

## 注視点からの奥行情報を考慮した視覚的注意モデル

## A Visual Attention Model with the Depth of Field Map

小川 貴弘 尾関 基行 岡 夏樹  
Takahiro Ogawa Motoyuki Ozeki 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

We propose a visual attention model in which a novel feature map called the “depth-of-field map” is introduced. The depth-of-field map functions similar to the depth of field effect of human vision by enhancing the saliency of the regions near the point of gaze in the direction of depth and reducing the gaze movements between regions widely separated from each other. In this study, we first counted the relative frequency of three types of human gaze movements (movement to adjacent objects at a similar depth, movement to adjacent objects in the same two-dimensional direction, and movement to other objects) in three situations (matching, memorizing, and no assigned task). The result shows that the percentage of gaze movement to adjacent objects at the same depth is more than chance and the proportions of these three types of gaze movements vary according to the task. We then implement the proposed model using Microsoft Kinect for Windows and input the scenes from the human experiment into the model. The result shows an occurrence distribution similar to that of human gaze movement according to a given task. It indicates that the proposed model could bring robotic gaze movement closer to that of a human.

## 1. はじめに

古くは計算論的神経科学の分野でヒトの脳をトップダウンに解明するために研究されてきた視覚的注意の計算モデルが、近年では知能ロボットにも応用されるようになってきた。このロボットへの応用では、視覚処理の計算負荷の低減や、複雑な背景から重要な物体の抽出、そして、人との共同注意などに役立つと考えられる。しかし、2次元情報しか用いない従来の視覚的注意モデルでは、奥行きが大きく異なっている領域間でも、2次的に隣接していれば注意が容易に移ってしまう。これにより、例えば、ロボットの目の前にある物体と遠くの壁に貼られたポスターを交互に見るといった望ましくない行動が生じる。

ヒトの目の場合、注視点と奥行きが大きく異なる領域はボケるため、それらの領域の顕著度（注意の引かれる度合い）は低くなり、奥行き方向に離れた領域への視線移動はある程度抑制されると考えられる。これは光学レンズにおける“被写界深度”と同様の効果とみなせるが、ヒトと同等のボケの効果を得るには、被写界深度が浅く、明るいレンズでなければならない。また、注目対象にピントを合わせ続けるためには、正確で高速なオートフォーカス機能も必要である。しかし、そのようなカメラは一般に大きくて高価であり、USBやIEEE1394といったPCと連携しやすいインタフェースを備えているものも少ない。

一方、Microsoft Kinect<sup>®</sup>の登場によって、人物の姿勢や動作、顔の向きや3Dモデルといった情報を利用することが格段に容易となった。Kinectは、人とコミュニケーションするロボットの視覚デバイスとしても有効であり、自作のロボットでは頻りに利用されるようになってきた。このKinectから取得できるシーンの奥行情報を用いて前述の問題を解決できれば、安価で利便性もよい。

そこで我々は、視覚的注意モデルに奥行情報を取り入れるこ

とで、奥行き方向に大きく離れた領域間を注視点が行き来するという問題を解決する手法を提案する。具体的には、現在の注視点の奥行き距離との差を視野全体に渡って計算し、その差に反比例した顕著度を与える。反比例の度合いはパラメータによって調整できる。この顕著度を従来の視覚的注意モデルに組み込むことにより、現在の注視点の奥行き距離から遠い位置にある領域には注意が向きづらくなる。これは生理学的に忠実なモデルではないが、振る舞いのレベルでは人に似通った視線移動を実現できると期待している。

奥行情報を利用した視覚的注意モデルはこれまでもいくつか提案されている。Ouerhaniらは、ヒトの視覚的注意において奥行きの違いによるポップアウトが生じることに着目し、周囲と奥行きの異なる領域に高い顕著度を与える手法を提案している [Ouerhani 00]。また、Penalozaらは、幼児の注意の振る舞いを参考にして、人の手前に当たる領域に高い顕著度を与える手法を提案している [Penaloza 12]。これらの研究はヒトの視覚的注意の知見に基づいたものであるが、いずれの手法も、奥行き距離が離れた領域間の視線移動を抑止することはできない。

