

視覚注意ピラミッド上での確率的な注意選択とシフト

Stochastic Selective Attention and its Shift on the Visual Attention Pyramid

渥美雅保

Masayasu Atsumi

創価大学工学部情報システム工学科

Information Systems Sci., Faculty of Eng., Soka University

This paper proposes a computational model of visual attention which performs stochastic attentional selection and shift on the visual attention pyramid that is computed for each frame of a video sequence. In this model, the visual attention pyramid is generated according to the rareness criteria by using features of contrast, orientation and motion energy on a Gaussian resolution pyramid. On this attention pyramid, stochastic attentional selection and shift is performed on mechanisms of dynamic maintenance of attentional IOR, local bottom-up and top-down spatial modulation of attention and global competitive modulation of attention. Experimental results show that this model achieves stochastic visual pop-out to artificial pop-out targets and stochastic attentional selection and shift, especially the whole-part attention shift and the motion-follow attention, in daily scenes.

1. はじめに

視覚注意は、膨大な情報を含む動的シーンから注目すべき物体や事象を含む領域を選択し深い情報処理に渡す役割を果たし、計算資源と時間の制約の中で効果的に実世界を認識するために、人間はもちろん実世界で活動するロボットやコンピュータビジョンシステムに必要不可欠な機能である。視覚注意の研究の歴史は古く多くの研究がなされているが、現在の研究に大きな影響を与えている研究に Treisman の特徴統合理論 [Treisman 80] や Wolfe の Guided Search Model [Wolfe 89], そして, Koch&Ullman の顕著性マップモデル [Koch 85, Itti 98] がある。

本論文では、主に顕著性に基づく注意に関して、従来の研究 [Itti 01] で扱われていない次のような特徴を持つ注意の計算モデル、即ち、動画の各フレームに対して計算された視覚注意ピラミッド上で確率的な注意の選択とシフト、特に人物や物体の全体部分関係に沿った注意のシフトを可能とする注意のモデルを提案する。具体的には、本モデルは次のような特徴を持つ。第1に、ガウシアン解像度ピラミッド [Burt 83] 上で計算された明度・彩度・色相コントラスト、方位成分、及びモーションエネルギーを用いて、希少性基準と呼ぶ顕著性競合基準に基づき視覚注意ピラミッドと名づける多レベル顕著性(注意)マップを計算する。第2に、視覚注意ピラミッド上での確率的な注意選択とシフトを、注意のIOR(Inhibition Of Return)[Klein 00]の動的管理、注意の局所空間的な調節、注意の大域的な競合による調節のもとに実現し、注意の集中と分散、及び全体部分関係に沿った注意のシフトを可能とする。また、第3に、顕著性に基づくボトムアップの注意と外生的なトップダウンの空間注意の融合を実現する。

そして、本注意のモデルに関して、視覚ポップアウト効果、及び実世界で注意を引くべく設計された物体や日常的な動的シーン内の人物や物体への選択的注意とそのシフト、特に全体部分関係に沿った注意のシフトと移動体への追従的注意の性能を、ビデオクリップを用いた実験により評価する。

