，運動方向が 24 種類，輝度変化が 6 種類の，合計 72 種類の正規化された特徴マップを統合することで最終的な顕著性マップを生成する．統合前には，各特徴マップにおいて， $(M - \bar{m})$  を全体に乘ず



ることで、正規化処理を行う。ここで、 $M$  はそのマップ全体の最大値であり、 $\bar{m}$  は全体の最大値を除いた局所最大値の平均である。この正規化処理には、少数の強いピークが存在するマップを全体的に増幅し、多数の同程度のピークが存在するマップを全体的に抑制する効果がある。

Harel らのモデル [7] は、用いる特徴画像は Itti らのモデルと同一であるが、第 2 ステージと第 3 ステージにおいてグラフベースのアプローチを採用している。特徴マップと最終的な顕著性マップはマルコフ連鎖グラフの平衡分布を計算することによって生成される。第 2 ステージにおいては、ノード間の相違度関数と距離関数を定義し、この 2 つの関数を互いに乗ずることで各ノードの重みを計算する。第 3 ステージにおいては、特徴マップ上の対応する位置の値と距離関数を乗ずることで各ノードの重みを得る。

### 3.3 被験者実験

第 3.2 節のように、既存の視覚的顕著性マップモデルは単純な低レベルの特徴画像を用いて顕著性マップを計算している。これらのモデルは既に撮影済みの画像や映像を人間に提示した場合の注意を推定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。一人称視点において既存の視覚的顕著性マップモデルを用いる有効性について評価するため、我々は以下に説明するような被験者実験を行った。

#### 3.3.1 被験者実験の手順

一人称視点における視覚的顕著性マップモデルの性能を評価するため、頭部に装着されたカメラから撮影された映像に対する視線計測結果と、顕著性マップの関係を分析する実験を行った。

実験では、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 [56] 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 3.2(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  $121^\circ$  であり、撮影される映像の解像度は  $640 \times 480$  である。

前節で述べた Itti らのモデル [6] と、Harel らのモデル [7] を代表的な視覚的顕著性マップモデルとして評価に用いた。実験は、室内で着席した状態、室内で自由に動き回る状態、屋外で自由に動き回る状態の、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 3.2(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．

人間の注意は実行中のタスクによって影響されるが，既存の視覚的顕著性マップモデルは，このような高レベルの注意のメカニズムを効果的に扱うことは出来ない．我々の研究の目的は，一人称視点において視覚的顕著性マップモデルを用いることの有効性を検討することであるため，着席の指示と自由な移動以外の特別なタスクを被験者に与えなかった．

#### 3.3.2 被験者実験の結果

視覚的注意の推定精度に各特徴がどの程度寄与しているかを評価するため，5 つの各特徴について，単一の特徴のみを用いて計算された顕著性マップ，及び，それらを組み合わせた顕著性マップと，視線計測結果の関係を分析した．図 3.3 に示す曲線は，実験に用いた全データセットを用い，顕著度の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．縦軸は検出率であり，顕著性マップにおいて顕著度が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，顕著性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．

図 3.3 (a) 及び (b) は前節で説明した 2 つの視覚的顕著性マップモデルを用いて生成した，5 つの各特徴ごとの顕著性マップの性能比較である．図 3.3 (c) 及び (d) は静的顕著性マップ，動的顕著性マップ，標準顕著性マップの，3 つの顕著性マップの性能比較である．静的顕著性マップは輝度，色，方向の静的特徴のみを用いて計算された顕著性マップであり，動的顕著性マップは運動方向，輝度変化の動的特徴のみを用いて計算された顕著性マップである．標準顕著性マップは全ての特徴を統合した，通常の顕著性マップである．図 3.3 (a) 及び (c) は Itti らのモデル [6] を用いた結果であり，図 3.3 (b) 及び (d) は Harel らのモデル [7] を用いた結果である．表 3.1 に示す AUC (Area under the curve, ROC 曲線の曲線下面積) は，図 3.3 (c) , (d) における 3 本の ROC 曲線の推定性能を表している．これらの結果から，これらの既存の視覚的顕著性マップモデルは，一人称視点における視覚的注意を偶然に比ると十分良く推定できているが，両モデルにお

表 3.1: 特徴を統合した顕著性マップの性能評価 (AUC, 曲線下面積) .

|                 | 静的特徴  | 動的特徴  | 全特徴   |
|-----------------|-------|-------|-------|
| Itti らのモデル [6]  | 0.764 | 0.590 | 0.742 |
| Harel らのモデル [7] | 0.809 | 0.677 | 0.765 |

いて動的特徴は性能に寄与しておらず，それどころか精度を低下させてしまっていることがわかる．

### 3.3.3 考察

実験結果は，一人称視点において，運動方向と輝度変化の動的特徴が明らかに予測性能を低下させているということを示している．しかしながら，実験において，動的な視覚的刺激が被験者の注意を引いた場面が複数回観測された．図 3.4 は被験者が移動物体（歩行者）を見ていたシーンの視線軌跡の例である．また，図 3.5 は輝度が急激に変化した物体（ノート PC のモニター）を含むシーンの視線軌跡の例である．これらのケースでは，被験者は動的な視覚刺激の変化に対して注意を向けているが，物体の移動，輝度の変化に対応する運動方向特徴，輝度変化特徴による顕著性マップでは，動的变化の無い他の位置に，高い顕著度が与えられていた．このことは，既存のモデルで提案された特徴が，一人称視点においては動的な視覚刺激を適切に扱うことができないことを示唆している．

図 3.6 に，自己運動を伴うシーンにおける，動的特徴による顕著性マップの例を示す．(a) と (b) は自己運動が小さい時のフレームであり，(c) と (d) は自己運動が大きい時のフレームである．(a) と (c) は運動方向特徴のみを用いて生成された顕著性マップであり (b) と (d) は輝度変化特徴のみを用いて生成された顕著性マップである．左列が入力画像，中列が各動的特徴による顕著性マップ，右列が入力画像に顕著性マップを重ね合わせたものである．

既に撮影済みの画像や映像を人間に提示するケースと異なり，我々の想定しているケースでは，自己運動の影響が生じる．人間の視覚システムが自己運動を補償する機能を有している [21] 一方で，既存の視覚的顕著性マップモデルはこのような補償のためのメカニズムを考慮していない．それゆえ，動的特徴による顕著性マップ（図 3.6）では，実世界では動的な変化が無いにもかかわらず，自己運動の影響により，，図 3.6 のように，視差

がや輝度が大きく異なる領域のエッジ等，視覚的な運動方向や輝度変化が自己運動により周囲と異なる領域に，高い顕著度が現れている．しかしながら，このような視覚刺激に対する注意は，自己運動の補償機能により抑制されるため，顕著性マップの性能低下の原因となっている．

#### 3.4 本章のまとめ

本章では，一人称視点から撮影された映像から計算した顕著性マップと視線位置の関係を実験的に分析することで，一人称視点における視覚的顕著性マップモデルを用いた視覚的注意推定の性能評価を行った．実験の結果，既存の視覚的顕著性マップモデルは，一人称視点においても視覚的注意を偶然よりよく予測できることを確認した．しかしながら，既存のモデルは人間の視覚が持つ，自己運動による影響を補償するメカニズムをモデル化できておらず，一人称視点においては，動的特徴が性能を低下させていた．それゆえ，既存のモデルは一人称視点映像を効率よく取り扱うためには，特に動的マップにおいて，改良の余地が十分残されているといえる．



(a)



(b)

図 3.2: 被験者実験環境 . (a) EMR-9 [56], 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  $121^{\circ}$  であり , 撮影される映像の解像度は  $640 \times 480$  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 .

