

視線分析手法の高度化とその応用

A Novel Eye Movement Analysis Method and Its Applications

江川 陽*¹

Egawa Akira

白山 晋*²

Shirayama Susumu

*¹ 東京大学大学院工学系研究科

Graduate School of Engineering, the University of Tokyo

*² 東京大学人工物工学研究センター

RACE, the University of Tokyo

Eye movement analysis is a popular method to analyze features of a visual image or human cognitive process. However, several issues exist in the conventional methods make quantitative analysis difficult. In this paper, we enhance an eye movement analysis method to overcome the difficulty. First, eye movement data is acquired, and mapped into the image that segmented by a saliency map model. And then a network is generated by regarding the segments as nodes, and a quantitative measure of importance for the parts of the image is calculated by the PageRank algorithm. Usefulness of our method is discussed through several experiments.

1. 緒言

人間の視線の動きを分析することにより、視覚認知処理の解明、あるいは視覚情報において誘目性が高い領域を抽出する手法に、視線分析がある。近年は、視線追跡装置の低価格化と信頼性の向上がすすみ、視線分析は様々な分野で利用されている [大野 02, 藤本 04, 高久 08]。

一般に視線追跡装置により測定される生の視線データは大量のノイズを含み、さらに個人差があるため、直接的に有用な情報を取得することは難しいとされる [Duchowski 07]。したがって、視線分析においては視線データに様々な処理がなされる。代表的なものに視線計測の対象となる視覚情報をいくつかの領域に分割し、その領域上に停留点を形成することで、どの領域が注視されたか(注視行動)と視野の移動を抽出する手法がある [大野 02, 藤本 04, 脇山 07]。領域の分割手法は、主として全体を矩形で分割する手法 [藤本 04, 江川 08b] と、オブジェクトごとに分割する手法 [脇山 07] に分けられる。前者には、矩形の大きさや位置に分析結果が大きく依存するという欠点がある [江川 08b]。後者では、オブジェクトの機械処理による抽出が難しく(例えば、図地分離の問題)、多くの場合、人間が明示的にオブジェクトの位置や境界を設定しなければならないという問題がある。

また、領域がどの程度注目されたか、あるいはどのように注目されたかを示す定量的な指標が少ないことも問題である。定量的指標の代表は、領域上で停留している注視点の個数にもとづく領域の注目度である。単純に加算する方法 [大野 02]、パルツェン窓関数を用いるもの [藤本 04]、ガウス分布を重ねる方法 [Špakov 07] などがある。これらの手法は、領域がどの程度注目されたかを知る上で有用であるが、視線の移動情報から得られる領域がどのように注目されたか(視線が領域間をどのように移動したか)という情報が棄却されてしまう。

本稿では、はじめに視覚情報の持つ生来的な性質を利用したセグメンテーションによって領域分割の問題を解決を試みる。次に、視覚情報の性質と視線の移動情報を利用して領域毎の注目度と領域間の視線の遷移確率を求める。これらを注目度と注目のされ方を示す定量的指標として提案する。いくつかの画像に対して提案手法を適用し、その応用について考察する。

2. 提案手法

2.1 入力画像の領域分割

2.1.1 Saliency map の作成

Itti らが提案している Saliency map モデル [Itti 98] は、人間の視覚認知過程のうち、特に視覚の注意を数理モデル化したものである。はじめに、このモデルにもとづき、視覚情報が有する人間が注意を集めやすい領域を saliency map として抽出する。具体的には、Itti らの方法にしたがい、はじめに、入力画像を、色、輝度、方向の3つの特徴量に分解し、各特徴量に対しガウシアンピラミッドを作成する。次に、視覚系における”中心-周辺差分処理”を模擬した処理を行い、3つの特徴量に対して局所的に周辺部より目立つ部分が抽出された顕著度マップを作成し、線形結合によって融合することで saliency map を構成する。図1に saliency map の例を示す。図1左図が入力画像、右図がこの画像から作成された saliency map である。Saliency map は、入力画像の各領域において、どれくらい注意を集めやすいかを定量化した画像である。この画像は、入力画像から一意に作成できる(ただし、いくつかのパラメータは存在する)。

ここで、saliency map を $s(l, m)$ とし、注意の集めやすさを saliency と呼ぶことにする。また、saliency は0から255までの値をとるものとする。図1右図で白くなっている点は saliency が高いことを示している。なお、本稿の手法では、saliency map は縦、横ともに元画像の $1/16$ の大きさになる。



図1: Source image (left) and its saliency map (right).