以下、2章で視覚注意ピラミッド、3章でその上での確率的な注意選択とシフトの機構、4章で実験結果について述べ、5章で本論のまとめを述べる。

連絡先: 渥美雅保, 創価大学工学部情報システム工学科, 〒192-8577 八王子市丹木町 1-236, matsumi@t.soka.ac.jp

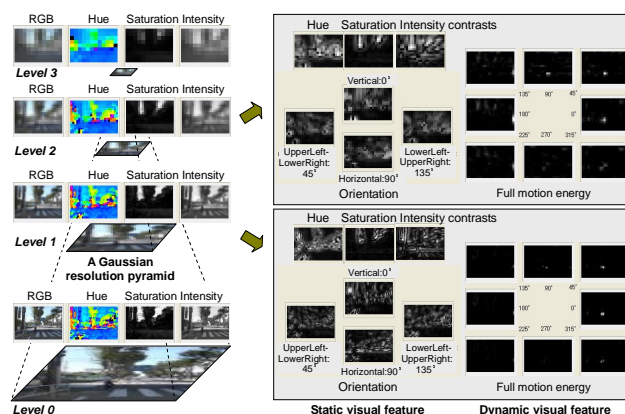


図 1: ガウシアン解像度ピラミッド上の視覚特徴量

2. 視覚注意ピラミッド

2.1 ガウシアン解像度ピラミッド上の視覚特徴量

人の視覚系には異なる時空間周波数に応答選択性を持つ並列知覚チャンネルが存在することが知られており、これら並列知覚チャンネルは視覚計算モデルとしてはガウシアン解像度ピラミッド [Burt 83] 上に実現される。そこで、本モデルでは、画像からガウシアン解像度ピラミッドを視覚特徴量の計算の基盤として求めて、その各レベルの各点の明度・彩度・色相、明度・彩度・色相の空間変化からそれらのコントラストと方位成分、及び明度の時空間変化からフルモーションエネルギー [Adelson 85] を、視覚特徴量として計算する。

図 1 にガウシアン解像度ピラミッド上の視覚特徴量の計算例を示す。明度コントラスト C_I , 彩度コントラスト C_S , 色相コントラスト C_H は、明度、彩度、色相をそれぞれ LoG (Laplacian of a Gaussian) フィルタを用いて畳み込むことにより計算される。ただし、色相の値 $[0, 2\pi)$ はそのカテゴリを表しているため、各点の周囲の点との相対色相値を求めて畳み込みを行う [Atsumi 07]。明度方位成分 O_I , 彩度方位成分 O_S , 色相方位成分 O_H は、明度、彩度、色相をそれぞれ $0^\circ, 45^\circ, 90^\circ, 135^\circ$ のガボールフィルタを用いて畳み込むことにより、これら 4 つの方位に関して計算され、各点の方位成分はそれらの最大値として求められる。ただし、色相方位成分に関しては各点

の周囲の点との相対色相値を求めて畳み込みを行う。フルモーションエネルギー M は、ガウシアン解像度ピラミッドの時系列に対して、各レベルの各点においてそれら明度を時空間ガボールフィルタを用いて畳み込むことに基づいて求められる。8つの方向 $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$ と2つの速度のフルモーションエネルギーを計算する。

2.2 多レベル顕著性マップ

ガウシアン解像度ピラミッドの各レベルにおいて計算された明度・彩度・色相コントラスト、方位成分、及びフルモーションエネルギーの特徴は、顕著性を競合的に表現するためにそれぞれ複数の次元により構成される。明度コントラストは、オン中心-オフ周辺型反応とオフ中心-オン周辺型反応に対応する2つの次元により構成される。彩度コントラストも同様に、オン中心-オフ周辺型反応とオフ中心-オン周辺型反応に対応する2次元で構成される。色相コントラストは、赤、黄、緑、シアン、青、マゼンタの6つの代表色相に対応する6次元で構成される。また、方位成分は、4方位に対応する4次元、フルモーションエネルギーは8方向、2速度の16次元で構成される。各特徴に関する顕著性は、その特徴を構成する次元間で希少性基準と呼ぶ基準に従い競合的に計算される。希少性基準は、同一特徴の異なる次元の領域が画像内にあるとき、より少なく現れる次元の領域がより目立つことを保証する基準である。