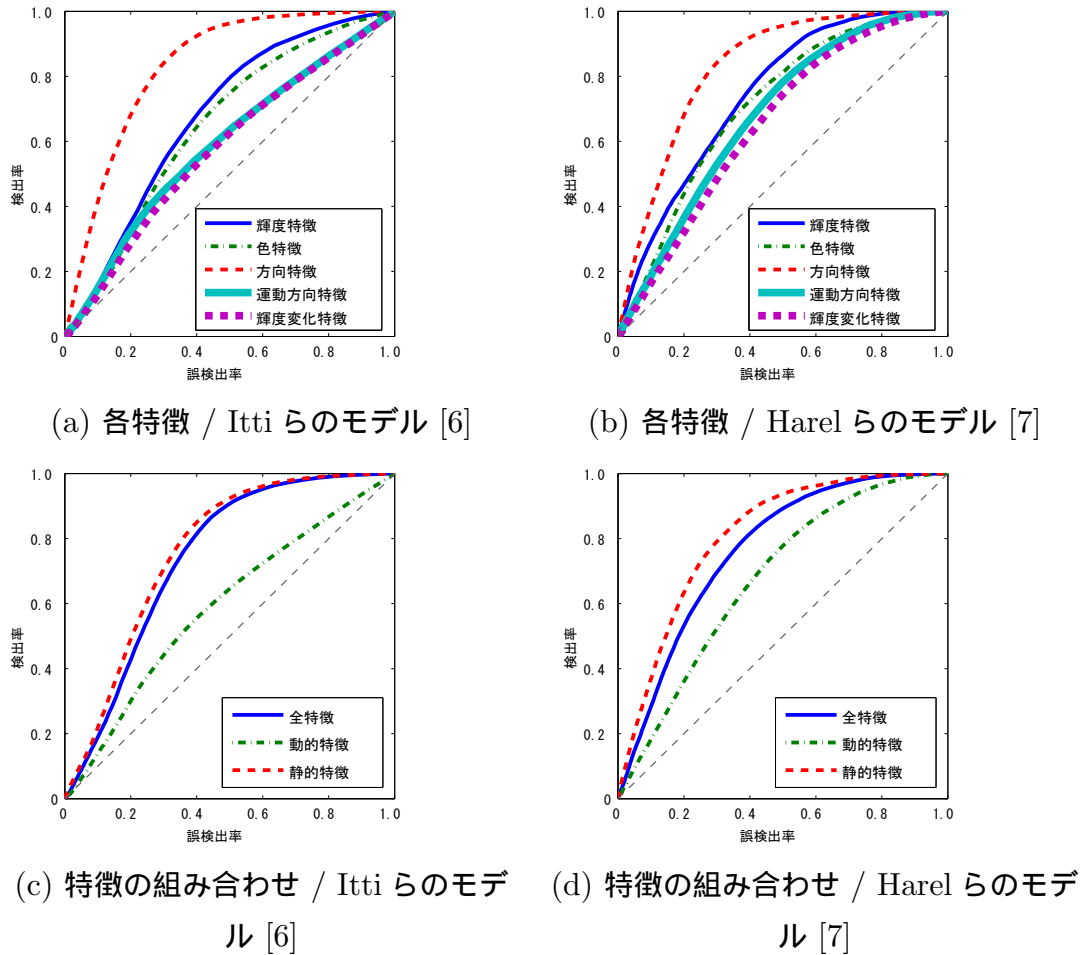


図 3.3: 各特徴ごとのマップの ROC 曲線 ((a) Itti らのモデル [6], (b) Harel らのモデル [7]) 及び、静的特徴，動的特徴，全ての特徴を用いたマップの ROC 曲線 ((c) Itti らのモデル [6], (d) Harel らのモデル [7]). ROC 曲線は，実験に用いた全データを用い，顕著度の閾値を最小から最大まで変化させることで算出した．縦軸は検出率であり，各マップにおいて顕著度が閾値よりも高い値を持つ領域に含まれた実際の視線位置の割合に対応する．横軸は誤検出率であり，各マップ中で閾値よりも高い値を持ち，視線が存在しない領域の割合を示している．

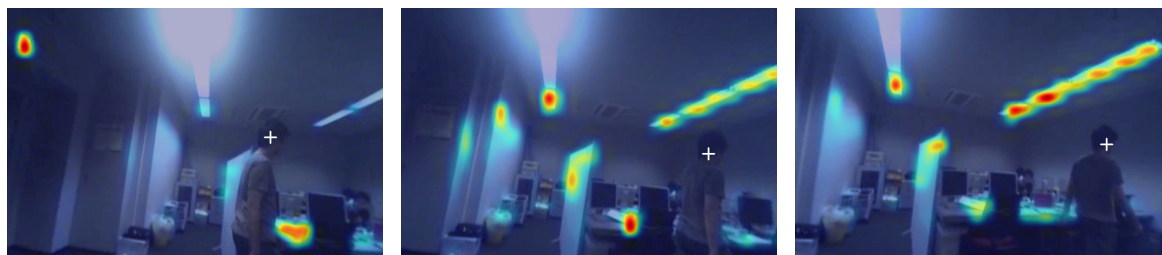


図 3.4: 被験者が移動物体（歩行者）を見ていたシーンの視線軌跡例．一人称視点映像に運動方向特徴のみによる顕著性マップを重ね合わせてある．白い十字は視線位置を示す．

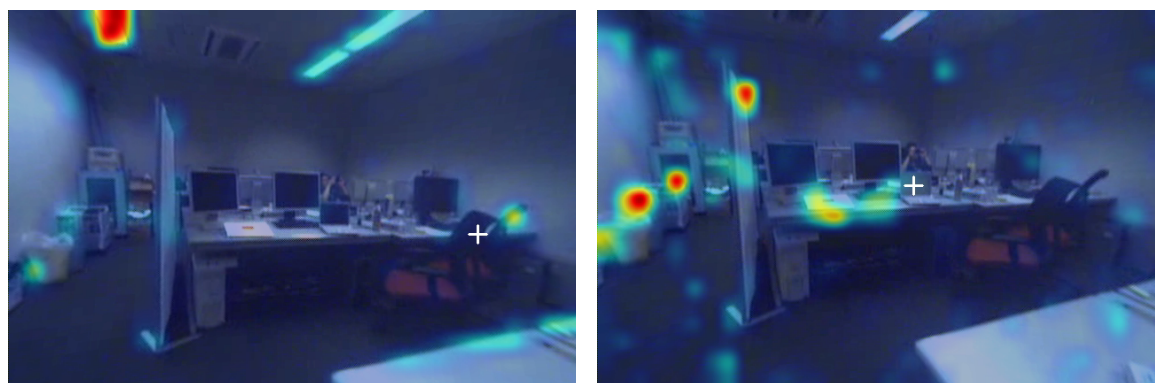
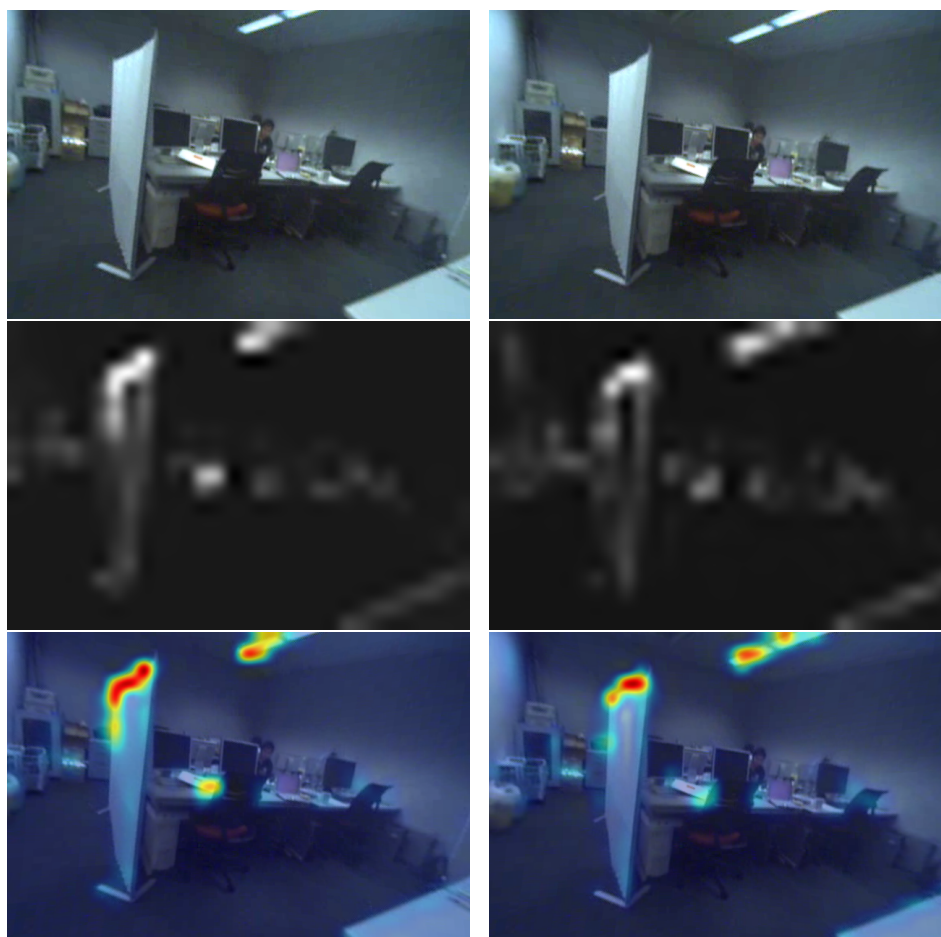


図 3.5: 輝度が急激に変化した物体（ノート PC のモニター）を含むシーンの視線軌跡例．一人称視点映像に輝度変化特徴のみによる顕著性マップを重ねてある．白い十字は視線位置を示す．





(a) 運動方向特徴

(b) 輝度変化特徴

図 3.6: 自己運動を伴うシーンにおける，動的特徴による顕著性マップの例．(a) は運動方向特徴のみを用いて生成された顕著性マップであり (b) は輝度変化特徴のみを用いて生成された顕著性マップである．上段が入力画像，中段が各動的特徴による顕著性マップ，下段が入力画像に顕著性マップを重ね合わせたものである．



## 第 4 章

# 自己運動と顕著性に基づく視覚的注意推定

### 4.1 はじめに

前章で述べたように，既存の視覚的顕著性マップモデルは，単純な低レベルの特徴画像を用いて顕著性マップを計算している．これらのモデルは既に撮影済みの画像や映像を人間に提示した場合の顕著性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．本章では，一人称視点映像から自己運動を推定し，顕著性と自己運動に基づいた視覚的注意推定モデルを提案し，評価実験によりその有用性を確認する．

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 [57]．そこで，本研究では注意を誘導する進行方向の情報として，FOE を用いた自己運動方向マップを提案し，これを顕著性マップと統合して，注意推定モデルを構築する．

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進行方向とフローの交点は一致しない (図 4.1)．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 [58–62]，独立して運動する対象物があった場

