

令和元年6月26日現在

機関番号：13501

研究種目：挑戦的萌芽研究

研究期間：2016～2018

課題番号：16K12459

研究課題名（和文）リーディングラインを考慮した時変顕著性マップの構築

研究課題名（英文）Generating Saliency Map for Images with Leading Lines

研究代表者

茅 暁陽（MAO, xiaoyang）

山梨大学・大学院総合研究部・教授

研究者番号：20283195

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：画像提示の初期段階における人間の注意を予測する顕著性マップに関する研究が注目を浴びている。我々は特に、画像内のある一点に線が収束するリーディングラインと呼ばれる構造に着目し、リーディングラインによる誘目効果を表現したマップの生成方法、及び視覚刺激の中心周辺差分と画像中央の誘目効果、そしてこれらの効果の相乗効果を反映した新しい顕著性マップの生成モデルを提案した。評価実験では、顕著性マップの評価に用いる代表的な指標において、既存手法と比較して我々のモデルが高い精度を示し、視覚注意の推定におけるリーディングラインの重要性を示すことができた。

研究成果の学術的意義や社会的意義

画像提示の初期段階における人間の注意を予測する顕著性マップはロボットビジョンや広告デザインをはじめ、様々な分野で注目されている。局所的な低レベルの画像特徴を利用したモデルに加え、近年では意味論的な情報を考慮したものやDeep Learningを利用したものなど、様々な顕著性モデルが提案された。一方で、これらのモデルは画像内に存在する大域的な構造による誘目効果を考慮していないかった。本研究の成果はロボットビジョン、広告デザイン、画像編集などへの応用が期待できるのみでなく、認知心理学分野においても多くの後続研究を誘発するポテンシャルの高さを有している。

研究成果の概要（英文）：Researchers have proposed a wide variety of saliency models, ranging from models that use local, low-level image features to recent approaches that incorporate semantic information and deep learning. However, these models do not account for the visual attention related to certain global structures evident in images. We focused specifically on “leading-line” structures, in which lines converge on a single point. We have conducted the experiments to investigate the visual attentions in images with leading line structure and proposed a new saliency model combining the low level feature of center-surrounding differences of visual stimuli, the semantic feature of center-bias and the structure feature of leading lines. Experimental results show that our model outperforms the existing models with all the representative evaluation metrics of saliency map for the images with leading line structures.

研究分野：ビジュアルコンピューティング

キーワード：顕著性マップ

## 様式 C-19、F-19-1、Z-19、CK-19（共通）

### 1. 研究開始当初の背景

顕著性マップ (saliency map) とは、視覚の初期段階 (200ミリ秒以下) において画像領域に対する「人の注意」の度合いを表したものである。米国南カリフォルニア大学のIttiら[1]は、2001年に初めて特徴統合理論に基づき、顕著性マップの計算論的モデルを構築した。彼らは、輝度、色、テクスチャ方向等の視覚特徴に関して、多重解像度中心周辺差分を求め、周囲と大きく異なる領域と特徴に非線形的に高い値を割り当て統合することで、画像の各画素の顕著度を表す顕著性マップを求める画像処理方式を開発した。近年、高解像度画像・映像の効率的処理や高品質CG画像のリアルタイム合成等のニーズの高まりを背景に、顕著性マップの構成とその応用に関する研究に大きな注目が寄せられている。事実本申請時、例えばACM DLだけでも1,568件の論文がヒットした。

その一方で、画家やプロの写真家が鑑賞する人の目を誘導する目的でしばしば意図的に配置されるリーディングライン (同一の点に収束する線群) が、対象画像内に存在する場合、視覚の初期段階であっても、ラインの収束先に無意識的に注意が向けられていることが知られている。このような場合に視線追跡結果と照合すると、Ittiら[1]の顕著性マップではこの視線誘導を的確に予測できていないことが分かった。従来のボトムアップアプローチに対して、画像領域全体に対する各領域の特異度や顔特徴などの高レベル情報を考慮した拡張モデルも種々提案されているが、リーディングラインのような大局的な構造をもつ特徴にも人の注意が向けられている事実を取り込む試みは、まだ報告されていなかった。