物体追跡の分野でも、奥行き距離を使った追跡物体検出の研究がいくつか行われている。Ozekiらは、複雑なシーンから有用な物体を選び出すため、ステレオカメラを使用して作業空間（作業台上）に含まれる物体のみを検出する方法を提案している [Ozeki 02]。Makiらは、追跡対象の奥行き距離が急激に変化しないことを仮定し、直前の検出位置から近い奥行き距離にある物体を検出する手法を提案している [Maki 96]。しかし、これらの研究は物体追跡が目的であり、視覚的注意モデルには統合されていない。視覚的注意モデルは、奥行き差の他にも様々な刺激（視覚的特徴）を統合的に扱っており、それぞれの特徴から得られた顕著度の度合いを調節する洗練されたインタフェースを提供する。

以降、本稿では、まず2章で奥行き方向に対する人の視線移動の調査結果を示し、続いて3章で提案方法について述べる。そして4章にて、2章の調査でみられた人の視線移動の傾向が提案手法によって再現できることを示す。

連絡先: 京都工芸繊維大学情報工学部門

〒606-8585 京都市左京区松ヶ崎橋上町

E-mail: ozeki@kit.ac.jp



図 1: 実験に使用したシーン

## 2. 人の視線移動の観察

本章では、提案手法で実現しようとしている次の二つの特徴が、人の視線移動においても観察されることを調べた結果について述べる。

- 直前の注視点位置に対して、2次元方向に近い領域だけでなく、奥行方向に近い領域にも視線が移りやすいこと
- 視線移動の傾向は与えられたタスクによって変化すること

本実験のためにテーブル上に物体を10個配置したシーンを用意した(図1)。各物体は約40cmの間隔で3列に配置され、列の間隔は約100cmとした。ただし、図1からわかるように、ある物体に一番近い物体は、同じ列の物体ではなく異なる列の物体のように見える。本実験の協力者は22歳~24歳の学生12名で、視線検出器を装着してもらった後、最前列の物体から1.5m離れた台座にあごを置いて視点を固定した。なお、各物体の顕著度の違いや物体同士の関係性によって特定の視線移動の確率に偏りが生じることを避けるため、実験協力者毎に使用する物体と配置は変更した。

実験協力者に指示した内容は次のとおりである。

**指示 1:** 実験準備のため、あごを台座に置いた状態でしばらく待機するよう指示する [タスクなし]

**指示 2:** テーブル上の物体の中から共通の関係を持つ組み合わせを探索するよう指示する [組合せタスク]

**指示 3:** テーブル上の物体の配置を記憶するよう指示する [記憶タスク]

**指示 4:** 実験の後処理のため、あごを台座に置いた状態でしばらく待機するよう指示する [タスクなし]

本実験には、「タスクなし」も含めて三つのタスクが含まれている。「組合せタスク」と「記憶タスク」は実験協力者の視線移動が少なくなるまで継続した。これらのタスクでの全実験協力者の平均継続時間は約50秒となった。一方、「タスクなし」での継続時間は30秒とし、2回の「タスクなし」のデータを合わせて分析に使用した。

データ分析では、まず、三つのタスクそれぞれについて各実験協力者毎に視線が0.2秒以上停留した位置を調べた。テーブル外の領域に視線が停留した場合は、その前後の視線移動を除外した。続いて、図2に示すように、その停留位置の移動パターン(視線移動パターン)が以下のいずれに当たるかを分類して生起回数を数えた。