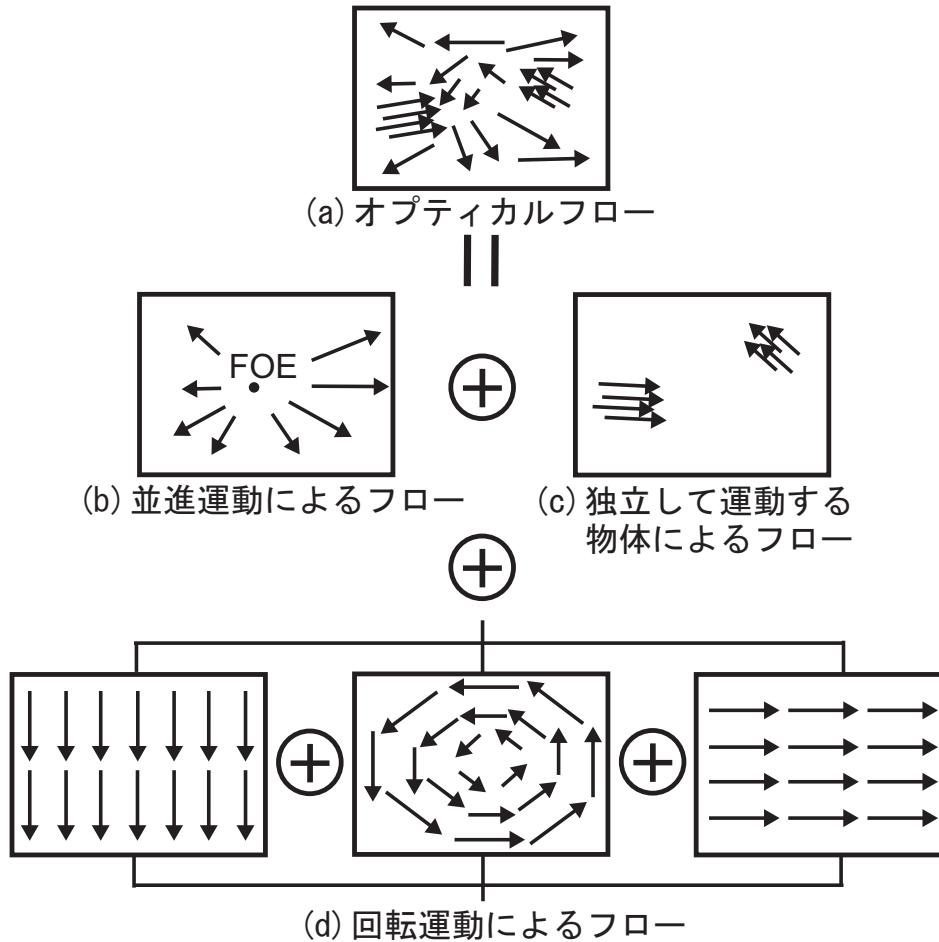


図 4.1: FOE とオプティカルフローの成分．並進運動によるフロー (b) は放射状の拡大運動となり，その交点 (FOE; Focus of Expansion, 動きの消失点, 動きの無限遠点) として進行方向の位置が求められる．オプティカルフロー (a) は，独立して運動する対象物 (c) や自己運動による回転運動 (d) によるフローを反映したものとなり，これらを含む場合には進行方向とフローの交点は一致しない．運動物体によるフローの除去や，フローの回転成分と並進成分の分離をすることができれば，並進運動のみによるフローを用いて FOE を求め，進行方向として得ることができる．

合にも進行方向は正しく知覚される [63]．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．

本章で提案する視覚的注意推定手法は，頭部に装着されたカメラ等から撮影された一人

称視点映像を入力とし、自己運動を考慮した動的マップを生成し、前章で述べた既存のモデルの静的マップと統合することで、視覚的注意を推定するマップを出力する。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 [64]、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、本研究では一人称視点映像を撮影するカメラは頭部に装着され、カメラの光軸は顔の正面方向を向いているとする。自己運動に基づく視覚的注意の推定には、カメラ間の幾何学的な関係と自己運動パラメータを利用するが、カメラ間の幾何学的な関係式の導出と 2 画像からの自己運動パラメータ推定については、付録 A を参照されたい。

## 4.2 自己運動と顕著性に基づく視覚的注意の推定

本節で説明する自己運動と顕著性に基づく視覚的注意推定マップ生成の流れを図 4.2 に示す。一人称視点映像から、自己運動に基づいて進行方向に対する注意を予測する FOE マップを生成し、低レベルな静的視覚特徴に基づいて生成される静的顕著性マップと統合することで、視覚的注意を推定するマップを生成する。静的顕著性マップの生成については、第 3.2 節で既に説明したため、本節では自己運動に基づいて FOE マップを生成する手順と、FOE マップと顕著性マップの統合について述べる。

### 4.2.1 自己運動パラメータからの前方運動判定

FOE の推定を行う前に、自己運動パラメータ  $R, t$  から、カメラ装着者が前進しているのか否かの判定を行う。並進ベクトル  $t$  は付録 A.1 のように得られているが、並進ベクトル  $t$  のスケールは不定であり、単純に大きさによる判定は行えない。しかしながら、並進ベクトル  $t$  は、 $x, y, z$  方向の成分の比率のわかる単位ベクトルとして得られており、本研究ではカメラの光軸（ $Z$  軸）と人の顔の正面方向は同一としている。そこで、前節の処理で自己運動による影響が大きいと判定され、並進ベクトル  $t$  の  $z$  成分が閾値  $z_0$  以上のフレーム間の運動において、装着者は前進していると考えられる。 $t$  が単位ベクトルであり、各成分の値の二乗がその方向成分の比率となることから、 $z_0^2 = 0.7$  となるように設定した。本研究ではカメラの光軸（ $Z$  軸）と人の顔の正面方向は同一と仮定しているが、それだけでは、画像中の進行方向の位置を決めることはできない。そこで、既に紹介した FOE を次項のように計算することで、画像中の進行方向の位置を求める。

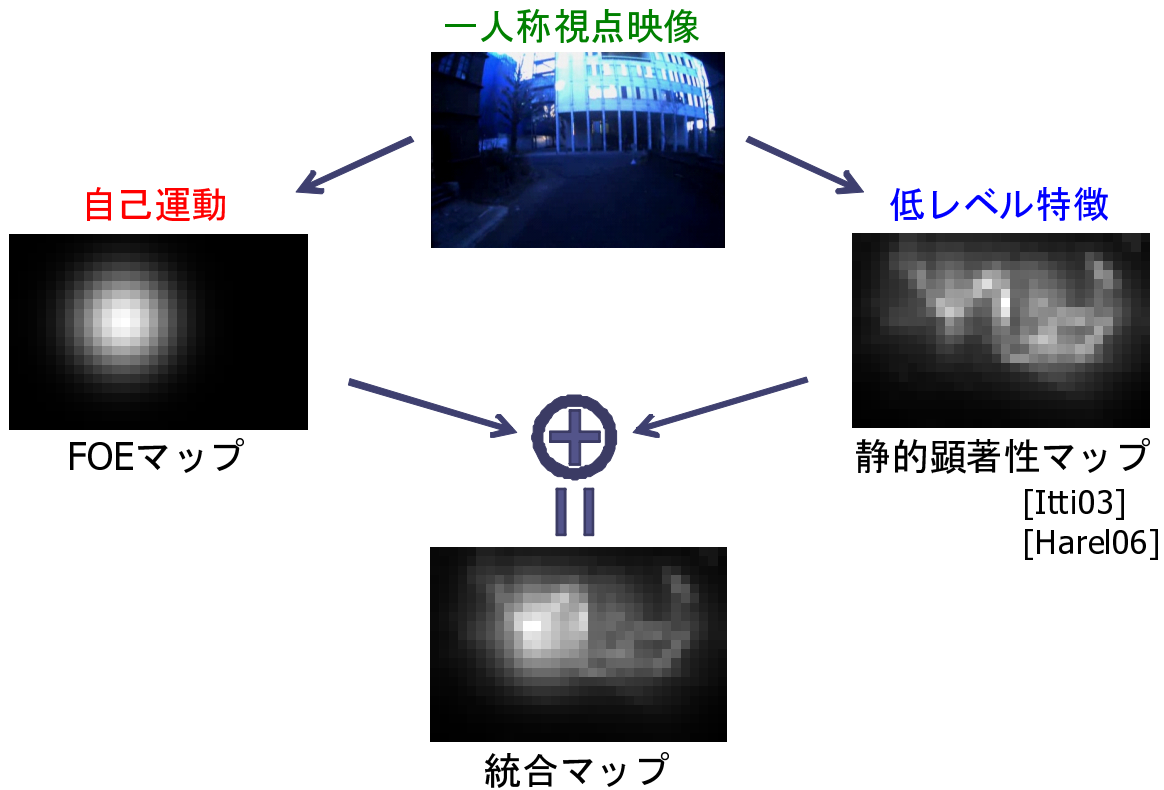


図 4.2: 自己運動と顕著性に基づく視覚的注意の推定．一人称視点映像から，自己運動に基づき進行方向に対する注意を予測する FOE マップを生成し，低レベルな静的視覚特徴に基づいて生成される静的顕著性マップと統合することで，視覚的注意を推定するマップを生成する．

#### 4.2.2 FOE の推定

まず，運動物体や，誤対応によるフローを除去する．これには前節において基礎行列  $F$  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 [65] 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，前節において得られた回転行列  $R$  (式 A.12) を利用することで，同様の処理を行う．図 A.1 における，

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  $m'_R$  とすると , 式 A.4, 式 A.5 と同様に , 次の関係式が得られる .

$$\tilde{m}'_R \sim A\tilde{x}'_R \quad (4.1)$$

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 .

$$\tilde{x}'_R \sim R\tilde{x}' + t_0 = R\tilde{x}' \quad (4.2)$$

式 A.5 , 式 4.1 , 式 4.2 より , 次の式が成り立つ .

$$\tilde{m}'_R \sim ARA^{-1}\tilde{m}' \quad (4.3)$$

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 4.3 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 . (a) は回転成分と並進成分を含んだフローであり , (b) は上述の手順により分離した , 並進成分のみのフローである . 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 . 本研究では , 人間の視覚処理と同様に , 並進運動によるフローの交点 (FOE) として画像中の進行方向位置を計算したが , これはエピポーラ幾何において , 画像平面と並進ベクトル  $t$  の交点であるエピポール (図 4.4) と一致する . すべてのエピポーラ線はエピポール  $e$  を通るため , エピポーラ方程式 (式 A.8) より ,  $F^T e = 0$  が成り立ち , これを用いて自己運動方向の画像中座標を計算することも可能である .

