Vol.2011-CG-145 No.17
Vol.2011-CVIM-179 No.17
2011/11/17

タスクを用いたトップダウン刺激による視覚注意システムの検討

山ノ井 大貴[†]　櫻井 大督[†]　高橋 成雄[†]

可視化の分野において，画像そのものや画像を見るための環境に何かしらかの条件を施すことで，見る人の視覚注意を意図的に特定領域に誘導するための手法が，近年盛んに研究されている．しかし，既存手法では画像の輝度や色といった人の低次視覚野で処理されるボトムアップ刺激のみを考慮した視覚注意の誘導にとどまっており，高次視覚野で処理されるトップダウン刺激については考慮されていなかった．そこで本研究では，トップダウン刺激のひとつであるタスクを用いて視覚注意を意図的に誘導するモデルを新たに構築し，その効果についても検証を行なっていく．

A Task-Based Computational Model for Directing Top-Down Visual Attention

Daiki Yamanoi[†] , Daisuke Sakurai[†] , and Shigeo Takahashi[†]

Directing visual attention from viewers by modulating subject images and/or imposing specific conditions on the viewers has been an important research theme especially in the area of visualization. However, conventional approaches usually modulate intensity and color of the visualization images by taking advantage of bottom-up visual processing in low-level vision only, and top-down processing in high-level vision has not been fully incorporated in the visualization techniques. This report presents a new model for intentionally directing visual attention from viewers by employing visual tasks as the means of affecting top-down visual processing. User studies were also conducted to demonstrate the effectiveness of the proposed model.

1. はじめに

人は網膜を通して外界の情報を理解する際に，網膜に映る画像情報すべてを瞬時に認識しているわけではなく，ある特定の画像特徴だけを選択的に理解して，外界の情報を理解している．これは，ある情報を可視化画像を通して人に理解させる場合でも同様であり，人は画像の特徴部分に対して選択的に視覚注意を向け，その内容を理解しようとする．そのため，重要な情報を正確に伝えるために，重要な画像情報に視覚注意を誘導する技術の開発は，可視化分野において近年重要視されている研究課題のひとつとなってきている．

人の視覚注意に影響を与える刺激は，ボトムアップ刺激とトップダウン刺激の２種類に分類することができる．ボトムアップ刺激は，色，輝度，線特徴の向きなど，人の視覚システムにおける低次視覚野によって処理される刺激のことを指し，これらの刺激がどのように我々の視覚注意に影響を与えるかは，ほぼその概要が明らかになっている[1]．一方，トップダウン刺激は記憶，経験，タスクといったような高次視覚野によって処理される刺激のことを指し，その詳細なメカニズムはいまだに重要な研究対象として日々研究が行われている．研究や可視化などの研究分野において，ボトムアップ指摘を用いて，ある特定の画像領域に人の視覚注意を効果的に誘導する研究は既に様々な研究が存在する[2,3,4]．これに対して，トップダウン刺激を用いて意図的に人の視覚注意を誘導するモデルは，ほとんどその研究の事例が存在しない．そこで本研究では，トップダウン刺激のひとつであるタスクを用いて，人の視覚注意を意図的に誘導するモデルを新たに構築し，検証実験を通して，その効果の評価を行っていく．

2. 関連研究

人間には，目に映る画像から選択的に重要な特徴に対して注視するメカニズムが備わっており，このメカニズムを計算機上でシミュレートして，Saliency map と呼ばれる画像上の人の視覚注意の分布として定式化する研究が盛んに行われてきている[1]．現在利用可能となっている Saliency map の計算アルゴリズムは，Itti ら[5]によって提案されたモデルが基礎となっている． そのモデルにおいて，Itti らは入力画像から輝度，色，方向などの特徴量を抽出し，各画素において注目画素とその周辺の画素の差をとる center-surround 機構をシミュレートし，特徴量マップを作成した．ここで，特徴量マップとは各成分での目立つ場所を定量化したものであり，求めた特徴量マップを正規化処理して，線形和で統合することにより Saliency map を作成している．この Itti らが提案する Saliency map の計算モデルは，人の視覚システムにおいてその処理