連絡先: *¹ 〒277-8568 千葉県柏市柏の葉 5-1-5 総合研究棟 560号室, egawa@race.u-tokyo.ac.jp

2.1.2 Saliency map のクラスタリング

Saliency map 上で, saliency が高い領域を抽出することでクラスタリングする. クラスタリングは, 以下の2つのステップからなる.

はじめに, saliency map 上で閾値よりも saliency が高い点を抽出し, 表1にしたがいグルーピングする. 擬似コードを表1に示す. K はクラスタのインデックスを表す. また, クラスタ K の重心を $(x_K, y_K) = \frac{1}{N_K} (\sum_{(l,m) \in C_K} l, \sum_{(l,m) \in C_K} m)$ とする. Saliency map 上の点 (l, m) が属するクラスタのインデックスを $c(l, m)$, (l, m) の8近傍に存在するクラスタの中で (l, m) とクラスタ重心の距離が最も近いクラスタのインデックスを返す関数を $\text{NearestCluster}(l, m)$ とする. δ は閾値であり, 本稿では saliency map 上の最大 saliency の半分としている. このステップで抽出されるクラスタを, コアクラスタと呼ぶことにする.

次に, コアクラスタに含まれる点の8近傍でコアクラスタに属していない点を, 最も近いコアクラスタに割り当て, 新たなクラスタを生成する. すべての点がいずれかのクラスタに属するまで同様の処理を行う.

このようにクラスタリングした saliency map を Saliency Cluster Map(SCM)と呼ぶ. SCMの例を, 図2に示す. 図2左図はSCMにおいてコアクラスタのみを示しており, 右図はコアクラスタ以外の領域も示している.

2.1.3 Saliency Cluster Map を利用した画像の分割

SCMを利用して画像のセグメンテーションを行う. SCMにおけるクラスタと対応する画像上の領域をセグメントとして抽出する. この際, SCMのコアクラスタに対応する領域をコアセグメントとして扱う. 図1のセグメンテーション結果の例を, 図3の左図に示す. セグメントの中でコアセグメントに対応する領域を濃い色, その他の領域を薄い色で示している.

表 1: Pseudocode for constructing the *core-clusters*

```

set 0 to  $c(l, m)$ 
 $K \leftarrow 0$ 
for  $l = 0$  to  $[(imax - 1)/16]$ :
  for  $m = 0$  to  $[(jmax - 1)/16]$ :
    if  $s(l, m) \geq \delta$  and  $c(l, m) = 0$  {
      if at least one of the eight neighbors of  $(l, m)$ 
      belongs to any cluster {
         $K \leftarrow K + 1$ 
         $c(l, m) = K$ 
         $(x_K, y_K) = (l, m)$ 
      }
      else {
         $k = \text{NearestCluster}(l, m)$ 
         $c(l, m) = k$ 
        update  $(x_k, y_k)$ 
      }
    }
  }
if  $s(l, m) < \delta$  {  $c(l, m) = 0$  }
    
```

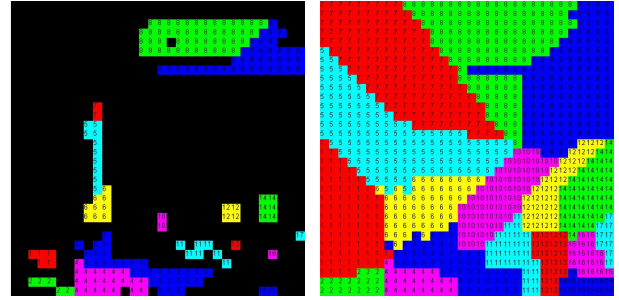


図 2: Saliency Cluster Map.

2.2 定量的指標

2.2.1 初期重要度マップの作成

Ittiらは, 視覚的注意 (FOA: Focus Of Attention) の移動モデルも提案している [Itti 98]. このモデルでは, FOA は saliency の高い領域から低い領域へ順番に移動する. 本研究では, Ittiらと同様に, この移動を視覚情報の持つ生来的な性質に起因する視線の移動と考え, 各コアセグメントにおける saliency の最大値を利用して, FOA の移動を計算し, 移動の順番に応じてセグメントの順位付けを行う. k 番目の FOA として選ばれたセグメントのうち, コアセグメントを R_k , コアセグメント以外の領域を R'_k という指標で表す ($k = 1, \dots, M$). 図3左図においては, 1番目と2番目の FOA として計算されたセグメントを示している.