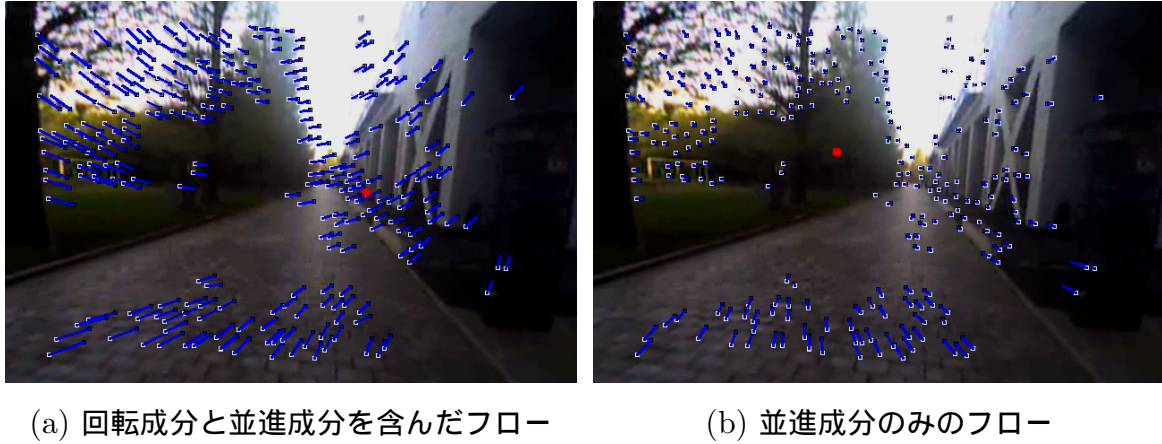


図 4.3: オプティカルフローからの並進成分の分離．前フレームの画像を回転させ，特徴点座標を求めることで，回転成分と並進成分を含んだフロー (a) から，並進成分のみのフロー (b) を得ることができる．各画像において，白い四角は現在のフレームの画像における特徴点座標であり，黒い四角は，(a) では前フレームの画像における特徴点座標，(b) では前フレームを回転させた画像における特徴点座標である．赤い丸は，各フローまでのユークリッド距離の合計が最小となる点として推定されたフローの交点を示す．

#### 4.2.3 複数フレームからの FOE マップ生成

本項では，前項で得られた FOE を用いて自己運動方向に対する注意を利用した FOE マップの生成する手順について説明を行う．前項では，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 フレーム間のみの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 [49] 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 [66] の「経験則」である，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

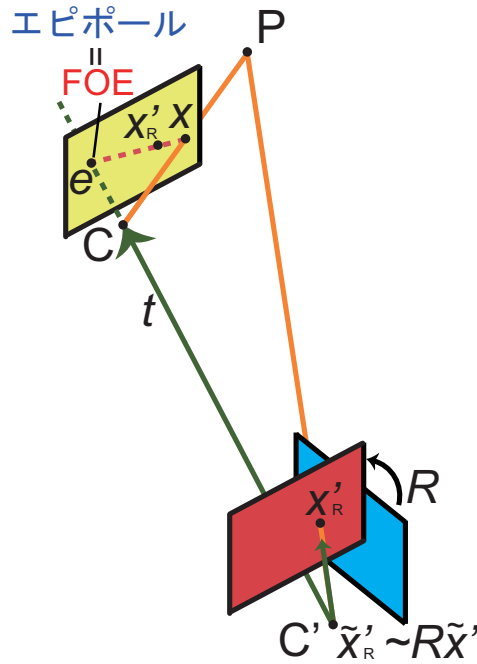


図 4.4: FOE とエピポール．エピポールは画像平面と直線  $CC'$ ，すなわち並進ベクトル  $t$  との交点である．したがって，エピポールは画像中における進行方向の位置を意味し， $xx'_R$  のような並進成分のみによるフロー直線から得られる FOE と一致する．

処理により求めた．

#### 4.2.4 FOE マップと静的顕著性マップの統合による注意推定マップ生成

上述のようにして得られた自己運動に基づく自己運動方向マップと，低レベルな特徴に基づく視覚的顕著性マップモデルによって生成された顕著性マップを組み合わせることによって，最終的な注意推定マップを生成する．前章の予備実験により，視覚的顕著性マップモデルの動的特徴は，一人称視点映像においては性能低下の原因となることが示されたので，静的特徴のみから生成された顕著性マップを用いる．また，本研究ではベースとする視覚的顕著性マップモデルとして，本研究では第 3.2 節で紹介した Itti らのモデル [6] と，Harel らのモデル [7] を用いる．各モデルの正規化処理によって正規化された FOE マップと，顕著性マップを 1:1 の割合で足し合わせ，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，最終的な注意推定マップを生成した．FOE マップと静的顕著性マップの統合による注意推定マップ生成例を図 4.6 に示す．それぞ



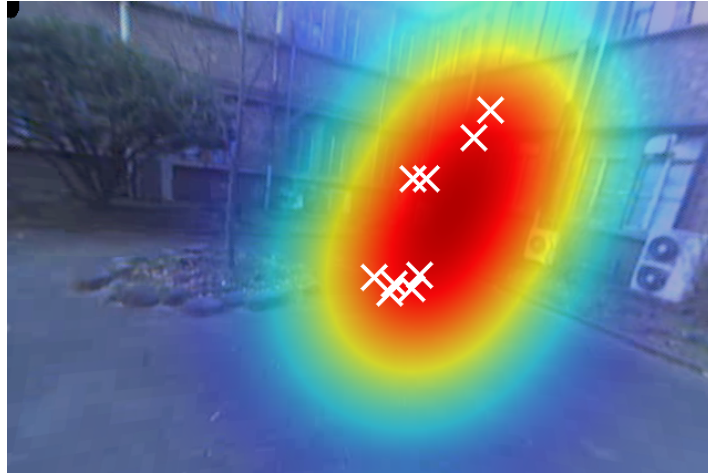


図 4.5: 複数フレームからの FOE マップ生成．一人称視点映像においては歩行中の頭部運動や身体運動の影響により，2 フレーム間のみの関係では人間が知覚している進行方向を正確に得ることはできない．ロバストに FOE マップを生成するため，現在のフレームの画像と，現在のフレームから  $k$  フレーム前 ( $k = 1, 2, \dots, n$ ) の画像を用いて自己運動パラメータと FOE の推定を行い，現在のフレームの画像中に FOE が検出されたすべてのフレームの FOE を用いる．得られた複数の FOE から，ガウシアンカーネルを用いた 2 次元カーネル密度推定により，FOE マップを生成する．図中の白い x 印は得られた FOE で，生成された FOE マップを現在のフレーム画像に重ね合わせた例である．

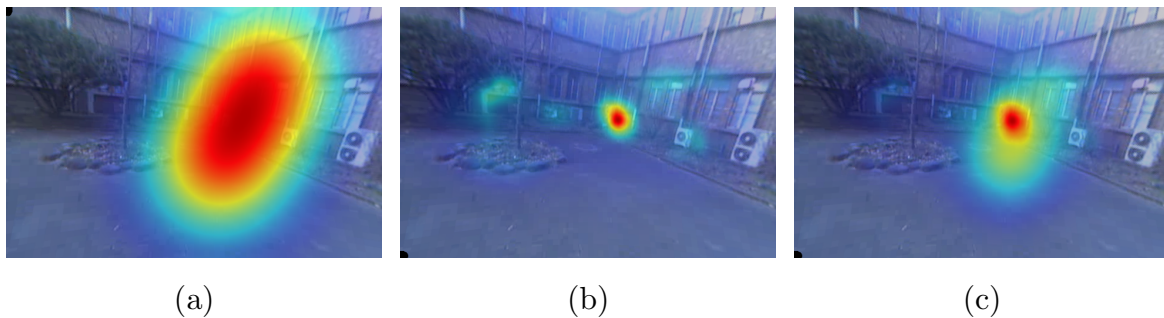


図 4.6: FOE マップと静的顕著性マップの統合による注意推定マップ生成例．それぞれ，(a) は自己運動に基づき生成された FOE マップ，(b) は静的な低レベル視覚特徴に基づき生成された静的顕著性マップ，(c) は FOE マップと静的顕著性マップを統合した注意推定マップを，入力フレーム画像に重ね合わせた例である．



表 4.1: Itti ら [6] の視覚的顕著性マップモデルとそれを用いた提案手法との性能比較 (AUC, 曲線下面積) .

|         | 顕著性マップ | 静的顕著性マップ | 提案手法  |
|---------|--------|----------|-------|
| 室内/着席   | 0.772  | 0.794    | 0.782 |
| 室内/移動   | 0.748  | 0.770    | 0.791 |
| 屋外/移動   | 0.708  | 0.730    | 0.831 |
| 全データセット | 0.742  | 0.764    | 0.801 |

表 4.2: Harel ら [7] の視覚的顕著性マップモデルとそれを用いた提案手法との性能比較 (AUC, 曲線下面積) .

|         | 顕著性マップ | 静的顕著性マップ | 提案手法  |
|---------|--------|----------|-------|
| 室内/着席   | 0.780  | 0.823    | 0.835 |
| 室内/移動   | 0.759  | 0.804    | 0.835 |
| 屋外/移動   | 0.769  | 0.809    | 0.858 |
| 全データセット | 0.765  | 0.809    | 0.841 |

れ, (a) は自己運動に基づき生成された FOE マップ, (b) は静的な低レベル視覚特徴に基づき生成された静的顕著性マップ, (c) は FOE マップと静的顕著性マップを統合した注意推定マップを, 入力フレームの画像に重ね合わせた例である.