† 東京大学
The University of Tokyo

Vol.2011-CG-145 No.17
Vol.2011-CVIM-179 No.17
2011/11/17

の概要がほぼ解明されている輝度，色，方向などの，人の低次視覚野で処理されるボトムアップ刺激が考慮されており，実際の注視点計測による人の視覚注意の分布を比較的よく再現することができた．しかしながら，人の視覚注意は記憶，経験，タスクなど人の脳の高次視覚野で処理される情報，すなわちトップダウン刺激にも大きな影響を受けることが知られている． Navalpakkam と Itti[6]は，[5]の Saliency map の計算モデルにおいて特徴量マップの線形和を計算する際に，特徴量マップごとの重みの値をトップダウン刺激の影響を考慮して変更することで，人の高次視覚野の働きを反映した Saliency map の計算アルゴリズムを新たに提案した．さらに Frintrop ら[7]は，Itti らが提案している特徴量マップをさらに細かく分け，それらの重み付け線形和を定義することで，トップダウン刺激とボトムアップ刺激の影響の割合を考慮に入れた Saliency map を計算するモデルを提案した．また，Itti らの手法では，人の生理学的な側面に忠実にモデル化を行うため，入力画像を数多くのスケールに分解し計算するため，計算量が大きくなる上に得られる Saliency map の解像度が低くなる問題点があったが， Frintrop ら[8]は解像度を高く保ちつつ Saliency map を短時間で作成できる integral images を用いた手法も提案している．さらに Zhao ら[9]は，GPU を用いて Saliency map を実時間で計算できる手法を提案した．

基本的に Saliency map の定式化は，与えられた入力画像に対して人の視覚注意の分布を予測するための計算アルゴリズムを提供する．これに対し，実現した視覚注意分布をあらかじめ Saliency map のかたちで与え，それに見合うように対象画像に変調を施す定式化は，この Saliency map 計算問題の逆問題としてとらえることができ，近年意図的に人の視覚注意を誘導する道具立てとして，数多くの研究が提案されるようになっている． Kim[2]らは，ボリュームレンダリングの可視化画像において，輝度や色といったボトムアップ刺激に変調を加えることで，特定の特徴領域に視覚注意を誘導するための効果的な手法を提案した．さらに，Su ら[3]は，ボケを利用した手法を提案し動画に適用した．Mendez ら[4]は，RGB 色空間を CIE L*a*b 色空間に変換し，画像の変調量を局所的な色合いとコントラストを考慮することで最小化する手法を提案した．しかし，既存手法ではボトムアップ刺激のみを考慮した視覚注意の誘導にとどまっており，トップダウン刺激については考慮されていなかった．

3. 提案手法

本手法は, 主に 3 つのステップを経て視覚注意を誘導できる適切なタスクを求める．ここで，タスクは，人の視覚注意を誘導するための指示であり，画像を見る際に，「A を探せ」，「A かつ B を探せ」，「A または B を探せ」のいずれかで与えられるものとする．まず，1) 初めに，様々な学習オブジェクト n（n はオブジェクト名）に対し，特徴量マップを重み付け線形和することにより top-down saliency map $S_{td(n)}$ を作成する．ここで，学習オブジェクト n は，上記のタスクの説明での A，B 候補にあたるものである．次に，2) task saliency map S_{task} を $S_{td(n)}$ 重み付け線形和として求める．このとき S_{task} は，入力として与えられた目標 saliency map との誤差が最小 2 乗になるように，各学習オブジェクト n の重み w_n を決定する．最後に，3) 求めた w_n の中で値が大きい上位 2 つの学習オブジェクト n をタスクオブジェクトとして採用し，合成することにより目標 saliency map で表された場所に視覚注意を誘導するための適切なタスク求める．ここで，タスクオブジェクトを 2 つ採用した理由は，1 つのタスクオブジェクトと比較し，より的確に目標 saliency map に近づけるからである．以下，それぞれのステップについて詳細に説明を行なう．

