

粒子フィルタを用いた視覚的注意モデルの検討

A Visual Attention Model Using a Particle Filter

尾関 基行* 井上 茉莉子 柏木 康寛 岡 夏樹

Motoyuki OZEKI, Mariko INOUE, Yasuhiro KASHIWAGI, and Natsuki OKA

京都工芸繊維大学 大学院工芸科学研究科

Graduate school of Science and Technology, Kyoto Institute of Technology

Abstract: A visual attention model based on human visual perception is essential for realizing a natural gaze control and information selection for robots. One of the most famous visual attention models is Itti, Ullman, & Koch's model, which has been used in many different visual tasks and has derived many related visual attention models. However, this model is insufficient to be utilized effectively in a HAI domain because it focuses on bottom-up (task-independent) processing. In this paper, we propose a top-down oriented model that enables a detailed top-down attention control by introducing a particle filter, and demonstrate our method by a brief experiment.

1 はじめに

人間は視覚的注意の機能によって、特に意識していなくても視野内の重要な情報に気づくことができる。また、意識的に注意を働かせることで、無数の候補の中から必要な情報だけを選び出すことができる。前者をボトムアップ性注意と呼び、後者をトップダウン性注意と呼ぶ。我々の注意がどこかに向くという機能は、これら2種類の注意の相互作用によって実現されている。

この視覚的注意の機能を計算機上でシミュレートすることができれば、ヒトの認知発達の構成論的分析[長井 04]をはじめ、知能ロボット[陳 03]や画像検索[Pang 08]、シーン分析[Siagian 07]など、様々な応用分野で役に立つ。ヒューマンエージェントインタラクション(HAI)においても、エージェントに生き物らしい振る舞いを与えたり[Breazeal 99][小嶋 01]、コミュニケーションを円滑化する[今井 01]など、期待される効果は大きい。

代表的な視覚的注意モデルとして、Treismanの特徴統合理論[Treisman 80]に基づいたKoch, Ullman, Ittiらによるモデル[Koch 85][Itti 01]が挙げられる。このモデルを以降ではItti-Kochモデルと呼ぶ。このItti-Kochモデルは応用研究に幅広く利用されており、また、ニューラルネットワークによる実装[久米 99]や確率モデルの導入[Pang 08][宮里 09]といったの派生モデルの研究も盛んに行われている。

一方、これら先行研究のモデルは、ボトムアップ性注意において人間の視覚をよく予測するものであるが、

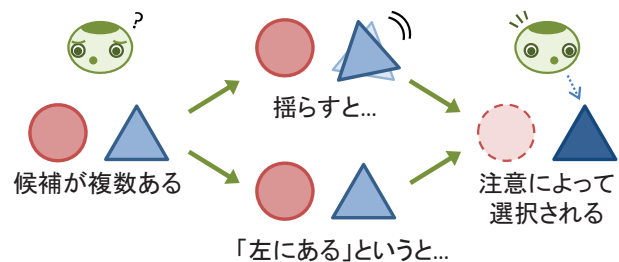


図 1: 視覚的注意による情報選択(上のルートはボトムアップ性注意, 下のルートはトップダウン性注意)

トップダウン性注意による注意制御には重点が置かれていない。シーン解析や画像検索などの用途ならそれで十分だが、HAIで有効に活用するにはトップダウン性注意による注意制御が細やかに行えることが望ましい。例えば、図1に示した視覚的注意による情報選択を考えよう。下のルートによる情報選択はトップダウン性注意によるものである。これを(視線制御ではなく)注意制御の枠組みで扱うには、現在の注視点の左側を重点的に知覚処理するといった制御ができなくてはならない。[Breazeal 99]や[Oliva 03]などではItti-Kochモデルに基づいたトップダウン性注意の研究もなされてはいるが、いずれも知覚処理(特徴抽出処理)よりも上位のレベルでの制御・工夫を行っている。

このような背景から、我々は、トップダウン性注意による細やかな注意制御を可能する視覚的注視モデルを構築している。我々のモデルの特徴は、ボトムアップ性注意の処理の部分に“粒子フィルタ”を導入した点である。粒子フィルタとは、観測値 y に基づいて、内部

