

## 顔・人体への誘目性を考慮した視覚探索モデルの提案

釣 大輔 石井 雅博 唐 政 山下 和也

富山大学大学院 〒930-0855 富山県富山市五福 3190 番地

E-mail: m0671117@ems.u-toyama.ac.jp, {ishii,ztang,kazuya}@eng.u-toyama.ac.jp

**あらまし** 画像から人間の視覚的な注意を推測する事はコンピュータビジョンの観点より様々な利点が存在し、より人間に近い推測を実現するため現在様々な計算モデルが提案されている。本研究では人間が視野内において顔や人体などの人的なパターンに対し高い誘目性をもつ点に着目し、これらを考慮した視覚的注意の探索モデルを提案する。

**キーワード** 視覚的注意、動画像処理、パターン認識

### A visual search model contains human body and face detection

Daisuke TSURI Masahiro ISHII Zheng Tang and Kazuya YAMASHITA

Graduate School of Science and Engineering Education, University of Toyama

3190 Gofuku, Toyama-shi, Toyama, 930-8555 Japan

E-mail: m0671117@ems.u-toyama.ac.jp, {ishii,ztang,kazuya}@eng.u-toyama.ac.jp

**Abstract** Guessing visual attention from the image has various advantages in the computer vision. The model of many calculations which aimed at guess like the human is proposed now. Human has high saliency in view for human patterns such as faces and the human bodies. We proposes the visual attention model that detects information on the face and the human body.

**Keyword** Visual attention, Motion analysis, pattern recognition

#### 1. まえがき

近年、コンピュータに人間的な処理を行わせる研究が盛んである。特に人間の視覚機能を再現し、動画像の認識をさせる試みがなされている。網膜は中心窩において最も高い解像度を持つ。人は視野内から必要な情報を選択して観察する必要がある、興味に従い視覚探索を行う。均一な解像度を持つカメラビジョンにおいてこの性質を実現するにあたり、現在まで様々な手法が考案されてきた。視認識における視覚的注意の重要性は非常に高く、その仕組みを普段使用しているシステムに利用する事で安全性・快適性の向上が期待される。

一般的な視覚的注意の算出方法として、Kochらによる誘目性の概念が用いられており、そのモデルとしてIttiらによる誘目性評価地図(saliency map)が作成された[1]。画像の色や輝度など、低次特徴をそれぞれ並列的に計算し、統合することにより評価地図を作成する。誘目性評価地図は網膜から得られる空間座標と対応しており、地図上で評価値の高い箇所はすなわち目を引く要素を多く持つ。すなわちボトムアップ性の注意を模したモデルとも言える。しかし、本来の誘目性は網膜から得られる情報より決定されるボトムアップ性の

注意と、記憶や経験から決定されるトップダウン性の注意の2点より構成されており、未解明の部分が多く、新たなモデルの作成が望まれている。

人は視野内の人間的な特徴に対し、高い誘目性を持ち、様々な画像の特徴から視野内の”人間らしさ”を持つ対象を検出する。これは反射的に行われる事が多く、従来のボトムアップ処理に加える事で、より人に近い処理結果が得られるものと思われる。そこで、本研究では人体検出・顔検出を処理に加えた誘目性算出方法を提案する。

また、一般に注意と注視点は独立して働くものであり、切り離して考えるべき事柄ではあるが、本研究では解析の便宜上、注意が最も向きやすい点を注視点とする。

## 2. 視覚探索モデル

### 2.1. モデルの概要

本研究では入力動画像に対し注意を予測するモデルを提案する。解析の流れを図1に示す。

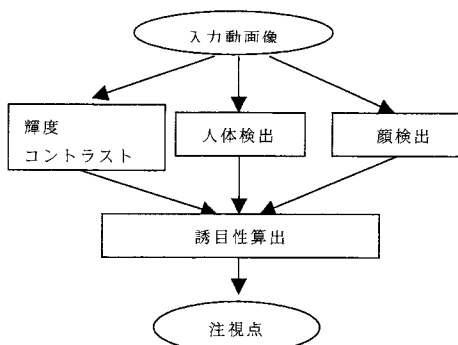


図1 注意モデルの流れ

本モデルによる誘目性評価値図の作成の為、入力動画像に対し初期段階として各要素の計算を行う。要素はそれぞれ①従来法による空間的・時間的輝度コントラストの抽出、②人体パターンの検出、③顔パターンの検出である。その後、それらを統合し、誘目性評価値図を作成する。

