

## 動画像による視覚的注意モデルの構築

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 従来、静止画像に対し顕著性の地図を作成する事で視覚的注意の特定を行っていたが、静止画像の連続である動画像に対してはあらゆる解析要素が加わり、注意の特定は難しい。本研究では静止画の解析を時間的方向に拡大するための方法として、人の注意量を流動的なものとし、時間変化を行うアルゴリズムを構築する。また、実際に動画像を用いた注意点予測を行い、本手法の有効性を検証する。

キーワード コンピュータビジョン、視覚的注意、動画像処理

## Model of visual attention for dynamic images

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** We were specifying visual attention by making the static image a saliency map. However, the dynamic images have a lot of analytical elements. It is difficult to specify attention. In this research, we constructed the algorithm into which attention changed along with time. We forecast notes by using the natural images, and verified the effectiveness of our algorithm.

**Keyword** Computer Vision, Visual attention, Motion analysis

### 1. まえがき

#### 1.1. 背景

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。特に人間の視覚を代行させるという目的より、動画像の認識をさせる試みは応用範囲も広く、必要不可欠なものである。

動画像の認識における問題の一つとして注視点決定をいかに行うかというものがある。人は注視点を決する際、事前にその対象に注意を向けるが、その場の誘目性や記憶・目的等、様々な要因から決定される。

注意推定方法として様々な手法が考案されてきたが、その最たるものは Itti らによる静止画像に対して複数の画像処理結果を統合した誘目性地図の算出手法である[1]。それは一見して目を引きやすい特徴を有する箇所を推定する方法であり、人の好みや記憶、特に時間的な視点移動の概念は含まれていない。その原因として、それらは非常に大きな個人差(ノイズ)を含むものであり、全ての要因の包含が困難であるからだ。

本研究ではこれに対し、人の注意を流体の移動と捉えた概念を導入し、実際の効果の検証を行う。

なお、本研究では既存の誘目性地図を元に時間経過

による注意移動のシミュレーションに主眼をおくものであり、前処理である誘目性地図(エネルギーマップ)の算出においては詳細の記述を省略する。

また、一般に注意と視点は多対一の関係であり、切り離して考えるべき事柄ではあるが、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. 視覚的注意モデル

#### 2.1. モデルの概要

エネルギーマップを元に、前フレームで算出された注意座標に対して、現在の注意座標を決定する。

決定に使用する要素は、(1)エネルギーマップ、(2)前フレームで決定した注意座標値、(3)前フレームまで場に向けられた累積の注意量である。(3)は注意が時間変化する要因として様々な脳の処理過程が考えられる中、特に復帰抑制についての効果を反映したものである。復帰抑制はすでに注意した空間位置に被験者が再び注意することを防ぎ、新奇な位置へ注意を向けやすくするメカニズムであり、時間経過により発生するものである。また、一度発生した復帰抑制はある時間でピークを迎えた後、時間経過により抑制量が減少するものと思われる。

本モデルでこれを実現する場合、エネルギーマップを元に注意箇所を決定するが、エネルギー値の高い場所に注意が集中するものと思われる。エネルギーマップに値の逆数を高さとした“山”を描いた場合、標高の低い、いわゆる“谷底”に注意が集中する。同じ場所に注意が集中し、時間変化してもそこに注意が留まり続けるなら、前述の復帰抑制により、注意は別の谷底へと移動せざるをえないだろう。

以上は谷底に降る雨に例える事ができる。山間に降った雨(図1①)は地表を流れた後(図1②)近傍にある谷底に堆積し(図1③)、谷底の見かけ上の標高は高くなる。標高が高くなると、次の機会にその谷底に降った雨が集中する確率は減少する(図1④)。新規に降る雨が注意の向きやすさすると、一度向けた注意箇所に対し次の注意の機会を阻害する意味をこめて堆積した雨は注意抑制量と置き換えることができる。