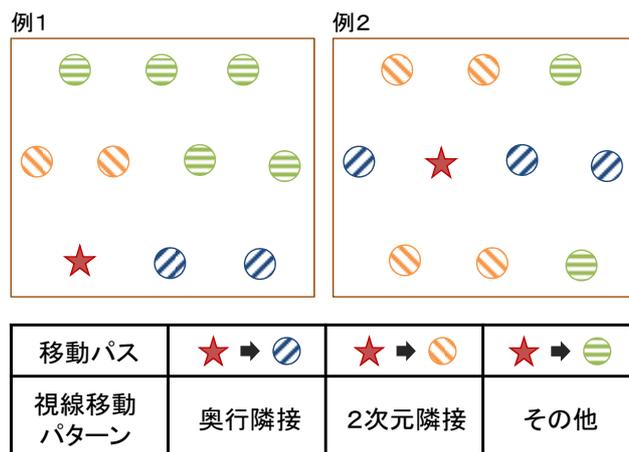


図 2: 視線移動パターンの分類

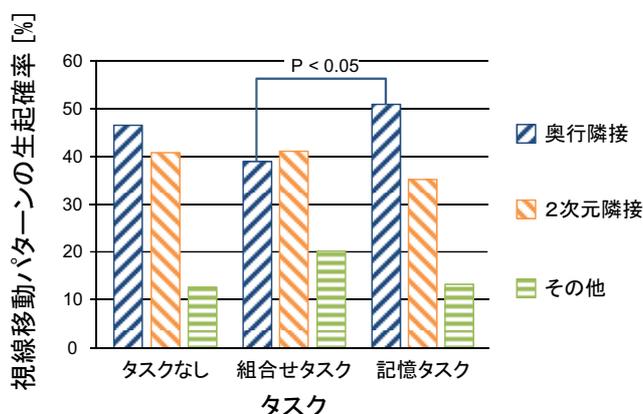


図 3: 視線移動パターンの生起確率 (人の場合)

- 奥行方向に隣接している物体への視線移動 [奥行隣接]
- 2次元方向に隣接している物体への視線移動 [2次元隣接]
- その他の物体への視線移動 [その他]

各視線移動パターンの全実験協力者の平均値を図3に示す。人の視線移動が物体間でランダムに生じると仮定すると[奥行隣接]の生起確率は26.7%になるが、実験で観察された[奥行隣接]の生起確率は26.7%より明らかに高い結果になっている。よって、奥行方向に隣接している領域の間には何らかのバイアスが加わっていると推測される。同じことが[2次元隣接]についてもいえるが、それについては中心窩を模したバイアス(顕著度の強調)を組み入れることで2次的に近い領域への視線移動を促進することが従来より行われている。本実験結果は、これと同様の機構が[奥行隣接]についても必要であることを示唆している。

次に、与えられたタスクの違いによって[奥行隣接]の生起確率が変化するか否かを調べた。分散分析の結果、タスクの違いによる効果は有意となり、TukeyのHSD法で多重比較を行ったところ、組合せタスクと記憶タスクの間で[奥行隣接]の生起確率に有意差( $p < 0.05$ )がみられた。この結果は、奥行方向に近い領域への視線移動を促すバイアスの大きさを調整させる必要があることを示唆している。このようなバイアスの調整は、視覚的注意モデルにおいては“トップダウン注意(制御)”と呼ばれる枠組みで扱われる(詳しくは次章で述べる)。