### 2. 研究の目的

従来の顕著性マップは、基本的に色やテクスチャ等の視覚特徴の中心周辺差分に基づくものであり、同一の点に収束する線群 (リーディングライン) 等の大局的構造をもつ特徴にも人間の目が無意識的に誘導される事実は未だ上手く取り込めていない。初期知覚においても、中心周辺差分効果とリーディングライン効果は、時間的差異がありながら、人の視線の動きを予測するうえでどちらも重要な指標となる。本研究はこの点に着目し、両者の特徴を統合した顕著性マップ構成のための新たな計算モデルを構築する。

### 3. 研究の方法

リーディングラインを含む画像を用いた実験の結果から、以下の視覚特徴が確認されている。

- リーディングラインの収束先周辺に視線が誘導されるが、中心周辺差分による効果も考慮する必要がある；
- リーディングラインと視覚刺激の中心周辺差分の高い位置が重なると、その位置に視線がより強く引き付けられる。すなわち、リーディングラインと視覚刺激の中心周辺差分との相乗効果を考慮する必要がある；
- 実験のタスクに関係なく画像中央に視線が引きつけられ、特にリーディングラインの収束先が画像中央にある場合は、視線が強く引き付けられる。

#### (1) 提案モデル

これらの視覚注意の特徴を踏まえ、本研究ではリーディングラインとそのほかの特徴との関係を明らかにするために、以下の3つの顕著性モデルを提案する。

##### ① リーディングラインによる効果と中心周辺差分による効果を統合した顕著性マップ

入力画像に対してリーディングラインによる顕著性マップ $M_l$ と中心周辺差分に基づく顕著性マップ $M_s$ を生成し、両マップを下記式(1)に従って統合することで、リーディングライン構造における誘目効果を考慮した顕著性マップ $M$ を生成する。

$$M = kM_s + (1-k)M_l \quad (1)$$

ここで $0 \leq k \leq 1$ は結果の顕著性マップにおける中心周辺差分顕著性マップとリーディングライン顕著性マップの重みを決定する係数である。本研究では、リーディングラインを含む画像において被験者の視線データを追跡する実験を実施し、記録した視線データから生成される注視マップ $M_n^E$ を教師データとして利用することで、係数 $k$ の最適値を決定する。すなわち、

$$k = \arg \min_k \sum_{n=1}^N |kM_{sn} + (1-k)M_{ln} - M_n^E| \quad (2)$$

$N$ は教師データとして利用する画像の枚数である。

##### ② リーディングラインと中心周辺差分の相乗効果も考慮した顕著性マップ

式(1)で示したモデルに中心周辺差分による顕著性マップ $M_s$ とリーディングラインによる顕著性マップ $M_l$ の積の項を加えることで、中心周辺差分とリーディングラインの相乗効果を反映し

たモデルとする.

$$M = k_1 M_s + k_2 M_l + (1 - k_1 - k_2) \sqrt{M_s M_l} \quad (3)$$

$k_1, k_2$ は前述したモデルと同様に, 被験者の視線データから生成された注視マップ $M^E$ を教師データとして, 下記式(4)により制約付き非線形多変数関数 ( $0 \leq k \leq 1$ ) の最小値を求めることで  $k = (k_1, k_2)$  の最適値の推定を行った.

$$k = \arg \min_k \sum_{n=1}^N (k_1 M_{sn} + k_2 M_{ln} + (1 - k_1 - k_2) \sqrt{M_{sn} M_{ln}} - M_n^E)^2 \quad (4)$$

### ③ 相乗効果とセンターバイアスの追加

画像の中央に近づくにつれて顕著となるようなセンターバイアスマップ $M_c$ を生成し, それを以下の式によりリーディングラインによるマップ $M_l$ と中心周辺差分によるマップ $M_s$ と統合することにより, 相乗効果とセンターバイアスの両方の効果を反映した.

$$M = k_1 M_s + k_2 M_l + k_3 M_c + (1 - k_1 - k_2 - k_3) \sqrt{M_s M_l M_c} \quad (5)$$

重み係数 $k = (k_1, k_2, k_3)$ は下記式(6)により, 相乗効果を追加したモデルと同様に制約付き非線形多変数関数 ( $0 \leq k \leq 1$ ) の最小値として求める.

