

# 画像の極値点情報に基づく誘目領域の推定

## Estimation of Salient Regions Based on Local Extrema of Images

丸田 英徳<sup>†1,†2</sup>, 石井 雅博<sup>†3</sup>, 正会員 佐藤 誠<sup>†4</sup>

Hidenori Maruta<sup>†1,†2</sup>, Masahiro Ishii<sup>†3</sup> and Makoto Sato<sup>†4</sup>

**Abstract** Estimating the salient regions of an image plays a key role in scene analysis and image understanding. We can also apply saliency-based image processing techniques to image compression, evaluation, and effective searching methods. One of the most difficult problems is estimating the regions before recognizing what is in the image – a problem that can be solved by accessing information of the low-level structures of objects in the image. This paper describes a method for estimating salient regions in images based on the distribution stability of local extrema of luminance during image blurring. Under blurring conditions, if an object's region has a more stable structure compared to another area, it must be more salient, so the saliency of these regions must be defined based on their stability for blurring. In the developed method, the local extrema of images are used to describe the complexity of the image's objects and background. Salient regions are estimated based on the stability of the local extrema for the blurring parameter. Experiments were conducted to compare the estimated result of salient regions and the psychophysical result obtained from the analysis of eye movement recordings. Results show that our method successfully extracts salient regions of natural images.

キーワード：誘目領域, 極値, 多重解像度, 尺度空間

### 1. ま え が き

画像全体を画一に処理するのではなく、重要度の高い部分を優先的に処理する技術は、膨大な画像情報の解析を必要とする広い分野で求められている。画像領域の重要度は、その部分がいかに注目すべき領域か、すなわち領域の誘目性の高さによって決定される。誘目性の高い領域の推定とその優先処理により、計算コストを大きく削減することが可能となる。また、画像の領域毎の誘目性を推定する技術が確立できれば、誘目性に基づいた画像領域の処理が可能

となり、画像の符号化や検索・蓄積手法は大きく進展し、ロボットビジョンやセンシングシステムなどの分野においても、より人に近いすぐれた視覚システムの実現が可能となる。これを実現するためには、画像にいかなる対象が含まれているかについての事前知識なしに、輝度値や色・領域面積などの物理的特徴のみを用いて、部分ごとの重要性を評価する必要がある。この機能を実現しているとされるシステムとして、人の視覚情報処理がある。

人は、まず画像の大局的な情報をとらえ、次に視点を様々な位置に移動させながら、徐々に細部の情報を獲得していく<sup>1)2)</sup>。このとき、ランダムな位置に視点移動するのではなく、目を引かれた領域に視点を移動させていく。つまり画像には、観察者の目を強く引く誘目性の高い領域と、そうでない領域とが混在している。視覚的注意とよばれるこのような機能について、これまでいくつかのモデルが提案されている<sup>3)-9)</sup>。これらのモデルにおいて、画像の輝度や色、方位などの物理的な要素に基づく特徴抽出とそれらの統合が、誘目領域の決定において重要な役割を持つことが報告されている。

これまでも、これらのモデルによる単純な視覚刺激に対する誘目性に関して多くの研究がなされ、さらに、これらの知見の応用を目指して、自然画像に対する誘目領域を推定しようとする研究が行われている。

横澤<sup>10)</sup>は、画像の多重解像度モデルを用いた視覚的注意

2009年10月28日受付, 2010年3月31日再受付, 2010年6月7日採録

†1 東京工業大学 大学院総合理工学研究科

(〒226-8502 横浜市緑区長津田町 4259, TEL 045-524-5050)

†2 長崎大学 情報メディア基盤センター

(〒852-8521 長崎市文教町 1-14, TEL 095-819-2219)

†3 富山大学 大学院理工学研究部

(〒930-8555 富山県富山市五福 3190, TEL 076-445-6750)

†4 東京工業大学 精密工学研究所

(〒226-8503 横浜市緑区長津田町 4259, TEL 045-524-5050)

†1 Graduate School of Science and Engineering, Tokyo Institute of Technology

(4259 Nagatsuta, Midori-ku, Yokohama, 226-8502, Japan)

†2 Information Media Center, Nagasaki University

(1-14, Bunkyo-machi, Nagasaki-shi, Nagasaki 852-8521, Japan)

†3 Graduate School of Science and Engineering for Research, University of Toyama

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

†4 Precision and Intelligence Laboratory, Tokyo Institute of Technology

(4259 Nagatsuta, Midori-ku, Yokohama 226-8503, Japan)

のモデルを提案した。このモデルにおいて、画像の輝度・色などの特徴から注意の位置の算出のための注意関数を、画像のガウシアン・ラプラシアンピラミッド<sup>11)</sup>で定義した。この際、色に関してはガウシアンピラミッド、輝度に関してはラプラシアンピラミッドの情報に基づいている。すなわち、色については色特徴、輝度については輝度に関する不連続部分(エッジ)情報を用いていることになる。Ittiら<sup>12)</sup>は、画像の輝度・色・方位成分について、それらの不連続部分が注意を強く引くと仮定し、画像の多重解像度表現に基づく誘目性の計算モデルを提案した。この手法は、少ないパラメータの設定のみで、多くの画像で顕著領域を良好に推定すると報告されている<sup>13)~16)</sup>。しかし、これらのモデルにおいて、不連続部分は、画像に含まれる対象の境界を示すものであり、周辺領域を代表するような点であるとは考えにくい。よって、それだけの情報では、複雑な対象や複数の対象を含むような画像の大局的な構造をとらえることはできず、良い推定結果が得られない場合もあり、不十分であると考えられる。竹内ら<sup>17)</sup>は輝度情報による情報理論に基づき、情報量の高さを誘目性の定義に用いた。しかし、情報理論の複雑さでは、画像の複雑さをうまくとらえることができないと考えられる。例えば、一様輝度分布を持つノイズ状の画像領域の情報量は高いが、画像中のこのような領域が注意を引くことは多くない。田中ら<sup>18)</sup>は、色相・彩度・領域の面積など複数の特徴量を用いて、誘目領域を抽出する手法を提案している。彼らは、主観的な心理実験の結果と比較し、高い精度で誘目領域を抽出することを示したが、多くのパラメータやそのヒューリスティックな決定、複雑な処理を必要とするモデルとなっている。

これらの手法の共通点は、ある画像特徴に着目しその特徴量の大小が注意を引く強度に一致する、と仮定していることである。ここで重要なのは、どのような画像特徴に着目するかであろう。前述したように、人はまず画像の大局的な情報をとらえ、それに続いて、視点位置を移動しながら詳細情報を獲得していく。そこで筆者らは、このような視覚情報の獲得行動を考慮した画像特徴に着目する必要があると考えた。この画像特徴は、画像の大局的構造を表現可能で、さらに必要に応じて詳細情報を提示可能な画像記述手法と高い親和性を持たなければならない。また、画像処理技術としての汎用性のためには、パラメータ選択の容易さや実装のしやすさが不可欠となる。

