

## 注視誘導のための視覚的顕著性に基づく画像加工

萩原 愛子<sup>†1</sup>      杉本晃宏<sup>†2</sup>      川本一彦<sup>†3</sup>

人間の活動を支援する情報システムを実現するためには、情報システムと人間との自然なインタラクションが重要である。とりわけ視線はその人の意図や注意を表しているため、視線を利用したインタフェースが注目されている。人の注視を遮ることなく、別の場所に自然と誘導することができれば、その利用価値は高いと考えられる。そのようなインタフェースを実現するために、本研究では与えられた画像の中で指定した場所が注視されやすくなるように、与えられた画像を加工する手法を提案する。提案手法では、まず入力画像の視覚的顕著性マップを計算する。そして、指定した領域の視覚的顕著性を上げ、領域外の視覚的顕著性を下げるように明るさと色の調整を繰り返すことで、領域内の顕著性が最大となるような画像を生成する。

### Saliency-Based Image Processing for Guiding Visual Attention

AIKO HAGIWARA,<sup>†1</sup> AKIHIRO SUGIMOTO<sup>†2</sup>  
and KAZUHIKO KAWAMOTO<sup>†3</sup>

The information system that assists human activities involves natural interface with human beings as its important part. Gaze information strongly reflects his/her interest or attention and thus gaze-based interface is promising. In particular, if we can smoothly guide his/her visual attention toward a target without interrupting his/her current visual attention, the usefulness of gaze-based interface will be highly enhanced. To realize such an interface, this paper proposes a method for editing an image, given a region in the image, to synthesize an image in which the region is most salient. Our method first computes a saliency map of a given image and then iteratively adjusts intensity and colors so that saliency inside a given region becomes high while that outside the region becomes low. This iteration is carried out until the saliency inside the region becomes highest over the image.

### 1. はじめに

我々は、人間の活動を支援する情報システムに囲まれている。事実、日常では携帯電話をはじめとして、街頭の音声案内や電子掲示板などといった情報システムを頻繁に利用している。このように、情報システムは、もはや、我々の生活に欠かせないものとなっている。誰もが日常的に利用するものであるからこそ、これらは人間との自然なインタラクションを通して利用可能であることが求められている。

情報システムを対象としたインタフェースの手段は主に手や音声や顔である。なかでも、視線を利用したインタフェースが特に注目されている。視線は他の手段に比べ、ユーザの意図や注意をよりよく反映すると考えられるからである。また、視線のインタフェースは、手が他の作業をしていても利用できるという面でも、今後様々な場面で活用することができる。たとえば、注視した場所から人の意図を推定し、作業を支援する手法が提案されている<sup>1)</sup>。そこでは、レゴブロックを組み立てる作業において、その人の注視先から次に行おうと意図している作業を推定し、組み立て作業をロボットが支援している。

視線を利用したインタフェースでは、操作部分や情報表示部分に視線を誘導する必要がある。その際、その人の行動を妨害して無理に視線を誘導するのではなく、自然に視線を誘導することが望ましい。人の視覚特性に基いた視線インタフェースがあれば、ユーザにかかる負担も少なくなり、情報システムをより容易に利用することが可能となる。また、将来的にはウェアラブルディスプレイと組み合わせた道案内や情報提供などにも応用できると考えられる。進行方向や有益な情報がある場所を文字や記号で明示するのではなく、そこに自然と視線が向くようにすることによってユーザビリティを高めることができると考えられるからである。

人が注視する対象を選択する主な要因は、ユーザの興味や意図であると考えられるが、必ずしもそのみではない。見る対象の様子によっても注視のしやすさは変わってくる。見る対象の様子を変化させるために視野内に刺激を呈示し、注視を誘導する方法がある<sup>2)</sup>。周辺視野内に刺激が呈示されると、現在の注視を中断し、視線を移動させて刺激部分を新たに注視

<sup>†1</sup> 千葉大学大学院融合科学研究科  
Graduate School of Advanced Integration Science, Chiba University

<sup>†2</sup> 国立情報学研究所  
National Institute of Informatics

<sup>†3</sup> 千葉大学総合メディア基盤センター  
Institute of Media and Information Technology, Chiba University

することになる。LED の点灯を利用した注視誘導システム<sup>3)</sup> もこれを利用している。しかしこういった方法は、警告としては有意義だが、現在注視している対象を強制的に変えるという面で人の行動を阻害する要因を含んだ注視誘導になっている。一方、アイコンタクトや指差しによって視線を誘導することも可能である<sup>4)</sup>。しかしこれは利用者が指差しといった指標を最初から注視していることが前提となっている。この指標部分への注視の誘導が確立できていない限り、利用することができない。

そこで、無意識かつ自然に注視を誘導するため、視覚的顕著性<sup>5)</sup> を利用する方法を考える。視覚的顕著性とは人間の注視の引きつけやすさを意味し、画像から単純な視覚特徴を抽出・統合することによって、人間が注意を向けやすい画像中の領域を視覚的顕著性マップとして求めることができる。ここでは、視覚的顕著性が高いほど、人間の注視を引きつけやすいとされている。ユーザの興味や意図といった要因を考慮しなければ、人間は視野内で最も視覚的顕著性の高い領域を注視すると考えられている。そこで本研究では、画像中で指定した領域の視覚的顕著性が最も高くなるように入力画像を加工する手法を提案する。提案手法では入力画像に対して、視覚的顕著性が指定領域内で高くなり、領域外で低くなるように明るさと色を調整することを繰り返す。そして、指定領域内の視覚的顕著性が画像内で最も高くなった時点での画像を出力する。これにより、注視を指定領域に自然に誘導することができる。