### 4.3 評価実験

本章で提案した注意推定マップの性能を評価するため, 以下のような実験を行った. 評価には, 第 3.3 節で得られた被験者実験データセットを用い, 第 3.3 節の性能評価と同様に, 注意推定マップによる視線推定性能を ROC 曲線により評価した. 提案手法による注意推定マップは, Itti らの視覚的顕著性マップモデル [6] を用いたものと, Harel らの視覚的顕著性マップモデル [7] を用いたものを別々に生成・評価し, それぞれ, 基になった

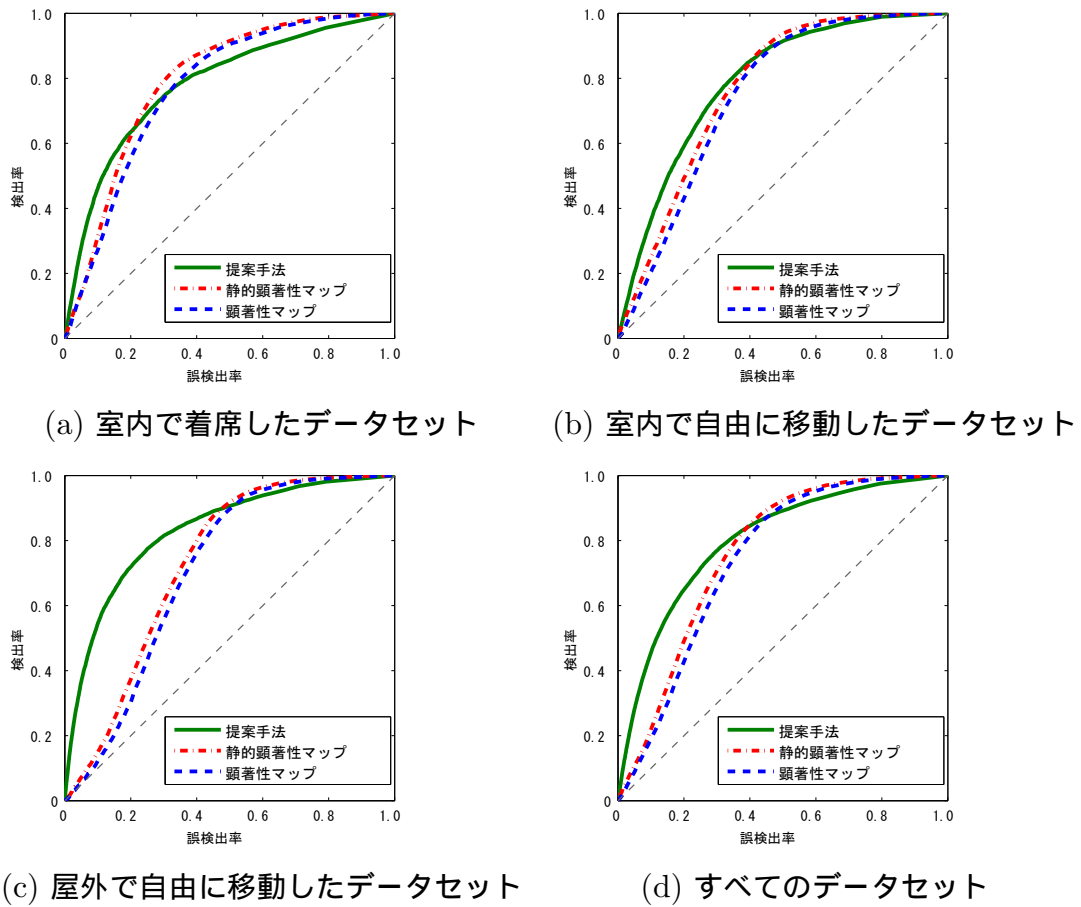


図 4.7: Itti らのモデル [6] を顕著性として利用した提案手法の性能評価．(a) は室内で着席したデータセット，(b) は室内で自由に移動したデータセット，(c) は屋外で自由に移動したデータセット，(d) はすべてのデータセットの性能を示している．ROC 曲線は，各データセットの注意推定マップと視線データを用い，顕著度の閾値を最小から最大まで変化させることで算出した．縦軸は検出率であり，各マップにおいて顕著度が閾値よりも高い値を持つ領域に含まれた実際の視線位置の割合に対応する．横軸は誤検出率であり，各マップ中で閾値よりも高い値を持ち，視線が存在しない領域の割合を示している．比較のため，提案手法によるマップと，それに用いた静的顕著性マップ，そして静的顕著性マップと動的顕著性マップを統合した通常の顕著性マップの性能を示した．

モデルと性能の比較を行った．具体的には，静的特徴のみを用いて生成された顕著性マップ，及び，全特徴を用いて生成された顕著性マップの ROC 曲線との比較を行った．それぞれのモデルを用いた結果について，被験者実験における，3 パターン（室内で着席した

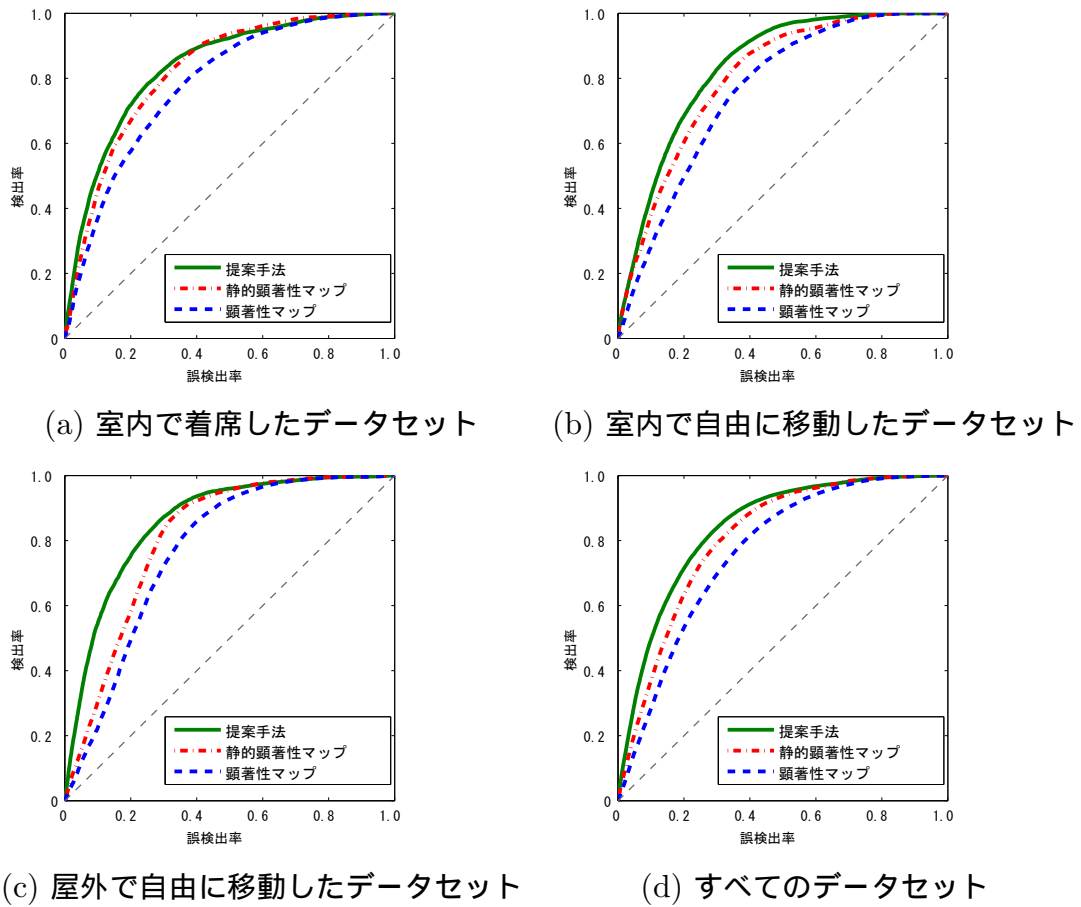


図 4.8: Harel らのモデル [7] を顕著性として利用した提案手法の性能評価．(a) は室内で着席したデータセット，(b) は室内で自由に移動したデータセット，(c) は屋外で自由に移動したデータセット，(d) はすべてのデータセットの性能を示している．ROC 曲線は，各データセットの注意推定マップと視線データを用い，顕著度の閾値を最小から最大まで変化させることで算出した．縦軸は検出率であり，各マップにおいて顕著度が閾値よりも高い値を持つ領域に含まれた実際の視線位置の割合に対応する．横軸は誤検出率であり，各マップ中で閾値よりも高い値を持ち，視線が存在しない領域の割合を示している．比較のため，提案手法によるマップと，それに用いた静的顕著性マップ，そして静的顕著性マップと動的顕著性マップを統合した通常の顕著性マップの性能を示した．