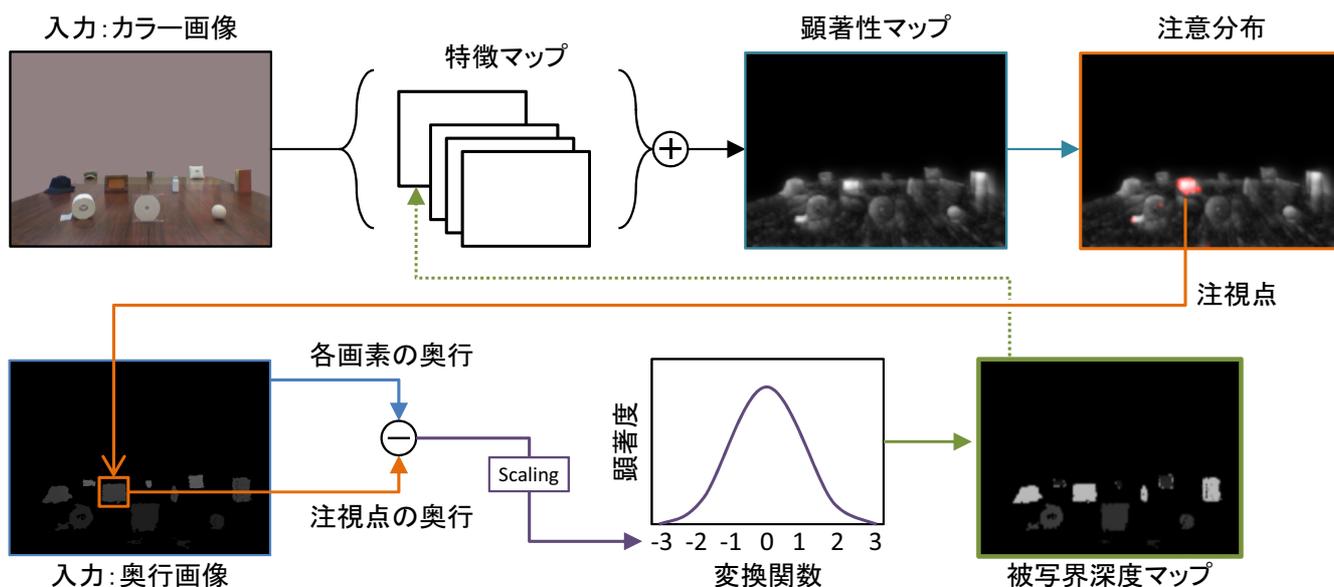


図 4: 提案手法の概要

### 3. 提案手法

本研究では、前章の視線移動パターンを再現することのできる新しい視覚的注意モデルを提案する。本章では、まず我々のモデルの基となっている Itti らのモデル [Itti 98] と Ozeki らのモデル [Ozeki 11] を概説した後、提案手法について述べる。

ヒトの視覚的注意の機能は、知覚刺激によって反射的に引き起こされる“ボトムアップ注意”と、そのボトムアップ注意の特定の成分や場所を選択的に強調する“トップダウン注意”に分けられる。さらにトップダウン注意は、特徴への注意、空間への注意（空間的注意）、オブジェクトへの注意の少なくとも三つに分類されることが一般的に知られている。ボトムアップ注意の代表的な計算モデルの一つに Itti らの“顕著性マップモデル”がある。このモデルでは、輝度や色など低レベルの特徴量に基づき、個々の特徴によって注意が引かれる強さ（顕著度）を画素値で表した“特徴マップ”を算出する。そして、各特徴マップに重みを掛け合わせて足し合わせ、正規化することで1枚の“顕著性マップ”を構成する。この顕著性マップモデルでは、特定の特徴マップの重みを変えることによって特徴へのトップダウン注意を実現する。

Ozeki らは、この顕著性マップ上に走らせた粒子フィルタ [樋口 05] を注意の向けられた領域（注意分布）と見做し、粒子フィルタのパラメータ調整によって空間的注意を実現する視覚的注意モデルを提案している。粒子フィルタは視野上を照らして動くスポットライトの役割を果たし、移動方向を与えると、顕著度の高い領域を辿りながら概ねその方向に移動していく。また、注意分布の広がりも動的に制御できるなど、従来の視覚的注意モデルよりも扱いやすい空間的注意の制御インタフェースを提供している。注視点は注意分布の値が最大となる位置として求められる。

本提案手法では、Ozeki らのモデルに“被写界深度マップ”と中心窩マップの二つの特徴マップを追加することで、奥行方向と2次元方向に対する注視点の移動しやすさを制御する。被写界深度マップの各画素には、現在の注視点との奥行距離に反比例した顕著度が与えられる。この特徴マップを顕著性マップに加えることで、現在の注視点と奥行距離に近い領域ほど視線が移りやすくなる。被写界深度マップ生成の概要を図4に