本研究では、自然画像を対象とし、これらの要求を満たす画像特徴として、画像の多重解像度表現における極値点およびその分布に着目する。画像の極値点とは、2次元の輝度空間において画素の輝度値が極値(極大値あるいは極小値)を取る点であり、その画素を含む近傍の小領域(blob)を代表するような点であり、物体の領域を表現する<sup>19)20)</sup>。筆者らは、輝度に関する極値点が多重解像度表現で他の極値点より安定であれば、より誘目性が高いと考えた。すなわち、輝度の極値点の位置の分布の安定性により、画像の

大局的な誘目性の情報を定義する。ここでいう極値の安定性とは、多重解像度表現における解像度パラメータの変化について、極値点はその位置を維持するかということである。すなわち、解像度パラメータのより広い範囲でその位置を維持する極値点ほど安定性が高いことになる。この画像の極値点の情報に注目し、人の視覚系の特徴、特に視機能の視野依存性を取り込んだ誘目領域の推定方法を提案する。また、提案手法により自然画像での誘目領域を推定できるか検証する。

誘目領域の推定を考える上で、周辺視における画像情報が重要となる。そこで本研究では、画像の多重解像度表現である尺度空間理論<sup>21)~24)</sup>を用いる。画像の多重解像度(尺度空間)表現における極値分布について、極値位置のぼけに対する安定性が高いものほど、より誘目性が高くなるようにぼけパラメータに関して統合を行い、誘目領域の推定に利用する。

本論文の構成は以下のとおりである。2.では、極値情報による誘目性の定義と持つべき性質について述べる。3.では、画像の尺度空間表現、およびそのうえでの極値分布の情報にもとづいた誘目性マップ(極値分布マップ)を、具体的に定義する。4.では、3.に基づく提案手法の評価を行う。視線計測データに基づく人間の誘目領域マップを評価基準として、提案手法とIttiらによるSaliency mapに関して類似度を比較する実験を行い、提案手法の有効性を検証する。提案する誘目性は、自然画像の輝度情報により定義される。一方、自然画像に対して定義されるSaliency mapは、誘目性を算出する手法として広く用いられており、誘目性の比較対象としては適切である。また、実験対象として、カラー自然画像およびそのRGB情報より算出されるグレースケール画像の2種類を用いた。

## 2. 極値による誘目性

画像は周辺視においてぼけており、視野中心から離れ、ぼけが大きくなるにつれ、より解像力の低い画像表現となる<sup>25)26)</sup>。人の視覚情報処理において、周辺視の情報を用いる目的は、周辺視における情報をもとに、もっとも解像力の高い処理が可能な中心視へ注目すべき領域を誘導することにあると考えられる。

本研究では、この視覚情報処理の特性を考慮し、ぼけ画像から得られる情報において、誘目性の高い重要な領域を選択するための画像特徴として、画像の輝度情報における極値点に注目する。

極値点はその近傍の画素と比べ、より目立つ点であり、それを含む近傍領域を代表するような点である。すなわち、画像領域の大局的構造は、その極値点の分布の情報から把握できる。また、尺度空間における極値点は、ぼけパラメータの変化にロバストに存在する場合、あるいはその位置が変化したり、極値点そのものが消滅する場合がある。筆者らは、尺度空間における極値点の分布や安定性によって、画

像の誘目領域の推定ができる考えた。

これまで、画像認識やコンピュータビジョンの分野においても、尺度空間におけるぼけ画像の極値点近傍の局所的な形態変化<sup>27)</sup>や、エッジ<sup>28)</sup>などの不連続な部分に注目した研究はなされているものの、局所的な特徴のぼけパラメータに関する変化の情報だけでは、画像の大局的な情報をとらえることができるとは考えにくい。そこで、尺度空間上の極値点の分布および極値点のぼけパラメータに対する安定性に基づいた誘目性について検討する。個々の極値点について、ぼけに対して安定に存在する極値点は、周囲のより不安定な極値点と比べて誘目性が高いと考えられる。例えば、平坦な画像領域には極値は存在せず、周期的な模様については、その周波数が高いほど極値が密に集合するが、その周波数に応じたぼけ処理の際に、極値点が安定に存在できないため、誘目性は低くなる。画像の多重解像度表現でのさまざまな解像度のぼけ画像において、つねに安定に観察されるぼけに安定な極値点は、そのほかの不安定な極値点に対してより誘目性が高いとする。すなわち、極値点のぼけに対する安定性により、その誘目性を定義する。

具体的に誘目性を算出するため、尺度空間による画像の多重解像度表現を用い、そのうえでの極値点の分布を求める。輝度値そのものではなく、極値点の位置情報を用いることは、その近傍において他と異なる点群についての情報を用いることに相当する。画像に存在する極値点の位置は、周囲の輝度値の分布によりその安定性が決定され、尺度空間表現により、極値の安定性による画像の大局的な誘目性の評価が可能となる。

一方、求めた極値点の分布について、解像度が低い（ぼけが大きい）ほど極値の位置の不確かさが大きくなると考えられる。そこで位置の不確かさを表現するため、その解像度に応じてぼけ処理を行う。このぼけ処理後の多重解像度表現された極値分布情報をぼけパラメータについて統合することにより、極値情報による誘目領域を推定する2次元マップを生成する。

提案手法では、RGB表現されたカラー画像について、人の3原色視感的な特性に基づいたNTSC加重平均法でグレースケール化した輝度情報を用い、その極値に基づいた誘目領域の推定を行う。通常の生活において、人の視覚に入力される画像はカラーであるが、グレースケール化した輝度情報だけからでも、シーンや対象の認識などが可能な場合も多い。これは、輝度情報に物体の知覚やシーン認識に十分な情報が含まれているからであると考えられる。実際、このような視覚特性を考慮し、輝度情報に対するカラー情報の削減による画像の圧縮符号化などの技術に利用されている<sup>29)</sup>。また、輝度情報のみから誘目領域の推定が可能となれば、グレースケール・カラー問わず既存画像コンテンツへの適用が可能となり、提案手法の応用の範囲は広いものとなる。

提案手法の比較対象として、自然画像の誘目領域推定の

最も有力な手法の一つであるSaliency mapを用いる。自然画像での誘目性が、輝度値の極値に基づく手法により説明できるか検証する。そのために、グレースケール画像およびカラー画像の2種類について提案手法とSaliency mapを比較する。また、比較の基準として、視線計測データから求めた人による誘目領域の抽出結果を用いる。提案手法とSaliency mapについて、人による誘目領域の抽出結果との類似度により有効性を検証する。

### 3. 尺度空間における極値情報による誘目性の定義

#### 3.1 画像の尺度空間表現

画像を多重解像度表現する手法として尺度空間理論を用いる。尺度空間において、画像の輝度情報 $f(x, y)$ の尺度空間表現は式(1)のように、 $\tau(> 0)$ をパラメータとしてもつガウス関数 $\mathcal{G}$ との畳み込みで表現される。

$$f(x, y; \tau) = f * \mathcal{G} \quad (1)$$

ただし、

$$\mathcal{G}(x, y; \tau) = \frac{1}{2\pi\tau} \exp\left\{-\frac{x^2 + y^2}{2\tau}\right\} \quad (2)$$

