

人物姿勢と注視対象配置制約を利用した 後ろ向き人物の注視領域推定

弓矢 隼大[†] 出口 大輔[†] 川西 康友^{‡†} 村瀬 洋[†]

名古屋大学 情報学研究科[†]

理化学研究所 情報統合本部 GRP[‡]

1. はじめに

人物の行動や意図の理解において、その人物が何を注視しているかは大きな意味を持つ。このような人物が何に注視を向けているかを推定する注視領域推定は、マーケティングにおける商品への興味度合いの調査といった様々な活用が期待される重要な技術である。このような背景から、画像中の人物の注視領域を推定する手法がいくつか提案されている。平山らは、対象となる人物の顔画像から顔の向き及び視線の向きを抽出することで注視領域を精度良く推定する手法を提案している [1]。しかし、図1に示すような、対象となる人物が後ろ向きの場合には顔画像が取得出来ず、注視領域を推定することができない。一方、人物の顔画像を正面から観測できないような場合において、その人物の姿勢と人物の注視領域の関係が調査されており、Kawanishi らは人物の姿勢と注視領域に何らかの関係性があることを報告している [2]。個人差は存在するものの、対象の位置によって頭の向きを変化させたり、対象が低い位置の場合は屈んだ姿勢を取るといったように、注視対象に合わせて姿勢の変化を観察することができる。このような人物の姿勢は後ろ向きであっても Azure Kinect^{*1} 等を用いることで取得可能である。一方、そこに存在する物体の配置や、各物体の大きさなどは、その人物がどれを見ているかを推定するのに重要な要素となる。そこで本発表では、

- ・ 物体の配置
- ・ 姿勢のとり方の個人差
- ・ 物体の大きさなどによる物体領域の大きさの違い

を考慮した注視領域推定手法を提案する。



図1 棚を注視している後ろ向きの人物の様子

2. 提案手法

提案手法は、図2に示すように学習処理と推定処理の2つに分けられる。

a. 学習処理

まず、Azure Kinect^{*1}によって取得可能な26個のカメラ座標系の3次元関節座標(以下姿勢情報)を首と腰の間の距離が一定になるように正規化処理をする。また、撮影時の注視物体領域を表すヒートマップを作成し、教師データとする。次に、物体の配置を考慮するため、このヒートマップを注視領域推定の中間表現として生成するネットワークを学習する。具体的には、3次元関節座標を入力とし、注視尤度を示すヒートマップを出力とする逆畳み込みニューラルネットワークである。

b. 推定処理

最初に、姿勢情報を正規化する。次に、姿勢のとり方の個人差を考慮するため、ある人物に着目した場合、同じ領域を注視する場合同じ姿勢を取るといった仮定の元に、推定対象人物毎に3次元関節座標をもとに k-means 法 [3] を用いて姿勢をクラスタリングする。次に、学習した生成器に3次元関節座標で記述された、そのクラスタ内の姿勢を入力し、ヒートマップを得る。そして、そのクラスタに含まれる全ての姿勢から生成されるヒートマップを重ね合わせて一つのヒートマップにする。以上の処理を全てのクラスタにおいて行う。

Estimating the gazing area of a backward-facing person based on the person's posture and gazing target placement constraints

[†] Graduate school of informatics, Nagoya University

[‡] GRP, R-IH, RIKEN

^{*1} Microsoft. Azure Kinect dk AI モデルの開発(2021/1/23)