$$k = \arg \min_k \sum_{n=1}^N (k_1 M_{sn} + k_2 M_{ln} + k_3 M_{cn} + (1 - k_1 - k_2 - k_3) \sqrt{M_{sn} M_{ln} M_{cn}} - M_n^E)^2 \quad (6)$$

## (2) マップの生成

Kongら[3]は, 画像中のオフロードの道を検出するために, 連続した直線や曲線として単純に求めることはできない場合が多いことを考慮し, 多くの画素でのテクスチャ方向がある点に向くときに, その点を消失点の候補とすることで道を検出する方法を提案した. 具体的には, ある消失点候補の画素Pについて, その画素を中心とする広域な半円の領域内の各画素Qについて, 画素Qでのテクスチャ方向と直線QPとのなす角 $\theta$ が与えられた閾値以内であれば, 画素Pに投票を行う. この投票処理を画像内の全画素に対して行い, 突出して高い票数を獲得した画素は多くのテクスチャ方向がその画素へ集中していることを意味するため消失点と考えられる. 本研究では, 投票処理によって得た投票マップにDoGフィルタをかけて非線形に強調したマップをリーディングラインの構造による誘目効果を表すマップとした. 一方, 中心周辺差分による顕著性マップについては, Ittiらのモデル[4]およびHarelらのGBVSモデル[5]を用い, HarelらのGBVSツール[6]によりマップ $M_s$ を生成した. Maratらは, 画像中央に分散 $\sigma$ が視野角 $10^\circ$ と一致するようなGaussianカーネルを置いた画像がセンターバイアスを表す顕著性マップとして有効であることを示した[7]. 本研究の実験環境を考慮し,  $\sigma$ を186画素としてセンターバイアスマップ $M_c$ を生成した.

## (3) 注視マップの生成

視線追跡実験により得た視線位置に半値幅 (HWHM) が中心窩の範囲 (視野角 $2^\circ$ ) と一致するような分散 $\sigma$ を持つGaussianカーネルを置くことで, 学習時の教師データとして利用する注視マップ $M^E$ を生成する. 今回の実験環境では視野角 $2^\circ$ での画面上の幅は37.5画素であったため,  $\sigma$ は32画素とした.

## 4. 研究成果

### (1) 学習結果と生成された顕著性マップ

重み $k$ の学習及びモデルの評価には, リーディングラインを含む画像110枚に対する被験者16人の視線を使用した. 10枚1セットとして110枚の画像を11セットに分け, 10セットを学習, 1セットを評価に使用し, 11-fold cross validationを行った. 重み $k$ についてどの提案モデルにおいてもリーディングラインマップの係数が最も大きく, リーディングラインを含む画像における視線推定において, リーディングラインの収束先周辺の情報が中心周辺差分が高い領域や画像中央よりも重要であったことが分かった.

直感的な比較を行うために, 提案手法と既存手法の各モデルにより生成された顕著性マップを図1に示す. 図1中のHumanは被験者の視線位置にGaussianカーネルを置いた注視マップである. Our1はリーディングラインと中心周辺差分の2つのマップを重みにより統合するモデル, Our2はOur1に相乗効果を追加したモデル, Our3はさらにセンターバイアスを追加したモデルであり, 括弧内は提案モデルにおいて中心周辺差分として使用した顕著性マップを表している. 図1より, 低レベルの画像特徴量の中心周辺差分を利用したIttiらのモデル[1]やHarelらのGBVSモデル[5], 画像分類のためのCNNを顕著性予測に転用したPanらのSalNet[8]と比較し, 提案モデルがより被