状態，室内で自由に動き回る状態，屋外で自由に動き回る状態）の各データセットを用いて計算した ROC 曲線と，全データセットを用いて計算した ROC 曲線を図 4.7，図 4.8 にそれぞれ示す．また，それらの推定性能を示す ROC 曲線の AUC (Area under the curve，

曲線下面積) を表 4.1, 表 4.2 に示す．図 4.7, 図 4.8 において, (a) は室内で着席したデータセット, (b) は室内で自由に移動したデータセット, (c) は屋外で自由に移動したデータセット, (d) はすべてのデータセットの性能を示している．また, 前章の評価と同じく, ROC 曲線は, 各データセットの注意推定マップと視線データを用い, 顕著度の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, 各マップにおいて顕著度が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．

視覚的顕著性マップモデルの動的特徴が自己運動の影響を扱えず, 性能を下げる要因となっているのに対し, 静的顕著性マップと自己運動方向マップを組み合わせた提案手法は, 静的顕著性マップと比較しても, 注意推定性能が向上していることがわかる．

#### 4.4 考察

図 4.7, 図 4.8 の (b), (c) 及び, 表 4.1, 表 4.2 のように, 両モデルにおいて, 室内で自由に動き回る状態では精度が若干向上しており, また, 屋外で自由に動き回る状態では, 明らかな精度の向上が見られる．特に, 屋外を自由に動き回る状態では, 精度の向上が非常に大きく, 屋外での一人称視点映像は室内に比べて方向転換が少なく, 並進運動が多かったため, FOE による注意推定が非常に有効に働いたと考えられる．

ただし, 図 4.7, 図 4.8 の (a) のように, 室内で着席した状態では, 視覚的顕著性マップモデルと比較して, 明確な精度の改善があるとはいえない．これは, 着席状態では自己運動は, ほぼ頭部の回転運動に限られ, FOE による注意推定マップが有効な並進運動が非常に少ないためである．また, 本来であれば回転運動のみの場合については FOE が検出されないのが理想であるが, 原理上並進ベクトル  $t$  の大きさを得ることができないため, 並進運動が非常に小さい場合でも FOE が検出されてしまい, 着席状態においてもわずかではあるが, 性能の上下が見られる．並進運動が小さい場合の FOE 検出を防ぐことは, 今後の課題といえる．また, 回転運動は並進運動に劣らず重要な注意行動であると考えられるため, 着席状態のように自己運動が頭部の回転運動のみの場合でも精度が向上するような回転運動に適応した注意推定手法についても今後の課題としたい．

## 第 5 章

# 結論

### 5.1 本論文のまとめ

本研究では，一人称視点における，視覚的顕著性マップモデルを用いた視覚的注意推定に関する検討を行い，顕著性と自己運動に基づいた新たな注意推定モデルの提案を行った．一人称視点から撮影された映像から，既存の視覚的顕著性マップモデルにより計算した顕著性マップと，視線位置の関係を実験的に分析することで，一人称視点における視覚的顕著性マップモデルを用いた視覚的注意推定の性能評価を行った．実験の結果，既存のモデルは人間の視覚が持つ，自己運動による影響を補償するメカニズムをモデル化できておらず，一人称視点においては，動的マップが性能を低下させていることを確認した．

また，一人称視点映像を効率よく取り扱うため，自己運動を考慮した動的マップを生成し，顕著性マップと統合することで一人称視点に適した注意推定を行う手法を提案した．提案手法では，自己運動と人間の視覚の関係の中で，非常に重要な運動方向への注意に着目し，運動方向に対する新たな注意推定マップモデルを構築した．複雑な 3 次元運動によるオプティカルフローから，一人称映像フレーム中の運動方向位置を推定するため，自己運動パラメータを求めた上で，回転運動成分を取り除き，並進運動のみからなるフローを用いて，FOE（動きの無限遠点）を推定した．提案手法は，既存の視覚的顕著性マップモデルよりも高精度に一人称視点における視覚的注意の推定が可能であることを，評価実験により確認し，提案手法の有用性を示した．

## 5.2 今後の課題と展望

本研究で提案した一人称視点における視覚的注意推定モデルは、我々人間の日常的な注意を扱うものであるにも関わらず、これまであまり研究されてこなかった分野である。したがって、複数の今後の課題があり、更なる発展が望まれる。本節では、今後の課題と、その解決策の方向性についての検討を述べる。

自己運動と視覚的注意の関係については、本研究で用いた運動方向知覚以外にも多数の研究がなされており、視覚的注意推定に有用と考えられる知見が存在する。たとえば、自由に移動でき、能動的に注視行動が可能な状況における視線・頭部運動・身体運動の関係として、視線が最初に移動し、その後頭部運動、身体運動の順で追従する性質があることが知られている [67]。本研究の提案手法では、並進運動に対する効果的な注意推定が可能であったが、並進運動が非常に少なく、回転運動の多い映像では、性能低下は起きないものの、明確な性能向上はみられなかった。しかし、上述の頭部回転運動が視線に追従して起きるといった知見を用いることで、回転運動に対しても、効果的な注意推定が期待できる。たとえば、現在以前のフレームだけでなく、後のフレームを用いて自己運動軌跡を解析することで、頭部回転運動に先行して回転方向に注意が向かうといった推定が可能になるだろう。ただし、未来のフレームを必要とするため、仮に処理が非常に高速であったとしても、リアルタイム処理は不可能となる。

本研究では、被験者実験（第 3.3 節）によって、既存の視覚的顕著性マップモデルは人間が持つ自己運動補償メカニズムを考慮しておらず、動的な特徴マップが性能を低下させていることを、明らかにした。しかし、図 3.4 や、図 3.5 のように、運動物体や時間的な輝度変化が、視覚的注意を誘導することもまた事実であり、自己運動を考慮したうえで、これらを正確に推定することは、重要である。運動物体に対する注意を推定するためには、移動カメラから撮影された映像における運動物体分離の手法 [68] などを応用することが有用だろう。また、自己運動を伴う一人称視点映像から、時間的な輝度変化領域を正確に推定するためには、3 次元空間の情報が必要となるだろう。本研究でも利用した、カメラの運動パラメータ（並進ベクトル  $t$ 、回転行列  $R$ ）を用いれば、特徴点对応による疎な 3 次元形状復元が可能となる。移動カメラ等により、自己位置推定と、環境地図作成を同時に行う技術は SLAM と呼ばれ、PTAM [69] 等、多数研究されているため、これらを応用することで、時間的な輝度変化による注意推定が一人称視点においても有効に活用できると期待できる。

更なる将来的な課題として、各フレームにおける画像中の注意の位置、すなわち、空間的な注意の推定だけでなく、あるフレームよりも、その前や後のほうが強く視覚的注意を向けやすい（または向けにくい）といった、時間的な注意（集中）の推定を組み合わせ、時空間的な注意推定があげられる。眼球運動を計測することで、集中・注意状態を推定する研究は存在し、それによる学習支援システムも提案されている [70]。しかし、一人称視点映像から、注意状態を推定するような研究は知られていない。時間的な注意推定が可能となれば、一人称視点映像から、注意が集中されたフレームのみを取り出した上で、本研究で提案したような空間的な注意推定を行うことで、行動理解や動的イベントの認識、行動支援において、より効果的な活用が期待できる。

## 付録 A

# 自己運動の推定

第 4.2 節では前方並進運動の判定と，並進運動のみによる FOE 推定のために，カメラの運動パラメータを用いた．本付録では，映像からカメラの運動パラメータを推定する手法について説明する．カメラの運動パラメータは，焦点距離以外の内部パラメータが既知であれば 2 フレーム間の点対応から計算することが可能である [71]．まず，自己運動推定の基本となる理論としてエピポーラ幾何を紹介し，次に，エピポーラ方程式に基づいて，自己運動パラメータを推定する手順を説明する．

### A.1 エピポーラ幾何

運動パラメータ推定の準備として以下に，同一のカメラで撮影された 2 画像間の幾何関係を定めるエピポーラ幾何について説明する [72]．2 つの画像の光学中心を  $C, C'$  としたとき，3 点  $P, C, C'$  を通る平面をエピポーラ平面，エピポーラ平面とそれぞれの画像面とが交わる線をエピポーラ線という．また，直線  $CC'$  がそれぞれの画像面と交わる点をエピポールといい，すべてのエピポーラ線はエピポールを通る．

ここで， $C'$  のカメラ座標系を基準とし，2 つの画像間の相対的な位置関係を表す回転行列と並進ベクトルを  $R, t$ ，点  $P$  のそれぞれの画像への投影点を正規化画像座標で  $x, x'$  とすると，図 A.1 に示すように，3 つのベクトル  $\tilde{x}', R\tilde{x}, t$  はすべてエピポーラ平面内に存在する（ただし，ベクトル記号の上の $\sim$ は，そのベクトル記号が表す座標の同次座標を意味する）．すなわち，この 3 つのベクトルからなるスカラー 3 重積は 0 となることから，次の式が成り立つ．

$$\tilde{x}'^T (t \times R\tilde{x}) = 0 \quad (\text{A.1})$$



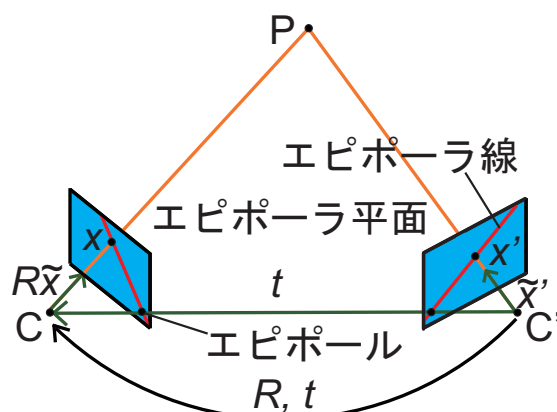


図 A.1: エピポーラ幾何とエピポーラ平面内の 3 つのベクトル．2 つの画像の光学中心を  $C, C'$  としたとき，3 点  $P, C, C'$  を通る平面をエピポーラ平面，エピポーラ平面とそれぞれの画像面とが交わる線をエピポーラ線という．また，直線  $CC'$  がそれぞれの画像面と交わる点をエピポールといい，すべてのエピポーラ線はエピポールを通る．2 つの画像間の相対的な位置関係を表す回転行列と並進ベクトルを  $R, t$  とし，点  $P$  のそれぞれの画像への投影点を正規化画像座標で  $x, x'$  とすると，図 A.1 に示すように，3 つのベクトル  $\tilde{x}', R\tilde{x}, t$  はすべてエピポーラ平面内に存在する．