<https://azure.microsoft.com/ja-jp/services/kinect-dk>

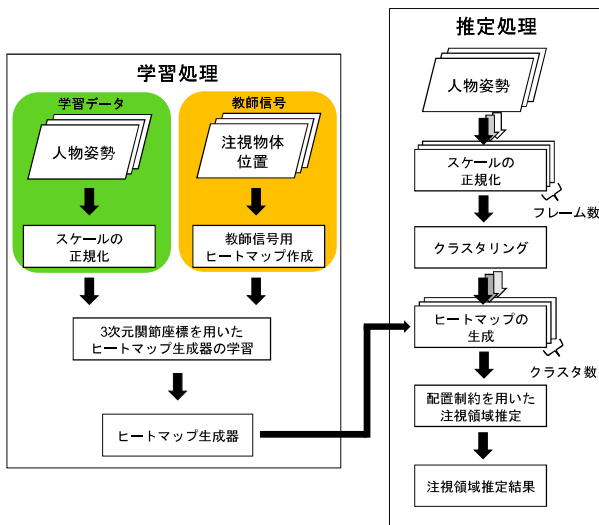


図2 提案手法の処理手順

最後に、物体の大きさによる違いを加味するため、得られたクラスタ毎のヒートマップから算出される各物体の配置領域に制限した尤度の平均値が最も高い物体領域を推定結果とする。

3. 実験方法

本実験では、生成したヒートマップから注視領域を推定する手法を比較した。なお、姿勢と注視領域を対応付けた公開データセットは存在しないため、独自に後ろ向き人物の3次元関節座標と注視領域を対応付けたデータセットを構築した。比較手法は、クラスタリングを用いず単フレームの姿勢から生成されたヒートマップのみから推定を行なう。実験では、全7人分のデータ(1,077 フレーム)から6人分を学習データ、1人分をテストデータとする交差検証を行ない、評価指標としては推定結果の正解率を採用した。

4. 実験結果

図3に、あるクラスタに属する姿勢のみから推定したヒートマップの例を示す。次に表1に注視領域推定結果の正解率を示す。提案手法によって、姿勢の個人差によるヒートマップの曖昧さを抑えることを期待したが、テストデータによって影響が異なった。ヒートマップ生成器のヒートマップの推定精度が影響していると考えられる。

5. むすび

本発表では、後ろ向き人物の姿勢情報から注視領域を推定する手法を提案した。個人差を加味する手法については今後ヒートマップ生成器と合わせて検討が必要である。

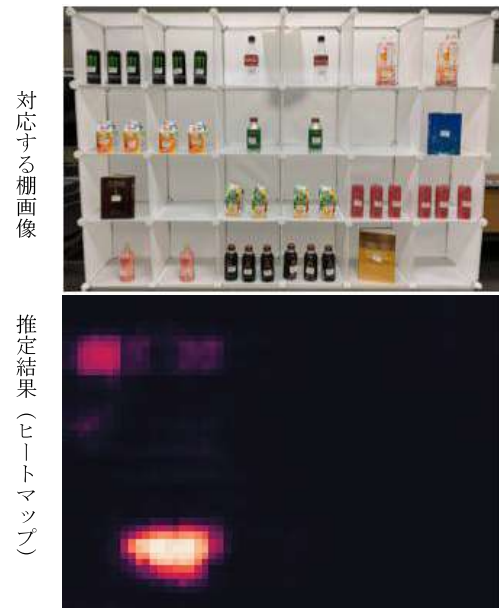


図3 生成したヒートマップと対応する棚の画像

表1:各テストデータにおける正解率(%)

テストデータ	1	2	3	4	5	6	7
比較	43.2	41.7	19.8	26.9	24.9	21.2	41.7
提案	57.5	33.3	25.8	24.1	25.0	20.0	40.0

謝辞 本研究の一部は科研費 (17H00745) による。

参考文献

- [1] Hirayama T, Sumi Y, Kawahara T, and Matsuyama T. Info-concierge: Proactive multi-modal interaction through mind probing. Proc. Of APSIA Annual Summit and Conference 2011.
- [2] Kawanishi Y, Murase H, Xu J, Tasaka K, and Yanagihara H. Which content in a booklet is he/she reading? Reading content estimation using an indoor surveillance camera. Proc. of ICPR 2018, pp.1731-1736.
- [3] MacQueen, James. "Some methods for classification and analysis of multivariate observations." Proc. of the 5th Berkeley symposium on mathematical statistics and probability. Vol.1. No.14. (1967)