3.1 Top-Down Saliency Map 作成

本紹介手法では，トップダウン刺激が目標オブジェクトを探す刺激と仮定し，主に 3 つのステップを経て top-down saliency map $S_{td(n)}$ を求める．初めに，1) Frintrop ら[8]の手法を用いて，ボトムアップ刺激のみを考慮に入れた bottom-up saliency map を作成する．次に，2) bottom-up saliency map を利用することにより学習モードで目標オブジェクトを学習する．最後に，3) 検索モードで学習オブジェクトが画像のどこに含まれているか検索し，$S_{td(n)}$ を求める．以下，それぞれのステップについて詳細に説明を行なう．

3.1.1 Bottom-up Saliency Map 作成

ここでは, Frintrop ら[8]が提案した, ボトムアップ刺激のみを考慮に入れた bottom-up saliency map の作成方法について紹介する．まず，入力画像を輝度成分 I，色成分 R, G, B, Y, 方向成分 O｛0°, 45°, 90°, 135°｝に分解する．それぞれの分解成分ごとに，各画素において注目画素とその周辺の画素の Difference of Gaussian を計算することで center-surround 機構をシミュレートし，特徴量マップ X_i を作成する．ここで，i は特徴成分の種類であり, 輝度成分として{I+,I-}, 方向成分として{0°, 45°, 90°, 135°}, 色成分として｛RG,GR,BY,YB｝を持つ．ここで，I+は注目画素を明かつその周辺の画素を暗としたものであり，I-は逆に注目画素を明かつその周辺の画素を暗としたものとする．また，色成分では，R と G、B と Y が 2 重反対の組み合わせとなっており，RG では，注目画素を R 成分かつその周辺の画素を G 成分としたものであり，GR, BY，YB も RG と同様の処理が行われる．

注視すべきこととして，人の視覚は目立つ場所が多くなるにつれて， 1 つの対象に注意を絞れなくなる．このことを考慮に入れて，本紹介手法では， 目立つ場所が特定の範囲でしか存在しない特徴マップの重みを大きく，目立つ場所が多数存在する特徴マップの重みを小さくする．この重み R_i は式(1)により与えられる．

$$R_i = 1/\sqrt{N} \tag{1}$$

ここで，N は，各特徴マップの輝度の最大値の半分を閾値として，閾値を超える極

Vol.2011-CG-145 No.17
Vol.2011-CVIM-179 No.17
2011/11/17

大値が何か所存在するかを表している．最後に求めた特徴量マップを重み付け線形和で統合することにより bottom-up saliency map が求まる．

(a) 入力画像　　(b) Bottom-up Saliency Map

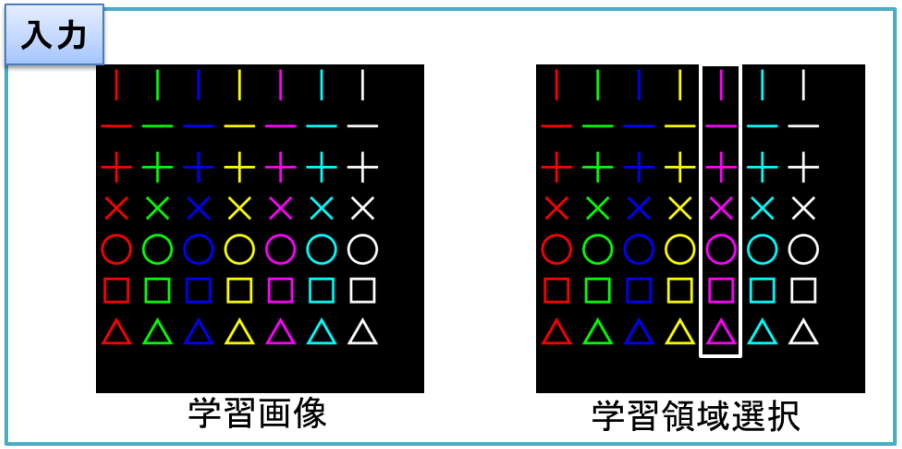
図 1　Bottom-up Saliency Map

3.1.2 学習モード

学習モードは，学習 object がどのような特徴を持つかを計算する．まず初めに，学習させたい領域を画像の中から選択する（図 2 の場合，白で囲った部分を学習領域とする）．次に，学習領域内で bottom-up saliency map の輝度の高い場所 HSR_{in} ，学習領域の周囲で bottom-up saliency map の輝度の高い場所 HSR_{out} を計算し求める．最後に下式(2)を用いて重み v_i を求める． v_i は HSR_{in} が HSR_{out} に対してどのような特徴を持つかを表している．