とする。画像の尺度空間表現における $\tau$ は、解像度を表すパラメータである。また、画像の尺度空間表現 $f(x, y; \tau)$ は $\tau$ が大きくなるほどぼけた低解像度の画像となる。極値点に関して、 $\tau$ の変化により、その位置の移動や生成・消滅が発生する。一般的には、 $\tau$ が大きくなると、ぼけに対して不安定な極値点は消滅し、その数は減少していく<sup>30)31)\*</sup>。

#### 3.2 極値点情報による誘目性の算出

本研究において、誘目性は画像の尺度空間表現における、極値点のぼけパラメータ $\tau$ の変化に対する安定性として表現される。視覚情報処理において、もっとも優れた解像力で対象を分析できる中心視への誘導は、低解像度情報からの誘目領域の推定により行われると考えられる。極値点による誘目性の推定についても、この特性を反映させる。すなわち、尺度空間表現された画像について、ある程度ぼけた画像から得られる極値情報を用いることを検討する。先に述べたように、 $\tau$ に対し、より安定な極値情報ほどより誘目性が高いとする。以下に、この条件をみたすための極値による誘目性の算出方法を提案する。

ぼけパラメータ $\tau$ の変化に対する極値点の安定性を、次のようにして定義する。まず、極大値について議論する。ある $\tau$ に対して、 $f(x, y; \tau)$ の極大点を算出する。その結果、 $(x, y)$ において $f(x, y; \tau)$ が極大であれば1、それ以外ならば0とする。すなわち、 $f(x, y; \tau)$ についての極大値に関する情報 $\mathcal{L}_{max}(x, y; \tau)$ を式(3)とする。

\*ただし、単調に減少していくわけではなく、周囲の極値点の分布次第では新たな極値点が生成されることもある<sup>32)33)</sup>。

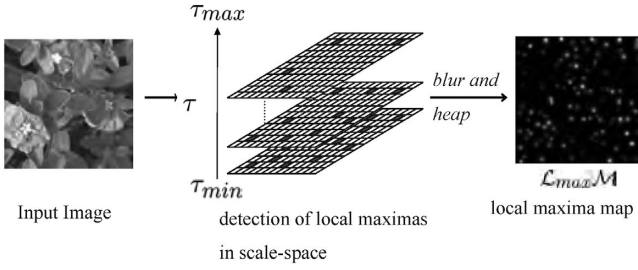


図 1 極大値分布マップの生成  
Extraction of local maxima map of the input image.

$$\mathcal{L}_{max}(x, y; \tau) = \begin{cases} 1 & (f(x, y; \tau) \text{ が極大のとき}) \\ 0 & (\text{それ以外}) \end{cases} \quad (3)$$

このとき、 $\tau$ の大きさに応じて、画像はぼけており、その極大値の位置情報には不確実性が生じると考えられる。この不確実性を、 $\mathcal{L}_{max}(x, y; \tau)$ と同じ $\tau$ の値を持つガウス関数と畳み込むことで表現する。ある極大値の位置に関する情報が、その不確実性の度合い、つまり $\tau$ の大きさに応じて、ぼけた状態で得られる。また、 $\tau$ に関して、最小の値 $\tau_{min}$ から最大の値 $\tau_{max}$ まで、間隔 $\Delta\tau$ で変化させ、 $\tau_{min}$ から $\tau_{max}$ までの $\mathcal{L}_{max}(x, y; \tau)$ を足しこむ。これは、次のような処理に相当する。

- (1) 視覚情報処理における周辺視のようなある程度ぼけた状態の画像から、そのぼけ具合に応じた誘目性情報を算出
- (2) さまざまなぼけ画像における個々の誘目性情報を、ぼけに関して足しこむことにより誘目性マップとして統合

この $\tau$ に関して足し込む操作により、 $\tau$ の変化に安定な極大値点がある位置の不確実性を考慮した上で、より高い誘目性をもつことになる。最終的な極大値による誘目性の定義は、式(2)を用いて、式(4)のような2次元マップになる。

$$\mathcal{L}_{max}\mathcal{M}(x, y) = \sum_{\tau=\tau_{min}}^{\tau_{max}} \mathcal{L}_{max}(x, y; \tau) * \mathcal{G} \quad (4)$$

本研究では、得られた $\mathcal{L}_{max}\mathcal{M}$ を極大値分布マップとよぶ。極大値分布マップ生成の概要を図1に示す。その定義から、極大値分布マップは入力画像と同じ大きさを持ち、その高さの情報が誘目性の指標となる。

極小値についても同様に、極小値に関する情報 $\mathcal{L}_{min}(x, y; \tau)$ 、および極小値分布マップ $\mathcal{L}_{min}\mathcal{M}(x, y)$ が定義される。

次に、極小値と極大値の両方を組み合わせた情報として、極小値分布マップと極大値分布マップを足し合わせ、極値分布マップ $\mathcal{L}(x, y)$ を式(5)のように定義する。

$$\mathcal{L}(x, y) = \mathcal{L}_{max}\mathcal{M}(x, y) + \mathcal{L}_{min}\mathcal{M}(x, y) \quad (5)$$

ここで定義された各マップにより、極値情報に基づく誘目性の高い領域の推定を行う。

#### 4. 評価実験

提案手法を用いた輝度情報による誘目領域抽出の一般的な有効性を検証するため、グレースケールおよびカラーの自然画像それぞれを用いた評価実験を行う。

人にグレースケールおよびカラーの自然画像を提示し、そのときの視線を計測し注目領域を算出する。このデータを基準として、提案手法によって推定される誘目領域の評価を行う。比較対象として、Ittiらのモデルで提案された誘目性を定義する2次元マップであるSaliency mapを用いる。

Ittiらのモデルでは、画像のピラミッド表現を用い、様々な解像度で、画像の輝度、色および方位に関する不連続部分を注目領域とし、これらの特徴に対応する特徴マップを濃淡情報として生成する。さらに、各特徴マップの不連続部分をフィルタリングすることで強調したうえで、最終的な誘目領域を定義する2次元マップ(Saliency map)として算出する。

以下、評価実験について詳細を述べる。

##### 4.1 評価実験で用いた画像

評価実験には、グレースケールおよびカラーの自然画像(480×480(pixel))をそれぞれ16枚ずつ用いた。これらの画像について、カラー画像は24ビット、RGBであり、グレースケール画像は式(6)によるNTSC加算平均法によりカラー画像のRGB情報から生成した。

$$I = 0.298912 \times R + 0.586611 \times G + 0.114478 \times B \quad (6)$$

実験に用いた画像例(グレースケール)を図2に示す。

##### 4.2 視線計測データの取得

人の注目位置は、ナック社製視線計測装置EMR-8を用いて計測した。視線計測におけるサンプリングレートは60fpsである。提示画像は、グレースケールおよびカラー画像それぞれについて、被験者から200cm離れたスクリーンに、視野角にして水平・垂直方向にそれぞれ24度となるように投影される。被験者は正常な視力をもつ16名の大学生であり、グレースケールおよびカラー画像それぞれ8名ずつについて、頭部を固定の上、投影された提示画像を20秒間自由観察するように指示された。視線計測データの取得環境を図3に示す。

後述する手法(4.3節)において、視線計測データから被験者の固視状態を判定し、視覚心理実験による注目領域の抽出を行う。その際必要となる、固視に関する情報を取得した。固視の継続時間は約100msから約400msを中心に分布し、また、長時間固視が継続する場合も存在する<sup>34)</sup>。一般的に、人の自然画像の自由観察において、その固視の継続時間は、自然画像中に含まれる対象の検出や分析に応じて変化する。また、対象に関する事前知識などの影響を



図 2 評価実験に用いた画像  
Images used in the experiments.