次に, FOA の移動を利用して重要度を計算する. セグメント R_k の重要度を i_k^1 とし, i_k^1 を以下の式で求める.

$$i_k^1 = \frac{1}{k \cdot \sum_{k=1}^M \frac{1}{k}}. \quad (1)$$

本稿では, FOA の移動と上式で求めた各セグメントの重要度を付与した図を初期重要度マップと呼ぶ. 図3右図に初期重要度マップの例を示す. 初期重要度マップは, 人間の視覚モデルである Saliency map モデルをもとに, 領域の注目度と注目のされ方を定量化したものである.

2.2.2 第2の重要度マップの作成

初期重要度マップは, Saliency map モデルという決定的なモデルをもとに作成されるので, 個人固有の見方は反映されていない. そこで, 注視行動の個人差が反映される定量的指標を検討する.

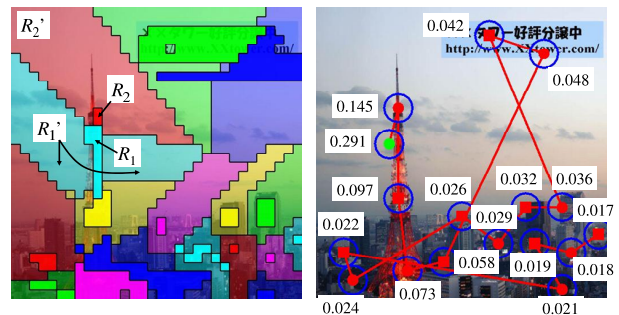


図 3: Example of image segmentation(left) and first Image Importance Map(right).

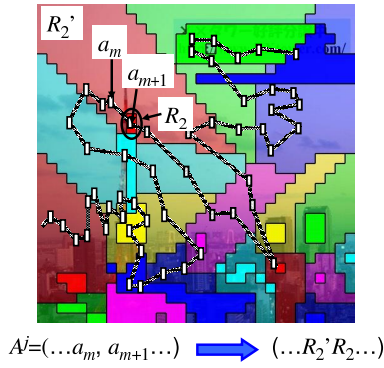


図 4: Schema of encoding the eye-movement dataset.

はじめに、視線計測装置を利用して、視覚情報に対する被験者の視線データを計測する。\$n\$ 番目の視線データを、\$a_n = (x_n, y_n, t_n)\$ と表す。\$(x_n, y_n)\$ は \$a_n\$ の座標であり、\$t_n\$ は計測された時間を表す。個人 \$j\$ の視線データの系列を、\$A^j = (a_1, a_2, \dots, a_N)\$ と表す。\$N\$ は視線データの総数とする。

第2に、2.1 節で述べたセグメント上に \$A^j\$ をマッピングし、視線データをセグメントの指標によって符号化する。図 4 に示すように、\$a_m\$ と \$a_{m+1}\$ がそれぞれ \$R_2'\$, \$R_2\$ 上に位置するとき、これらの視線データは \$R_2'R_2\$ として符号化される。

第3に、コアセグメント \$R_k\$ の重心をノード \$v_k\$、セグメント間の視線の動きをリンク（アーク）とすることで、符号化された視線データからネットワークを形成する。本稿では、この際に、コアセグメント外のセグメントに位置するものをコアセグメントに置き換えた。この点は今後の課題である。例えば、\$R_2R_4R_1R_1R_1R_2R_4R_3\$ と符号化された視線データからは、\$v_2 \to v_4 \to v_1 \to v_1 \to v_1 \to v_2 \to v_4 \to v_3\$ というネットワークが形成される。リンクの数を重みとするループも含む隣接行列 \$\mathbf{A}\$ は以下となる。

$$\mathbf{A} = \begin{matrix} & v_1 & v_2 & v_3 & v_4 \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{matrix} & \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \end{matrix} \quad (2)$$

最後に、ネットワークの各ノード（セグメント）の重要度を PageRank アルゴリズムを利用して算出する。はじめに、視線のネットワークの隣接行列を行ベクトルによる推移確率行列に変換する。上記の隣接行列 \$\mathbf{A}\$ からは、以下の推移確率行列 \$\mathbf{T}\$ が得られる。

$$\mathbf{T} = \begin{matrix} & v_1 & v_2 & v_3 & v_4 \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{matrix} & \begin{pmatrix} 2/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \end{matrix} \quad (3)$$

この推移確率行列を利用して、各ノードの重要度を計算する。ノード \$v_k\$ の PageRank を \$i_k^2\$、全体を \$\mathbf{p} = (i_1^2, \dots, i_k^2, \dots, i_M^2)\$ とすると、視線ネットワークの各ノードの PageRank は以下のように求められる。

$$\mathbf{p}^{\nu+1} = \mathbf{T}^t \mathbf{p}^\nu \quad (4)$$

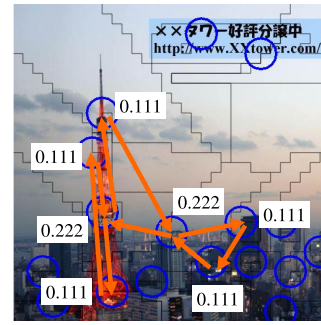


図 5: Second Image Importance Map.