## 2. 人間の視覚特性

### 2.1 網膜神経細胞と受容野

人間の眼に入った光は網膜上の視神経細胞によって電気信号に変換される<sup>6)</sup>。そしてその信号は脳に伝達される前に、網膜神経節細胞 (Retinal Ganglion Cell; 以下 RGC と略記) と呼ばれる神経細胞によって処理が施される。この RGC には受容野と呼ばれる領域がある。

受容野にはオン (on) 領域とオフ (off) 領域がある。オン領域では明るさが増加する変化に対して正の反応を、明るさが減少する変化に対して負の反応を示し、オフ領域ではこの逆となる。これらオン領域とオフ領域が同心円状に形成され、中心と周辺の光の受け方によってそれぞれの挙動が変化する。中心がオンで周辺がオフのオン中心型と、中心がオフで周辺がオンのオフ中心型の 2 種類がある。

図 1 のように中心のみに光による刺激が与えられた場合、オン中心型が強く反応する。一方で、オフ中心型は周辺のみへの刺激に対して強く反応する。また、オン中心型オフ中心型と

もに、受容野全体を覆うような反応には、中心と周辺が抑制し合い弱い反応しか示さないとされている。この仕組みにより、光の強度の差やエッジの検出が行われている。受容野は単純に光の刺激が強ければ大きく反応するのではなく、変化に対して最も大きな反応を示す。

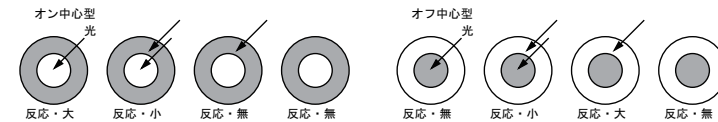


図 1 オン中心型とオフ中心型の刺激に対する反応

### 2.2 中心視と周辺視

人間の視野には中心視と周辺視という 2 種類がある<sup>6)</sup>。中心視は視線の先の物体の細かい形や色を認識することに優れているのに対して、周辺視は、形状把握は中心視には劣るが、周辺視内の物体の位置や運動に対して素早く反応する。また、中心視の範囲が狭いのに対して、周辺視の範囲は広い。

人は目を向け、注視することで視線の先の物体から多くの情報を得る。そのため、人の注意を向けたい対象がある場合は、単に周辺視で見えるように視野に入れるのではなく、注視させることが重要である。人間は視線を向けている今の場所から次の場所へ視線を移すとき、周辺視の情報に基づいて高速に次の視線の先を選択しているとされている<sup>2)</sup>。そこで、人が注視している先を検出し、そこから周辺視内の画像を変化させることで次の注視先を誘導することができると考えられる。たとえ注視の最終誘導目的が周辺視の外にあったとしても、注視を誘導し、移動した注視先における周辺視の中でまた誘導する、ということを繰り返していけば、最終的に目的の場所に注視を誘導することは可能である。

## 3. 視覚的顕著性マップの計算モデル

視覚的顕著性とは、人間の注視の引きつけやすさを示す指標である。そして、これを画像の形で表したものが視覚的顕著性マップである。視覚的顕著性を求める計算手法は広く研究され、様々な手法がある<sup>5)</sup>。画像から単純な視覚特徴を抽出・統合することで視覚的顕著性マップを求める仕組みは Koch と Ullman<sup>7)</sup> によって提案され、その後、Itti<sup>8)</sup> によって計算モデルが確立された。この Itti のモデルはその後の研究で幅広く用いられており、本研究でもこのモデルに基づいて視覚的顕著性マップを求める。Itti, Koch の手法は画像の輝

度, 色相成分, エッジの向きの 3 要素を用いて視覚的顕著性マップを計算する.

エッジの向きは対象物の形状を反映した特徴であり, これに調整を加えると画像中で物体を移動させるという変化を引き起こしてしまう可能性がある. これは視野内に急激な変化をもたらす刺激を提示することになり, 本研究の目的である自然な注視誘導には適していない. そこで, 今回の画像加工は輝度と色相の操作に限るとする. そのため, 輝度と色相のみを用いて視覚的顕著性マップを計算する. 図 2 にその概要を示す.