示す。

次章の実験で使用したシステムでは Kinect を用いて提案手法を実装した。被写界深度マップの算出手順を以下に示す。

1. Ozeki らの手法によってカラー画像から注視点を求める。
2. 現在の注視点との奥行距離の差を全面素に渡って計算する。なお、本実装では、奥行差が $-2.6\text{m} \sim 2.6\text{m}$  (Kinect から得られる値は $-2600 \sim 2600$ ) の範囲から外れる画素については被写界深度マップの顕著度を0とした。
3. 奥行差が $-3 \sim 3$ の範囲に収まるように $3/2600$ を掛けてスケールし、これを正規分布の確率密度関数によって被写界深度マップの顕著度に変換する。

なお、中心窩マップの算出では、奥行差の代わりに注視点からの2次元距離を使用する。その他の処理は被写界深度マップの生成過程と同様である。

被写界深度マップには“depth-of-focus”と呼ばれるトップダウン注意制御のためのパラメータがある。depth-of-focus は写真でいう焦点深度と同様の意味を持ち、ピントの合う範囲、つまり顕著度が強調される範囲を表す。奥行差から顕著度への変換に正規分布を用いる場合、depth-of-focus は正規分布の分散を表す。例えば、depth-of-focus を小さくするほど正規分布の裾野が狭くなり、注視点から奥行方向に離れた位置にある領域の顕著度が小さくなる (図5)。

### 4. 実験

本実験では、2章で調べた人の視線移動パターンの生起分布と同様の傾向が提案手法によって再現できることを確認する。入力したデータは2章の実験で使用したシーンと同様のもの (Kinect で撮影したカラー画像と奥行画像のセット) であり、物体の配置を変えた6セットを用意した (各データの長さは1分間)。各データセットを提案手法に入力して注視点を取得し、2章の実験と同様に三つの視線移動パターンに分類した。なお、本実験では、人の視線移動に近い結果が得られるよう depth-of-focus の値を徐々に変更しながら実験を行った。

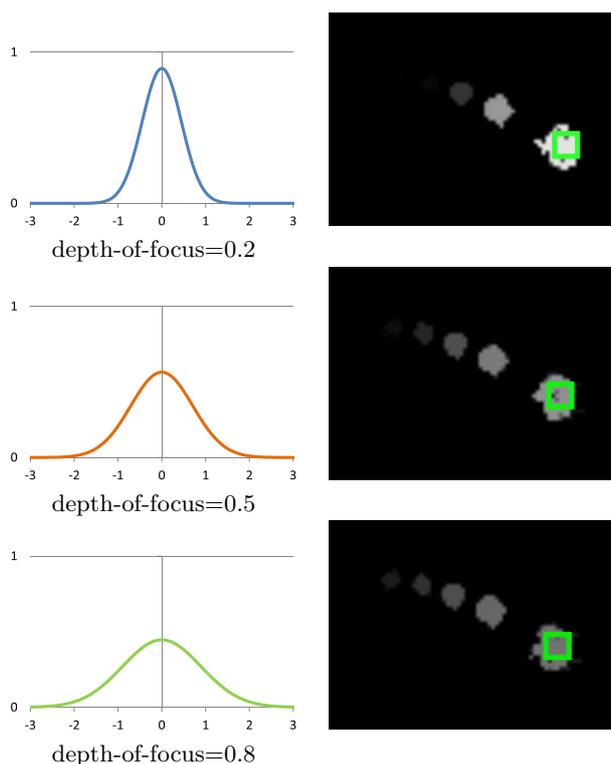


図 5: トップダウン注意制御

3種類の depth-of-focus の値における視線移動パターンの生起分布を図 6 に示す。各グラフは 6 セットの平均値である。depth-of-focus が 0.2 のとき、人を対象とした実験での記憶タスクの生起分布をよく再現していることがわかる。同様に、パラメータが 0.3 でタスクなし、0.4 で組合せタスクの分布をよく再現している。本実験では、中心窩マップのパラメータは固定し、被写界深度マップの depth-of-focus のみを使ってトップダウン注意制御を行ったが、[2次元隣接] や [その他] の生起確率もそれに合わせて変化し、人の場合と概ね似通った分布形状が得られた。