また、堆積した雨は時間経過により地面に吸収される。これは堆積した注意抑制量が時間経過により消滅してゆく働きに置き換えることができる。

そのための処理方法として本研究ではエネルギーマップを仮想の地面、注意を仮想の降雨、注意抑制量を仮想の堆積した雨としたモデルを提案する。

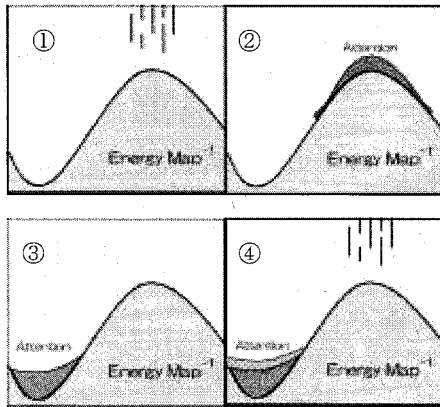


図1 モデルによる注意変化の概念

## 2.2. エネルギーマップの算出

本研究では服部らによる notE+S モデルを使用する。

従来の誘目性地図算出方法は、対象の画像のエッジや色相、彩度などの要素を経験的に選択し、抽出結果を統合したものである。これに対し notE+S モデルでは注意決定処理に使用する脳処理の経路をモデル化した上で、それぞれの処理に対する重み付けを心理実験によって取得し、解析に用いている。

他の誘目性地図算出手法と比較し、このモデルの採

用理由として、時間的処理を考慮した効果が導入しやすい点、入力画像の輝度情報のみという点が挙げられる。

なお、処理の便宜上このモデルで算出した誘目性地図の事をエネルギーマップと呼称する。図2にエネルギーマップとその変換元となる画像を白黒濃淡で示す。マップ中の白領域がエネルギーの高い点である。

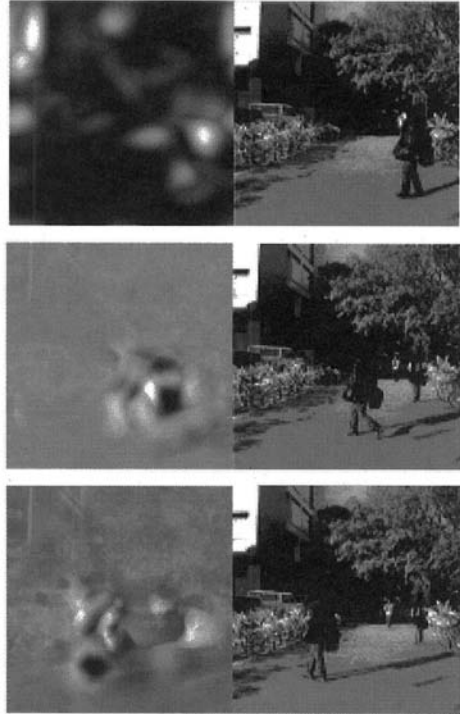


図2 notE+S モデルによるエネルギーマップ

## 2.3. 空間探索

前フレームで求めた注意座標より現在のフレームに雨を降らし堆積を待つ。その上で最も水が堆積した箇所を求め、そのフレームの注意座標とする。

フレーム毎の処理対象となるのは2次元のエネルギーマップであるが、本研究では処理のイメージ上、マップの各セルの値を高さとした3次元の場として処理を行う。

エネルギーマップの各セル値を[0 1]の範囲で正規化し、式(1)を適用する。

$$Emap'(x, y) = 1 - Emap(x, y) \quad \dots (1)$$

これにより標高0(谷底)が最もエネルギーマップとして値の高い点となる。

次に前フレームで決定した注意座標に一定量の雨を

降らす。その後、水を含んだセルは高低差に応じ移動を行う。移動先はセル8近傍より最も標高（エネルギーマップの標高と水の堆積高さの和）が低いセルである。水が移動する際、標高差により流水の勢いが変化する点を考慮して、雨粒の移動元の高さを  $h_1$ 、移動先の高さを  $h_2$  としたときの移動量  $M$  を式(2)に示す。

