

日常生活空間における長期的な人物追跡を伴う 人-場所間の関係性の推定

中西 航† 角所 考†

関西学院大学 理工学研究科†

1 はじめに

オフィス等の日常生活空間において、知能ロボットなどが、特定の人物に物を届ける、在席状況を把握する、といった人や場所に応じた支援を行うには、人-場所間の関係性を把握している必要がある。このような関係性は、日頃その空間で過ごしている者ならば、互いの行動から自然に把握できるものであることから、ロボットの場合も、対象空間をカメラで継続観測し、各場所への各人物の滞在率を把握することで同様の関係性を推定できると期待される。

このためには対象空間内での継続的な人物追跡が必要であり、コンピュータビジョンの分野で古くから提案されてきた様々な手法[1][2]が利用できるが、それらの多くは同一人物に対する位置や速度の連続性、色の不変性等を仮定しているのに対し、日常生活空間内の人物には速度を得るための移動が少ない上、日をまたいだ追跡も必要となるため、服装も変化し、色の不変性も期待できない。その一方、各場所への各人物の滞在率には人-場所間の関係性に応じた偏りがみられると考えられるため、これに着目して追跡結果を補正しつつ滞在率を推定することで、人-場所間の関係性を推定することを試みる。

2 提案手法

2.1 近接セグメント

映像中の t フレーム目に人物検出処理を適用して得られた i 番目の領域の重心 p_t^i の位置を p_t^i とする。このとき、従来研究と同様に近接フレーム間での同一人物の位置は近いと考え、これらを満たす重心の集合を同一セグメントとする。具体的には、(1)式で定義する $\delta(p_t^i, p_{t'}^j)$ が閾値以下であるような任意の $p_t^i, p_{t'}^j (t' < t)$ に対して $p_t^i \in O_{right}(p_{t'}^j)$, $p_{t'}^j \in O_{left}(p_t^i)$ と定め、互いに $O_{left}(p_t^i) = \{p_{t'}^j\}$, $O_{right}(p_{t'}^j) = \{p_t^i\}$ を満たす任意の $p_t^i, p_{t'}^j$ を要素とする集合 P (“近接セグメント”と呼ぶ)を求める。

$$\delta(p_t^i, p_{t'}^j) = \alpha(t - t') + \beta \|p_t^i - p_{t'}^j\| \quad (1)$$

2.2 連結候補セグメント

上の P の要素の中で、それぞれ最小と最大のフレーム番号を持つものを p_{min}, p_{max} として、 $O_{left}(P) = O_{left}(p_{min})$, $O_{right}(P) = O_{right}(p_{max})$ と定めると、 $O_{left}(P), O_{right}(P) \notin P$ より、これら

には複数の重心が含まれるため、これらを P と同一人物の軌跡の構成要素の候補として、その中から一つを選択し、 P と連結することで、入室から退室までの一連の追跡軌跡の候補 (“連結候補セグメント”と呼ぶ)を得る。

2.3 同一人物セグメント

上で得られる複数の連結候補セグメントの集合を C で表すと、その中の任意の連結候補セグメント $C \in C$ に対し、そこに含まれる重心位置に基づく各場所への滞在率を各成分とするベクトル $s(C)$ (“滞在率ベクトル”と呼ぶ)を求める。このベクトルは、同一人物に対応する連結候補セグメント間では類似していると考えられるため、それらのコサイン類似度に基づいて、 C を互いに類似した複数の集合に分類することによって、日をまたいだ場合も含めた同一人物の軌跡 (“同一人物セグメント”と呼ぶ)を求める。

これは次の2つのステップにより実現する。

(1) 初期セグメントの作成

まず、 C の中で、同一フレームから得られた重心を要素に含む連結候補セグメントの集合を考え、そこに含まれる連結候補セグメント数が最大となる(このときの数を N とする)フレームに対する連結候補セグメントそれぞれを初期の同一人物セグメントとし、 H_1, \dots, H_N で表す($\mathcal{H} = \{H_1, \dots, H_N\}$)。

(2) 連結候補セグメントの分類

次に、 C の中で、上の H_1, \dots, H_N いずれにも含まれない任意の連結候補セグメント C を、任意の同一人物セグメント $H_n \in \mathcal{H}$ の中で、 $s(C)$ と $s(H_n)$ のコサイン類似度が最小となる H_n に要素として追加する。この結果変化した H_n を \tilde{H}_n で表す。

2.4 連結候補セグメントの修正と再分類

2.2で作成したそれぞれの連結候補セグメントには、同一人物のものではない近接セグメントが混在している可能性がある。一方で、それぞれの人物の各場所への滞在率には人-場所間の関係性に応じた分布がみられるはずである。そこで、2.3(3)で H_n に追加された任意の連結セグメントを C として、 $s(C)$ と $s(\tilde{H}_n - C)$ とのコサイン類似度が増加するように C 中の近接セグメントを入れ替える。これは、 C を2.2で述べた方法で作成した際に要素となった近接セグメント P に対し、 $O_{left}(P), O_{right}(P)$ の中から選択したセグメントを