明度コントラストのオン中心-オフ周辺型とオフ中心-オン周辺型の2つの次元は、LoGフィルタを用いた畳み込みにおいてそれぞれ正符号と負符号により表され、コントラストの強さは畳み込み値の絶対値により表される。そこで、畳み込み値が正のとき、オン中心-オフ周辺次元のコントラストはその値、オフ中心-オン周辺次元のコントラストは0、畳み込み値が負のとき、オン中心-オフ周辺次元のコントラストは0、オフ中心-オン周辺次元のコントラストはその値の絶対値、畳み込み値が0のときは2つの次元ともコントラストは0とする。そして、オン中心-オフ周辺次元とオフ中心-オン周辺次元のコントラストの最大値 C_{Ion}^*, C_{Ioff}^* と、ある希少性閾率 $\theta_R (0 \leq \theta_R \leq 1)$ に対して、各次元の値がそれぞれ $\theta_R \times C_{Ion}^*, \theta_R \times C_{Ioff}^*$ 以上の点の個数 n_{Ion}, n_{Ioff} を求め、各次元の希少性を単調減少関数 $r(\cdot)$ を用いて $r(n_{Ion})$ と $r(n_{Ioff})$ で定める。この関数 $r(\cdot)$ を希少性関数と呼び、本論では $r(n) = 1/\sqrt{n}$ を用いる。このとき、各点 p の明度コントラストに関する顕著性 $S_{IC}(p)$ は、2つの次元の値をそれぞれの希少性関数値で重み付けした和を求め、それらの最大値がコントラストの最大値 $\max(C_{Ion}^*, C_{Ioff}^*)$ となるように正規化した値として求められる。

彩度コントラストの2つの次元の値も同様に表され、彩度コントラストに関する顕著性 S_{SC} も同様に計算される。また、方位、モーションエネルギーに関する顕著性 S_{OR}, S_{MO} も、それぞれ、4次元、16次元の値を各次元に対して求めた希少性で重み付けをして和を取り、それぞれの最大値で正規化することにより求められる。一方、色相コントラストに関する顕著性 S_{HC} は、同様の考え方に基づいて、赤、黄、緑、シアン、青、マゼンタの6つの色相次元の希少性を求め、各点の色相コントラストをその色相に最も近い次元の希少性により重み付けした後、同様に正規化することにより求められる [Atsumi 07]。

静的顕著性 $S_S(p)$ は、各点 p の明度顕著性 $S_{IC}(p)$ 、彩度顕著性 $S_{SC}(p)$ 、色相顕著性 $S_{HC}(p)$ 、方位顕著性 $S_{OR}(p)$ から、重み $w_i (i = IC, SC, HC, OR)$ 、ここで $w_{IC} + w_{SC} + w_{HC} + w_{OR} = 1$) を用いて次のように計算される。

$$S_S(p) = \sum_{i=IC,SC,HC,OR} (w_i \times S_i(p)) \quad (1)$$

また、動的顕著性 $S_D(p)$ はモーションエネルギー顕著性から $S_D(p) = S_{ME}(p)$ と求められ、これら静的顕著性と動的顕著性が重み $w_S, w_D (w_S + w_D = 1)$ により合成されて顕著性

$$S(p) = w_S \times S_S(p) + w_D \times S_D(p) \quad (2)$$

が求められる。これら顕著性を各レベルの各点で計算することにより、多レベル顕著性マップ SP が求められる。このマップ、及びこれをIOR、局所的空間注意、大域的競合抑制で調節したマップを視覚注意ピラミッドと呼ぶ(図2)。

3. 確率的な注意選択とシフト

3.1 注意確率空間

視覚注意ピラミッド上に確率空間を定義する。まず、視覚注意ピラミッドの各レベルの各点への注意は互いに識別可能である、即ち、異なるレベルの空間的に包含関係にある点のそれぞれに注意は別々に向けると仮定する。そして、視覚注意ピラミッドのあるレベル $n (n = 0, \dots, N-1; n=0$ は元画像のレベル) のある点 $p(n, x, y)$ の注意度(顕著性)を $a(n, x, y)$ とするとき、その点への注意確率 $prob(n, x, y)$ を

$$prob(n, x, y) = \frac{a(n, x, y)}{\sum_{n=0}^{N-1} \sum_{x,y} a(n, x, y)} \quad (3)$$

により定義する。

ところで、IOR[Klein 00]は、注意が最近注意した場所へ戻ることを一時的に抑制する。そこで、注意ピラミッド上でIORが課せられる点の注意度は0とする。また、顕著性に基づく注意は、局所空間的な調節と大域的競合による調節を受けると仮定する。局所空間注意としては、注意に伴いその注意点の周りに生ずるボトムアップの空間注意と、意図的または他者による空間指示等で生ずるトップダウンの空間注意を扱う。