*連絡先: 京都工芸繊維大学 大学院工芸科学研究科
〒606-8585 京都市左京区松ヶ崎橋上町
E-mail: ozeki@kit.ac.jp

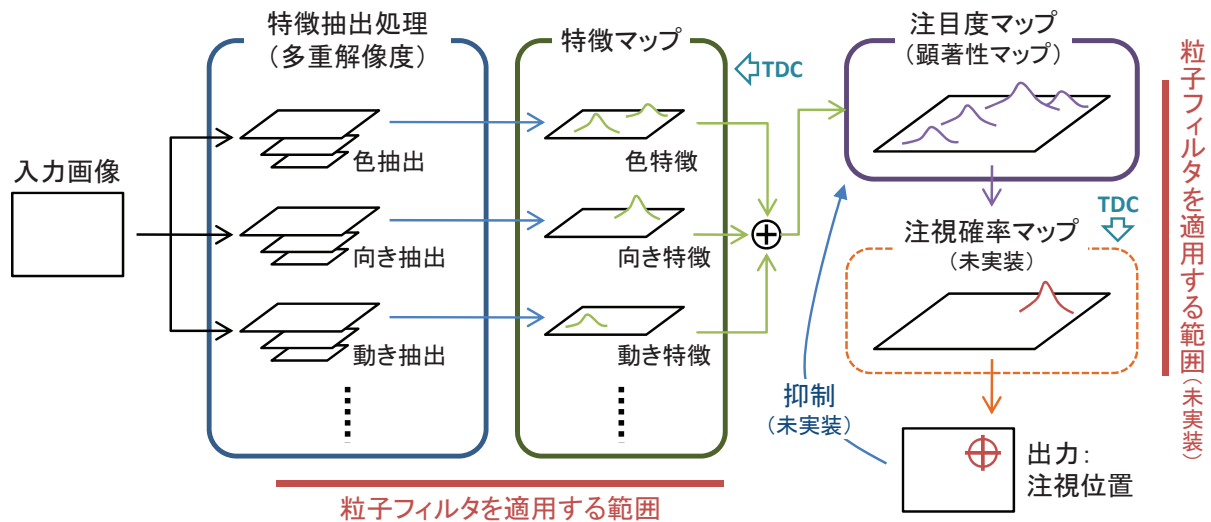


図 2: 視覚的注意モデルの概要 (TDC はトップダウン注意制御の略)

状態 x の分布を逐次推定する確率的手法である。ここで、観測値 y を外部からの知覚刺激、状態 x を視野内における注目位置とし、条件付確率 $P(y|x)$ を位置 x における注目度とすることで、Itti-Koch モデルに適用できる。粒子フィルタにはいくつかのパラメータがあり、それらをトップダウン性注意に基づいて調整することで、注意の振る舞いを細かく調整することができる。

以下、本稿では、2 章で本研究の全体構想とその位置づけを示し、3 章に粒子フィルタの概説を挟んで、4 章で全体構想の一部（特徴抽出処理の部分）の具体的な実装方法を説明する。最後に 5 章にて、実装したモデルを用いた実験の結果とその考察を述べる。

2 視覚的注意モデル

本研究で実現しようとする視覚的注意モデルの概要を図 2 に示す。このモデルは Itti-Koch モデルを本研究の目的に合わせて簡略化したものである。

まず、エージェントの目（カメラ）から入力された画像に対して複数の特徴抽出を並列的にを行い、それぞれの特徴マップ（feature map）を算出する。特徴マップとは、視野の各位置における特徴の強さ（注意の引かれる度合い）を画素値とした画像である。その後、各特徴マップの線形和をとることで顕著性マップ（saliency map）を構成する。ただし、本モデルの顕著性マップは厳密には Itti-Koch モデルのものとは異なるため、顕著度を注目度と呼び、顕著性マップを注目度マップと呼ぶことにする。以上の処理をボトムアップ処理と呼ぶ。

注目度マップにおいて一定以上の注目度を得た知覚情報は、知識や情動、意図などに基づいた上位の処理（これをトップダウン処理と呼ぶ）への入力となる。ボ

トムアップ処理が外部からの刺激のみを処理するのに対し、トップダウン処理は、内部にある情報とボトムアップ処理から受け取った入力を組み合わせて処理する。トップダウン処理の結果はボトムアップ処理に影響を及ぼし、その結果として注目度マップや注視位置にも影響を及ぼす。図 2 の TDC と示した部分にトップダウン性注意制御が働く箇所を示す。