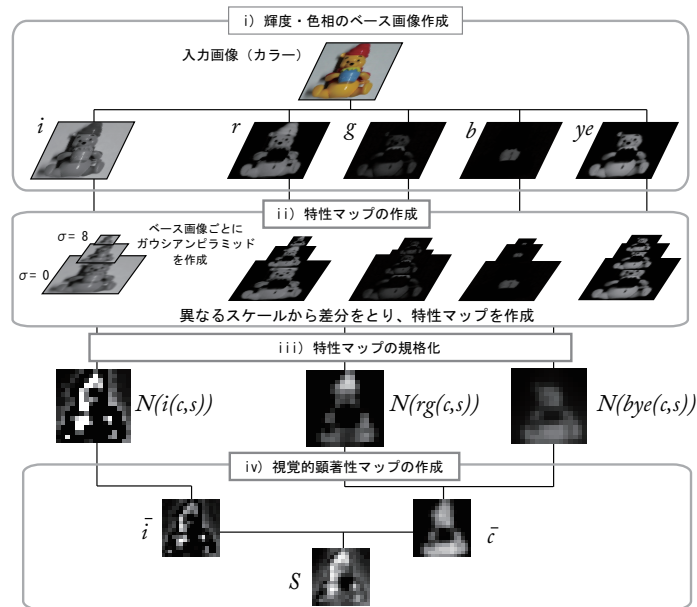


図 2 視覚的顕著性マップの計算モデルの概要

### 3.1 輝度・色相のベース画像作成

入力画像から, 輝度の大きさを表す画像  $i$ , 色相の大きさを表す画像  $r, g, b, ye$  の 5 枚のベース画像を作成する. 入力画像中の点  $(x, y)$  の  $R, G, B$  それぞれの値を  $R_{xy}, G_{xy}, B_{xy}$  と記す.  $i, r, g, b, ye$  の各点の値は次のように定義する.

$$i_{xy} = \frac{R_{xy} + G_{xy} + B_{xy}}{3},$$

$$r_{xy} = \frac{R_{xy} - (G_{xy} + B_{xy})}{2},$$

$$g_{xy} = \frac{G_{xy} - (R_{xy} + B_{xy})}{2},$$

$$b_{xy} = \frac{B_{xy} - (R_{xy} + G_{xy})}{2},$$

$$ye_{xy} = \frac{(R_{xy} + G_{xy})}{2} - |R_{xy} - G_{xy}| - B_{xy}.$$

$i_{xy}$  は画素値の平均という白黒画像としてとらえたときの明るさを表す. また,  $r_{xy}, g_{xy}, b_{xy}, ye_{xy}$  はそれぞれ赤, 緑, 青, 黄という 4 色それぞれが各点でどれだけの強さをもつかを表す. たとえば, ある 2 点において一方の  $R, G, B$  が  $(R, G, B) = (100, 100, 100)$ , もう一方が  $(R, G, B) = (100, 0, 0)$  であるとする. このとき, 輝度  $i_{xy}$  の値は前者の方が高く, 後者が低くなる. しかし, 色相  $r_{xy}$  は, 前者が低く後者が高い値を示すことになる. なお, 人間は輝度の低い領域に対しては色彩を知覚することができないため, ある点において  $r_{xy}, g_{xy}, b_{xy}, ye_{xy}$  の値が  $i$  の最大値の  $\frac{1}{10}$  以下であった場合は, その  $r_{xy}, g_{xy}, b_{xy}, ye_{xy}$  の値を 0 にする. また, 負の値になった場合もその  $r_{xy}, g_{xy}, b_{xy}, ye_{xy}$  の値を 0 とする.

### 3.2 特性マップの作成

5 枚のベース画像それぞれをガウシアンフィルタにより平滑化する, そして 2 次元配列において, 1 行おき, 1 列おきに読み飛ばすことでダウンサンプリングを行う. この平滑化とダウンサンプリングの処理を行うことでガウシアンピラミッドを作成する. 入力画像のスケールを  $\sigma = 0$  として, この平滑化とダウンサンプリングを 1 度行う度にスケール  $\sigma$  が 1 つ大きくなる.  $\sigma = 0, \dots, 8$  と 9 段階の異なる画像  $i(\sigma), r(\sigma), g(\sigma), b(\sigma), ye(\sigma)$  を作成する. ここで,  $i(\sigma)$  はスケールが  $\sigma$  の  $i$  についてのガウシアンピラミッド画像を示す.  $\sigma = 8$  となる画像の大きさは  $\frac{1}{256}$  まで縮尺されることになる.

異なるスケールの画像に対して, 小さい方の画像を大きい方の画像と同じ大きさになるようバイリニア補間<sup>9)</sup>によって拡大した後, 対応するピクセルごとに差分をとる演算を  $\ominus$  と記す. 輝度に関する特性マップ  $i(c, s)$  と色相に関する特性マップ  $rg(c, s), bye(c, s)$  を次のようにして求める<sup>\*1</sup>.

\*1 本稿では, 2 枚の画像間で行う  $+$ ,  $-$  の計算は, 画像間に対応するピクセルごとに和または差を求めることを意味する.

$$i(c, s) = |i(c) \ominus i(s)|, \quad (1)$$

$$rg(c, s) = |(r(c) - g(c)) \ominus (g(s) - r(s))|, \quad (2)$$

$$bye(c, s) = |(b(c) - ye(c)) \ominus (ye(s) - b(s))|. \quad (3)$$

ここでは  $\sigma = s, c (c \in \{2, 3, 4\}, s = c + \delta, \delta \in \{3, 4\})$  としている。  $s$  と  $c$  の組み合わせには 6 通りあるため、輝度に関して 6 種類、色相に関して 12 種類の特性マップが作成されることになる。