受け、対象および観察者により継続時間の変動の大きさは異なると考えられる。これらの要因を考慮し、本研究においては、自由観察における固視の判定について、継続時間の上限は特に制限しなかった。すなわち、取得したデータにおいて、視線が視角 2 度以内の円形領域内に 100ms 以上停留したとき、被験者はその位置を固視したと判断した。グレースケールおよびカラーの各画像について、それぞれ 8 名づつの被験者の異なる視線計測データから得られた固視に関する統計情報（総回数、平均回数（±SE）、固視の平均時間（±SE））を表 1 に示す。

#### 4.3 視線計測データによる誘目領域の定義

被験者に画像を提示し、その視線移動の計測データから

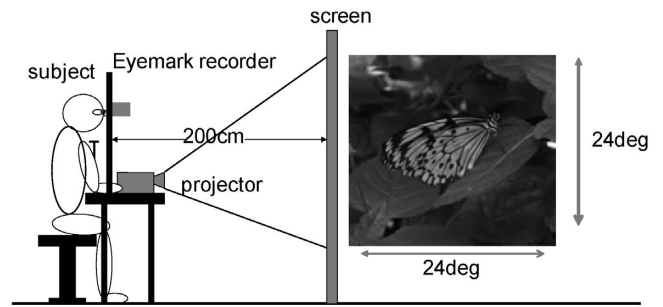


図 3 視線計測取得のための実験環境  
Experimental environment.

表 1 視線計測データから得られた固視に関する統計情報  
Statistical information about fixations of eye-movement data.

画像 (No.)	固視の総回数 (回)		固視の平均回数 (回, ±SE)		固視の平均時間 (msec, ±SE)	
	gray- scale	color	gray- scale	color	gray- scale	color
1	367	294	46±4.04	37±4.55	362±13.16	474±32.89
2	361	408	45±3.55	51±3.26	276±10.73	333±11.11
3	399	386	50±3.82	48±1.36	259±8.17	331±13.18
4	418	420	52±3.94	53±2.85	246±6.04	309±8.83
5	394	399	49±2.06	50±2.11	248±8.07	323±10.94
6	379	397	47±4.64	50±3.47	256±12.28	341±14.99
7	407	419	51±3.65	52±2.56	238±6.53	305±10.88
8	384	455	48±4.57	57±2.61	242±8.20	290±8.56
9	385	389	48±2.55	49±3.15	309±10.27	344±14.96
10	444	428	56±1.38	54±3.06	273±8.95	320±12.15
11	438	386	55±2.37	48±2.66	262±6.86	341±13.20
12	407	380	51±2.45	48±2.25	283±10.30	355±15.44
13	421	381	53±1.97	48±3.33	270±9.39	329±12.31
14	408	406	51±1.22	51±1.47	268±8.38	324±11.39
15	395	414	49±2.90	52±2.80	270±10.26	314±11.70
16	384	405	48±2.58	51±3.99	278±8.80	334±12.06

提示画像における注目領域を 2 次元マップとして定義する手法として Fixation map がある<sup>35)</sup>。人間の視覚において、対象の詳細な認識に充分な解像力をもつ中心視は、視角にして約 2 度といわれている<sup>36)</sup>。ゆえに視線計測データにおける固視の推定精度は視角 2 度の不確定性を持つ。この特性を反映させるため、計測された固視の位置と回数の情報に、この不確定性に相当する分散を持つガウシアンを畳み込む。すなわち、Fixation map は式 (7) で表わされる、提示画像と同じサイズをもつ 2 次元マップとなる。

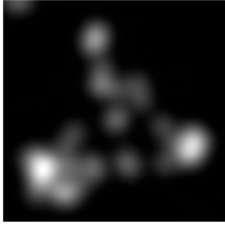
$$F(x, y) = \int N(\xi, \eta) \cdot \frac{1}{(\sqrt{2\pi}\sigma)^2} \exp\left\{-\frac{(x-\xi)^2 + (y-\eta)^2}{2\sigma^2}\right\} d\xi d\eta \quad (7)$$

ここで、 $N(\xi, \eta)$  は、位置  $(\xi, \eta)$  で発生したすべての被験者の固視の総数であり、 $\sigma$  が視角 2 度に相当する。

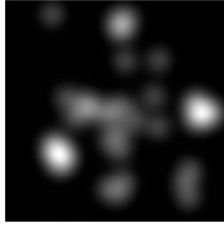
本研究で用いる Fixation map は、固視の回数により誘目性を定義するが、固視の継続（停留）時間は考慮されていない。それは、固視の判定の基準と同様、人の自然画像の自由観察による誘目性の判断では、対象の知識情報などトップダウン的な影響があるので、停留時間が短いまたは長い場合など、いろいろな時間が予測されるためである。その結果、単純に停留時間の範囲を制限すると、誘目性の



(a) Original image



(b) Observer 1 (DT)



(c) Observer 2 (RH)

図 4 被験者間の Fixation map の差異の例: (a) 原画像 (b) 被験者 1(DT) の Fixation map (c) 被験者 2(RH) の Fixation map

Example of the difference of Fixation map between two observers: (a) Original image (b) Fixation map of observer 1 (DT) (c) Fixation map of observer 2 (RH).