空間注意を、その中心と広がりにより定義する。レベル l の点 $p(l, x_l, y_l)$ を中心とする空間注意のそのレベルでの広がりを σ_l 、 $p(l, x_l, y_l)$ に対応するレベル $n (n = 0, \dots, N-1)$ の空間注意の中心点を $p(n, x_n, y_n)$ とするとき、視覚注意ピラミッドの各レベル n の各点 $p(n, x, y)$ への空間注意の影響 $\phi(n, x, y)$ を次のガウス型関数により与える。

$$\phi(n, x, y) = \lambda^n \exp\left(-\frac{(x-x_n)^2 + (y-y_n)^2}{2 \times \left(\frac{\sigma_l}{2^n - 1}\right)^2}\right) \quad (4)$$

ここで、 λ は空間注意のレベル間乗数で、 $\lambda > 1$ のとき、部分より全体に注意が向けられる傾向を表す。空間注意による注意度 $a(n, x, y)$ の調節は $\phi * a$ により与えられる。ここで、 $*$ は $\sigma_n (= \sigma_l / 2^{n-l})$ が定める空間注意の広がりの範囲内の各点での $\phi(n, x, y)$ と $a(n, x, y)$ の積を表す。

大域的競合抑制による注意の調節は、IORと空間注意による調節がなされた視覚注意ピラミッドに対して、そのすべてのレベルの点の注意度の最大値のある比率 γ 未満の注意度の点への注意度を0とすることにより実現される。この比率を競合抑制閾率と呼び、競合抑制を $\psi(\gamma)$ により表す。

以上により、注意確率空間は、

$$APS := (AP(SP, Q_{IOR}, (\phi_B, \phi_T), \psi(\gamma)), prob) \quad (5)$$

により表される。ここで、 $AP(SP, Q_{IOR}, (\phi_B, \phi_T), \psi(\gamma))$ は、多レベル顕著性マップ SP を、IORが働く点の集合(キュー) Q_{IOR} 、式(4)で与えられるボトムアップ空間注意 ϕ_B またはトップダウン空間注意 ϕ_T 、及び競合抑制 $\psi(\gamma)$ により調節した視覚注意ピラミッドで、 $prob$ は AP 上で式(3)により定義される注意確率である。

3.2 確率的な注意選択とシフトの機構

動画像においては、画像の連続フレーム、即ち視覚注意ピラミッドの系列に対して、確率的な注意選択とシフトが遂行される。そのためには、IOR、及び注意に伴うボトムアップの空間注意と外生的なトップダウンの空間注意は動的に管理される必要がある。まず、IORの動的管理に関して、IORはそれが働く点の顕著性が変わらない限りあるフレーム数に渡って継続するが、IORが働く点の顕著性がある大きさ（IOR解除閾値）以上に変化した場合に解除されるとする。次に、空間注意の動的管理に関して、トップダウンの空間注意は顕著性に基づく注意に対して非同期的に外部から与えられ一定時間継続し、また、注意に伴うボトムアップの空間注意と排他的に作用する、即ち、トップダウンの空間注意が働く間ボトムアップの空間注意は作用しないとする。ここで、注意に伴うボトムアップの空間注意は、注意の占有率に基づいて確率的に生起される。今、注意がレベル l の点 $p(l, x, y)$ に向けられ、その点の注意度が $a(l, x, y)$ とする。このとき、注意占有率 η を、レベル l の全注意度 $a(l)$ に対して $\eta = a(l, x, y)/a(l)$ で与え、ボトムアップに空間注意が作用する確率 p を次の式により求める。

$$p = \frac{1}{1 + \exp\left(-\frac{\eta - \theta_s}{T_s}\right)} \quad (6)$$

ここで、 θ_s は空間注意生起確率が 0.5 となる注意占有率を与える空間注意の変移閾値、 T_s は注意占有率と空間注意生起確率の間の依存関係を規定する温度である。ボトムアップの空間注意は次の注意点の選択に対してのみ作用する。