輝度に関する特性マップ  $i(c, s)$  は異なるスケールの画像から差分をとることで求まる。一方、色相に関する特性マップ  $rg(c, s), bye(c, s)$  は、人間の視覚における反対色の性質から、対になる  $r$  と  $g$ 、 $b$  と  $ye$  それぞれの差を求めてから、異なるスケールの画像から差分をとることで求まる。なお、 $rg(c, s), bye(c, s)$  はそれぞれ反対色と差が大きく、かつ異なるスケールとも差が大きい点が大きな値をもつことになる。

式 (1), (2), (3) にある異なるスケールの画像から差分をとる処理により、周囲に比べて差異のある画素が高い値を示すことになる。これは 2.1 節に記した周囲と異なる刺激に対して強く反応する人間の受容野の働きとよく似た働きを示す。小さな受容野はスケールが小さく細かい画像、大きな受容野はスケールが大きく粗い画像に対応する。受容野の大きさは様々なので、数種類の  $s$  と  $c$  の組み合わせによって受容野を表している。

### 3.3 特性マップの規格化

視覚的顕著性を考える上では、周囲と異なる刺激が重視される。特性マップ上で、大きな値をもつ点に対して、他の特性マップと統合していくうちにその値が小さくならないように規格化を行う。この規格化によって、ノイズの蓄積による特性マップ上での値の増加を抑制することも可能となる。次に記す規格化演算  $N$  を全ての特性マップに施す。

---

#### Algorithm 1 規格化演算 $N$

---

- 1: 特性マップの数値が  $[0, M]$  の範囲内におさまるようにする。
  - 2: 特性マップ上の最大値  $M$  を除いた、極大値の平均  $\bar{m}$  を求める。
  - 3:  $(M - \bar{m})^2$  を全体に乗じる。
- 

特性マップ上のある点が他の点と比べて非常に大きな値であれば、その特性マップは重要な特徴を示しているため、 $(M - \bar{m})^2$  の値は大きくなる。したがって、大きな値を乗じるこ

とで、その特性マップの重要度を上げることができる。一方、均一な刺激が散乱しているような画像であれば、多数の点が類似した値をもち、たとえその値が大きかったとしても  $(M - \bar{m})^2$  の値は小さくなる。したがって、 $(M - \bar{m})^2$  を乗じることで、その特性マップの重要度を下げることができる。特性マップ  $n$  に規格化演算  $N$  を施した後の画像を  $N(n)$  と記す。

### 3.4 視覚的顕著性マップの作成

正規化演算を施した特性マップを重ね合わせ、最終的な視覚的顕著性マップを作成する。ここで、スケールが異なる 2 つの画像に対して、画像の大きさが小さい方の画像を大きい画像と同じ大きさになるようにバイリニア補間によって拡大した後、対応するピクセルごとに和を求める演算を  $\oplus$  と記す。輝度に関する特性マップを統合した画像  $\bar{i}$ 、色相に関する特性マップを統合した画像  $\bar{c}$  を次のように定義する。

$$\bar{i} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(i(c, s)),$$

$$\bar{c} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [N(rg(c, s)) + N(bye(c, s))].$$

輝度、色相成分について統合した  $\bar{i}, \bar{c}$  にそれぞれ再度規格化演算  $N$  を施して、その 2 枚を足し合わせることで視覚的顕著性マップが求まる。

$$S = \frac{1}{2}(N(\bar{i}) + N(\bar{c})).$$

視覚的顕著性マップにおける各点の値を  $S$  値と呼ぶ。この  $S$  値が高ければその点（領域）の視覚的顕著性が高いということになる。

## 4. 視覚的顕著性に基づく画像加工

### 4.1 視覚的顕著性の制御

視覚的顕著性マップにおいて、 $S$  値が高い領域は視覚的顕著性が高い。視覚的顕著性が高ければ人の注視をひきつけると考えられる。このことから、与えられた領域の  $S$  値が増加するように画像を加工すれば、加工後はその領域に人の注視が向けられやすい画像になると考えられる。つまり、 $S$  値を制御する画像加工方法があれば、人の注視を任意の場所に引き込めることになる。

画像を加工する際に、画像のピクセルごとに輝度・色相のみを変化させることとし、物体を動かすといった変化は考えないこととする。たとえば、画面中央にボールが写っている画

像を加工する場合、ボール自体を動かしたり大きさを変化させることはせず、ボールや周囲の輝度、色相のみを変化させる。画像に動きを与えることで視覚的顕著性を変化させる方法は Abdollahian<sup>10)</sup> が提唱しているが、本研究では静止画における視覚的顕著性の制御に限定する。

与えられた領域  $D$  の  $S$  値を増加させるためには、指定領域  $D$  内の  $S$  値を上げると同時に、領域外の  $S$  値を下げる必要がある。  $S$  値を上げるにはその点の特徴量の強度を強めればよい。逆に、  $S$  値を下げるにはその点の特徴量の強度を弱めればよい。ここでいう特徴量とは、  $S$  値の計算で用いる輝度や色相である。効率よく領域内の  $S$  値を上げ、領域外の  $S$  値を下げるには画像全体を均一に加工するのではなく、その領域に応じた加工が必要である。