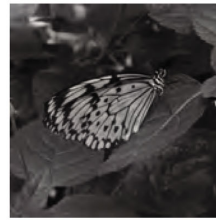
判断を行っている場合を逃す危険性が考えられる。これらを考慮し、Fixation map では、停留時間の情報は用いず、固視の回数だけの情報より誘目性が定義される。

被験者間の Fixation map の差異に関する例として、ある画像についての任意の被験者 2 名のそれぞれの Fixation map を図 4 に示す。

#### 4.4 評価手法

本研究では、誘目性マップについて、Fixation map とより類似性の高いものほど、正しい誘目領域の推定が可能であると判断する。なお、極値に基づく誘目性マップの生成では、先述と同様 NTSC 加重平均法 (式 (6)) によりカラー画像をグレースケール化した画像  $I$  を用いる。極値点は、その近傍として周囲の 8 画素 (8-近傍) を用いて算出した。また、すべての画像について、3.2 の極大値・極小値・極値分布マップの算出に必要なパラメータ  $\tau_{min}$ ,  $\tau_{max}$ ,  $\Delta\tau$  をそれぞれ、 $\tau_{min} = 20(pixel)$ ,  $\tau_{max} = 120(pixel)$ ,  $\Delta\tau = 0.5$  とした。

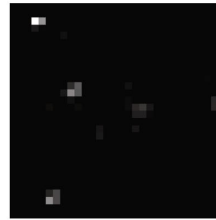
また、Saliency map は次のようにして求めた<sup>12)16)</sup>。画像の多重解像度表現をガウシアンピラミッドにより算出する。ピラミッドの階層は 8 レベルとした。各解像度毎に輝度、色 ( $RG$ ,  $BY$ ), 4 方位 (輝度情報に  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  のガボールフィルタを施したもの) に関する特徴マップを求め、これらを解像度について統合し最終的な Saliency map を算出する。なお、入力画像がグレースケールの場合は、Saliency map の算出には色に関する特徴マップは用いていない。具体的な算出には、SaliencyToolbox<sup>15)</sup> を用いた。実験で用いた画像は、 $480 \times 480(pixel)$  であり、この画像から算出される Saliency map の大きさは  $30 \times 30(pixel)$  となる。一方、極大値・極小値・極値分布マップは、その定義



(a) Original image



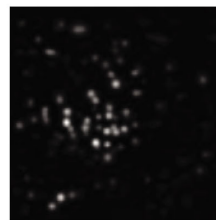
(b) Fixation map



(c) Saliency map



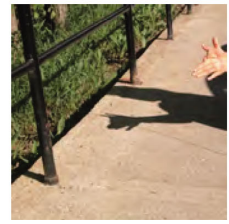
(d) local maxima map



(e) local minima map



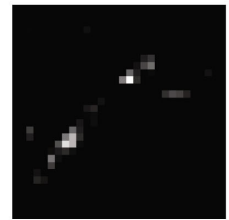
(f) local extrema map



(a) Original image



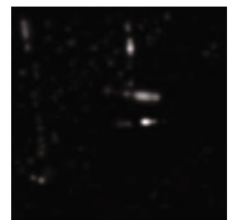
(b) Fixation map



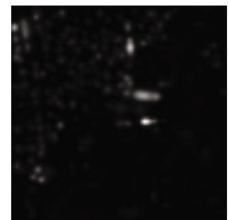
(c) Saliency map



(d) local maxima map



(e) local minima map



(f) local extrema map

図 5 各マップの算出例: (a) 原画像 (グレースケールおよびカラー画像), (b) Fixation map, (c) Saliency map, (d) 極大値分布マップ, (e) 極小値分布マップ, (f) 極値分布マップ

Examples of maps: (a) Original image (gray-scale one and colored one), (b) Fixation map, (c) Saliency map, (d) local maxima map, (e) local minima map, (f) local extrema map

よりそれぞれもとの画像と同じ大きさとなる。

以上により得られたグレースケールおよびカラー画像の各マップの例を、図 5 に示す。



誘目性の基準となる Fixation map と比較するため、次のような正規化を行い、類似度を定義する<sup>37)</sup>。まず、Fixation map および極大値・極小値・極値分布マップについて、Saliency map と同じ大きさとなるよう 2次元マップをそれぞれ縮小する\*。次に、それぞれを  $30 \times 30 = 900$  次元のベクトルとみなし、そのノルムが 1 となるように規格化する。規格化された、Fixation map, Saliency map, 極大値・極小値・極値分布マップのベクトル表現をそれぞれ  $\hat{F}$ ,  $\hat{S}$ ,  $\hat{L}_{max}\mathcal{M}$ ,  $\hat{L}_{min}\mathcal{M}$ ,  $\hat{L}$  とする。Fixation map を基準にそれぞれのベクトル表現を用いて、類似度を内積により式 (8)–(11) のように定義する。

$$Sim(\hat{F}, \hat{S}) = \sum_{(i,j)} \hat{F}(i,j) \cdot \hat{S}(i,j) \quad (8)$$

$$Sim(\hat{F}, \hat{L}_{max}\mathcal{M}) = \sum_{(i,j)} \hat{F}(i,j) \cdot \hat{L}_{max}\mathcal{M}(i,j) \quad (9)$$

$$Sim(\hat{F}, \hat{L}_{min}\mathcal{M}) = \sum_{(i,j)} \hat{F}(i,j) \cdot \hat{L}_{min}\mathcal{M}(i,j) \quad (10)$$

$$Sim(\hat{F}, \hat{L}) = \sum_{(i,j)} \hat{F}(i,j) \cdot \hat{L}(i,j) \quad (11)$$

なお、各ベクトルの成分は 0 以上であり、ノルム 1 に正規化されているため、類似度は 0 以上 1 以下の値となる。これらの類似度を用いて、グレースケールおよびカラー画像の場合において誘目領域の推定結果を評価する。

#### 4.5 実験結果

4.4 で定義した類似度の算出結果をグレースケール画像については図 6 に、カラー画像については図 7 に示す。

グレースケール画像の結果 (図 6) について、実験に用いた 16 枚のグレースケール画像のうち、極大値分布マップについては 12 枚、極小値マップについては 14 枚、極値分布マップについては 13 枚の画像について、Saliency map より高い類似度を示した。また、類似度の平均 ( $\pm$ SD) が Saliency map・極大値分布マップ・極小値分布マップ・極値分布マップについてそれぞれ、 $0.204(\pm 0.150)$ ,  $0.293(\pm 0.132)$ ,  $0.390(\pm 0.096)$ ,  $0.418(\pm 0.132)$  となった。類似度の平均のもっとも高い極値分布マップと従来の Saliency map について、 $t$  検定を行ったところ、有意な差が認められた ( $t(15) = 3.8856, p < 0.0015$ )。

一方、カラー画像の結果 (図 7) について、実験に用いた 16 枚のカラー画像のうち、極大値分布マップについては 12 枚、極小値および極値分布マップについては 13 枚の画像について、Saliency map より高い類似度を示した。また、類似度の平均 ( $\pm$ SD) が Saliency map・極大値分布マップ・極小値分布マップ・極値分布マップについてそれぞれ、 $0.211(\pm 0.123)$ ,  $0.287(\pm 0.151)$ ,  $0.380(\pm 0.112)$ ,  $0.415(\pm 0.161)$  となった。類似度の平均のもっとも高い極

値分布マップと従来の Saliency map について、 $t$  検定を行ったところ、有意な差が認められた ( $t(15) = 3.5197, p < 0.0031$ )。

これらの結果より、グレースケールおよびカラー画像の両方において、輝度・色・方位成分を用いた Saliency map よりも平均的に高い類似度が得られていることがわかる。すなわち、画像の不連続情報よりも輝度値に関する極値情報のほうが、より有効に誘目領域を推定できるといえる。