ここで、\$\nu\$ は反復回数を示す。このように符号化された視線データからネットワークを作成し、ノードに相当するセグメントに PageRank を付与したものを第2の重要度マップと呼ぶ。第2の重要度マップの例を、図5に示す。図中の矢印は視線データから作成したネットワークであり、数字は PageRank の値であり、これが各セグメントの重要度になる。第2の重要度マップは Saliency map モデルという視知覚モデルのもとで分割された領域の重要性を、注視された領域の移動という観点から定量化したものである。

2.2.3 重要度マップの統合

初期重要度マップと第2の重要度マップを統合し、領域毎の重要度を算出する。これを統合重要度と呼び、セグメント \$R_k\$ の統合重要度を \$i_k^s\$ とし、全体を \$\mathbf{p} = (i_1^s, \dots, i_k^s, \dots, i_M^s)\$ で表す。統合重要度を、biased PageRank アルゴリズム [Kamvar 03] にもとづき、次式で求める。

$$\mathbf{p}^{\nu+1} = \omega \mathbf{T}^t \mathbf{p}^\nu + (1 - \omega) \mathbf{i}, \quad (5)$$

ここで、\$\omega (0 \le \omega \le 1)\$ は統合パラメータである。本稿では、\$\omega = 0.5\$ としている。

3. 実験と考察

提案手法は、“Itti らの Saliency map モデルにより分割された領域に人間は注意を向けやすい” という仮説にもとづいている。Saliency map モデル自体は、Itti らや Parkhurst らによる被験者実験によってその有効性が確認されている [Parkhurst 02]。分割された領域に対する注意の向けやすさについては、文献 [江川 08a] において6名の被験者実験を行い、概ね50%以上の確率で saliency の高い領域に視線が向くことを示した [江川 08a]。そこで本稿では、上記の仮説が成り立つと考え、一連の重要度マップを作成する。

はじめに、図1の左図の画像に対して視線追跡実験を行い、重要度マップを作成した。初期重要度マップを、図3右図に示す。また、2名の被験者 A, B の視線データから作成した第2の重要度マップを図6に示す。図7はそれぞれの統合重要度マップである。

画像を Saliency map モデルをもとに分割することで、人間が注意を向けやすい領域が機械的に抽出される。矩形による分割や、オブジェクトごとの分割などの従来手法よりも、より人間の認知的特性に合致した手法であると考えられる。

一連の重要度マップにおいては、FOA の移動軌跡と視線の領域間の推移確率をもとに各領域に定量的な重要度を付与している。したがって、重要度を比較することにより、潜在的な重要度と被験者間の見方の違いが顕在化できる。

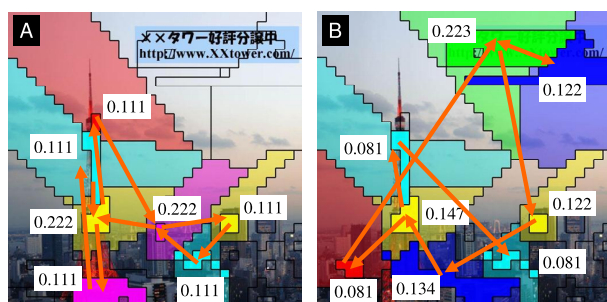


図 6: Second Image Importance Maps.

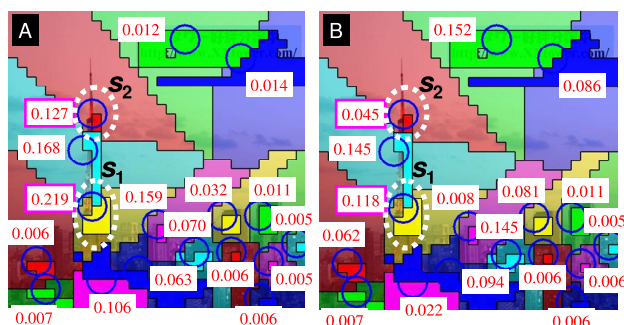


図 7: Integrated Image Importance Maps.