領域に応じた加工の違いを、簡単のため、白黒画像の例で説明する。図 3 のように同じ

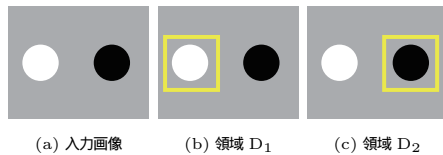


図 3 加工方法の変化説明参考図

入力画像であっても、異なる領域  $D_1, D_2$  を指定した場合、それぞれに必要な加工は異なる。  $D_1$  を指定した場合、領域内は視覚的顕著性を上げるために、白丸はより明るくするべきである。一方、領域外の黒丸は、視覚的顕著性を下げるために、周囲と同じ明るさに近づけ、周囲との差が減るようにするべきである。  $D_2$  が指定された場合、領域内の黒丸はより暗く、領域外の白丸は周囲の明るさに近づけるべきである。この例からわかるように、領域の内外と周囲とを比べた際の明るさの違いによる必要な加工は、表 1 に示すように整理することができる。

表 1 領域と明るさの違いによる加工

	領域内 (=顕著性を上げる)	領域外 (=顕著性を下げる)
周囲より明るい	輝度を上げる	輝度を下げる
周囲より暗い	輝度を下げる	輝度を上げる

本手法では、画像を加工するときは表 1 の関係を考慮し、画素ごとの変化量と強度係数と

いう 2 つの要素を設定する。画素ごとの変化量は、周囲と比べた明るさをふまえて輝度をどれだけ変化させるか決定する要素である。また、強度係数は、領域の内か外かをふまえて輝度や色相の変化の増減を決定する要素である。

点  $(x, y)$  における更新前の画素値を  $R_{xy}, G_{xy}, B_{xy}$ 、更新後の画素値を  $R'_{xy}, G'_{xy}, B'_{xy}$  と記す。提案する画像加工の更新手順を Algorithm 2 に記す。ここで  $Q_{Rxy}, Q_{Gxy}, Q_{Bxy}$  は画素ごとの変化量である。また、  $p_{xy}$  は一度の更新で行う変化の度合いを決定する強度係数である。本手法では一度の加工で大幅に変化させるのではなく、少しずつ画像を更新させて画素値を操作する。指定領域が画像全体で最も高い  $S$  値を示すまで、更新と更新後の  $S$  値の計算を繰り返す。

---

#### Algorithm 2 画素ごとの変化量と強度係数による更新の手順

---

- 1: 強度係数  $p_{xy}$  計算
  - 2: 画素ごとの変化量  $Q_{Rxy}, Q_{Gxy}, Q_{Bxy}$  計算
  - 3: 更新  $\alpha'_{xy} = \alpha_{xy} + p_{xy} Q_{\alpha xy} (\alpha = R, G, B)$
- 

#### 4.2 画素ごとの変化量 $Q_{Rxy}, Q_{Gxy}, Q_{Bxy}$

画素ごとの変化量  $Q_{Rxy}, Q_{Gxy}, Q_{Bxy}$  は、視覚的顕著性マップの計算過程において  $S$  値が輝度や色相のどの特徴量の影響を受けているかを判別し、  $R, G, B$  それぞれに適応して値を定める。たとえば、周囲と比べ赤みが強い点は、赤の色相が  $S$  値に大きな影響を与えているため、  $Q_{Rxy}$  は大きな値をもつ。

画素ごとの変化量は、  $S$  値計算から逆算的に、次のように定義する。

$$Q_{Rxy} = z(i, x, y)q_i + z(r, x, y)q_r - z(g, x, y)q_g - z(b, x, y)q_b + z(y, x, y)q_{ye},$$

$$Q_{Gxy} = z(i, x, y)q_i - z(r, x, y)q_r + z(g, x, y)q_g - z(b, x, y)q_b + z(y, x, y)q_{ye},$$

$$Q_{Bxy} = z(i, x, y)q_i - z(r, x, y)q_r - z(g, x, y)q_g + z(b, x, y)q_b + 0.$$

ここで、  $q_i, q_r, q_g, q_b, q_{ye}$  は各点の輝度と色相がもつ特徴の大きさの割合を表す。また、  $z(\beta, x, y) (\beta = i, r, g, b, ye)$  はその座標が示す明るさまたは色相が周囲と比べて大きい小さいかによって異なる符号をもつ関数である。ベース画像の座標  $(x, y)$  における値を  $\beta_{xy}$  とし、ベース画像全体の  $\beta$  の平均を  $\beta_{ave}$  とする。  $z(\beta, x, y)$  は次のように定義する。

$$z(\beta, x, y) = \begin{cases} 1 & (\beta_{xy} > \beta_{ave}) \\ -1 & (\beta_{xy} \leq \beta_{ave}) \end{cases}$$

