

視線位置の確率分布モデルを用いた誘目度の推定  
 Estimation of saliency using a stochastic model of gaze

伊東 孝幸<sup>†</sup> 松宮 一道<sup>‡</sup> 栗木 一郎<sup>†</sup> 塩入 諭<sup>†</sup>  
 Takayuki Itoh Kazumichi Matsumiya Ichiro Kuriki Satoshi Shioiri

### 1. はじめに

我々は視覚を通じて外界に関する非常に多くの情報を得ている。脳の多くの領域は視覚情報処理に関連しているが、視野全体の情報量は膨大であり、我々の脳の処理能力ではその全て均等に処理することは不可能であると考えられている。実際に我々は視野の中から重要であると思われる領域を抽出し、その領域について優先的に処理している。たとえば人ごみの中で人探しをする場合、木々や建物、空などの情報を無視し、人間を優先的に処理する。視覚的注意は、そのような選択過程に働く情報抽出のためのメカニズムである。

視覚的注意は入力情報の目立ち具合によって制御されるボトムアップ的注意と、記憶や意思などの高次な要素によって決定されるトップダウン的注意に分けることができる。このうちトップダウン的注意は、脳の高次な領域の処理である上、個人個人によって異なるため、予測することが非常に難しい。一方、ボトムアップ的注意については初期視覚の特性さえモデル化すれば、入力情報のみによって決まると考えることができる。このような考えに基づきボトムアップ的注意を予測するモデルが提案されている。その代表例は Koch らのもので、入力画像の低次の特徴(色、傾き、動きの方向など)を特徴ごとに並列に抽出し、それらを単一の誘目性マップ(Saliency Map)に統合して、誘目性マップの情報をもとに注意方向を決定するというものである[1]。Itti らは Koch らの理論をもとに視覚注意モデルを計算機上に実装し、動画像への適用や視線移動の考慮などを行っている[2][3]。

我々は、視野に依存した資格特性の大きな変化を考慮するため、Itti らの視覚的注意モデルに周辺視の視野特性を取り入れた [4]。視覚的注意を予測するにあたって、視野特性は重要な要素となる。例えば、テレビ画面の右側で小さな虫が飛び始めた場合、右側を注視していた人はそれに気づき、注意を向けるが、左側を注視していた人はそれに気づかない可能性もある。このような現象は人間の視野の不均一から発生するものであり、視野特性を考慮しない注意モデルでは予測することが不可能な現象である。

過去の視覚的注意モデルにおいて視野特性はほとんど議論されていないが、その理由の 1 つは視線位置に関する問題である。視野特性を考慮するためには“注視点”を決める必要があるが、同じ視覚シーンを見ている場合でも、観察者により、また同一観察者でも状況に応じてさまざまな場所を注視し得るため、注視点の移動を一意に決定することは不可能である。モデルの視線予測位置と人間の注視する位置が異なった場合、モデルが視野特性を考慮していると、モデルと人間の“網膜像”が異なってしまう事態となる。視野特性を考慮するモデルでは、注視位置が正しく推定されていることが前提となる。

我々のモデルでは周辺視の視野特性を考慮するため、視野位置ごとのコントラスト感度特性に基づいた特徴抽出を行う。このモデルにおいて、注視位置が既知の条件では、視野特性を取り入れることにより予測能力が大幅に向上することが確認された。しかしながら、長時間で視線移動を伴う条件では、その優位性は認められなかった。人間とモデルの視線位置の違いの影響であると考えられる。

本研究ではこの問題を解決するために、注視点位置を確率的に予測し、その分布に基づいた誘目度の推定を行うシステムを提案する。

### 2. モデル

本モデルの概略を図 1 に示す。モデルでは、視線位置の確率分布  $E(t)$  を考える。視線位置の確率分布は、ある点  $x$  に視線が向けられる確率をマップ化したものである。実験的には、動画像を多数の被験者に見せたときの視線位置から求めることが出来るが、今回は視覚的注意モデルの出力から視線位置分布の推定を行う。視線と注意は別の方向に向けることができるが、一般的には、両者には高い相関がある。本研究では、視線は注意と同一方向に向けられるものと仮定し、誘目性マップの値から視線の存在確率を導出する。すなわち式 1 により視線位置確率を算出する。

$$\begin{aligned} E(t) &= \{e(t, \mathbf{y})\}_{\mathbf{y} \in I} \\ S(t) &= \{s(t, \mathbf{y})\}_{\mathbf{y} \in I} \quad (1) \\ e(t, \mathbf{y}) &= \frac{1}{C} s(t-1, \mathbf{y}) \end{aligned}$$