#### 4.6 考察

実験結果より、提案手法は輝度に関する極値情報のみを用いているにも関わらず、グレースケールおよびカラー画像の違いに関係なく、輝度・色・方位成分を用いた Saliency map より平均的に高い類似度が得られている。このことより、輝度に関する情報は、誘目領域の推定に関して有効なものであることを示している。また、全般的に極値分布マップは、極小値・極大値分布マップと比べて高い類似度をもつことがわかる。極大 (明部) と極小 (暗部) の組合せにより、注目領域の複雑さをよりうまく抽出することができるためと考えられる。また、極大値分布マップと極小値分布マップについて、極小値分布のほうがより類似度が高いことがわかる。複雑な構造をもつ自然画像については、明部だけに注目するのではなく、暗部にも誘目情報が充分含まれていると考えられる。

さらに、一部の画像については、明暗の組合せを表現している極値分布マップより、同様に極小値分布マップのほうが類似度が高いものがある。特に、極値分布マップは Saliency map の結果と比較して、グレースケール画像においては、画像 No. 1, No. 9, No. 12 については低い類似度を示している。また、カラー画像においては、画像 No. 11 がわずかに低い類似度を、画像 No. 9 および No. 12 については低い類似度を示す結果となった。これらの画像の輝度情報に共通することとして、

- (1) 背景画像の輝度値変化がとほしい
- (2) 含まれる対象が直線や円といった単純な要素で構成される

ことがある。ほぼ様な輝度値の分布を持ち、極値点がほぼ周期的な規則性をもった分布となる、あるいは極端に安定なごく少数の極値点が出現するような画像において、その極値点は、非常に安定なごく一部の極値点と、不安定ですぐに消滅してしまう極値点とに大別される。よって、輝度値の極値情報に基づく各マップは、安定な極値点近傍の非常に限られた領域にのみ誘目領域を抽出してしまう。

一方、Saliency map は、これらの単純な構造や対象の自然画像では高い類似度を示しているものの、その他の比較的複雑な自然画像では、輝度・色・方位など複数のチャンネルの情報を用いているにもかかわらず、視線計測データから算出された人の注目領域をうまく抽出することができないと言える<sup>38)39)</sup>。

また、Fixation map を基準とした場合、画像に含まれ

\* 縮小には MATLAB® Image Processing Toolbox の `imresize` 関数を用いた。

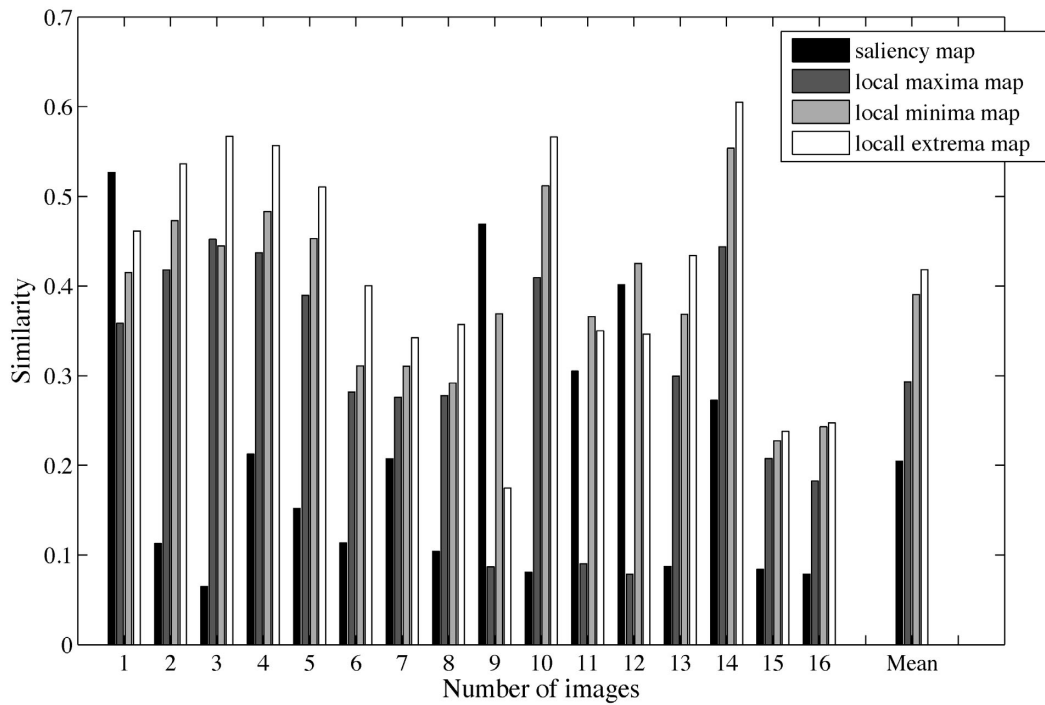


図 6 グレースケール画像に対する Fixation map を基準とした類似度の算出結果  
The result of the similarity of each map using gray-scale images.

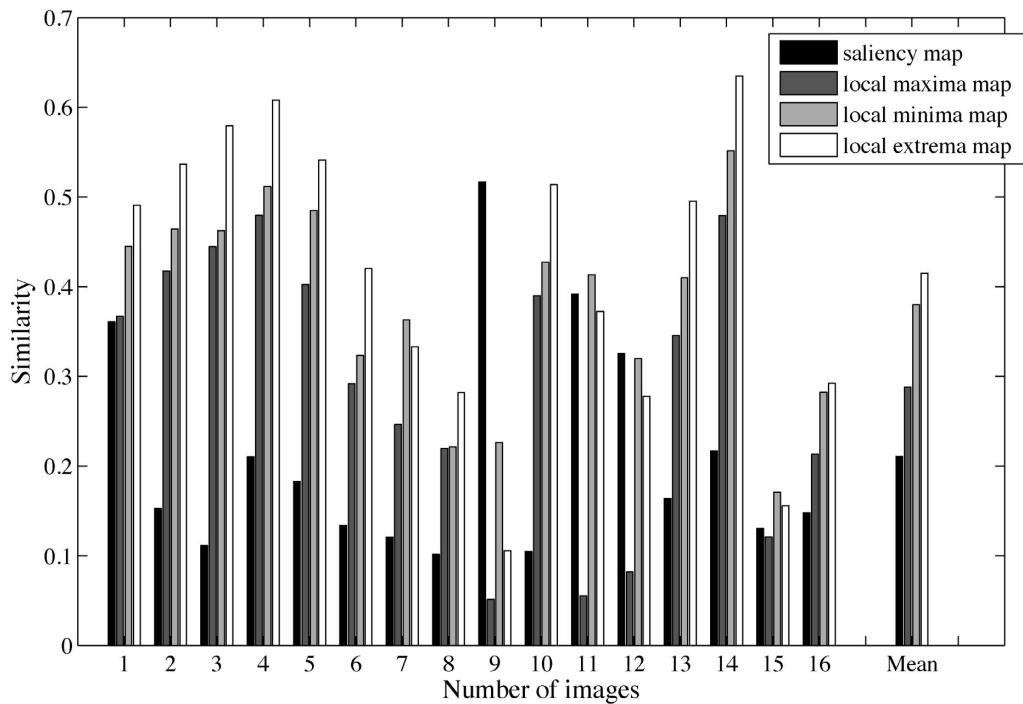


図 7 カラー画像に対する Fixation map を基準とした類似度の算出結果  
The result of the similarity of each map using color images.

