

RGB-D カメラを用いた棚からの物の出し入れ行動認識手法の評価

野村 拓矢[†] 岡 哲資[†] 鈴木 俊孝[‡][†]日本大学生産工学部数理情報工学科[‡]日本大学大学院生産工学研究科数理情報工学専攻

1 はじめに

行動認識の研究は古くから行われている¹⁾。近年、RGB-Dカメラの性能向上によって、色情報と同時に、高い精度の距離情報を取得できるようになった。これを用いて、人間の行動を認識し、生活に役立てる研究が行われている。例えば、摂食行動を認識し、ログをとることで健康管理に役立てる研究²⁾や、調理のかき混ぜ行動を認識し調理補助に役立てる研究³⁾がある。

本研究では、棚からの物の出し入れ行動の認識に取り組んだ。まず、RGB-Dカメラを用いて棚からの物の出し入れ行動を認識するシステムを開発した。次に、開発したシステムの評価を行い、実用性、誤認識の原因、改善点を検討した。店舗などでこの行動を認識し、人物を追跡することができれば、マーケティングや万引きの防止に役立てることができる。

2 行動認識システム

行動認識システムは真下を見下ろす形で RGB-D カメラを天井に設置し、視野角内に入ってきた人を追跡する。さらに、追跡している人が「棚から物を取り出す・戻す」行動を取った際に、画面に「took」もしくは「returned」と表示する。また、システムは追跡部とイベント検出部に分かれている。イベント検出部は棚接近検出部、棚交差検出部、物体把持検出部、行動判定部に分かれている。

追跡部では人領域の抽出と追跡を行う。初めに、150フレームの間、フレーム $t-1$ の距離画像の欠損ピクセルをフレーム t で取得できていた場合に上書きし、背景画像 $B(i,j)$ を作成する。以降、毎フレーム取得する距離画像 $I(i,j)$ の欠損部分は $B(i,j)$ を用いて補間する。次に、 $I(i,j)$ を式(2.1)によって二値化し、前景画像 $F(i,j)$ を作成する。

$$F(i,j) = \begin{cases} 255, & (\text{if } I(i,j) \geq 1000 \text{ and } I(i,j) \leq 2000 \text{ and } I(i,j) - B(i,j) \geq 50) \\ 0, & (\text{otherwise}) \end{cases} \quad (2.1)$$

ここで、255 が人領域であり、0 はそれ以外である。さらに、人領域の分裂を防ぐため $F(i,j)$ の全てのピクセルに対してその上下左右それぞれ5ピクセル

離れた点の値を参照し、全て255なら $F(i,j)$ を255とすることで連結処理を行う。最後に、 $F(i,j)$ に対してラベリングを施し、取得した人領域の重心を用いて追跡を行う。ここで、人領域は1人の人であると仮定し接触した複数の人ではないものとする。

イベント検出部では、追跡している人の行動を認識する。まず棚接近検出を行い、棚周辺に人が存在した場合、棚交差判定を行う。検出は、 $F(i,j)$ の棚周辺ピクセル、棚領域ピクセルを走査することで行う。交差していなければ1フレーム前に棚と交差していたかどうかを判定し、していれば行動判定、していなければ物体把持判定を行う。

物体把持検出部では、人の物体把持を検出する。初めに RGB-D カメラから取得した色画像を距離画像サイズ(512×424)にリサイズする。その後、リサイズした色画像の値を RGB から HSV に変換し、 $F(i,j)$ の人領域と対応する点以外の値を0にした人領域 HSV 画像 $H(i,j)$ を作成する。次に、棚に最も近い人領域ピクセルから上下左右15ピクセル以内を手先領域とし、その中の人領域ピクセル数を求める。また、 $H(i,j)$ の手先領域内の人領域に対して肌色検出を行い、肌色ピクセル数を求める。求めた人領域ピクセル数と肌色ピクセル数から、手先領域における人領域の肌の割合(肌色率)を計算し、閾値45.0%以上は0(非把持)、未満は1(把持)とし仮把持状態を求め、仮把持状態リングバッファ $r(n)$ ($n=0\sim4$) に格納する。最後に、 $r(n)$ に格納された過去の仮把持状態のうち0が半数以上なら0、そうでなければ1として現在の把持状態を決定する。