ただし  $E(t)$ ,  $e(t, \mathbf{y})$  は視線位置確率分布およびそのピクセル値、 $S(t)$ ,  $s(t, \mathbf{y})$  は誘目性マップおよびそのピクセル値、 $\mathbf{y}$ ,  $t$ ,  $1/C$  はそれぞれ座標、時間、正規化定数である。

人間の視野は不均一であることから、時刻  $t$  における誘目性は、その瞬間の網膜像に依存する。言い換えれば、時刻  $t$  における入力画像および視線位置に依存する。我々のこれまでの研究では、時刻  $t$  において点  $\mathbf{x}$  を注視していたときの誘目性マップ  $S_f(t, \mathbf{x})$  を導出するモデルを構築した[4]。このモデルは、注視点からの距離とコントラスト感度に基づいて特徴抽出を行い、誘目性を計算するものである。本研究ではこのモデル出力  $S_f(t, \mathbf{x})$  および視線位置の確率分布  $E(t)$  を用い、誘目度の計算を行う。視線位置の確率分布を用いて算出した誘目性マップを  $S(t)$  とすると、それは式 2 により算出される。

$$S(t) = \sum_i e(t, \mathbf{x}_i) S_f(t, \mathbf{x}_i) \quad (2)$$

<sup>†</sup> 東北大学大学院情報科学研究科 GSIS Tohoku University

<sup>‡</sup> 東北大学電気通信研究所 RIEC Tohoku University

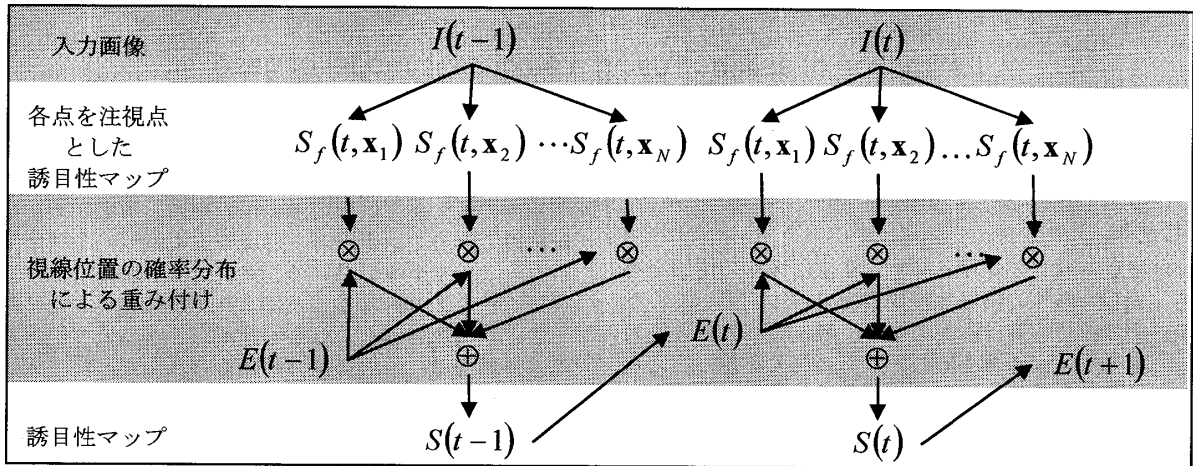


図1 視線位置の確率分布を用いた誘目度の推定

人間はさまざまな場所を注視するが、この計算はそのあらゆる視線を模擬したものだけとすることができる。例えば40%の人が画面左上(例えば  $x_1 = (0.1, 0.1)$ )を、60%の人が右下( $x_2 = (0.9, 0.8)$ )を見るような画像の場合、モデルは左上を注視した場合の誘目性マップ  $S_f(t, x_1)$  と右下を注視した場合の誘目性マップ  $S_f(t, x_2)$  の両方を計算し、それらの重み付け加算を行う。