### 2.2 輝度コントラストからの視覚的注意算出

動画像を解析する際、saliency map では空間的な情報からの誘目性評価地図算出を行う。これを動画像の解析に適用した場合、時々刻々と変化する網膜からの情報の変化に正確な対応がなされない場合がある。この問題の解決の為、本研究では塩入らによる動的注意移動モデル[2]を用いる。この手法では輝度コントラストの空間的・時間的変化を特徴とし、連続した画像に対しての反応を再現している。

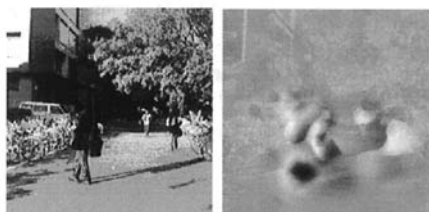


図2 輝度コントラストによる注意モデル

3名の人物がそれぞれの方向を向いて移動する動画像を入力したモデルのあるフレームの結果を図2に

示す。空間的なコントラストであるエッジの抽出と、時間的なコントラストであるフレーム間の輝度変化を抽出している。動きの大きい箇所に対して高い評価値を得る事が可能であり、中央付近を移動している3名に対し特に強い反応が出ている。

### 2.3 人体検出

足から頭まで全身の確認が可能な物体を人体として検出を行った。動画像において連続したフレーム間の差分を取得し(図3左)、抽出した動物体の縦横比から人体に近い物体を検出した(図3右)。



図3 フレーム間差分による人体の検出

### 2.4 顔検出

画像中より顔パターンの検出を行う。検出は画像処理ライブラリである openCV[3]により提供されている機能である face detection を用いる。検出結果の一例を図4に示す。

face detection は Haar-like 特徴による学習済みのパターンから入力画像中の適合率の高い領域を抽出する手法であり、パターンのサイズ変化に強く、パターンマッチング法に比べ高速である。誤検出領域も存在し、本来ならば除外すべき項目ではあるが、これについて人が顔以外の、顔らしい特徴に対しても注意が引かれてしまう点と照らし合わせ、検出領域として等しく扱う。



図4 顔パターンの検出

## 2.5 誘目性算出

前述の過程で得られたデータを足し合わせ、誘目性評価値図を作成する。輝度コントラストの情報と顔・人体検出の結果の関係について現在まで知見は発表されておらず、本研究では経験的に最良の結果となる重み付けを行い足し合わせ、誘目性評価地図を作成した。

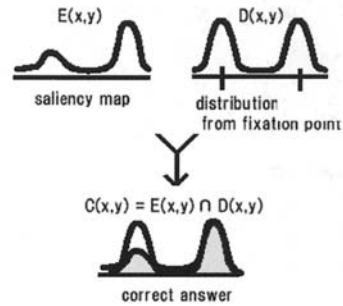


図4 正答率算出処理

正答率算出の概要を図4に示す。

注視点分布図に対し評価地図が満たす評価値を抽出し、平面全ての値を足し合わせたものを正答率とする。被験者計測による全ての被験者の注視点を通じた場合正答率は1に近づく。

## 3 評価実験

### 3.1 評価の概要

モデルの有効性を示すため、評価実験を行う。

本研究では顔検出・人体検出結果を誘目性評価地図に加える事で従来法による注意予測結果より結果の向上を狙う。そのため検出要素の組み合わせによりいかなる予測結果になるかを検証する。モデルによる注意予測の結果と実際の被験者注意計測結果を比較し、予測精度を計測する。

誘目性評価地図上で評価値の高い箇所は、実際に注視がなされる確率の高い箇所であると言える。しかし、すべての人が同じ対象を注視するのは極めて稀である。これはトップダウン性の注意が人の経験や記憶を反映しており、個人差が発生するためである。そのため、評価値の特に高い1点をモデルの解とするのは良い方法ではない。特定の対象への評価の偏りを防ぐため、人がその状況でとりうる注意の分布率を示すのが好ましい。

### 3.2 評価方法

被験者計測による注視点の分布図を評価性地図と同じ空間座標で作成し、誘目性評価地図での注意の領域と比較した。評価地図上で示す領域が、分布図を全て満たした場合に高い正答率を得るものとした。

比較を行うため、モデルによる評価地図と被験者計測による注視点の分布図の総和をそれぞれ1に正規化する。注視点の分布図は被験者N名の注視点計測結果から作成する。それぞれの注視点より半径 $2^\circ$ の円を範囲とした体積 $1/N$ の正規分布として与え、被験者全員の合計を算出した。