動画像に対する確率的な注意選択とシフトの手続きを示す。トップダウン空間注意を保持する変数を T で表し、トップダウン空間注意が与えられていない場合 $T := nil$ とする。また、1フレーム内の選択可能な注意点の数が与えられるとする。そして、初期値として、 $T := nil$ 、 Q_{IOR} を空、現在の注意点 $A := nil$ とし、連続的な動画像の毎画像フレームに対して次の手続きを実行する。

[procedure 注意確率空間上での確率的注意選択・シフト]

- Step1 画像から多レベル顕著性マップ SP を計算する。
- Step2 Q_{IOR} から IOR 継続フレーム数上限を越えた点、及び IOR 解除閾値以上に顕著性が変化した点を取り除いた後、 Q_{IOR} 内の点の注意度を 0 とする。
- Step3 トップダウン空間注意 ϕ_T が外生的に与えられた場合、 $T := \phi_T$ とする。すでに与えられているトップダウン空間注意がある場合は、それが継続時間を越えたならば $T := nil$ とする。
- Step4 注意カウントを 0 とする。
- Step5 トップダウン空間注意 ϕ_T が働いているならば、多レベル顕著性マップ SP を ϕ_T で調節し、さもなければ、現注意点の注意占有率に基づいて確率的にボトムアップ空間注意 ϕ_B を生成し SP を ϕ_B で調節する。その後、競合抑制 $\psi(\gamma)$ による調節を行う。
- Step6 多レベル顕著性マップ SP を調節した視覚注意ピラミッド AP が定める注意確率空間 APS において、注意点 $p(l, x_l, y_l)$ を確率的に選択するとともに注意占有率を計算する。 $A := p(l, x_l, y_l)$ とし、また、 $p(l, x_l, y_l)$ を Q_{IOR} に追加する。
- Step7 注意カウントを 1 増やす。注意カウントが、与えられたフレーム内注意数に達したならば終了する。そうでない場合は、Step5 へいく。

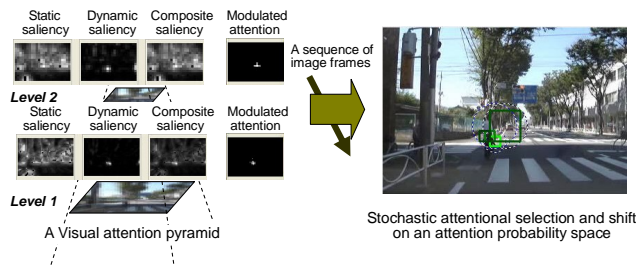


図 2: 視覚注意ピラミッド上の確率的な注意選択とシフト (図 2~6 で、小、中、大の四角形は視覚注意ピラミッド上のレベル 1, 2, 3 の注意点を、その明るさは注意の順序を表す。円は空間注意を表す。)

本手続きにより、高い注意点から低い注意点への選択的注意の確率的シフト、注意の集中と分散、及び全体部分関係に沿った注意のシフトが実現される。図 2 に視覚注意ピラミッド上での注意の選択とシフトの例を示す。

4. 実験結果

4.1 実験枠組

視覚注意ピラミッド上での確率的な注意選択とシフトの性能を評価するための実験を、視覚ポップアウト対象、実世界で注意を引くべく設計された物体、及び日常的な動的シーン内の人物や物体を含むビデオクリップを用いて行った。