### 3. 結果

入力動画および、出力された誘目性マップを図2に示す。図中の丸印は、入力動画を被験者(4名)に見せたときの注視位置である。図のように多くの入力動画において、モデル出力と被験者の注視位置に一致が見られた。

本研究では誘目性の計算において視線位置の確率分布を用いた。この効果を見るため、視線位置の情報を用いなかった場合として、視線位置の確率が画面内で一様とした場合との比較を行った。この結果を図3に示す。見た目では大きな変化は確認できないが、両者の差分をとると(図2下)、違いがあることが確認できる。図より、視線位置の確率分布を用いた場合、一様な分布を用いたものと比べて、注視点付近の誘目度が強くなっていることが確認できる。

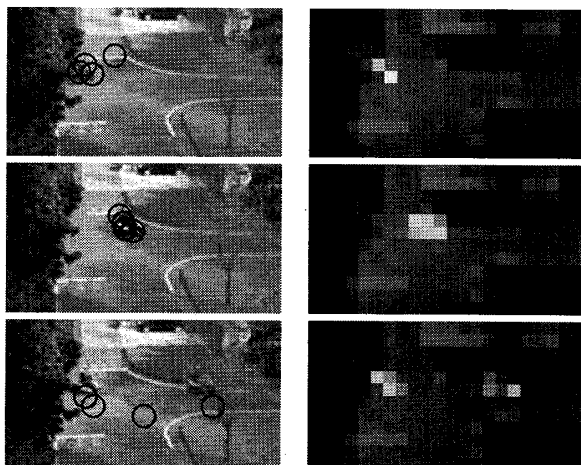


図2 入力動画および出力された誘目性マップ(例)

左: 入力画像(Oは被験者の視線位置)

右: 誘目性マップ

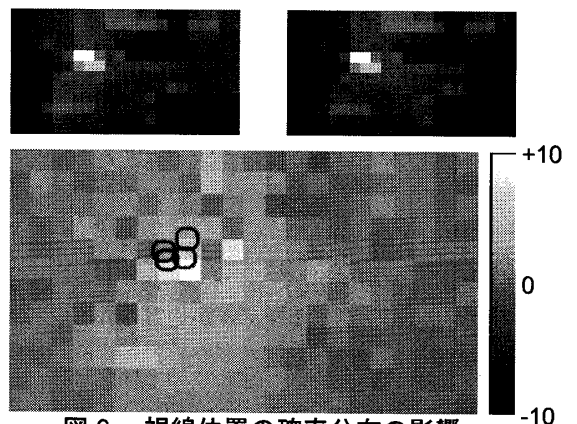


図3 視線位置の確率分布の影響

左上: 視線位置の確率分布を使用して算出した誘目性

右上: 視線位置確率分布を使用しない場合の誘目性

下: 2つの誘目性マップの差分(Oは被験者の注視位置)

### 4. まとめ

本研究では視線位置の確率分布に基づいた誘目度の推定を行った。視線位置の確率分布を用いることで、人間のさまざまな場所への注視を模擬することができた

#### 参考文献

- [1] Koch C and Ullman S, "Shifts in selective visual attention: towards the underlying neural circuitry," Hum Neurobiol., Vol. 4, pp. 219-227 (1985)
- [2] Itti L, Koch C, Niebur E, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, pp. 1254-1259 (1998)
- [3] Itti L, Dhavale N, Pighin F, "Realistic Avatar Eye and Head Animation Using a Neurobiological Model of Visual Attention," Proc. SPIE 48th Annual International Symposium on Optical Science and Technology, Vol. 5200, pp. 64-78 (2003)
- [4] 伊東孝幸, 塩入論, 栗木一郎, 松宮一道, "周辺視の視野特性を考慮した視覚的注意モデルの構築", Vision, vol. 20, no. 3, pp. 171 (2008)