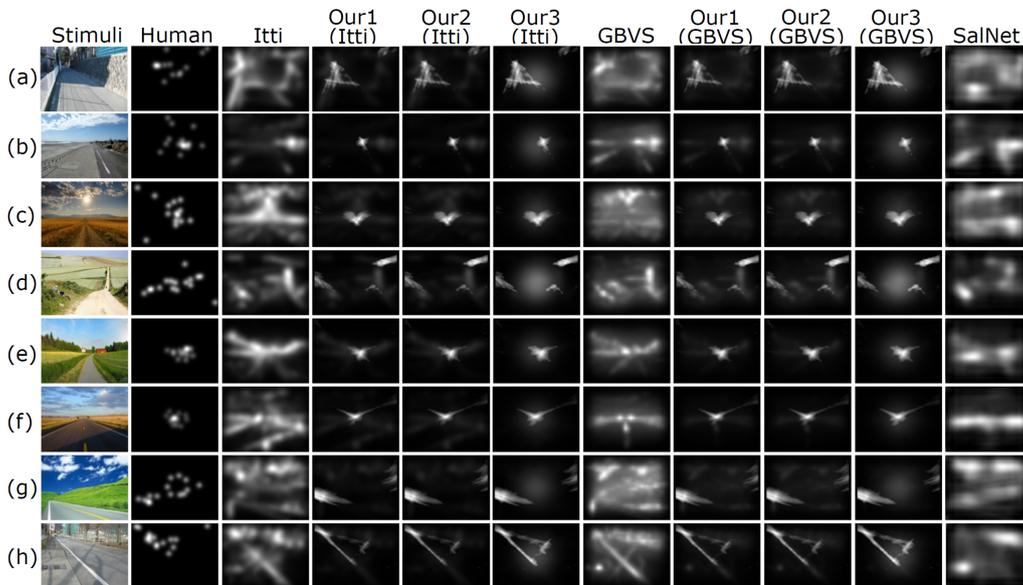


図1. 提案モデルにより生成されたリーディングラインを含む画像の顕著性マップ

験者の視線に近い予測ができていることが見てとれる. 図1(a)の電柱や(c)の太陽のように画像内に顕著な物体がある場合, 既存手法ではこれらの領域がより目を引くと予測しているが, 実際の視線はリーディングラインの収束先周辺に集まっている. また, 図1(b)の画像において, 既存手法では道のラインを顕著としているが, 道などの複数のラインが収束する場所は顕著としていない. これらのことから, 最新のDeep Learningを用いた手法も含め, 既存手法は物体やエッジそのものに対する視覚注意をモデリングできる一方で, これらの組み合わせによってできる「構造」は正しくモデリングできていないと考えられる. 提案手法による顕著性マップはリーディングラインにより引きつけられた視線を予測できており, 顕著性モデルを設計する上で, リーディングラインのような「構造」を考慮することでより正確に人の視覚注意をモデリングできることを示唆している. 提案手法による顕著性マップ同士を比較すると, リーディングラインと中心周辺差分の2つのマップを重みにより統合するモデル(Our1)と相乗効果の追加したモデル(Our2)では視覚的にあまり変化が見受けられない. これは前述した重みの学習結果から, Our2における相乗効果の項の係数が非常に小さく, 最終的な顕著性マップにあまり寄与しなかったためである. 図1(e)および(f)から, 図1(f)のような画像内にリーディングラインのみが存在する場合にはその収束先周辺に視線が集中する一方で, 図1(e)のように, リーディングラインの収束先よりもさらに少し先に顕著な物体が存在する場合, 注視はその物体の領域により強く引き付けられることが見てとれる. 提案手法で我々がリーディングラインによるマップ $M_l$ を生成する際に用いた投票マップは, 消失点よりも手前の領域もある程度顕著とすることができるが, 消失点よりも先の領域はあまり顕著としていないため, リーディングラインによるマップ $M_l$ と中心周辺差分によるマップ $M_s$ との相乗効果を計算する際に収束先とその少し先にある物体の領域との相乗効果を上手く反映できず, これによって学習時に相乗効果の重要度が低くなってしまったと考えられる. さらに, 図1(d)や(g)のようなリーディングラインが画像中央以外に収束する画像において, Our1やOur2では画像中央に分布している被験者の視線の位置を上手く予測できていないのに対し, Our3ではそれらの視線の位置をより顕著とすることができており, センターバイアスの追加によって画像中央に引きつけられた視線を予測できるようになったことが分かる. Tatler[2]は, 画像内の特徴領域の分布に関わらず人間の視線が画像中央に引きつけられることを実験により明らかにしており, リーディングラインを含む画像においてもその知見は有効であると考えられる.