ここまでにも示唆してきたように、本研究の目的は、ヒトの脳に忠実なシミュレータをつくらうとするものではない。我々が目指しているのは、HAI で手軽に利用できるモデルの構築であり、トップダウン性注意を容易に組み込めることが重要である。以下に、David Marr の三つの水準 [Marr 87] に対応させて、本研究の考え方を簡単にまとめる。

計算理論： ヒトの視覚的注意を目標とはするが、脳神経のレベルで模倣するのではなく、表面的な振る舞いのレベルで模倣する。本研究の計算理論の厳密な定義は今後の課題とするが、HAI での応用に際して適度に扱いやすいモデルを検討していく。

表現とアルゴリズム： Itti-Koch モデルを参考にしつつ、HAI での応用に際して不要な部分は省略する。例えば、ヒトの知覚処理は低レベル～高レベルまで階層的に行われているが、本研究では一つの層でこれを行うことで実装を容易にする。

ハードウェアによる実現： ボトムアップ処理ではコンピュータビジョンのための画像処理手法を用いる。また本研究では、特徴マップと注視確率マップ（図 2 の右側中央）の算出に粒子フィルタを用いる。トップダウン処理はベイジアンネットワークを用いて実現することを予定している。

3 粒子フィルタ

粒子フィルタについて、樋口の解説 [樋口 05] を参考に、本研究で必要となる箇所に焦点を絞って説明する。なお、以下の解説では状態 x と観測値 y の次元を省略しているが、これらはスカラーでもベクトルでも構わず、また次元数が一致している必要もない。

時刻 t の観測時系列を $y_t (t = 1, \dots, T)$ とし、その時刻の内部状態を x_t とする。観測可能であるのは y のみであり、 x は観測できない潜在変数である。このとき、これらの確率変数間の関係を次のような一般状態空間モデルで表現する。

$$x_t = F_t(x_{t-1}, v_t) \quad [\text{システムモデル}]$$

$$y_t = H_t(x_t, w_t) \quad [\text{観測モデル}]$$

ここで、 v_t を《システムノイズ》、 w_t を《観測ノイズ》と呼び、それぞれ密度関数 $q(v)$ 及び $r(w)$ に従う白色雑音とする。

ここでの目的は、この状態空間モデルに基づいて、時刻 $1 \sim t$ までの観測値 $y_{1:t}$ が与えられたときの状態 x_t の条件付分布 $p(x_t | y_{1:t})$ を推定することである¹。粒子フィルタでは、この推定しようとする条件付分布からサンプリングされた多数の実現値とみなせる粒子 ($X_{t|t} = \{x_{t|t}^{(i)}\}_{i=1}^N$) を用いて、次の式によって条件付分布を近似する。

$$p(x_t | y_{1:t}) \cong \frac{1}{N} \delta(x_t - x_{t|t}^{(i)})$$

ここで、 $\delta(\cdot)$ はデルタ関数、 N は《粒子数》である。この計算は、 $y_{1:t}$ の下で発生した粒子を状態空間全体に渡って数え上げ、合計で 1 となるように正規化することで条件付分布 $P(x_t | y_{1:t})$ を近似している。

この粒子の集団 $X_{t|t}$ を逐次的に求めるのが粒子フィルタの具体的な処理である。粒子フィルタのアルゴリズムを以下に示す。また、その概念図を図 3 に示す。

1. $i (i = 1, \dots, N)$ について、乱数 $x_{0|0}^{(i)} \sim p_0(x)$ を生成する。 $p_0(x)$ は x の初期分布を表す。
2. $t = 1, \dots, T$ について、次の (a) ~ (c) を実行する。
 - (a) $i (i = 1, \dots, N)$ について、次の (i) ~ (iii) を実行する。
 - (i) 乱数 $v_t^{(i)} \sim q(v)$ を生成する。
 - (ii) $x_{t|t-1}^{(i)} = F_t(x_{t-1|t-1}^{(i)}, v_t^{(i)})$ を計算する。これは図 3 の「予測」に当たる。
 - (iii) 尤度 (重み) $l_t^{(i)} = R_t(y_t | x_{t|t-1}^{(i)})$ を計算する。これは図 3 の「尤度計算」に当