$$v_i = m_{i(HSR_{in})} / m_{i(HSR_{out})} \tag{2}$$

ここで，m_i は特徴成分 i の平均値である．また，今回の我々の手法では簡単な色や形を学習させているので，色の学習オブジェクトを学習する時は色成分のみを，形の学習オブジェクトを学習する時は方向成分のみを学習させ，それ以外の成分は考慮に入れないようにした．こうすることで，学習画像から得られる関係ない成分を除外できるからである．図 2 は，上記学習モードにおいて紫色を学習させた結果を示している．図 2 をみると，紫色は RG，BY の特徴成分を多く含み，逆に GR，YB の成分はほぼ含まれていない特徴であることが分かる．また，今回は色を学習させたので，方向成分と輝度成分を考慮に入れなかった．



特徴成分	I+	I−	0°	45°	90°	135°	RG	GR	BY	YB
重み	−	−	−	−	−	−	3.5287	4E−14	2.7089	1E−13

図 2　学習モード（紫色を学習）

3.1.3 検出モード

ここでは，先ほど求めた重み v_i を用いて，オブジェクト n の top-down saliency map $S_{td(n)}$ を求める． $S_{td(n)}$ は，以下の式(3),(4),(5)に定義されるように，目標物のもつ特徴を強調する excitation map E と目標物以外が強調されるのを抑制する inhibition map I の差から構成されている．ここで，excitation map E とは，学習オブジェクトが周りに比べが大きい値をもつ特徴成分だけで作成され，inhibition map I は，学習オブジェクトが周りに比べが小さい値をもつ特徴成分だけで作成される．図 2 の学習オブジェクトが紫色の場合は，excitation map E の特徴成分として RG，BY が採用され，inhibition map I の特徴成分として GR，YB の成分が採用される．

Vol.2011-CG-145 No.17
Vol.2011-CVIM-179 No.17
2011/11/17

$$E = \sum_i (v_i \times X_i) \qquad \forall i : v_i > 1 \tag{3}$$

$$I = \sum_i ((1/v_i) \times X_i) \qquad \forall i : w_i < 1 \tag{4}$$

$$S_{td} = E - I \tag{5}$$

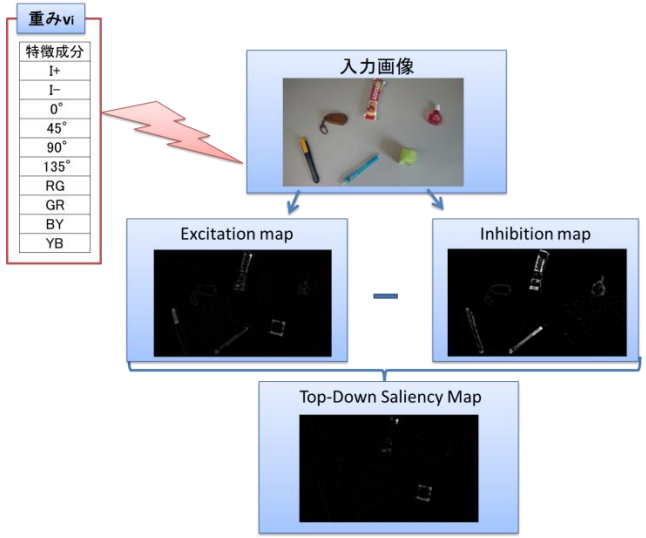


図 3　Top-Down Saliency Map $S_{td(n)}$ の作成方法（学習オブジェクトは緑色の箱）

3.2 タスクの選択方法

ここでは，シーン画像の各学習オブジェクト n に対して top-down saliency map $S_{td(n)}$ を作成し，それらを重み付け線形和する事により，task saliency map S_{task} を作成する．

$$S_{task} = \sum w_n \times S_{td(n)} \tag{6}$$

w_n は学習オブジェクト n の重みの値である，w_n は，S_{task} と誘導場所を表す目標 saliency map と比較し，差が最小 2 乗になるように選択する．ここで w_n が大きな値で割り当てられた学習オブジェクト n は，目標 saliency map で表された場所に注視を誘導する際，重要な役割を担っている．今回は，w_n のうち最も値が大きい学習オブジェクトをタスクオブジェクト 1, 2 番目に値が大きい学習オブジェクトをタスクオブジェクト 2 としてタスク採用する．