これらを比較し、正答率の計算を行う。評価地図を $E(x,y)$ と注視点分布図を $D(x,y)$ としたとき正答率 $R$ は以下の式で計算される。

$$C(x,y) = \min(E(x,y), D(x,y)) \quad \dots (1)$$

$$R = \iint C(x,y) dx dy \quad \dots (2)$$

### 3.3 取得環境

動画像を被験者に提示し、注視点移動を眼球運動計測装置により計測した。計測装置はnac社製EMR-8を台に固定し使用、画角 $44 \text{ deg}$ 、距離 $2.0 \text{ m}$ 、初期注視点は動画像の中心で、呈示時に特定のタスクは課さないものとする。被験者数は7名である。

実験装置の略図を図5に示す。

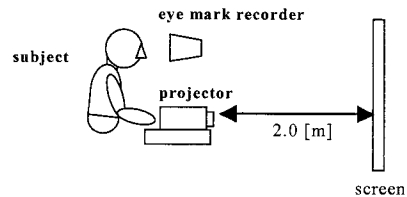


図5 実験装置の略図

### 3.4 呈示動画像

顔・人体情報が注視点予測に与える影響を検討するため、検出可能な情報を含む動画像に対しての有効性を検証する。

呈示動画像は顔・人体の情報を含む自然動画像を5点選択した。動画像は全てサイズ $256 \text{ pixel} \times 256 \text{ pixel}$ 、 $30 \text{ fps}$ 、 $450 \text{ frame} (15 \text{ sec})$ 、輝度画像になるように加工した。

評価に際し動画像に顔・人体検出処理を行い、動画像の全再生時間中に占めるパターン検出した時間の割合を表1に示す。

表1 呈示動画像における顔パターン・人体パターンの検出割合

movie No	Face detect Rate(%)	Body detect Rate(%)
1	99	0
2	75	0
3	50	0
4	0	60
5	10	30

### 3.5 評価

従来法と比較し、顔・人体検出により作成した誘目性評価地図の注意予測率を評価する。従来法(SM)、顔検出(FD)、人体検出(BD)をそれぞれ組み合わせた時の、各呈示動画像に対する注意予測率を図6に示す。

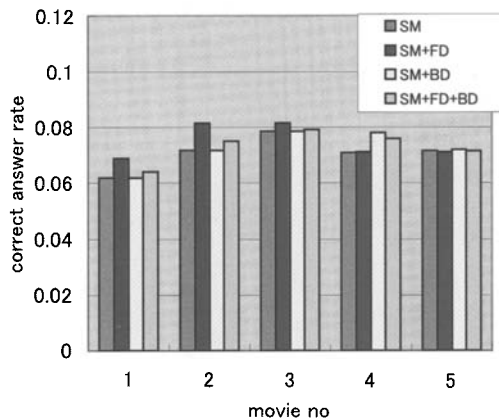


図6 呈示動画像に対する注意予測率

SM+FD、SM+BDモデルについて、それぞれが検出可能な要素を含む動画像に対し、予測結果の向上が見られる。SM+FD+BDモデルはSM+FD、SM+BDモデルと比較し、全ての呈示動画像に対し低い予測結果となっているが、本モデルにおける両特徴同時に検出可能な割合が低いためにこのような結果になったものと思われる。顔検出・人体検出についてより検出率の高いアルゴリズムに変更する事で結果改善が期待される。

### 4 あとがき

本研究では人間にとって特に誘目性の高い、顔・人体などの情報の検出を考慮するモデルを提案した。特徴の検出が可能な動画像に関し、予測の正答率の向上

が見られ、輝度や色などによる低次の特徴とは異なる誘目性の一要因として可能性が示された。

顔や人体はパターン認識を行い抽出するため、低次による特徴のみを扱うボトムアップ的な処理ではなく、記憶や経験により行われるトップダウン的な処理であると考えられる。一般に注意のトップダウン処理は個人の記憶や経験に依存する事が多く、決定的なモデル算出方法が解明されていない。しかし、本研究によって示された結果では、顔・人体に関してほとんどの被験者に対して明らかな誘目性を示す結果となっており、ボトムアップ処理による従来モデルの一要素として組み込み可能ではないかと思われる。

本研究で提案したモデルは一部パラメータとして経験的な手法により最良解を求めて使用しているため、心理学的根拠に乏しい。今後はそれらを心理学実験などの手法により決定していき、更なる結果の向上に努めたい。

### 参考文献

- [1] L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", IEEE Trans. Pattern Anal. Mach. Intell. 20(11) (1998) 1254-1259
- [2] 服部和成, "視覚的注意移動モデルの動画像への適用" Mar.2006 千葉大
- [3] Intel, URL: <http://www.intel.com/technology/computing/openCV>