¹ この操作はフィルタリングと呼ばれる。 $p(x_t | y_{1:t-1})$ を推定する場合は一期先予測と呼ばれる。

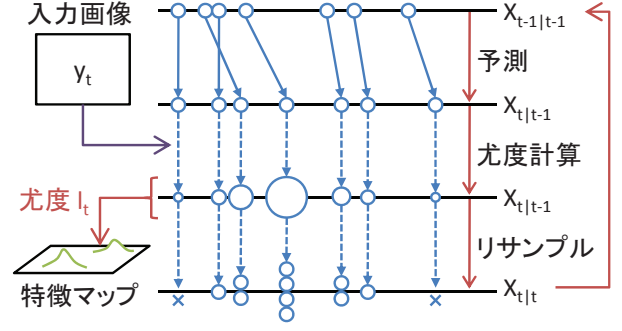


図 3: 粒子フィルタのアルゴリズムの概念図

たる。 R_t は y_t を観測した下での x の尤度関数である。 R_t は観測モデル H_t に基づいて設定する。

- (b) $L_t = \sum_{i=1}^N l_t^{(i)}$ を求める。

- (c) $X_{t|t-1} = \{x_{t|t-1}^{(i)}, \dots, x_{t|t-1}^{(N)}\}$ からそれぞれ $x_{t|t-1}^{(i)}$ を $l_t^{(i)} / L_t$ に比例する割合で復元抽出し、 $X_{t|t} = \{x_{t|t}^{(i)}, \dots, x_{t|t}^{(N)}\}$ を生成する。これは図 3 の「リサンプル」に当たる。

なお、上記アルゴリズムの手順 2 の (c) において、すべての粒子を復元抽出するのではなく、一部を一樣分布に従って生成することによって、粒子が局所的な状態に集中してしまうことを回避する方法がある。この割合をここでは《ランダム粒子率》と呼ぶ。

4 提案モデル

4.1 粒子フィルタの適用

図 2 の「粒子フィルタを適用する範囲」と記した部分に前章で述べた粒子フィルタを適用する。状態 x を入力画像上の座標、観測値 y を入力画像から求めた特徴量、尤度 l を抽出しようとする特徴の注目度とする。粒子フィルタは抽出する特徴の数だけ並列に起動され、それらの粒子フィルタの尤度 l を座標毎に足し合わせることで注目度マップを構成する。

システムモデル F_t によって、粒子のダイナミクスを決めることができる。本実験では以下の等速運動モデルを用いた。

$$x_t = x_{t-1} + d_t + v_t$$

ここで、 d_t は《移動量》を表し、《システムノイズ》 v_t は正規分布に従う白色雑音とした。リサンプリングされた各粒子の位置 (状態値) は、予測時に《移動量》の値だけ移動され、更に《システムノイズ》によって微調整される。

尤度関数 R_t は、抽出する特徴量によって個別に設定する。本実験で用いた赤色（緑色）を抽出する粒子フィルタでは、粒子 $x_{t|t-1}^{(i)}$ の位置を中心とした一辺 7 画素の矩形領域の平均色を $c_{x^{(i)}}$ 、検出対象とする赤色（緑色）を c_M として、以下のように尤度関数を定義した。

$$l_t^{(i)} = \frac{1}{\sqrt{2\pi}w_t} \exp\left(-\frac{(c_{x^{(i)}} - c_M)^2}{2w_t^2}\right)$$

ここで、《観測ノイズ》 w_t は尤度関数における標準偏差の役割を果たしている。

4.2 トップダウン性注意制御

以上の解説において《・》で表していたものを粒子フィルタのパラメータと呼ぶことにする。これらのパラメータを調整することで、特徴抽出処理は以下のように振る舞いを変える。

粒子数：注意を向ける特徴が制御される。《粒子数》が多い処理ほど注目マップへの影響も大きくなる。[Breazeal 99] では、この項目に相当するパラメータのみがトップダウン性注意から制御されている。

移動量：注意を向ける方向性が制御される。与えられた《移動量》に応じて、図 3 の「予測」の際に粒子が移動される。《移動量》がゼロの場合は同じ位置に注意を向け続ける。

システムノイズ：注意を払う空間的な範囲が制御される。《システムノイズ》が大きいほど粒子の予測位置のばらつきが大きくなるため、広い範囲に渡って特徴抽出処理を行うことになる。その代わり、粒子の密度が低くなる。これは Itti-Koch モデルの多重解像度による特徴抽出処理において、どの解像度に重点を置くかを制御していることに相当する。

観測ノイズ：注意を払う特徴量の範囲が制御される。《観測ノイズ》が大きいほど幅広い範囲の特徴量にまで尤度が付与される。その代わり、尤度の分布のピークが鈍る。これは Itti-Koch モデルの顯著度の算出におけるゲインを制御していることに相当する。