前章で述べたように、depth-of-focus の値を小さくするほど、現在の注視点と奥行距離が近い領域に視線が移りやすくなる。本実験の設定において、これは格子状に置かれた物体を横方向に走査していくことになる。depth-of-focus の値が大きくなるにつれて横方向走査の傾向が弱まる。これを人を対象とした実験結果に当てはめると、記憶タスクでは横方向に物体を確認していくような方策がとられ、組合せタスクでは見かけ上近くにある物体を走査していく方策がとられたと推測できる。

## 5. まとめ

本研究では、注視点からの奥行距離に反比例した顕著度を持つ“被写界深度マップ”を導入することで、現在の注視点から奥行が大きく離れた領域に注意が移動することを抑制できる新しい視覚的注意モデルを提案した。本手法はヒトの視覚的注意を生理学的に忠実にモデル化したものではないが、従来の視覚的注意モデルに特徴マップを追加するだけで実現でき、また、安価な Kinect を用いて実装できる。トップダウン注意制御もパラメーター一つで可能であり、タスクに応じて変化する人の視線移動パターンを本手法によって再現することも確認した。

今後は、コミュニケーションロボットに提案モデルを搭載

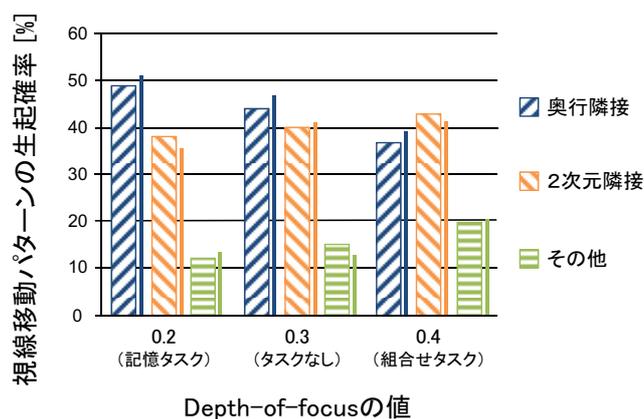


図 6: 提案手法による視線移動の割合

し、人とのコミュニケーションにおいて、実際に望ましい対象に注意を向けることができることを確認していく予定である。また、本稿では奥行距離から顕著度への変換に正規分布を用いたが、実際に利用していく中でより良い変換関数を検討していきたい。

## 参考文献

- [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)
- [Itti 98] Itti, L., Koch, C., and Niebur, E.: A model of Saliency based visual attention for rapid scene analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, pp. 1254–1258 (1998)
- [Maki 96] Maki, A.: A computational model of depth-based attention, *Proceedings of 13th International Conference on Pattern Recognition*, Vol. 4, pp. 734–739 (1996)
- [Ouerhani 00] Ouerhani, N.: Computing visual attention from scene depth, *Proceeding 15th International Conference on Pattern Recognition*, Vol. 1, pp. 375–378 (2000)
- [Ozeki 02] Ozeki, M., Itoh, M., Nakamura, Y., and Ohta, Y.: Tracking Hands and Objects for an Intelligent Video Production System, *Proceedings of 16th International Conference on Pattern Recognition*, Vol. 3, pp. 1011–1014 (2002)
- [Ozeki 11] Ozeki, M., Kashiwagi, Y., Inoue, M., and Oka, N.: Top-Down Visual Attention Control Based on a Particle Filter for Human-Interactive Robots, *Proceedings of 4th International Conference on Human System Interaction*, pp. 188–194 (2011)
- [Penaloza 12] Penaloza, C. I., Mae, Y., Ohara, K., and Arai, T.: Using Depth to Increase Robot Visual Attention Accuracy during Tutoring, *IEEE International Conference on Humanoid Robots - Workshop of Developmental Robotics* (2012)