さらに、

$$\boldsymbol{E} = \boldsymbol{t} \times \boldsymbol{R} \quad (\text{A.2})$$

と置き換えることで，最終的に，2つの画像における投影点の正規化画像座標  $x, x'$  の間には次の関係式が成り立つ．ただし，ベクトル  $t$  と行列  $R$  の積  $t \times R$  は， $t$  と  $R$  の各列とのベクトル積を列とする行列である．

$$\tilde{\mathbf{x}}'^{\text{T}} \mathbf{E} \tilde{\mathbf{x}} = 0 \quad (\text{A.3})$$

ここで、 $E$  は  $3 \times 3$  の行列で、基本行列と呼ばれる．続いて、上で述べた関係を一般的な画像座標を用いて表す．それぞれの画像への投影点の画像座標を  $m, m'$  とすれば、カメラの内部パラメータ  $A$  を用いて次のように表される．ただし、 $\sim$  はその両辺が定数倍の違いを許して等しいことを意味する．

$$\tilde{m} \sim A\tilde{x} \quad (\text{A.4})$$

$$\tilde{m}' \sim A\tilde{x}' \quad (\text{A.5})$$

式 A.1, 式 A.4, 式 A.5 により, 次の関係式が得られる.

$$\tilde{m}'(A^{-1})^T t \times R A^{-1} \tilde{m} = 0 \quad (\text{A.6})$$

さらに,

$$F = (A^{-1})^T t \times R A^{-1} \quad (\text{A.7})$$

と置き換えることで, 最終的に, 2 つの画像における投影点の画像座標  $m, m'$  の間には次の関係式が成り立つ.

$$\tilde{m}' F \tilde{m} = 0 \quad (\text{A.8})$$

ここで,  $F$  は  $3 \times 3$  の行列で, 基礎行列と呼ばれる. ここで, 式 A.3, 式 A.3 は, 2 つの画像上の投影点の間の拘束条件 (エピポーラ拘束) が存在することを示しており, 式 A.3 をエピポーラ方程式と呼ぶ. また, 2 つの画像上の投影点の対応から得られる情報は, このエピポーラ方程式 A.3 を満たすような基礎行列  $F$  に集約される.

## A.2 運動パラメータの推定

運動パラメータ推定のため, まずは 2 つの画像の対応点のうち 8 点を用いて, 8 点アルゴリズム [73] により基礎行列  $F$  を推定する. 安定な特徴点对応を得るため, 本研究では KLT 特徴点追跡法 [74, 75] により, 過去  $n$  フレームにおいて連続して追跡できた特徴点を用いた. 8 点アルゴリズムを用いる際, RANSAC [76] 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  $A$  と, 得られた基礎行列  $F$  を用いて, 運動パラメータ (並進ベクトル  $t$ , 回転行列  $R$ ) は, 次の手順で計算することができる [77, 78].

1. 式 A.2, 式 A.7 より, 基本行列  $E$  は次式により計算できる.

$$E = A^T F A \quad (\text{A.9})$$

2. 式 A.2 より, 並進ベクトル  $t$  は, 基本行列  $E$  の全ての列に直行する ( $E^T t = 0$  が成り立つ) ため, 対称行列  $EE^T$  の最小固有値に対する単位固有ベクトルとして,  $t$  を計算する.

3. 固有ベクトル  $t$  は符号が不定であるため，対応点の組の正規化画像座標  $x_\alpha, x'_\alpha (\alpha = 1, \dots, 8)$  に対し，次の条件が成り立てば， $t$  の符号を反転させる．

$$\sum_{\alpha=1}^8 |t, x_\alpha, Ex'_\alpha| < 0 \quad (\text{A.10})$$

ただし， $|a, b, c|$  はベクトル  $a, b, c$  のスカラ三重積である．

4. 行列  $-t \times E$  を次のように特異値分解する．

$$-t \times E = U \Lambda V^T \quad (\text{A.11})$$

ここで， $\Lambda$  は特異値を大きい順に対角成分に並べた対角行列であり， $V, U$  は直交行列である．

5. 式 A.2 の関係より，回転行列  $R$  を次のように計算する．

$$R = U \text{diag}(1, 1, \det(UV^T)) V^T \quad (\text{A.12})$$

ただし， $\text{diag}(a, b, c)$  は対角要素が  $a, b, c$  の対角行列を表す．

上述のように，運動パラメータ  $t, R$  が得られたが，ここで並進ベクトル  $t$  は，上記の手順 2 において単位ベクトルとして得られている．これは基礎行列  $F$  に定数倍の不定性があるためである．符号の不定性に関しては，手順 3 において  $E$  と適合させることができるが，大きさは不定のままである．

## 参考文献

- [1] 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.
- [2] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Transactions on Graphics (SIGGRAPH)*, Vol. 26, , July 2007.
- [3] M. Nishiyama, T. Okabe, Y. Sato, and I. Sato. Sensation-based photo cropping. In *Proceedings of the seventeen ACM international conference on Multimedia*, MM '09, pp. 669–672. ACM, 2009.
- [4] L. Marchesotti, C. Cifarelli, and G. Csurka. A framework for visual saliency detection with applications to image thumbnailing. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 2232–2239. IEEE, 2009.
- [5] Y.Y. Chang and H.T. Chen. Finding good composition in panoramic scenes. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 2225–2231. IEEE, 2009.
- [6] L. Itti, N. Dhavale, F. Pighin, et al. Realistic avatar eye and head animation using a neurobiological model of visual attention. In *SPIE 48th Annual International Symposium on Optical Science and Technology*, Vol. 5200, pp. 64–78, 2003.
- [7] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *Advances in Neural Information Processing Systems (NIPS)*, Vol. 19, pp. 545–552, 2006.
- [8] M.I. Posner. Orienting of attention. *The Quarterly journal of experimental psychology*, Vol. 32, No. 1, pp. 3–25, 1980.
- [9] B. Khurana and E. Kowler. Shared attentional control of smooth eye movement and perception. *Vision Research*, Vol. 27, No. 9, pp. 1603–1618, 1987.
- [10] 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.
- [11] 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.
- [12] A.M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, Vol. 12, No. 1, pp. 97–136, 1980.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] L.F. Costa. Visual saliency and attention as random walks on complex networks. *ArXiv Physics e-prints*, 2006.
- [17] W. Wang, Y. Wang, Q. Huang, and W. Gao. Measuring visual saliency by site entropy rate. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 2368–2375. IEEE, 2010.
- [18] 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.
- [19] L. Itti. Quantitative modelling of perceptual salience at human eye position. *Visual Cognition*, Vol. 14, No. 4, pp. 959–984, 2006.
- [20] 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.
- [21] I.P. Howard. The optokinetic system. *The Vestibulo-ocular Reflex and Vertigo*, pp. 163–184, 1993.
- [22] L. M. Ward. Attention. *Scholarpedia*, Vol. 3, No. 10, p. 1538, 2008.
- [23] D.E. Broadbent. *Perception and communication*. Pergamon Press, 1958.
- [24] D. Parikh, C. Zitnick, and T. Chen. Determining patch saliency using low-level

- context. In *European Conference on Computer Vision (ECCV) : Part II*, pp. 446–459. Springer, 2008.
- [25] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 2106–2113. IEEE, 2009.
- [26] A. Treisman. Features and objects in visual processing. *Scientific American*, Vol. 255, No. 5, pp. 106–115, 1986.
- [27] A. Treisman, P. Cavanagh, B. Fischer, V.S. Ramachandran, and R. von der Heydt. Form perception and attention: Striate cortex and beyond. *Visual perception: The neurophysiological foundations*, pp. 273–316, 1990.
- [28] L. Elazary and L. Itti. Interesting objects are visually salient. *Journal of Vision*, Vol. 8, No. 3, pp. 1–15, 2008.
- [29] Y. Fu, J. Cheng, Z. Li, and H. Lu. Saliency cuts: An automatic approach to object segmentation. In *International Conference on Pattern Recognition (ICPR)*, pp. 1–4. IEEE, 2008.
- [30] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof. Saliency driven total variation segmentation. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 817–824. IEEE, 2009.
- [31] K. Fukuda, T. Takiguchi, and Y. Ariki. Automatic segmentation of object region using graph cuts based on saliency maps and adaboost. In *IEEE International Symposium on Consumer Electronics (ISCE)*, pp. 36–37. IEEE, 2009.
- [32] G. Wang and D. Forsyth. Joint learning of visual attributes, object classes and visual saliency. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 537–544. IEEE, 2009.
- [33] Y. Pritch, E. Kav-Venaki, and S. Peleg. Shift-map image editing. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 151–158. IEEE, 2009.
- [34] N.T.N. Anh, W. Yang, and J. Cai. Seam carving extension: a compression perspective. In *Proceedings of the seventeen ACM international conference on Multimedia*, pp. 825–828. ACM, 2009.
- [35] T. Ren, Y. Liu, and G. Wu. Image retargeting using multi-map constrained region warping. In *Proceedings of the seventeen ACM international conference on Multimedia*, pp. 853–856. ACM, 2009.