ランダム粒子率：視野全体への注意の向け方が制御される。《ランダム粒子率》が高いほど、一様散布される粒子が増えるので、画像全体に渡って特徴を探索することになる²。その代わり、一箇所に注意を集中することが難しくなる。

²粒子フィルタのランダム散布の考え方はヒトの視覚モデルから乖離しすぎているため、将来的にはランダム散布をサッカードの仕組みに置き換える予定である。

5 実験

前節で述べたパラメータ調整によって、注目度マップ上の注意が想定どおりに制御されることを簡単な実験によって確かめた。実験のために、赤色と緑色に注意を向けるボトムアップ処理を実装した。なお、本実験では、トップダウン性注意とパラメータ調整の対応は人手で行った。

実施した五つの実験の概要を以下に示す。¹ 内には言葉で表したトップダウン性注意を記している。

実験 1 「赤色に注意を向けよ」：赤と緑のオブジェクトそれぞれに注意が同程度に分かれている状態から、赤に反応する《粒子数》を増やし、緑に反応する《粒子数》を減らすことで、赤のみに注意を向ける。これは Itti-Koch モデルでも顯著性マップの線形和の重みを調整することで実現される。

実験 2 「右に注意を向けよ」：二つの赤いオブジェクトのうち、左側のオブジェクトのみに注意が向いている状態から、右方向の《移動量》を増加させることで、右側のオブジェクトに注意を向ける。

実験 3 「広い範囲に注意を払え」：オブジェクト上の一点に注意が集中している状態から、《システムノイズ》を増加させることで、そのオブジェクト全体に注意が広げる。

実験 4 「似通った色にも注意せよ」：二つの似通った色のオブジェクトの一方にのみ注意が向いている状態から、《観測ノイズ》を増加させることで、もう一方のオブジェクトにも注意を向けるようになる。

実験 5 「全体に注意を払え」：注意が向いていたオブジェクトを奥に移動したことによって注意（粒子）が空を泳いでいる状態から《ランダム粒子率》を増加させることで、奥の小さなオブジェクトを発見させる。

実験の結果を図 4～図 8 に示す。左側の画像は入力画像上に粒子を上書きしたもの、右側の画像は注目度に応じて透過率を上げて表示したものである。これらの結果より、粒子フィルタのパラメータを調整することによって、注意の振る舞いのある程度制御できることがわかる。

ただし、実験 2 と実験 4 においてはパラメータの設定に微妙な加減を要した。実験 2（図 5）では、右側のオブジェクトに注意が移った瞬間に《移動量》をゼロに戻さないと、右側のオブジェクトの更に右側に粒子が移動してしまう。この問題を回避するには《移動量》が与えられていても、注目度の高い箇所に粒子が吸い付くような効果を与えなくてはならない。これは x_t で

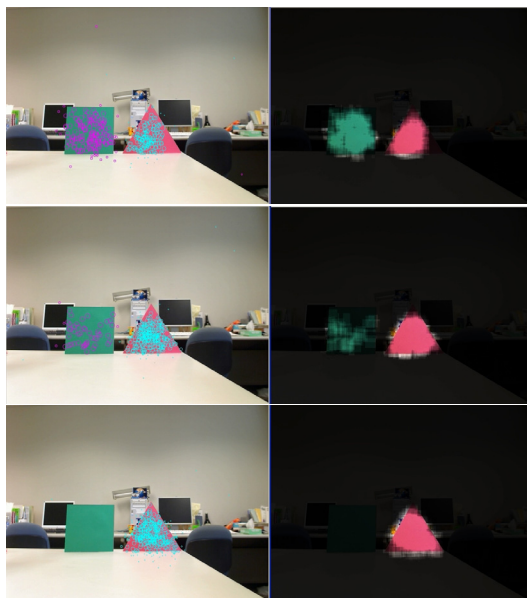


図 4: 実験 1「赤色に注意を向けよ」



図 5: 実験 2「右に注意を向けよ」

の注目度が x_{t-1} よりも大きいときに粒子が更新されやすくなるマルコフ連鎖モンテカルロ型の粒子フィルタ (MCMC 粒子フィルタ) を用いることで実現できる。実験 4 (図 7) では《観測ノイズ》が実験設定値より少し大きくなっただけで、赤とは程遠い色にまで注意を向けるようになってしまう。これは色の類似度 (距離) の精度が悪いことが原因であり、マハラノビス距離を用いることなどで改善できる。