ある点の特徴が周囲との輝度の差によるものであれば、 $q_i$  が大きな値をもち、色相によるものであれば、ベース画像  $r, g, b, ye$  に対応した  $q_r, q_g, q_b, q_{ye}$  が大きな値をもち、そこで、 $q_i$  は輝度と色相という 2 つの割合を比べ、次式のように定義する。

$$q_i = \frac{N(\bar{i})}{N(\bar{i}) + N(\bar{c})}.$$

これに対して  $q_r, q_g, q_b, q_{ye}$  は、色相の割合を求める。まず、 $r$  と  $g$ 、 $b$  と  $ye$  という反対色のペアで比較し、最後にペアとなっていた色同士を比較する。

$$q_r = \frac{N(\bar{c})}{N(\bar{i}) + N(\bar{c})} \frac{N(rg)}{N(rg) + N(bye)} \frac{N(r^*)}{N(r^*) + N(g^*)},$$

$$q_g = \frac{N(\bar{c})}{N(\bar{i}) + N(\bar{c})} \frac{N(rg)}{N(rg) + N(bye)} \frac{N(g^*)}{N(r^*) + N(g^*)},$$

$$q_b = \frac{N(\bar{c})}{N(\bar{i}) + N(\bar{c})} \frac{N(bye)}{N(rg) + N(bye)} \frac{N(b^*)}{N(b^*) + N(ye^*)},$$

$$q_{ye} = \frac{N(\bar{c})}{N(\bar{i}) + N(\bar{c})} \frac{N(bye)}{N(rg) + N(bye)} \frac{N(ye^*)}{N(b^*) + N(ye^*)}.$$

ここで、視覚的顕著性マップの計算において、 $rg, bye$  がどれだけの割合をもっていたのかを逆算的に求めるために、式 (2),(3) より次のように定義する。

$$rg = rg(2, 5),$$

$$bye = bye(2, 5).$$

さらに、式 (2),(3) で  $r$  と  $g$ 、 $b$  と  $ye$  の 2 種類の色相から特性マップを作成していることから、 $r$  と  $g$ 、 $b$  と  $ye$  がどれだけの影響度をもっていたのかを逆算的に求めるために、式 (1) の形を利用して次のように定義する。

$$r^* = |r(2) \ominus r(5)|,$$

$$g^* = |g(2) \ominus g(5)|,$$

$$b^* = |b(2) \ominus b(5)|,$$

$$ye^* = |ye(2) \ominus ye(5)|.$$

#### 4.3 強度係数 $p_{xy}$

加工前の画像の  $S$  値を全ピクセルに対応させて与える。こうして、 $S$  値の高い点は強度係数の絶対値が大きく、低い点は小さくなる。ただし、 $S$  値を下げるべきである領域外の  $(x, y) \notin D$  であれば、そこでは与えられた  $S$  値を負の値とする。最後に全体にガウシアンフィルタをかけ平滑化し、各ピクセルがもつ値をそのピクセルの強度係数  $p_{xy}$  と定義する。

## 5. 実験

### 5.1 実験手順

入力画像に領域を指定し、提案手法を用いた加工手法によって、領域内の  $S$  値が最大になった画像を得るに必要な更新回数や加工後の画像を検証する。

- (i)  $512 \times 512$  のカラー画像 (図 4(a), 図 5(a), 図 6(a))\*<sup>2</sup> を入力とし、視覚的顕著性マップを作成し、全体の  $S$  値を計算する。
- (ii)  $S$  値を上げる領域  $D$  とサンプル座標点を指定する。
- (iii) Algorithm 2 を適用する。その際、更新のたびに視覚的顕著性マップを求める。サンプル座標点とその 8 近傍における平均の  $S$  値を計算する。更新回数を  $k$  で表記し、50 回更新する。

### 5.2 結果

入力画像とその顕著性マップ、サンプル座標点の位置と領域、更新後の画像とその顕著性マップ、更新回数  $k$  に対する  $S$  値の変化を図 4,5,6 に示す。

実験結果を見ると全ての結果において領域内の点の  $S$  値は単調に増加している。しかし、領域内が最も高い  $S$  値を示すまでにかかる更新回数には差があった。

Img1 と Out1a, Out1b を見比べると、同じ入力画像であっても指定領域が異なれば、更新後の画像の見えは異なることがわかる。ここで、D1a の方が D1b よりも元の  $S$  値が高い領域を指定している。これより、同じ入力画像であれば、指定領域の  $S$  値が最初から高かった方が、少ない更新回数で指定領域が最も高い  $S$  値を示すといえる。

Img1, Img2 の結果から領域外の点に関しては  $S$  値は減少していく傾向にあることが読み取れる。ただし、Img3 の実験では領域外の  $S$  値が一時的に増加した。これより、画素ごとの変化量と強度係数による画像加工によって、画像全体で最も高い  $S$  値を示す点を領域内の点とすることが可能になったといえるが、まだ改良すべき点が残っている。

### 5.3 考察

Img1 を用いた実験では、サンプル座標点  $(x_3, y_3)$  の変化を見ると、更新を始めてからすぐに  $S$  値が急速に減少している。一方、同じく領域外の点であっても、Out1a における  $(x_2, y_2)$  のように元から  $S$  値が低い点については、 $S$  値はほとんど変化していない。これは、領域外で  $S$  値の高い点に対してのみ、その特徴の強さを大きく減らすように指定する