## (2) 定量的な評価

MIT Saliency Benchmark [9]で用いられる, AUC (Area Under the ROC curve), NSS (Normalized Scanpath Saliency), SIM (Similarity), CC (Correlation Coefficient), EMD (Earth Mover's Distance) の5つを用いて, 提案モデルと既存モデルの比較を行った. すべての指標において, 提案モデルがリーディングライン効果を考慮しないIttiらの手法及びGBVSのいずれより良い評価値を得ており, AUCとCCを除く3つの指標では3つの提案モデルのいずれも有意水準1%で有意差が確認できた. CCでもすべての提案モデルに有意差が確認でき, リーディングラインを考慮することでより人の視線に近い推定が可能になったことが分かる. センターバイアスまで考慮したOur3がもっとも精度がよく, 視線位置の検出率の向上においてはセンターバイアスの追加が大きく貢献していると考えられる. また, Our1およびOur2の評価結果においてAUCでは従来の中心周辺差分を用いた手法よりも有意に高い評価値が得られなかった一方, SIMとCC, EMDの分布に基づ

いた指標ではどちらの提案モデルも有意に高い評価値を得ている。これは、リーディングラインの収束先のみではなく、各画素における線の収束度がリーディングラインによる誘目効果を表現しているとしてリーディングラインに基づくマップを生成したことで、マップ内の顕著な領域にある程度の広がり生まれ、実際の視線の分布に近い領域の推定を可能にしたと考えられる。さらに、近年研究が盛んに行われているDeep Learningを用いた手法であるPanらのSalNet[8]との比較において、Our1~Our3のすべての提案モデルが、AUC以外の4つの指標でSalNetよりも有意に良い評価値を獲得しており、AUCにおいても、GBVSモデルによるマップを中心周辺差分による効果として使用した提案モデルでは有意に高い評価を得られていた。これは、SalNetを含む多くのDeep Learningを用いた顕著性モデルがボトムアップの特徴とトップダウンの要因の2つのレベルに着目してネットワークを設計しており、リーディングラインのような特殊な構造を考慮していないこと、また、あまりリーディングラインを含まない画像のデータセットによって学習を行っているために、リーディングライン構造に引きつけられた注視を上手く学習できていないためであると考えられる。SalNetとHarelらのGBVSとを比較すると、GBVSの方がSalNetよりも良い評価値を得ていることから、今回実験に使用したリーディングラインを含む画像のデータセットにはオブジェクトや動作、イベントなど高レベルな要因があまり含まれておらず、意味論的な情報における顕著性を学習することを目的としたデータセットであるSALICON[10]を用いて訓練されたSalNetでは注視を上手く予測できなかったと推測できる。実験結果から、本研究で提案実装したモデルはリーディングラインを含む画像に対して有効であることが言える。

#### 参考文献

- [1] L. Itti, C. Koch, and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.20, no.11, pp.1254-1259, 1998.
- [2] B.W. Tatler, The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions, *Journal of vision*, vol. 7, no. 14, pp. 4-4, 2007.
- [3] H. Kong, J.Y. Audibert, and J. Ponce, Vanishing point detection for road detection, In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 96-103, June 2009.
- [4] L. Itti, and C. Koch, Feature combination strategies for saliency-based visual attention systems, *Journal of Electronic imaging*, vol. 10, no. 1, pp. 161-170, 2001.
- [5] J. Harel, C. Koch, and P. Perona, Graph-based visual saliency, In *Advances in neural information processing systems*, pp. 545-552, 2007.
- [6] J. Harel, A Saliency Implementation in MATLAB: <http://www.klab.caltech.edu/harel/share/gbvs.php>
- [7] S. Marat, A. Rahman, D. Pellerin, N. Guyader, and D. Houzet, Improving visual saliency by adding 'face feature map' and 'center bias', *Cognitive Computation*, vol. 5, no. 1, pp. 63-75, 2013.
- [8] J. Pan, E. Sayrol, X. Giro-i-Nieto, K. McGuinness, and N.E. O'Connor, Shallow and deep convolutional networks for saliency prediction, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 598-606, 2016.
- [9] T. Judd, F. Durand, and A. Torralba, A benchmark of computational models of saliency to predict human fixations, 2012.
- [10] M. Jiang, S. Huang, J. Duan, and Q. Zhao, Salicon: Saliency in context, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1072-1080, 2015.