実験で用いたパラメータは次のとおりである。ガウシアン解像度ピラミッドのレベル数は 5 で、このうち、視覚注意ピラミッド (注意確率空間) はレベル 1 から 4 に対して計算される。顕著性の計算において、希少性閾率は 0.75、顕著性の合成重みは $(w_{IC}, w_{SC}, w_{HC}, w_{OR}) = (0.25, 0.1, 0.4, 0.25)$ 、 $(w_S, w_D) = (0.05, 0.95)$ である。確率的な注意選択とシフトに関して、IOR 継続フレーム数上限は 2、IOR 解除閾値は 0.1、空間注意のレベル間乗数は 1.25、ボトムアップ空間注意の標準偏差 σ_l はどのレベルに対しても 8、その生成確率を定める θ_s と T_s はそれぞれ 0.25、0.1、競合抑制閾率は 0.8、フレーム内注意数は 4 である。トップダウンの空間注意は、その効果を観測するため 4.4 の最後の実験でのみ継続時間 4 で外生的に作用させた。

4.2 確率的な視覚ポップアウト

明度・色相・方位ポップアウト実験を 25 フレームからなる静止ビデオクリップを用いて行った。図 3 にポップアウトの様子を示す。各フレームで 4 つの注意点が選択されるため注意の数は合計 100 である。明度ポップアウトでは、100 注意点のうち 47% の注意が白丸に向けられ、25 フレームのうち 88% の 22 フレームで少なくとも 1 つの注意が白丸に向けられた。また、白丸とそれを囲む領域との部分全体関係に沿った注意のシフトも観測された。これより、1 つの白丸にほぼ常に注意を集中させながらその周りの 9 個の黒丸にも注意を分散させていることがわかり強いポップアウトが確かめられた。色相、方位に関しても同様のポップアウトが観測された。

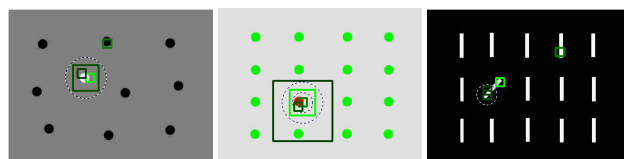


図 3: 確率的な明度 (左)、色相 (中)、方位 (右) ポップアウト。

4.3 注意を引くべく設計された物体への注意

注意を引くべく設計された物体としての交通標識を含む 25 フレームからなる静止ビデオクリップを用いた実験結果を示す。図 4 は、多レベル顕著性マップと選択的注意のシフトの例である。100 注意点のうち 84% の注意が 2 つの交通標識のいずれか、または 2 つを含む領域に向けられ、25 フレームのうち 92% の 23 フレームで少なくとも 1 つの注意が交通標識に向けられた。その際、下の標識から 2 つを含む領域へ、そして上の標識へとといった部分から全体、全体から部分への注意のシフトが観測された。また、車に対しても少数であるが注意が向けられ、その際、車全体とその部分の間で注意のシフトが観測された。これらより、注意を引くべく設計された交通標識に注意が高い確率で向けられること、及び全体部分関係に沿った注意のシフトが起こることが確かめられた。



図 4: 注意を引くべく設計された交通標識に対する注意

4.4 日常的な動的シーンにおける注意

手話と交差点のビデオクリップを用いた実験結果を示す。手話ビデオ^{*1}は 30 フレームからなり、計 120 注意点のうち、動いている右手に 80%、顔に 20% の注意が向けられた。特に、30 フレームのうち右手が動いている 17 のフレームでは、68 注意点中 91% の注意が右手の動きに追従して向けられ、右手への注意は 17 フレームのすべてにわたった。また、全体部分関係に沿った注意のシフトは、右手とその部分、顔とその部分の間において全シフトの 26% で観測された。図 5 に視覚注意ピラミッドとその上での選択的注意のシフトの例を示す。

交差点ビデオは 70 フレームからなり、左から右に進む人と自転車、右から左に進む 2 台の自転車が順に登場し、これら 4 つの移動体に対して、計 280 注意点のうち 87% の注意が、また、70 フレームのうち 97% の 68 フレームで少なくとも 1 つの注意が、移動体を断続的に追跡する様式で向けられた。その他、登場人物が少ないときには、交通信号、道路前方の空間、右側の建物と並木を含む情景にもときどき注意が向けられた。図 6(a) に選択的注意のシフトの例を示す。また、図 6(b) には、トップダウン空間注意を交通信号に向けたときの注意の様子を示す。注意が交通信号に向けられることが観察される。