\*2 (C)2011 BIZUTART PHOTOGRAPH <http://www.bizutart.com/photograph/>



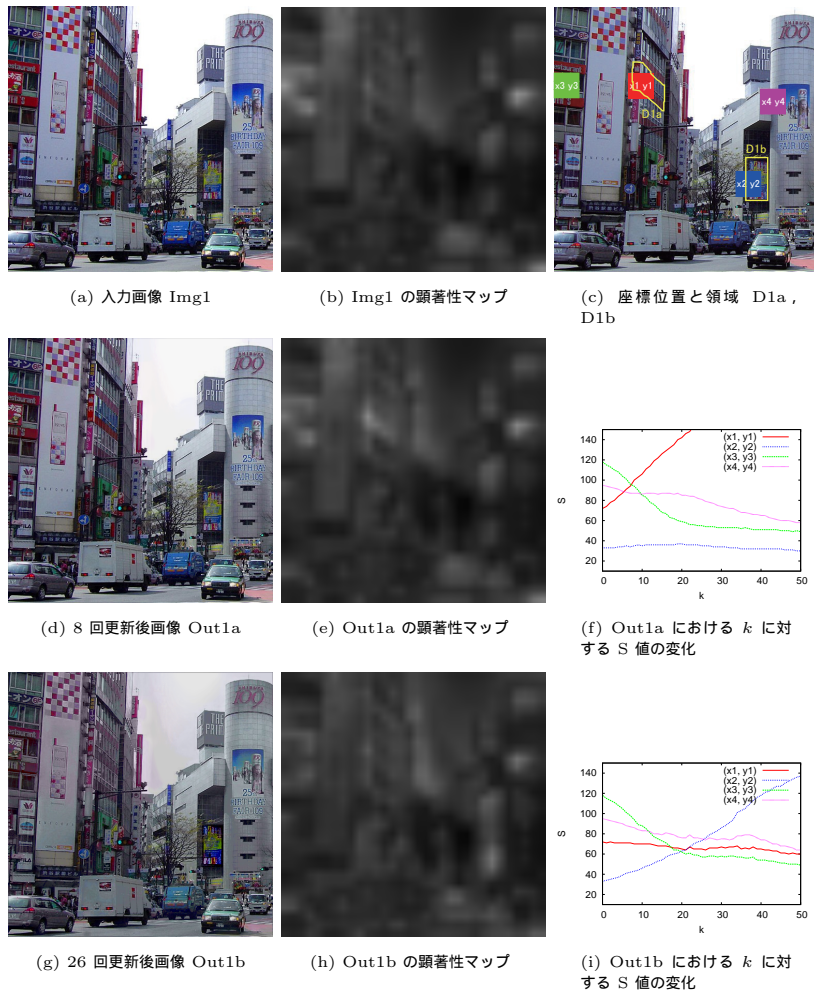


図 4  $Img1$  と  $D1a$ ,  $D1b$  による更新結果

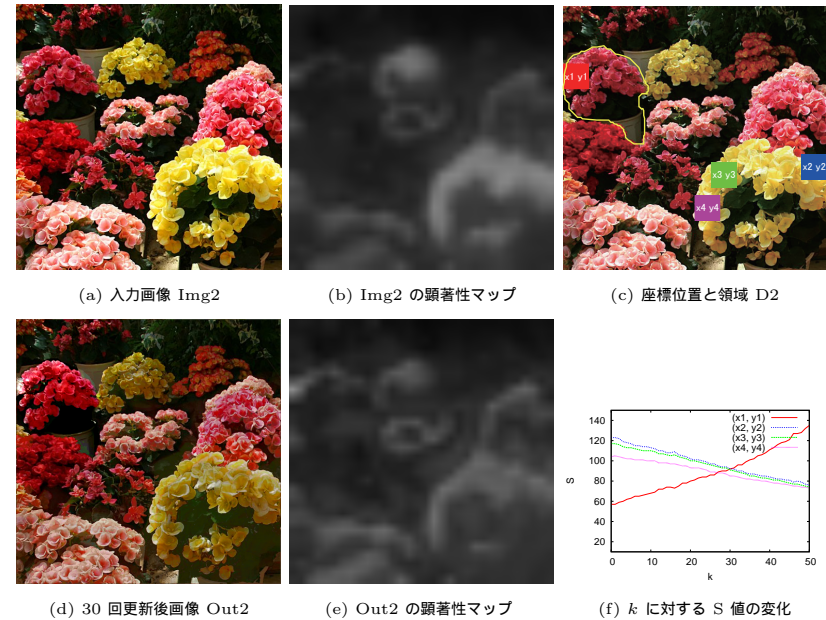


図 5  $Img2$  と  $D2$  による更新結果

強度係数が、意図通りに作用した結果だといえる。

$Img2$  を用いた実験では、 $S$  値は増減いずれの場合も単調に変化しているが、これはこの画像の特性によるものだと考えられる。花弁とそれ以外の明るさの差が大きいため、輝度が  $S$  値に大きく影響を与えていると考えられる。実際更新前後の画像を見比べても、色相の変化は気づけづらく、ほとんど輝度の変化によって  $S$  値を制御しているように見える。

$Img3$  を用いた実験によって、本手法の問題点が見つかった。 $D3$  のような指定領域の  $S$  値が低だけでなく、その領域がもつ特徴と周囲との差が少ない場合、更新を始めてすぐに  $S$  値が増加しないという問題である。 $Img1$  の  $D1b$  と比べると、 $D1b$  も元の  $S$  値は低い。しかし、こちらは領域の周囲には見られない青や黄緑といった色相の特徴をもっていたため、更新を始めてすぐに  $S$  値が増加したと推測できる。一方で、 $D3$  はほぼ後ろの壁と同化しているところから、特徴の強度を強めても急激に  $S$  値が増加しなかったと考えられる。