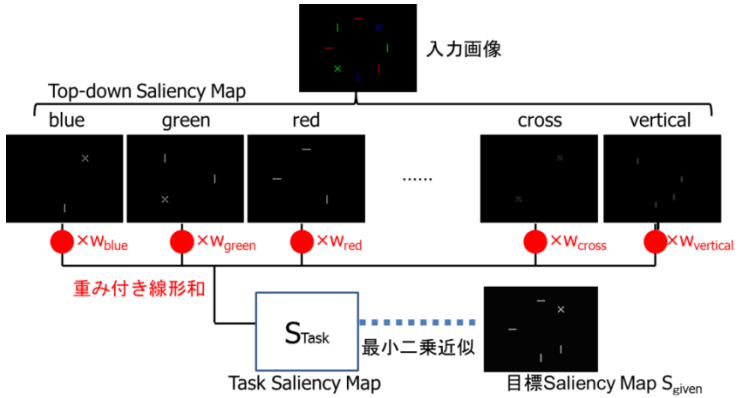


図 4　Task Saliency Map S_{task} の作り方

3.3 タスクの合成と出力

出力タスクとしてどれを選択するかは，タスクオブジェクト 1 の top-down saliency map $S_{td(n)}$ と，2 つのタスクオブジェクトの $S_{td(n)}$ を and で合成した画像と，or で合成した画像を作成し，どの画像が目標 saliency map との差が最小 2 乗になるかにより決定される．

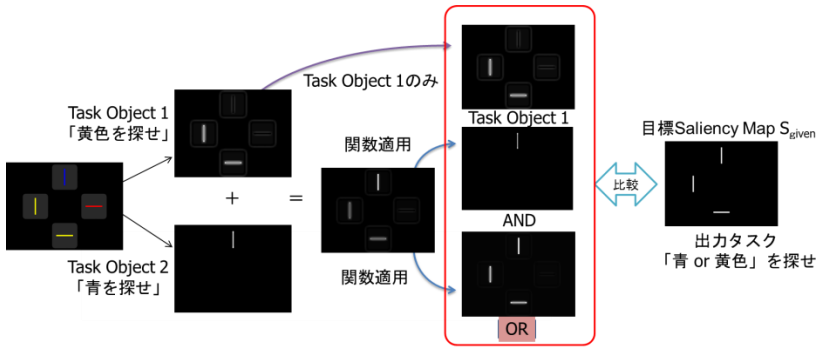


図 5　最適なタスクの選び方の例

ここで，2 つの $S_{td(n)}$ を求めた重み w_n を利用し，重み付け線形和することで task object saliency map $S_{task(two)}$ を作成する．and や or で合成した画像はこの $S_{task(two)}$ を利用することで作成できる．$S_{task(two)}$ は 2 つのタスクオブジェクトから強調される R_{double}，どちらか片方のタスクオブジェクトから強調される R_{single}，どちらからも強調されない R_{zero} の 3 つの領域が存在する．and による合成の場合は R_{double} のみを考え，or に