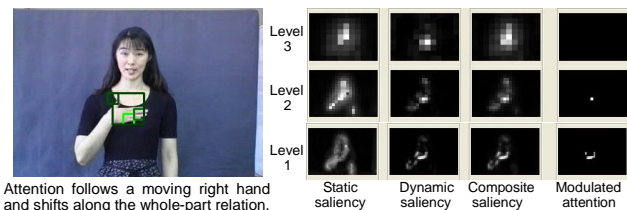


図 5: 手話をする人への選択的注意とそのシフト

これらより、人物の動きや色相的に目立つ顔に高い確率で注意が向けられ、それらが全体部分関係に沿い、また動きに追従してシフトすること、複数の移動体が存在する場合にはそれらの動きに断続的に追従する様式で注意が向けられること、及びトップダウンの空間注意の効果が確認された。

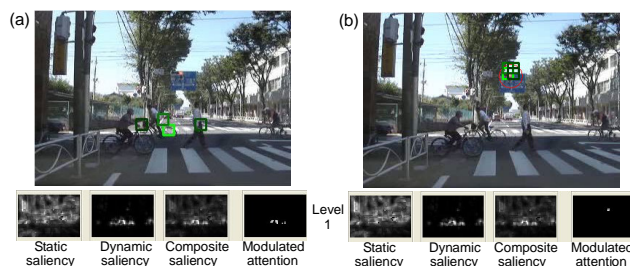


図 6: 交差点での選択的注意とそのシフト

5. おわりに

本論文では、動画上で計算される視覚注意ピラミッド上で確率的な注意の選択とシフトを遂行する注意のモデルを提案した。そして、実験により、視覚ポップアウト効果、及び実世界で注意を引くべく設計された物体や日常的な動的シーン内の人物や物体への、全体部分関係に沿い、また動きに追従的な選択的注意のシフトを確かめ、本モデルの有効性を示した。

謝辞 本研究の一部は科学研究費補助金(課題番号 18500121)の支援のもとに実施された。

参考文献

- [Adelson 85] Adelson, E. H., Bergen, J. R.: Spatiotemporal Energy Models for the Perception of Motion, *J. Opt. Soc. Am. A*, Vol.2, No.2, pp.284-299 (1985)
- [Atsumi 07] Atsumi, M.: Stochastic Attentional Selection and Shift on the Visual Attention Pyramid, *Proc. of the 5th International Conference on Computer Vision Systems*, CD-ROM, 10p. (2007)
- [Burt 83] Burt, P. E. and Adelson, E. H.: The Laplacian Pyramid as a Compact Image Code, *IEEE Trans. on Communications*, Vol.31, No.4, pp.532-541 (1983)
- [Itti 98] Itti, L., Koch, C., and Niebur, E.: A Model of Saliency-based Visual Attention for Rapid Scene Analysis, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.20, No.11, pp.1254-1259 (1998)
- [Itti 01] Itti, L. and Koch, C.: Computational Modelling of Visual Attention, *Nature Reviews Neuroscience*, Vol.2, No.3, pp.194-203 (2001)
- [Klein 00] Klein, R. M.: Inhibition of Return, *Trends in Cognitive Sciences*, Vol.4, No.4, pp.138-147 (2000)
- [Koch 85] Koch, C. and Ullman, S.: Shifts in Selective Visual Attention, *Human Neurobiology*, Vol.4, No.4, pp.219-227 (1985)
- [Treisman 80] Treisman, A. M. and Gelade, G.: A Feature Integration Theory of Attention, *Cognitive Psychology*, Vol.12, pp.97-136 (1980)
- [Wolfe 89] Wolfe, J. M., Cave, K., and Franzel, S.: Guided Search, *Journal of Experimental Psychology*, Vol.15, pp.419-433 (1989)

*1 <http://www.nttdata.co.jp/profile/shuwa/p061402.html>