統合重要度マップにおいて重要度が高い領域は、その領域の初期重要度が高く、かつ被験者に注視された領域である。よって、このような領域はその画像において重要な領域であり、被験者の見方の特徴が表れる領域と考えられる。例えば、図 7 の s_1 で示す部分の被験者 A と被験者 B の統合重要度は、0.219 と 0.118 と異なり、A の重要度が高い。この近傍のセグメントの重要度を考え合わせると、被験者 A は s_1 を含む近傍を繰り返し見ていたこと、およびどの近傍を見ていたかを推測できる。中間的な重要度の領域は、初期重要度が高いことを考慮しなければならない。例えば、図 7 の s_2 に対しては、初期重要度が 0.145、統合重要度は被験者 A が 0.127、被験者 B は 0.045 である。図 6 から被験者 B はこの領域を注視していなかったことがわかるが、重要度が A の 1/3 程度あることから s_2 は重要ではないという判断はできない。

また、初期重要度が低いけど統合重要度が高い領域については、色や輝度などの観点からは目立たないが、何か重要な情報があったために被験者が関心を持ち注視したのではないかと、などの推測が可能である。

以上のように提案手法を用いることで、画像において重要な領域の抽出、ならびに被験者の見方を定量的に比較することが可能となる。特に、重要領域を抽出する際に画像の見え方という要素と被験者固有の要素のどちらがより影響しているか判定できる点が重要である。

我々は、画像を抛り所として作業する操業者へのインタビューから、熟練者と初心者が見方の違いが、特に比較という行為に現れることを確認している。上述のように、提案手法は領域間の比較を定量的に評価することができる。したがって、両者の作業中の視線データを分析すれば、熟練者固有の見方がより詳細に抽出できると考えられる。

4. 結言

本稿では、従来の視線分析手法がもつ、領域分割の任意性、あるいは恣意性という問題の一解決法を、Saliency map モデルを利用したセグメンテーションによって示した。

また、視覚情報の持つ生来的な性質と視線計測によって領域毎の注目度と領域間の視線の遷移確率を定量的に求める方法を提案した。これによって従来手法では難しかった領域の注目のされ方が定量的に評価できるものと考えられる。また、被験者実験の結果から提案手法の応用面を考察し、被験者固有の意図や関心の詳細な分析に利用可能であることを示唆した。

参考文献

- [Duchowski 07] Duchowski, A.: *Eye Tracking Methodology*, Springer, second edition (2007)
- [Itti 98] Itti, L., Koch, C., and 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)
- [Kamvar 03] Kamvar, S., Haveliwala, T., Manning, C., and Golub, G.: Exploiting the Block Structure of the Web for Computing Pagerank, *Stanford University Technical Report 2003-17* (2003)
- [Parkhurst 02] Parkhurst, D., Law, L., and Niebur, E.: Modeling the role of salience in the allocation of overt visual attention, *Vision Research*, Vol. 42, pp. 107–123 (2002)
- [Špakov 07] Špakov, O. and Miniotas, D.: Visualization of Eye Gaze Data using Heat Maps, *Electronics and Electrical Engineering*, Vol. 74, No. 2, pp. 55–58 (2007)
- [江川 08a] 江川陽, 白山晋: Saliency Map を利用した画像コンテンツにおける視線誘導, 第 4 回デジタルコンテンツシンポジウム (2008)
- [江川 08b] 江川陽, 白山晋: 注目点からの効率的・効果的な画像コンテンツの再構成手法, 映像情報メディア学会誌, Vol. 62, No. 9, pp. 1443–1452 (2008)
- [高久 08] 高久 雅生, 寺井 仁, 江草 由佳, 齋藤 ひとみ, 三輪 眞木子, 神門 典子: Web 情報探索における視線データの予備的分析, 情報知識学会誌, Vol. 18, No. 2, pp. 181–188 (2008)
- [大野 02] 大野健彦: 視線から何がわかるか - 視線測定に基づく高次認知処理の解明, 認知科学, Vol. 9, No. 4, pp. 565–576 (2002)
- [藤本 04] 藤本武司, 砂山渡, 山口智浩, 谷内田正彦: 視線行動の可視化による着眼スキル伝達支援, 人工知能学会論文誌, Vol. 19, No. 3B, pp. 174–183 (2004)
- [脇山 07] 脇山孝貴, 吉高淳夫, 平嶋宗: 注目の検出に基づいた興味モデルの作成と絵画推薦, 情報処理学会論文誌, Vol. 48, No. 3, pp. 1048–1057 (2007)