$$M = d \cdot (h_1 - h_2) \quad \dots (2)$$

ここで係数  $d$  は定数である。全てのセルに対しこの処理を行い、堆積が安定するまで処理を継続する。

## 2.4. 空間探索の時間的拡張

算出したセルに対する水の堆積量を次フレームのエネルギーマップに反映する。水の堆積量はすなわち注意抑制の堆積量であり、次フレームにこれを持ち越す事で抑制効果を維持する。前フレームで求めた堆積量を次フレーム処理で同様に堆積させるが、前フレームとの標高の違いから水が標高の低い箇所へ移動を行うので安定状態になるまでそれを待つ。これにより異なるフレームの同じオブジェクトに対し、注意の抑制が保持される。この様な働きを持つ抑制量を過去数フレーム分、専用の堆積キューに蓄える事で時間変化による累積的な注意抑制効果を実現される。

また、注意抑制が減少する効果に対しては堆積キューの時間経過による最も古いフレームの堆積の排出によって実現する。以上より過去数フレーム分の抑制効果を反映した後、現在フレームへの新規の注水を行う。フレーム間処理の概略図を図3に示す。

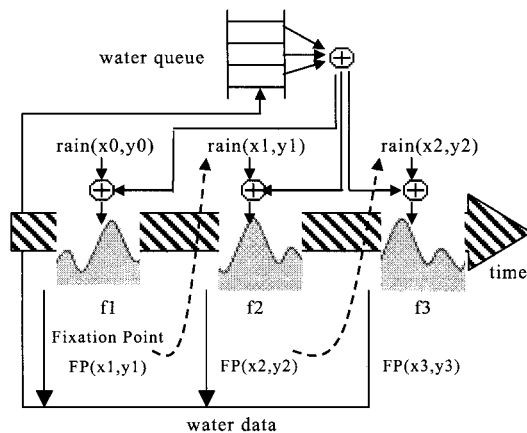


図3 フレーム間処理概略

## 3. 実験

### 3.1. シミュレーション

空間解像度  $256 \times 256$  pixel、時間解像度 30 frame/sec の5秒間のグレースケール化した自然動画に対し、本研究で考案したアルゴリズムを適用した。1秒毎の注視点推移の結果を図4に示す。プログラムに与える初期値として、1フレーム目直前に注視している座標は画像の中心(128,128)とした。また、一様に全ての箇所が見渡せる状態を想定し、プログラム上の視角を  $42^\circ$  とした。

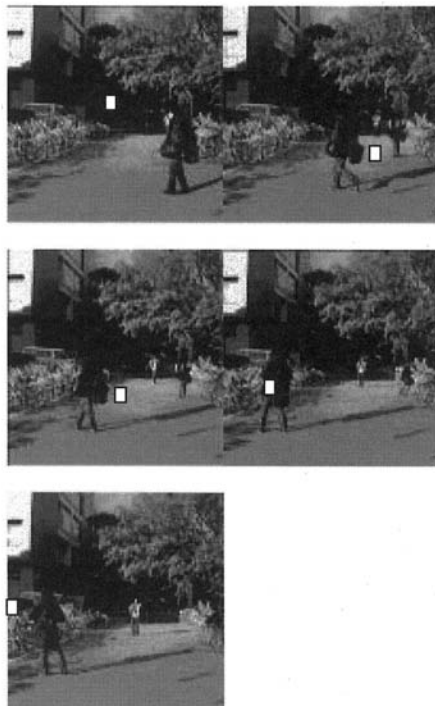


図4 モデルによる注視点算出処理結果

動画は校舎内の駐輪場付近を3名の人が移動している模様で、それぞれ手前を右から左の駐輪場へ横切る人(A)、中心付近を手前に向かって走る白い服を着た人(B)、右側付近をかすかに移動する黒い服を着た人(C)が確認できる。また、画像右上の木々が風でかすかに揺れている。