Vol.2011-CG-145 No.17
Vol.2011-CVIM-179 No.17
2011/11/17

よる合成 の場合は R_{double} と R_{single} の両方を注視すべき場所と決定する．しかし，S_{task} の結果だけでは，and と or の区別が不可能であったため，今回はシグモイド関数により各画素を変換する手法を採用した．シグモイド関数は式(7)で表せ，c の値を調整することにより，画素値が変化する．

$$Sig_c(x) = \frac{1}{1+e^{\frac{-12(x-c)}{1-2\times|c-0.5|}}} \tag{7}$$

結合方法が and である場合は c の値を高くし，$S_{task(two)}$ のなかで高い値を持つ R_{double} のみを強調するように値をとる．また，結合方法が or である場合は c の値を低くし，R_{double} と R_{single} どちらも強調できるように値をとる．今回は，いくつかの画像で注視点計測結果と様々な c の値で作成したシグモイド関数適応結果の比較実験を行い，合成画像を作成する際に使用する最適な c の値を求めた．比較実験のイメージ図を図 6 に示した．比較する際には，両者の注視率の差に着目する．注視点計測結果に対しては，範囲内に入った注視点観測数を全注視点観測数で割った値を注視率と定義し，関数適応結果に対しては，範囲内のピクセル値合計を全ピクセル値合計で割った値を注視率と定義した．図 7 は，2 つの画像での注視点計測の結果とシグモイド関数適応結果の比較実験の結果である．図 7 の縦軸は注視率の差であり，横軸はシグモイド関数 c の値である．画像(1)では，結合方法が and である場合は c=0.35，結合方法が or である場合は c=0.6 が注視率の差が小さくなり最適な c の値であることが分かる．また，画像(2)では，結合方法が and である場合は c=0.05，結合方法が or である場合は c=0.7 が注視率の差が小さくなり最適な c の値である．いくつもの画像で比較実験を行い最適な c の値を平均したところ表 1 の値が求まり，合成画像を作る際，表 1 の値採用することにした．