この原因は画素ごとの変化量と強度係数の両方の仕組みに原因があるといえる。画素ごと

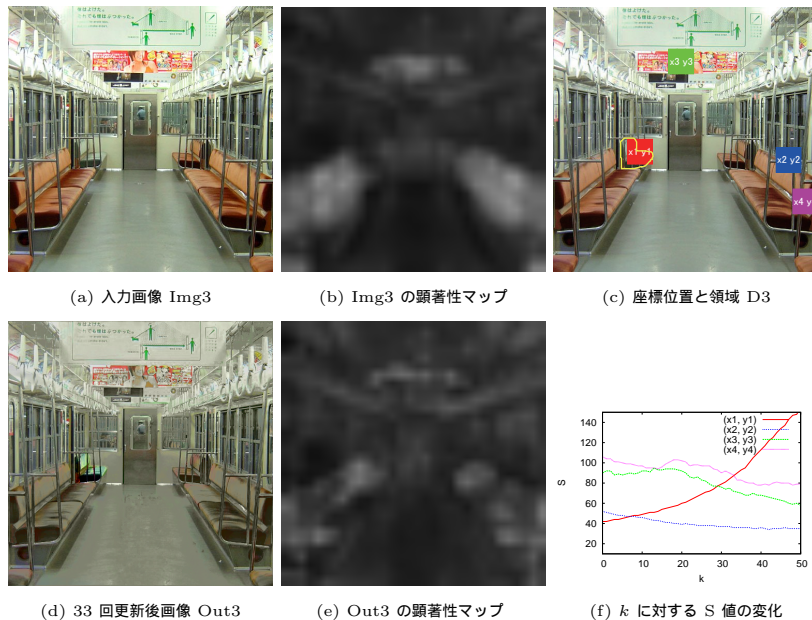


図 6 Img3 と D3 による更新結果

の変化量は、元の特徴を判別して決定しているため、その領域が最初からもっている色相を強めることしかできない。壁と同じような色を強めるよりも、他にない色（この画像でいえば青など）を加えるといった加工の方が、更新してすぐに S 値が増加しただろうと推測できる。これより、画像の中に見られない色を加えるという手法も今後検討していく必要があるといえる。また、強度係数は元の S 値によって絶対値が決まるため、最初の S 値が低いと、そもそも領域内にかかる画素値の変化が少ないという影響を与えている。強度係数に関しては領域外であれば S 値に応じて変化に差をつける必要がある。しかし、領域内には均一に変化を与える方が効率的に S 値が増加する可能性もある。今後、現手法と比較して検証する必要がある。

## 6. おわりに

人の注視を遮ることなく、自然と視線を誘導する視線インタフェースの実現を目指して、

本研究では、視覚的顕著性に基づく画像加工を提案した。見る対象の様子によって注視のしやすさは変わるという発想に基づき、注視の引き込みやすさを示す指標として、視覚的顕著性マップの計算モデルを利用した。そして、画像中で指定した領域の視覚的顕著性が最も高くなるように輝度と色相を画素ごとに調整するという処理を繰り返し、入力画像を加工した。また、画像を用いた実験によって視覚的顕著性の変化を検証し、提案手法の有効性を確認した。今後は、より効率的に視覚的顕著性を変化させるよう手法の改良方法を検討する。それと同時に、人間の実際の注視選択との関連性を検証する予定である。

謝辞 本研究の一部は、JST CREST「共生社会に向けた人間調和型情報技術の構築」領域採択課題「日常生活空間における人の注視の推定と誘導による情報支援基盤の実現」により実施した。

## 参考文献

- 1) 小川原光一, 崎田健二, 池内克史. 視線運動からの意図推定に基づいたロボットによる行動支援. *インタラクション*, pp. 103–110, 2005.
- 2) 緒方康匡, 内川恵二. 第一サッカーを誘導するための視覚刺激条件. *ITE technical Report*, Vol.33, No.17, pp. 57–60, 2009.
- 3) 猪目博也, 饗庭絵里子, 下斗米貴之, 刀 V 隆史, 長田典子. Led 警光灯の視認性向上: 目立ちやすさの感性指標に基づく点滅パターン評価法. *映像情報メディア学会技術報告*, Vol.34, No.10, pp. 89–92, 2010.
- 4) 鈴木祐也, 葛岡英明, 山下淳, 山崎敬一, 山崎晶子, 久野義徳. 無言のロボットによる注意誘導の研究. *ヒューマンインタフェースシンポジウム 2007*, pp. 615–618, 2007.
- 5) L.Itti and CKoch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, Vol.2, pp. 194–230, 2001.
- 6) 内川恵二, 稲葉敬三. *視覚 I 視覚系の構造と初期機能*. 朝倉出版, 2007.
- 7) C.Koch and S.Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, Vol.4, No.4, pp. 219–227, 1985.
- 8) 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.
- 9) コンピュータグラフィックス編集委員会編. *コンピュータグラフィックス*. CG-ARTS 協会, 2007.
- 10) G.Abdollahian, C.M. Taskiran, Z.Pizlo, and E.J. Delp. Camera Motion-Based Analysis of User Generated Video. *IEEE Transactions on Multimedia*, Vol.12, No.1, pp. 28–41, 2009.