

トップダウン注意制御による人とエージェントの共同注意

Shared Attention Between Humans and Agents by the Top-down Attention Control

尾関 基行* 柏木 康寛 井上 茉莉子 岡 夏樹

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

京都工芸繊維大学

Kyoto Institute of Technology

Abstract: We introduce a method that enables an agent to follow human gaze (facial direction). This is realized by using our visual attention model based on a particle-filter, which has a spotlight-type top-down attention control interface. In this paper, we briefly explain our attention model and how to realize agent's shared attention by the top-down control API. In comparison with other typical types of methods, our method can be broadly applicable to different types of scenes.

1 はじめに

相手の視線の先を追ってその注目対象を見つけることは、人とエージェントの円滑な対話を実現するために不可欠な能力である。この共同注意の機能をエージェントが備えることにより、対話相手は、現在どの対象について話しているかを自然なやりかたでエージェントに知らせることができる。

本研究では、我々が提案してきたスポットライト型視覚的注意モデル [1] のトップダウン注意制御インタフェースを用いて、人とエージェントの共同注意を実現する。ロボットなどのエージェントに共同注意を実装する方法は様々提案されているが、本手法は注意のスポットライトを粒子フィルタで表現しているところに特色があり、物体追跡における粒子フィルタの頑健性は視線追従でも有効である。

本稿では、我々の視覚的注意モデルと共同注意の実現方法について概要を述べ、本手法の特徴を示す四つのシーンに対する共同注意（視線追従）の例を示す。

2 提案手法

視覚的注意は視野上を照らすスポットライトに喩えられる。本手法では、この注意のスポットライトとその制御を状態空間表現の枠組みで捉える。提案モデルの概要を図 1 に示す。スポットライトの照射領域（注意の分布）は、システムモデルで定義されるダイナミクスと、外部からの視覚刺激に基づいて逐次更新され

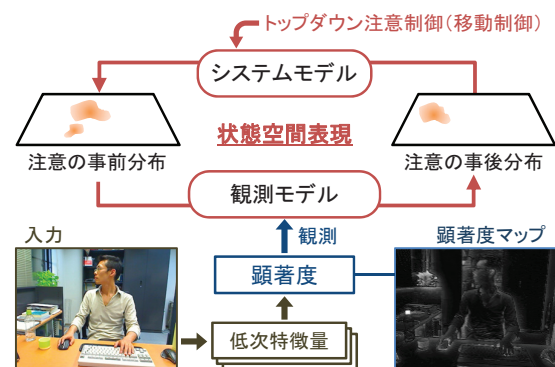


図 1: 視覚的注意モデルの概要

る。視覚刺激はモデルから観測される前に顕著度と呼ばれる「目立つ度合い」に変換されるが、本実験では赤色に近いほど顕著度が高くなるようにしている。

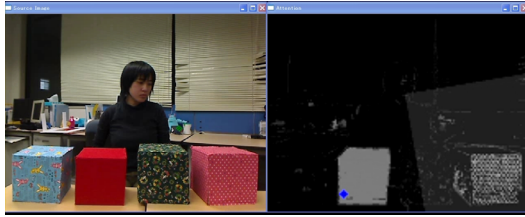
本手法では図 1 の実装に“粒子フィルタ”を用いる。粒子フィルタは、推定する分布を粒子の集まりとして近似する手法であり、分布の母数だけではなく分布の形状まで取得できる。本研究では、粒子フィルタから得られた分布を注意のスポットライトとする。本手法の詳細については文献 [1] を参照されたい。

本モデルには、より上位の処理系統から注意の振る舞いを制御するための API がある。この API を使った注意の制御を“トップダウン注意制御”と呼ぶ。注意のスポットライトを移動するには、以下のシステムモデルのパラメータを制御する。

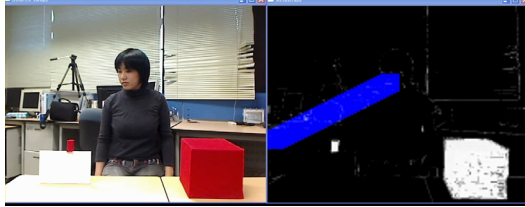
$$x_{t|t-1}^{(i)} = x_{t-1}^{(i)} + m + v_t^{(i)}$$