	cの値
and	0.75
or	0.25

表 1　シグモイド関数 c の値

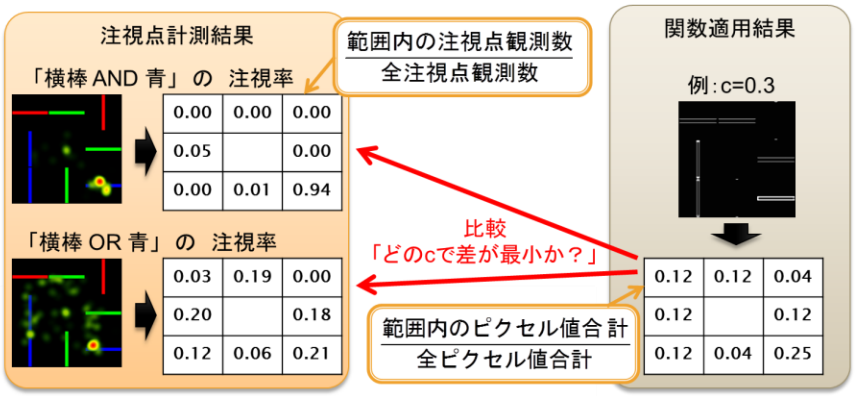


図 6　注視点計測の結果とシグモイド関数適応結果の比較実験のイメージ図

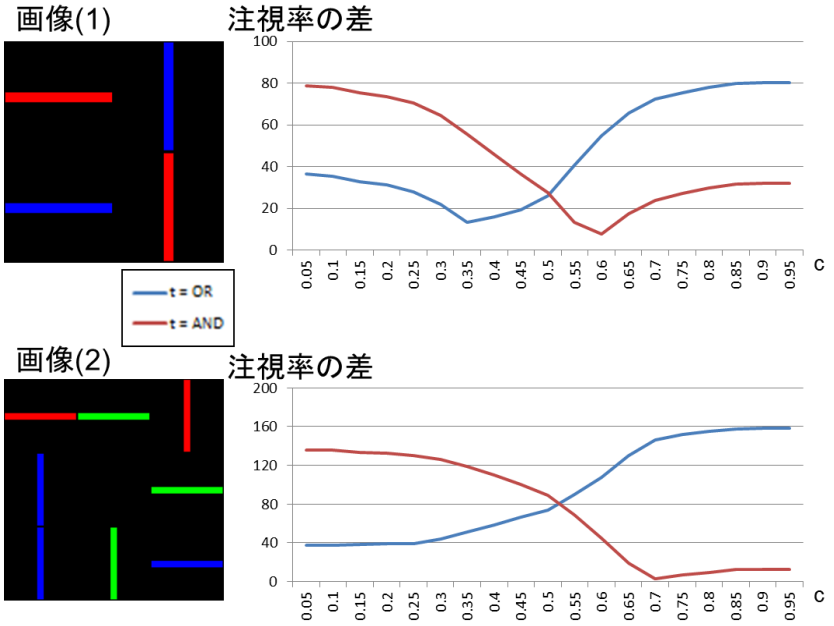


図 7　注視点計測の結果とシグモイド関数適応結果の比較実験の結果
縦軸：注視率の差　横軸：c の値

Vol.2011-CG-145 No.17
Vol.2011-CVIM-179 No.17
2011/11/17

4. 実験結果

実験結果として，代表的な実験結果を図 9 に示した．今回は，図 2 の学習画像から赤色，青色，緑色，黄色，紫色，水色，白色，縦棒，横棒，十字架，バツ，円，四角形，三角形を学習オブジェクトとして学習した．画像(a)～(e)は，注視先を誘導するために適切なタスクを求めることに成功した例である．これからそれぞれの画像の出力タスクの結果について説明と考察を行う．

まず，画像(a)，画像(b)の出力タスクは，タスクオブジェクト 1 のみで生成された例である．画像(a)のタスクオブジェクト 2 は黄色であったが，入力画像には黄色いオブジェクトしかないので不必要とされた．画像(b)のタスクオブジェクト 2 は三角形と求まったが，出力タスクに採用すると注視先が変化してしまうため不必要とされた．

画像(c)の出力タスクは 2 つのタスクオブジェクトを合成したものであり，注視先を誘導するにあたって適切なタスクを求めることに成功している．

画像(d) と(e)は，入力画像として，対象物が簡単な写真を採用した例である．画像(d)は，タスクオブジェクト 1 のみで生成された例で，画像(e)は，タスクオブジェクトを合成した例であり，いずれの結果も適切なタスクを求めることに成功している．

画像(f)は，2 つのタスクを合成するだけでは適切なタスクと言えず，3 つ以上のタスクオブジェクトが必要とされている．出力タスクを求める際，入力画像と目標 saliency map によって，タスクオブジェクトがいくつ必要か考慮に入れる必要がある．

(a) 画像(g)　　(b) $S_{td(円)}$

図 8　画像(g)の Top-Down Saliency Map $S_{td(円)}$

図 8 を見ると，円の形を｛0°，45°，90°，135°｝の 4 つの方向成分だけでは，正確に円成分を表すことができていないことが分かる．同じように画像(h)でも，4 つの方向成分だけでは，正しく学習オブジェクトを表すことができなかった．

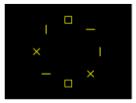
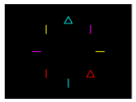
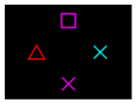
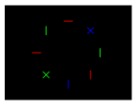
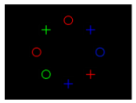
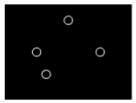
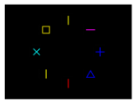
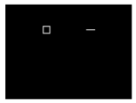
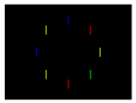
画像	入力画像	目標Saliency Map	出力タスク
(a)			バツを探して下さい
(b)			赤色を探して下さい
(c)			赤色または四角形を探して下さい
(d)			三角形を探して下さい
(e)			横棒または緑色を探して下さい
(f)			赤色または三角形を探して下さい
(g)			赤色または緑色を探して下さい
(h)			黄色または紫色を探して下さい
(i)			黄色かつ緑色を探して下さい

図 9　出力タスクの結果

Vol.2011-CG-145 No.17
Vol.2011-CVIM-179 No.17
2011/11/17

画像(h)では，「四角形または紫色を探して下さい」と言ったタスクが理想的である．理想のタスクを求められなかった原因は，学習オブジェクトが四角形である $S_{td(四角形)}$ にある．作成された $S_{td(四角形)}$ を図 10 に示した．