6 まとめ

本研究では、粒子フィルタを用いることで詳細なトップダウン性注意制御を可能とする視覚的注意モデルを提案し、粒子フィルタのパラメータの調整によって、いくつかのトップダウン性注意制御が行えることを示した。ただし、現時点では 2 章で述べた全体構想の一部を実装した状態であり、実験で明らかになった問題の解決も含めて、今後取り組むべき課題が多く残っている。主な課題を以下に挙げる。

- 実験で述べた《移動量》の問題などを解決するため、特徴抽出処理の粒子フィルタを MCMC 粒子フィルタに変更する。
- 注目度マップ上に粒子フィルタを乗せ、そこで算出される尤度を画素値とした注視確率マップを構成する。これにより、注目度マップと注視する位置を決定するためのマップを分ける。
- 中心・周辺拮抗型受容野 (ポップアウトの原理)

の仕組み、飽きによる注意の抑制 (inhibition-of-return)、サッカードの各機能を実装する。

- ボトムアップ処理における粒子フィルタのパラメータとトップダウン性注意制御の対応関係をベイジアンネットワークによって表現する。

謝辞

本研究は立石科学技術振興財団 研究助成「粒子フィルタを用いた注意の相互制御による注視動作モデルの検討 (平成 21 年度)」の支援による。

参考文献

- [陳 03] 陳 彬, 目黒 光彦, 金子 正秀: 知能ロボットにおける環境認識のための視覚的注意点の選択, 日本ロボット学会 第 21 回学術講演会, Vol. 3K19, (2003)
- [Breazeal 99] Breazeal, C. and Scassellati, B.: A Context-Dependent Attention System for a Social Robot, *International Joint Conference on Artificial Intelligence*, pp. 1146–1153 (1999)
- [樋口 05] 樋口 知之: 粒子フィルタ, 電子情報通信学会誌 解説, Vol. 88, No. 12, pp. 989–994 (2005)
- [今井 01] 今井 倫太, 小野 哲雄, 石黒 浩, 中津 良平, 安西 祐一郎: ロボットからの発話: 自発的発話生成のための注意の表出機構の実現, 情報処理学会論文誌, Vol. 42, No. 11, pp. 2618–2629 (2001)
- [Itti 01] Itti, L. and Koch, C.: Computational Modelling of Visual Attention, *Nature Reviews Neuroscience*, Vol. 2, No. 3, pp. 194–203 (2001)