る対象によっては、視線の停留位置が特定の場所に集中してしまうために類似度の計算に影響を与えることが考えられる。さらに、画像処理的な問題として、提案手法では極値の算出は8-近傍に基づいている。しかし、この近傍処理

は画像のぼけパラメータの増加に応じて、より広い近傍から処理することにより精密な算出が可能と考えられる。これらの要因も、Saliency map に比べて、一部の画像についての極値分布マップの類似度が低い値となったものと考え



られる。

以上より、一般的に、提案手法は、複雑な自然画像の構造を表現することが可能であり、自然画像の誘目領域の抽出について有効であると考えられる。また、実験結果は限られた画像集合に対するものであるため、今後さらに多くの画像について検証を行うことが必要であると考えられる。

## 5. む す び

本研究では、画像の輝度値の極値情報による誘目領域の推定手法を提案した。尺度空間理論を用いた画像の多重解像度表現により、視覚情報処理の特性を考慮し、極値情報のぼけパラメータに対する安定性を誘目性の高さとして定義した。提案手法の有効性を示すために、視覚心理実験により取得した視線計測データから得られる誘目領域の情報を基準とした比較実験を行った。現在広く用いられている Ittiらのモデル (Saliency map) と比較し、極値情報による推定手法の有効性を示した。すなわち、カラー情報の部分情報である輝度情報のみからでも、カラー情報すべての情報を用いている Saliency map より推定精度において優れていることを示した。

今後の課題として、まず、輝度値の極値情報以外の情報の利用とそれらの統合があげられる。画像に含まれる対象の物理的特徴 (色, テクスチャ, 形など) は誘目領域の推定に有効であり輝度値の極値情報と相補的に組合せることにより、物体の位置および大きさに関する有効な情報を抽出することができると考えられる。これらの特徴に関しては、今後の検討課題である。また、これらの特徴の組合せについても、単純な足し合せではなく、優先的な重み付けなどにより、誘目領域の優先順位などについて検討することが必要であろう。つぎに、誘目領域の優先順位についてさらに検討する必要がある。優先順位の決定については、視線の時系列的順序も考慮に入れるのが妥当であろう。本研究においては、視線計測データの固視の時系列的な順序は考慮されていない。誘目領域の時系列的な特徴を考慮し、明部と暗部あるいはその組合せの処理についての優先度などを決定できると考えられる。

同時に、自然画像以外の画像への適用も重要な課題となる。本研究における誘目性の基準となる極値の安定性は、画像中の対象およびそれらの組合せに依存する。また、文字の読み<sup>40)</sup> などでは、トップダウン的な処理により誘目性や視線移動が支配される。顔<sup>41)</sup>、人工物<sup>42)43)</sup> や合成された画像<sup>44)</sup> などでは、知識情報などの影響により、固視する位置が特定された領域に集中したり、停留時間が長くなる場合がある。これらの対象に対しては推定手法がうまく機能しない場合がある。本研究では、心理実験による誘目領域の判定基準として Fixation map を用いている。誘目性を、視線計測データにより定義する場合、その停留位置情報のみならずその停留時間情報を考慮することは重要である。ここで、画像中の対象の分析を行う場合、その対象に依

存した処理時間が必要となると考えられる。画像構造と固視の回数や停留時間の関係性については、視覚情報処理における高次の処理過程 (事前知識などによるトップダウン的な情報処理) が深く関与していると考えられる<sup>45)46)</sup>。例えば、複雑な構造物が存在し、それを認識することが難しい場合など、対象の画像構造の違いにより、停留時間および停留回数変動する場合がある可能性が考えられる。提案手法では、画像構造をその極値の分布により表現しており、ぼけパラメータについての安定性が高く、複雑な輝度値変化の構造をもつような対象での誘目性が高くなる。提案手法に基づく誘目性の高い領域では、固視の回数が増え固視の停留時間が長くなる可能性がある。よって、停留時間を考慮しない Fixation map と提案手法との類似度の結果に影響が出ていることが予想される。今後は、これらの要因を考慮し、停留時間の情報を取り入れた場合の心理実験による誘目領域との比較と提案手法の改善についても検討していく予定である。

また、自然画像の誘目領域の推定は、画像中の物体の位置と大きさの推定とも関連する。これらは、コンピュータビジョンにおける物体認識の本質的な課題であり<sup>47)48)</sup>、これらの解決への貢献が期待できる。

最後に、本研究の一部は膝附賢史氏の協力によって実施された。

## 〔文 献〕

- 1) 日本視覚学会 (編): “視覚情報処理ハンドブック”, 朝倉書店 (2000)
- 2) J.M. フィンドレイ, I.D. ギルクリスト: “アクティヴ・ビジョン-眼球運動の心理・神経科学”, 北大路書房 (2006)
- 3) A. Treisman, A. Gelade: “A Feature-Integration Theory of Attention”, *Cognitive Psychology*, **12**, pp. 97-136 (1980)
- 4) C. Koch, S. Ullman: “Shifts in selective visual attention: towards the underlying neural circuitry”, *Human Neurobiology*, **4**, pp. 219-227 (1985)
- 5) J. Wolfe, K. Cave, S. Franzel: “Guided Search: An Alternative to the Feature Integration Model for Visual Search”, *Journal of Experimental Psychology: Human Perception and Performance*, **15**, pp. 419-433 (1989)
- 6) L. Itti, C. Koch: “Computational modelling of visual attention”, *Nature Reviews Neuroscience*, **2**, pp. 194-203 (2001)
- 7) Z. Li: “A saliency map in primary visual cortex”, *Trends in Cognitive Sciences*, **6**, pp. 9-16 (2002)
- 8) 加納克則, 中内茂樹, 白井支朗: “高速連続呈示課題による色に対する視覚的注意の解析”, *信学技報 (NC)*, **101**, pp. 17-24 (2002)
- 9) J. Wolfe, T. Horowitz: “What attributes guide the deployment of visual attention and how do they do it?”, *Nature Reviews Neuroscience*, **5**, pp. 495-501 (2004)
- 10) 横澤一彦: “多解像度モデルによる視覚的注意と視覚探索の分析”, *認知科学*, **1**, pp. 64-82 (1994)
- 11) P. Burt, E. Adelson: “The Laplacian Pyramid as a Compact Image Code”, *IEEE Trans. Communications*, **31**, pp. 532-540 (1983)
- 12) L. Itti, C. Koch, E. Niebur: “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis”, *IEEE Trans. Pattern Anal. & Mach. Intell.*, **20**, pp. 1254-1259 (1998)
- 13) 三橋哲雄, 矢野澄男, 畑田豊彦: “画像と視覚情報科学”, コロナ社 (2009)
- 14) F. Shic, B. Scassellati: “A Behavioral Analysis of Computational Models of Visual Attention”, *International Journal of Computer Vision*, **73**, pp. 159-177 (2007)
- 15) D. Walthers, U. Rutishauser, C. Koch, P. Perona: “Selective visual attention enables learning and recognition of multiple objects in cluttered scenes”, *Computer Vision and Image Understanding*, **100**, pp. 41-63 (2005)
- 16) D. Walthers, C. Koch, “Modeling attention to salient proto-objects”, *Neural Networks*, **19**, pp. 1395-1407 (2006)