- 
- [36] Stephan Kopf, Johannes Kiess, Hendrik Lemelson, and Wolfgang Effelsberg. Fscav: fast seam carving for size adaptation of videos. In *Proceedings of the seventeen ACM international conference on Multimedia*, MM '09, pp. 321–330. ACM, 2009.
- [37] Stella X. Yu and Dimitri A. Lisin. Image compression based on visual saliency at individual scales. In *Proceedings of the 5th International Symposium on Advances in Visual Computing: Part I*, ISVC '09, pp. 157–166, Berlin, Heidelberg, 2009. Springer-Verlag.
- [38] N. Dhavale and L. Itti. Saliency-based multifoveated MPEG compression. In *Proceedings of the 7th International Symposium on Signal Processing and Its Applications*, Vol. 1, pp. 229–232. IEEE, 2003.
- [39] L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, Vol. 13, No. 10, pp. 1304–1318, 2004.
- [40] X. Sun, H. Yao, R. Ji, and S. Liu. Photo assessment based on computational visual attention model. In *Proceedings of the seventeen ACM international conference on Multimedia*, MM '09, pp. 541–544, New York, NY, USA, 2009. ACM.
- [41] J. You, A. Perkis, M. M. Hannuksela, and M. Gabbouj. Perceptual quality assessment based on visual attention analysis. In *Proceedings of the seventeen ACM international conference on Multimedia*, MM '09, pp. 561–564, New York, NY, USA, 2009. ACM.
- [42] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision (IJCV)*, Vol. 42, No. 3, pp. 145–175, 2001.
- [43] TE Starner. Wearable agents. *IEEE Pervasive Computing*, Vol. 1, No. 2, pp. 90–92, 2002.
- [44] 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 helthcare. In *International Workshop on Content-Based Multimedia Indexing, 2008. (CBMI 2008.)*, pp. 101–108. IEEE, 2008.
- [45] L. Sun, U. Klank, and M. Beetz. EYEWATCHME 3D Hand and object tracking for inside out activity analysis. In *IEEE Computer Society Conference on*

- Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. First Workshop on Egocentric Vision.*, pp. 9–16. IEEE, 2009.
- [46] 日本視覚学会. 視覚情報処理ハンドブック. 朝倉書店, 2000.
- [47] J. J. Gibson and L. Carmichael. *The perception of the visual world*. Houghton Mifflin Boston, 1950.
- [48] M. Fukuchi and C. Koch. The focus of expansion acts as a cue for visual attention. *Journal of Vision*, Vol. 8, No. 6, p. 880, 2008.
- [49] 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. 137, 2009.
- [50] Sébastien Hillaire, Anatole Lécuyer, Gaspard Breton, and Tony Regia Corte. Gaze behavior and visual attention model when turning in virtual environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology, VRST '09*, pp. 43–50, New York, NY, USA, 2009. ACM.
- [51] C.A. Rothkopf, D.H. Ballard, and M.M. Hayhoe. Task and context determine where you look. *Journal of Vision*, Vol. 7, No. 14, 2007.
- [52] C.A. Rothkopf and D.H. Ballard. Image statistics at the point of gaze during human navigation. *Visual neuroscience*, Vol. 26, No. 01, pp. 81–92, 2009.
- [53] D.H. Ballard and M.M. Hayhoe. Modelling the role of task in the control of gaze. *Visual cognition*, Vol. 17, No. 6, pp. 1185–1204, 2009.
- [54] A.M. Derrington, J. Krauskopf, and P. Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology*, Vol. 357, No. 1, pp. 241–265, 1984.
- [55] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and CH Anderson. Overcomplete steerable pyramid filters and rotation invariance. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 222–228. IEEE, 1994.
- [56] nac Image Technology Inc. EMR-9. <http://www.nacinc.com/products/Eye-Tracking-Products/EMR-9/>, 2008.
- [57] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice Hall, 5 1982.
- [58] 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.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] WH Warren. The state of flow. *High-Level Motion Processing: Computational, Neurobiological and Psychophysical Perspectives*, pp. 315–358, 1998.
- [63] 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.
- [64] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*, 2nd ed. Cambridge Univ Pr, 2003.
- [65] 向井利春, 大西昇. 2次元動画像中の特徴点の追跡方法. 特開 2000 - 339473 号, 2000.
- [66] B W Silverman. *Density Estimation for Statistics and Data Analysis*. London : Chapman and Hall, 1986.
- [67] 鈴木利友, 須貝成芳, 岡崎甚幸. 迷路空間における移動方法と注視行動の関係に関する研究 : 能動的探索歩行と車椅子による受動的移動の比較を通して. 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. 102, No. 44, pp. 13–18, 2002-05-03.
- [68] Yasuyuki Sugaya and Kenichi Kanatani. Multi-stage optimization for multi-body motion segmentation. *IEICE Transactions on Information and Systems*, Vol. E87-D, pp. 1935–1942, 2004.
- [69] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pp. 225–234. IEEE, 2008.
- [70] 中村亮太, 井上亮文, 市村哲, 岡田謙一, 松下温. 「ghost-tutor」 : 個人の学習ペースを考慮した学習支援システム. 情報処理学会論文誌, Vol. 47, No. 7, pp. 2099–2106, 2006-07-15.
- [71] 金谷健一, 松永力. 基礎行列の分解: 焦点距離の直接的表現. 情報処理学会研究報告, pp. 120–7, 2000.
- [72] 奥富正敏, 小沢慎治, 清水雅夫, 堀修, 江尻正員, 北本朝展, 後藤道子, 斎藤英雄, 佐藤洋一, 千葉直樹ほか. デジタル画像処理. CG-ARTS 協会, 2004.

- 
- [73] 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.
  - [74] C. Tomasi and T. Kanade. Detection and tracking of point features. *Carnegie Mellon University Technical Report CMU-CS-91-132*, 1991.
  - [75] J. Shi and C. Tomasi. Good features to track. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 593–600. IEEE, 1994.
  - [76] 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.
  - [77] 金谷健一. 画像理解 3次元認識の数理 . 森北出版, 1990.
  - [78] K. Kanatani. *Statistical optimization for geometric computation: theory and practice*. Elsevier, Amsterdam, 1996.

# 発表文献

## 国際会議

- [1] K. Yamada, Y. Sugano, T. Okabe, Y. Sato, A. Sugimoto and K. Hiraki: “Can Saliency Map Models Predict Human Egocentric Visual Attention?,” in Proc. International Workshop on Gaze Sensing and Interactions(IWGSI2010), Queenstown, New Zealand, Nov. 2010.

## 国内会議（査読なし）

- [2] 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顕著性マップモデルの性能評価,” ヒューマン情報処理研究会 (HIP), Okinawa, Feb. 2011.(発表予定)

## 修士論文に含まれない発表文献

### 論文誌

- [3] 山田健太郎, 山崎俊彦, 相澤清晴: “シルエット抽出とモデル生成の相互フィードバックによる Time-Varying Mesh の高精度化,” 電子情報通信学会論文誌 D, vol. J93-D, no, 8, pp. 1533-1543, Aug. 2010.

## 国際会議

- [4] T. Yamasaki, K. Yamada, K. Aizawa: “Time-Varying Mesh Generation Based on Iterative Feedback between Silhouette Extraction and Geometry Modeling,” in Proc. APSIPA ASC 2009 : Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, pp.502–508, Sapporo, Oct. 2009.

## 国内会議（査読あり）

- [5] 山田健太郎, 山崎俊彦, 相澤清晴: “シルエット抽出とモデル生成の相互フィードバックによる Time-Varying Mesh の欠損補填,” 画像の認識・理解シンポジウム (MIRU2009), OS13-1 pp.375–382, Matsue, July, 2009.

# 謝辞

本研究を行うにあたり，たくさんの人にお世話になりました．佐藤洋一教授には視覚的顯著性というテーマを与えていただき，終始丁寧にご指導いただきました．心から感謝しています．また，海外での学会発表の機会を与えていただき，これまでにない貴重な経験をすることができました．研究室での生活においても理想的な環境を作るためにご尽力いただき，有意義な研究生活を送ることができました．

本学総合文化研究科の開一夫教授には，視覚心理学に関わる本研究を行うにあたって非常に重要となる，認知心理学分野の研究のご紹介や，貴重なご助言を多数頂き，誠に感謝しています．

研究を進めるにあたり，多くの貴重なご助言を与えてくださった国立情報学研究所の杉本晃宏教授に深く感謝いたします．

本研究室の岡部孝弘助教には，研究をはじめとする様々な面で何度も相談にのって頂き，大変お世話になりました．深くお礼申し上げます．

本研究室特任助教の菅野裕介氏には，頻繁に研究方針のご指摘・ご助言を頂いただけでなく，論文の細かな添削や，研究室での活動まで，熱心にご指導いただきました．本当にありがとうございました．

本研究室研究員の小林理弘氏，杉村大輔氏には研究室での日常的な活動や，研究の相談まで，幅広く支援していただきました．厚くお礼申し上げます．

最後に共に研究生活を送ってきた仲間であるだけでなく，実験にも協力して頂いた，佐藤研究室の皆様に変感謝しています．

2011 年 2 月 9 日

山田 健太郎