ここで、 $x^{(i)}$ は個々の粒子の位置、 m は移動距離であ

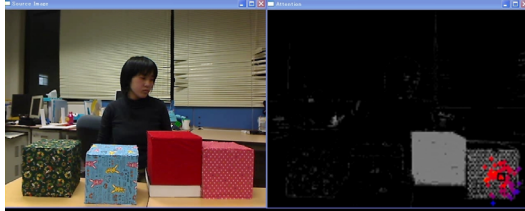
*連絡先：京都工芸繊維大学 情報工學部門
〒 606-8585 京都市左京区松ヶ崎橋上町
E-mail: ozeki@kit.ac.jp



シーン A：右端の物体に注目（結果はフィルタ型）



シーン B（結果は単純スポットライト型）



シーン C：右端の物体に注目（結果は提案手法）

図 2: 実験用シーンと各手法の結果

り、そこに白色ガウスノイズ v_t を加えたダイナミクスとなっている。

共同注意は、目前の人の顔の向きを求め、その方向に粒子群を移動するよう m を制御することで実現できる。人の注視対象に到達したか否かは、各粒子の重み（観測によって計算される）の総和によって判断する。ただし、現在の方法では、人の視線と顔の向きが一致していること、及び、エージェントの視野外には注目対象が無いことを前提としている。

3 実験

エージェントと相対した人（注視者）が机の上の物体のいずれかを見るシーンに対して、提案手法及び典型的な比較手法を適用した結果を示す。

比較手法は次の 2 種類を用意した。(1) フィルタ型：注視者の顔の位置から顔の向きに扇上に広がったバイアスを顕著度に足し合わせた上で、視野全体の中から顕著度が最大となる箇所を求めて注視点とする。(2) 単純スポットライト型：注視者の顔の位置から顔の向きに沿って走査しながら一定範囲の顕著度の総和を求め、その値が一定以上になったところを注視点とする。いずれも提案手法と同じく顕著度 + トップダウン注意制御の構成であり、視覚的注意のメタファとしても代表的なものである。図 2 右列に提案手法を含めた三つの

表 1: 対面者の視線追従の正解率

| シーン | 提案手法 | スポット | フィルタ |
|-----|------|------|------|
| A | 100% | 100% | 0% |
| B | 67% | 50% | 100% |
| C | 100% | 43% | 0% |
| D | 93% | 57% | 100% |

例を示す。

適用するシーンは 4 種類用意した。図 2 左列にシーン A ~ C を示す。シーン A では、2 番目に目立つ物体が注目されており、1 番目立つ物体は飛び抜けて顕著度が高い。このような場合にはフィルタ型の手法は失敗することが多い。シーン B は注目対象が小さく、本手法も含めてスポットライト型の手法が失敗しやすい場合である。シーン C とシーン D は、1 番目立つ物体と 2 番目に目立つ物体が隣接している場合であり、いずれも右端の物体を注視者は見ている。最も目立つ物体が、シーン C では手前に、シーン D では奥に配されている。フィルタ型の手法では、バイアスが両方の物体を含むため、目立たないほうの物体が注目されると必ず失敗する。また、注視者の視線上に手前の物体が大きくかぶさるため、単純なスポットライト型の手法も失敗しやすい。

各シーンにつき 30 回ずつ試行した結果を表 1 にまとめる。正解率は、注視者が注目した物体を正しく特定できた割合を示す。その結果、シーンの説明のところで述べた仮説とおりの傾向がみられた。提案手法は、比較手法よりも正解率の低い場合もあるが、0% になることもなかった。単純スポットライトとの性能の違いは、主に粒子フィルタの性質に由るものと考えられる。

4 まとめ

本稿では、我々が提案してきたスポットライト型視覚的注意モデルのトップダウン注意制御を用いて共同注意を実現し、二つの典型的な手法と比較した。その結果、提案手法は比較手法に比べて、より幅広いシーンに対応できることが示唆された。

ただし、比較手法と同様、提案手法でも注視点を決める際の閾値の決定が難しく、シーンに応じて調整する必要がある。この問題について我々は、注目物体と見做した物体すべてを上位の処理システムに送り、上位の処理システムで視線追従を続けるか否かを決定する枠組みを考えている。

参考文献

- [1] M. Ozeki, Y. Kashiwaga, M. Inoue, and N. Oka : Top-Down Visual Attention Control Based on a Particle Filter for Human-Interactive Robots, Proc. of 4th International Conference on Human System Interaction, P188-194 (2011)