シミュレーション結果より、プログラムより求めた注視点(白四角)は主にAをやや遅れて追従している他、最終的に人を離れ、左に位置する駐輪場の上を移動している。Aは他2名に比べエネルギーマップ上で常に高いエネルギーを保持し続け、抑制による注視対象排除の効果を得るに至らなかった事がわかる。

最終的に駐輪場の上で待機し続けた理由として、駐輪場の空間周波数が高く、顕著性が高く評価され、かつ直前の注視領域に近かった事が考えられる。この事より現在のモデルでは注視のターゲットとなるオブジェクトを移動するためにある程度の距離制限がかかっており、離れたオブジェクトを探索するためには更なるアルゴリズムの改善が求められる。

流体を模した注意は誘目性地図の地形に沿い移動を行うが、流体の量に限りがあるため、雨が降った場所と距離の近い谷間を注視点として判定するが多い。今回モデルに使用した条件では提示内容全てが視界内に存在し、注視点の移動は一瞬で行われる以上、特にモデルの移動能力に対して考察を行う必要がある。

### 3.2. 実測値との比較

本モデルで算出した座標値と被験者計測実験を行い得た座標値とを比較する。

被験者に眼球運動計測装置を取り付け、注視点の座標を取得した。提示する動画は 3.1 で用いた。記憶によるノイズ発生を抑えるため、初回計測の値を評価に用いる。評価の指標として、座標間距離を用いる。座標(Ax,Ay)と(Bx,By)の距離を求める算出式を式(3)に示す。

$$dist = \sqrt{(Ax - Bx)^2 + (Ay - By)^2} \quad \dots(3)$$

上式を用い、モデル算出した注意座標と計測座標の距離を求めた。

また、本モデルで時間による抑制効果を実現したが、その有無による効果の違いを評価するための比較対象として、エネルギーマップの各フレーム最大値を持つ座標を抽出し、同じく計測座標との距離を求めたものを図5、平均を図6に用意した。以後、モデル算出結果を抑制効果有り条件、エネルギーマップの最大値を持つ座標を抑制効果無し条件と呼ぶ。

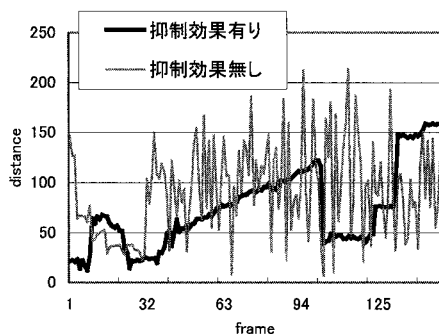


図5 抑制効果の導入結果

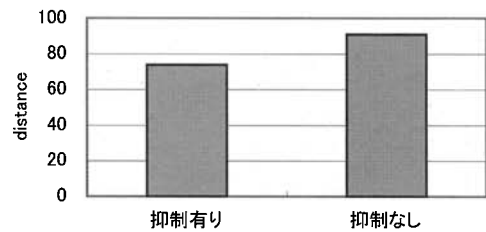


図6 抑制効果の導入結果・平均

図5について抑制効果の有無による各フレーム実測値との距離をそれぞれグラフの太線・細線で示す。抑制効果無し条件に比べ、抑制効果有り条件では座標移動の散乱を抑えており、図6に示す結果と比較して実測に近い解を算出している。

30フレーム以後の大幅な誤差発生については3.1項でも示した通り、モデルのオブジェクト選択を行わない性質が関与しているが、これに関しては個人差が大きいものと予想され、本モデルの概念とはまた別の、選択アルゴリズムの導入が必要である。

### 4. むすび

本研究では人の注意を流体の移動と捉えた注意移動の概念を導入し、効果の検証を行った。

顕著性地図に対して抑制量の時間的変化を加えた結果、注意の散乱を抑える効果は見られたものの、オブジェクト選択に関しては一考の余地がある。

実験では前処理として notE+S モデルを使用した、更なる適応の為に前処理の検討も必要である。

### 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