- 17) 竹内義則, 大西昇, 杉江昇: “情報理論に基づいたアクティブビジョンシステム”, 信学論, **J-81-D-II**, pp. 323-330 (1998)
- 18) 田中昭二, 井口征士, 岩館祐一, 中津良平: “画像領域の物理的特徴に基づく誘目度評価モデル” 信学論, **J-83-A**, pp. 576-588 (2000)
- 19) T. Lindeberg: “Scale-space behaviour of local extrema and blobs”, *Journal of Mathematical Imaging and Vision*, **1**, pp. 65-99 (1992)
- 20) T. Lindeberg: “Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention”, *International Journal of Computer Vision*, **11**, pp. 283-318 (1993)
- 21) 飯島泰蔵: “パターン認識理論”, 森北出版 (1989)
- 22) J. Koenderink: “The structure of images”, *Biological Cybernetics*, **50**, pp. 363-370 (1984)
- 23) T. Lindeberg: “Scale-space theory in computer vision”, Kluwer Academic Publisher (1994)
- 24) B. ter Haar Romeny: “Front-end vision and multi-scale image analysis”, Kluwer Academic Publisher (2003)
- 25) J. Mandelbaum, L. L. Sloan: “Peripheral visual acuity”, *American Journal of Ophthalmology*, **30**, pp. 581-588 (1947)
- 26) M. Millodot: “Foveal and extra-foveal acuity with and without stabilized retinal images”, *The British journal of physiological optics*, **23**, pp. 75-106 (1966)
- 27) A. Kuijper, L. Florack, M. Viergever: “Scale Space Hierarchy”, *Journal of Mathematical Imaging and Vision*, **18**, pp. 169-189 (2003)
- 28) P. Perona, J. Malik: “Scale-space and edge detection using anisotropic diffusion”, *IEEE Trans. Pattern Anal. & Mach. Intell.*, **12**, pp. 629-639 (1990)
- 29) 酒井善則, 吉田俊之: “映像情報符号化”, オーム社 (2001)
- 30) T. Lindeberg: “Effective scale: A natural unit for measuring scale-space lifetime”, *IEEE Trans. Pattern Anal. & Mach. Intell.*, **15**, pp.1068-1074 (1993)
- 31) Y. Leung, J. Zhang, Z. Xu: “Clustering by Scale-Space Filtering”, *IEEE Trans. Pattern Anal. & Mach. Intell.*, **22**, pp. 1396-1410 (2000)
- 32) J. Babaud, A. P. Witkin, M. Baudin, R. O. Duda: “Uniqueness of the Gaussian kernel for scale-space filtering”, *IEEE Trans. Pattern Anal. & Mach. Intell.*, **8**, pp. 26-33 (1986)
- 33) 佐藤潤一, 佐藤誠: “尺度空間フィルタリングに基づく画像パターンの局所構造解析”, 信学論, **J-74-D-II**, pp. 1715-1722 (1991)
- 34) M. Castelhano, M. Mack, J. Henderson: “Viewing task influences eye movement control during active scene perception”, *Journal of Vision*, **9**, pp. 1-15 (2009)
- 35) D. Wodding: “Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps”, *Behav. Res. Methods Instrum. Comput.*, **34**, pp. 518-528 (2002)
- 36) S. Palmer: “Vision Science: Photons to Phenomenology”, MIT Press (1999)
- 37) N. Ouerhani, R. von Wartburg, H. Hugli, R. Muri: “Empirical Validation of the Saliency-based Model of Visual Attention”, *Electronic Letters on Computer Vision and Image Analysis*, **3**, pp. 13-24 (2004)
- 38) L. Itti, C. Koch: “Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems”, *Proc. SPIE Human Vision and Electronic Imaging IV (HVEI'99)*, pp. 473-82 (1999)
- 39) L. Itti, C. Koch: “A saliency-based search mechanism for overt and covert shifts of visual attention”, *Vision Research*, 2000, **40**, pp. 1489-1506 (2000)
- 40) A. Yarbus: “Eye movements and vision”, Plenum Press (1967)
- 41) G. Kun, et.al.: “Longer fixation duration while viewing face images”, *Experimental Brain Research*, **171**, pp. 91-98 (2006)
- 42) A. Torralba, A. Oliva, M. Castelhano, J. Henderson: “Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search”, *Psychological Review*, **113**, pp. 766-786 (2006)
- 43) J. Brockmole, J. Henderson: “Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements”, *The Quarterly Journal of Experimental Psychology*, **59**, pp. 1177-1187 (2006)
- 44) L. Jansen, S. Onet, P. König: “Influence of disparity on fixation and saccades in free viewing of natural scenes”, *Journal of Vision*, **9**, pp. 1-19 (2009)
- 45) A. Oliva, A. Torralba: “The role of context in object recognition”, *Trends in Cognitive Sciences*, **11**, pp. 520-527 (2007)
- 46) L. Itti, P. Baldi: “Bayesian surprise attracts human attention”, *Vision Research*, **49**, pp. 1295-1306 (2009)

- 47) T. Lindeberg: “Feature detection with automatic scale selection”, *International Journal of Computer Vision*, **30**, pp. 79-116 (1998)
- 48) T. Kadir, M. Brady: “Scale, Saliency and Image Description”, *International Journal of Computer Vision*, **45**, pp. 83-105 (2001)



まるた ひでのり  
丸田 英徳 1998年, 九州大学理学部卒業. 東京工業大学精密工学研究所技官を経て, 現在長崎大学情報メディア基盤センター助教, 東京工業大学大学院総合理工学研究科在学中. 画像認識, コンピュータビジョンなどの研究に従事.



いしい まさひろ  
石井 雅博 1995年, 東京工業大学大学院博士課程修了. ヨーク大学視覚研究センター博士研究員, 東京工業大学精密工学研究所助手を経て, 現在富山大学大学院理工学研究部准教授. 知覚情報処理, ヒューマンインタフェース, ソフトコンピューティングなどの研究に従事. 博士 (工学).



きとう まこと  
佐藤 誠 1973年, 東京工業大学工学部電子物理卒業. 1978年同大学大学院博士課程了, 同大学工学部助手, 同大学精密工学研究所助教を経て, 現在同大学精密工学研究所教授. パターン認識, 画像処理, ヒューマンインタフェース, パーチャルリアリティなどの研究に従事. 工博. 正会員.