交換することで実現する。ただし、これによりCの要素が変化するため、再度2.3(2)を実行する。これを繰り返すことでコサイン類似度が最大となるものを選択する。

2.5 場所の抽出

2.3で滞在率を算出するには、フレーム中で同一場所に対応する範囲を特定しておく必要がある。それぞれの場所には、各人物の訪問による重心位置の集積が見られると考えられることから、全フレームから得られた全ての重心位置を画素毎に集計し、その個数を画素値とする濃淡画像にウォータシェッド法を適用して得られた領域をそれぞれ異なる場所とした。

3 実験結果

提案手法の有効性を確かめる実験を行った。オフィス空間の観測映像の例として、筆者らの研究室を魚眼カメラで撮影した全方位映像5分を用いた。人物検出にはYolo v5 [3]を用いた。

3.1 場所の抽出

2.5の処理結果を図1に示す。図中、緑の線で囲まれた各領域が、それぞれの場所を表している。図より、各座席が一つの領域として得られている一方、通路部分については、様々な位置に様々な人物が同時に位置するため、細かな領域に分かれてしまった。



図1. 場所の抽出結果

3.2 人-場所間の関係性

図1で得られた各場所に基づく2.1~2.4の処理結果のうち、図1の①、②の場所に対する各人物の滞在率を図2に示す。図の横軸は各人物を表し、縦軸は最も長く滞在した人物の時間で正規化したときの各人物の滞在率を表す。青色の棒グラフが本手法で得られた滞在率であり、橙色の棒グラフは2.4での修正を行わなかった場合の滞在率である。

場所①は人物Aの座席であるが、映像を目視で確認したところ、人物Bも時折訪ねていた。これに対して、図(a)では、修正ありの場合は、これを反映した滞在率が得られているのに対し、修正なしの場合は、人物の取り違えにより、別人物Iの滞在率が高くなってしまっている。また場所②は共同利用スペースであるが、映像中では、人物K, F, Dらの訪問が見られた。これに対して図(b)では、修正ありの場合は、やはりこれを反映した滞在率が得られているのに対し、修正なしの場合は、人物Kの滞在率は高いものの、人物

Eが二本の棒グラフで表されていることからわかるように、人物の取り違えにより、同一人物が別の2人の人物とみなされてしまっている。

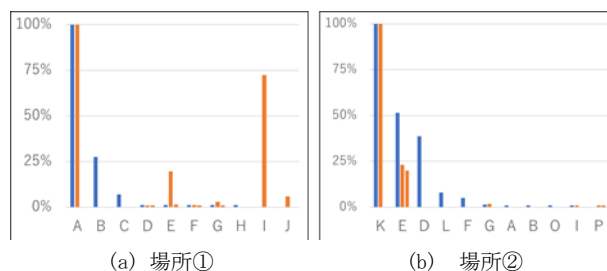


図2. 場所①、②への滞在率
(青色：2.4での修正あり、橙色：修正なし)

以上の結果より、滞在率の偏りを手掛かりとすることで、追跡結果を正しく修正できる可能性が示唆される。

ただし、本手法でも正しい滞在率が得られなかった場所も存在した。例えば場所③では、人物位置がカメラから遠く、得られる人物領域が小さい上、人物が常時カメラに背を向ける姿勢をとるため、人物検出結果に頻繁に検出漏れが生じ、その結果、2.1で得られるべき近接セグメント間の間隔が広がりすぎて互いに連結候補セグメントの候補に含まれない近接セグメントが増えてしまったことが、その原因として考えられる。

4 まとめ

日常生活空間での人-場所間の関係性の把握のために、対象空間のカメラ映像から各場所への各人物の滞在率の把握を試みた。このための人物追跡において、各場所への各人物の滞在率の偏りを利用して追跡結果を修正しつつ滞在率を推定した。実際のオフィス空間の観測映像を用いた実験結果から、上のような追跡結果の修正により、得られる滞在率が改善することを確認した。一方、人物検出に漏れが多い場所では、十分な改善が見られなかったため、今後は、これに対応可能な手法の改良について検討する必要がある。

参考文献

- [1] A. Bewley, et al.: "Simple online and realtime tracking," Int. Conf. Image Processing (ICIP) (2016).
- [2] N. Wojke, et al.: "Simple online and realtime tracking with a deep association metric," Int. Conf. Image Processing (ICIP) (2017).
- [3] G. Jocher, et al: YOLO v5 (2020). <https://github.com/ultralytics/yolov5>.