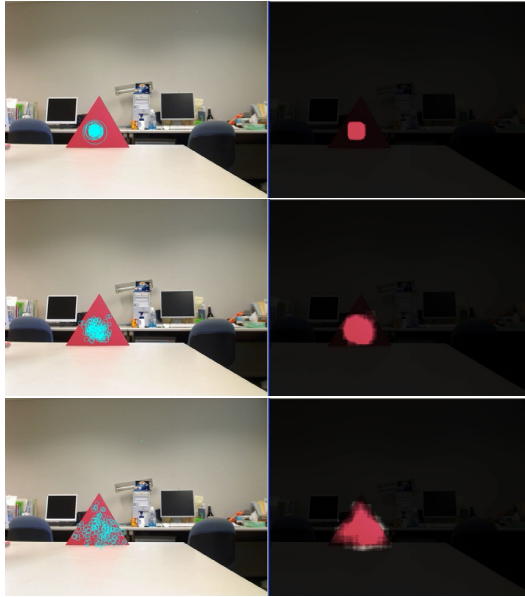


図 6: 実験 3「広い範囲に注意を払え」

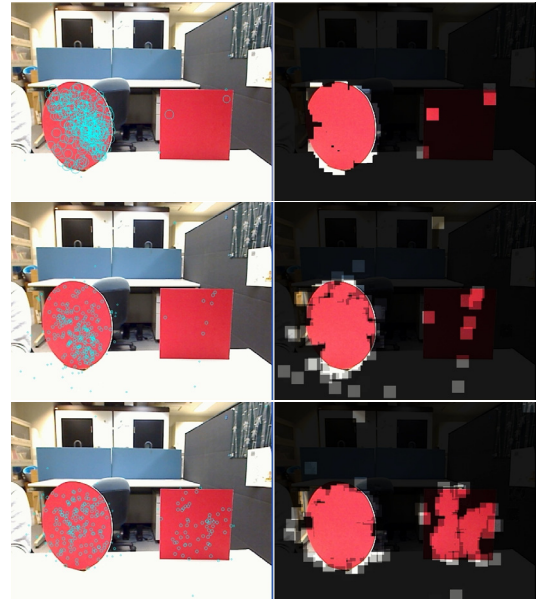


図 7: 実験 4「似通った色にも注意せよ」

- [Koch 85] Koch, C. and Ullman, S.: Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry, *Human Neurobiology*, Vol. 4, pp. 219–227 (1985)
- [小嶋 01] 小嶋 秀樹, 高田 明: 社会的総合行為への発達のアプローチ—社会のなかで発達するロボットの可能性—, 人工知能学会誌特集「社会的相互行為」, Vol. 16, No. 6, pp. 812–818 (2001)
- [久米 99] 久米 博士, 長名 優子, 萩原 将文: 特徴統合理論に基づく視覚ニューラルネットワークモデル, 電子情報通信学会 技術報告, Vol. NC98-187, pp. 261–268 (1999)
- [Marr 87] Marr, D.: ビジョン —視覚の計算論と脳内表現—, 産業図書 (1987)
- [宮里 09] 宮里 洸司, 木村 昭悟, 高木 茂, 大和 淳司, 柏野 邦夫: MCMC-based particle filter を用いた人間の注視行動の実時間推定, 画像の認識・理解シンポジウム, pp. 143–150 (2009)
- [長井 04] 長井 志江, 細田 耕, 森田 章生, 浅田 稔: 視覚注視と自己評価型学習の機能に基づくブートストラップ学習を通じた共同注意の創発, 人工知能学会論文誌, Vol. 19, No. 1, pp. 10–19 (2004)
- [Oliva 03] Oliva, A., Torralba, A., Castelano, M. S., and Henderson, J. M.: Top-Down Control of Visual Attention in Object Detection, *International Conference on Image Processing*, Vol. 1, pp. 253–256 (2003)
- [Pang 08] Pang, D., Kimura, A., Takeuchi, T., Yamato, J., and Kashino, K.: A Stochastic Model of Selective Visual Attention with a Dynamic Bayesian Network, 画像の認識・理解シンポジウム, pp. 1500–1505 (2008)
- [Siagian 07] Siagian, C. and Itti, L.: Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 2, pp. 300–312 (2007)
- [Treisman 80] Treisman, A. M. and Gelade, G.: A Feature-Integration Theory of Attention, *Cognitive Psychology*, Vol. 12, No. 1, pp. 97–136 (1980)

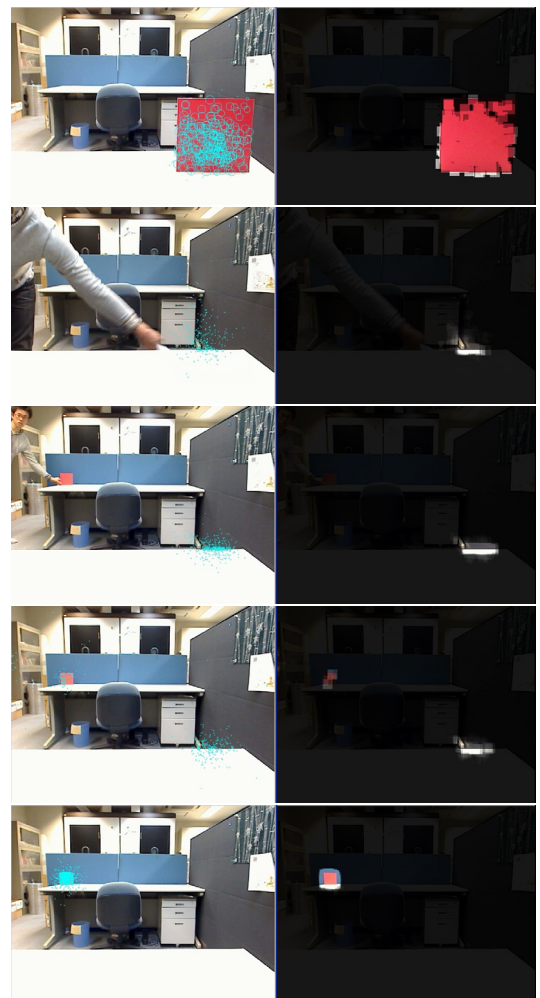


図 8: 実験 5「全体に注意を払え」