(a) 画像(h)　　(b) $S_{td(四角形)}$

図 10　画像(h)の Top-Down Saliency Map $S_{td(四角形)}$

図 10 を見ると $S_{td(四角形)}$ は，四角形以外の成分まで強調されていることが分かる．この理由は，学習モードで四角形を学習した際，0°成分と 90°成分の値が大きくなり，検出モードで，この 2 つの成分を含むオブジェクトが強調されるからである．今回は，0°成分と 90°成分を含む横棒，縦棒，十字架のオブジェクトまで一緒になって強調されてしまった．他のオブジェクトまで強調されてしまったため，四角形はタスクオブジェクトとして採用されなかった．

画像(h)の出力タスクは，「黄色かつ緑色を探して下さい」と言った矛盾した結果が求まった．この原因は大きく 2 つ考えられる．1 つ目の原因は，画像(h)に対し実験で求めたシグモイド関数の c の値では不適切であること．2 つ目の原因は，目標 saliency map と 2 つのタスクオブジェクトで作成した task object saliency map の差が最小 2 乗になるように計算する近似方法だけでは，出力タスクを求めるには不十分であること．これらの原因に対して出力タスクを求める方法を改善する必要がある．

5. まとめと今後の課題

本研究では，トップダウン刺激のひとつであるタスクを用いて視覚注意を意図的に誘導するモデルを提案した．

実験結果より，簡単な画像（現実画像も含む）では，注視先を誘導するにあたって，適切なタスクを求めることができた．しかし，様々なオブジェクトを持つ複雑な画像等では，出力タスクは求められたが，必ずしも適切なタスクであるとは言えない．この問題に対し，方向成分の追加，タスクオブジェクトの数やシグモイド関数の c の値を入力画像と目標 saliency map により自動調整，目標 saliency map と 2 つのタスクオブジェクトで作成した task object saliency map の比較方法を新たに検討，新たに Itti らの確率モデル[10]を取り入れ目標 saliency map からトップダウン刺激を予測等といった解決策があげられる．

また，出力タスクされたタスクに対し，どの程度精度が高いものなのか，注視点計測により検証する必要がある．

参考文献

1) S. Frintrop, E. Rome, and H. Christensen, “Computational Visual Attention Systems and Their Cognitive Foundations: A Survey,” *ACM Transactions on Applied Perception*, Vol. 7, No. 1, pp. 1-39, 2010.

2) Y. Kim and A. Varshney, “Saliency-guided Enhancement for Volume Visualization,” *IEEE Transactions on Visualization and Computer Graphics*, Vol. 12, No. 5, pp. 925-932, 2006.

3) Z. Su and S. Takahashi, “Real-Time Enhancement of Image and Video Saliency Using Semantic Depth of Field”, In *Proceedings of International Conference on Computer Vision Theory and Applications*, pp. 370-375, 2010.

4) E. Mendez, S. Feiner, and D. Schmalstieg, “Focus and Context in Mixed Reality by Modulating First Order Salient Features”, In *Proceedings of the 10th international Conference Smart Graphics2010* , pp. 232-243, 2010.

5) L. Itti, C. Koch, and E. Niebur, “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.

6) V. Navalpakkam and L. Itti, “Modeling the Influence of Task on Attention,” *Vision Research*, Vol. 45, No. 2, pp. 205-231, 2005.

7) S. Frintrop, G. Backer, and E. Rome, “Goal-directed Search with a Top-down Modulated Computational Attention System,” In *Proceedings of the Annual Meeting of the German Association for Pattern Recognition*, Vol. 3663, pp. 117-124, 2005.

8) S. Frintrop, M. Klodt, and E. Rome, “A Real-Time Visual Attention System using Integral Images,” In *Proceedings of International Conference on Computer Vision Systems*, pp. 21-24, 2007.

9) H. Zhao, X. Mao, X. Jin, J. Shen, F. Wei, and J. Feng, “Real-Time Saliency-Aware Video Abstaction,” *The Visual Computer*, Vol. 25, No. 11, pp. 973-984, 2009.

10) L. Itti and P. Baldi, “Bayesian Surprise Attracts Human Attention,” *Vision Research*, Vol. 49, No. 10, pp. 1295-1306, 2009.