#### 5. 主な発表論文等

〔雑誌論文〕(査読有, 計5件)

- [1] J. Li, J. Yu, L. Xu, C. Chang, X. Mao, and J Hu, A Cascaded Algorithm for Image Quality Assessment and Image Denoising Based on CNN for Image Security and Authorization, *Security and Communication Networks*, vol. 2018, pp. 1-13, Article ID 8176984, July 2018.
- [2] I. Mochizuki, M. Toyoura, X. Mao, Visual Saliency Prediction for Images with Leading Line Structure, *The Visual Computer (Proc. CGI)*, vol. 34, no. 6-8, pp. 1031-1041, June 2018.
- [3] C. Li, J. Xu, J. Li, X. Mao, Suggesting the Appropriate Number of Observers for Predicting Video Saliency with Eye-Tracking Data, *ACM International Conference Proceeding Series, Computer Graphics International 2018*, pp. 233-237, June 2018.
- [4] M. Toyoura, T. Tanaka, A. Sugiura, X. Mao, Improving Eye Tracking Accuracy by Head

Motion History, *IEEEJ Transactions on Image Electronics and Visual Computing*, vol. 5, no. 2, pp. 92-98, Dec. 2017.

- [5] M. Toyoura, N. Abe, X. Mao, Scene-Aware Style Transferring Using GIST, *Transactions on Computational Science*, vol. 30, pp. 29-49, Nov. 2017.

〔学会発表〕（査読有，計 7 件）

- [1] 野村 芽久美, 中山 雅紀, 藤代 一成, LinDA : 漫画背景画像の半自動生成に向けた線分特徴量抽出と分類, 映像表現・芸術科学フォーラム 2019, vol. 43, no. 9, pp. 315-318, 2019 年 3 月.
- [2] 野村 芽久美, 中山 雅紀, 藤代 一成, LinDA : 漫画背景画像の半自動生成に向けた線分特徴量抽出, 情報処理学会第 81 回全国大会, 6ZC-01, 2019 年 3 月.
- [3] 豊浦 正広, 川村 由基生, 西口 敏司, 茅 暁陽, 村上 正行, 熟練教師の注視を再現する全周授業映像提示, 日本教育工学会 全国大会, 2018 年 9 月.
- [4] 望月 一生, 豊浦 正広, 茅 暁陽, リーディングラインを考慮した顕著性マップの拡張, *Visual Computing/グラフィクスと CAD 合同シンポジウム*, Article 4, 2017 年 6 月.
- [5] T. Sawada, M. Toyoura, and X. Mao, Auto-Framing Based on User Camera Movement, *Computer Graphics International*, pp.1-6, Article 18, June 2017.
- [6] M. Toyoura, T. Tanaka, A. Sugiura, and X. Mao, Eye Tracking by Head Motion History, *International Workshop on Image Electronics and Visual Computing*, pp. 1-5, Article 4B-3, Mar. 2017.
- [7] M. Toyoura, N. Abe, and X. Mao, Painterly Image Generation Using Scene-Aware Style Transferring, In *2016 International Conference on Cyberworlds*, pp. 73-80, Sep. 2016.

## 6. 研究組織

### (1) 研究分担者

研究分担者氏名： 藤代 一成

ローマ字氏名： (FUJISHIRO, Issei)

所属研究機関名： 慶応義塾大学

部局名： 理工学部

職名： 教授

研究者番号 (8 桁)： 00181347

研究分担者氏名： 豊浦 正広

ローマ字氏名： (TOYOURA, Masahiro)

所属研究機関名： 山梨大学

部局名： 大学院総合研究部

職名： 准教授

研究者番号 (8 桁)： 80550780

### (2) 研究協力者

研究協力者氏名： 行場 次朗

ローマ字氏名： (GYOBA, Jiro)

所属研究機関名： 東北大学

部局名： 文学研究科

職名： 教授

研究者番号 (8 桁)： 50142899

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。