行動判定部では、追跡している人の行動を判定しシステム画面に表示する。行動の判定には、棚に触れる直前の把持状態と、棚に触れ、棚から離れた直後の把持状態の組み合わせを用いる。棚に触れる前に持っていて、離れた時に持っていなかったら戻した、その逆なら取り出した、と判定される。

3 実験方法

行動認識システムを評価するために RGB-D カメラ視野角内で棚接近検出実験、棚交差検出実験、物体把持検出実験、行動判定実験を行った。物体把持検出実験は21歳の男性1名が行い、その他の実験は18歳から23歳までの男女それぞれ12名が行った。

物体把持検出実験では、物を持ち、手首を0度から180度まで回転させる。これを10種類のものについて行い、45度間隔で物体把持検出状況を確認

Evaluation of an RGB-D Camera-based Method for Recognition of Taking Objects off a Shelf and Returning Them

[†]Takuya Nomura [†]Tetsushi Oka [‡]Toshitaka Suzuki

[†]Mathematical Information Engineering, College of Industrial Technology, Nihon University

[‡]Mathematical Information Engineering, Graduate School of Industrial Technology, Nihon University

する。また、物を持たない状態と、比較的肌の色に近い 12 種類の色の折り紙でラッピングした箱 (W126×D126×H59mm)でも同様の実験を行った。使用した物は書籍(大), 書籍(小), 箱, 350ml 缶, 500ml ペットボトル, リンゴ, ペン, マグカップ, 発泡スチロール容器, スナック菓子である。使用した色は赤, 橙, 黄橙, 山吹, 黄, クリーム, 牡丹, ピンク, 薄ピンク, ペールオレンジ, 茶, 黄土である。

棚接近・棚交差・行動判定実験で行ったタスクの流れを説明する。初めに、棚周辺領域に向かって 10 回手を伸ばす。次に棚に向かって 10 回手を伸ばす。次に棚から物を取り出す・戻す行動を 10 回行う。次に物を持った状態で棚に手を伸ばし、持ったまま手を戻す行動を 10 回行う。最後に物を持たない状態で棚に手を伸ばし、そのまま手を戻す行動を 10 回行う。

4 実験結果

棚接近・棚交差検出実験の検出率は男女ともに 100%であった。また、全行動の推定認識率は 88.02%, その 95%信頼区間は 85.94~90.06%であった。行動判定実験において実際に行った行動とその判定結果を表 1 に示す。ここで、行動①は取り出す行動、②は戻す行動、③は物を持った状態で棚に手を伸ばし持ったまま手を戻す行動、④は物を持たない状態で棚に手を伸ばしそのまま手を戻す行動である。物体別物体把持検出実験の結果、350ml 缶, 500ml ペットボトル, リンゴが 0 度の時正しく把持検出されず、非把持となった。また、ペンは 0 度, 45 度の時非把持となった。逆に、素手は 90 度, 135 度, 180 度の時把持になった。色別物体把持検出実験の結果、ペールオレンジは 45 度以外すべて非把持となった。また、茶色は 90 度の時非把持となった。それ以外の物体、色はすべて正しく検出された。

表 1 各行動と判定結果

判定結果	実際の行動			
	行動①	行動②	行動③	行動④
行動①	216	0	0	4
行動②	0	206	0	51
行動③	24	33	240	2
行動④	0	1	0	183

5 考察

本研究で開発した行動認識システムは、18 歳から 23 歳までの男女において、物体が手に隠れておらずペールオレンジ色以外の場合に、マーケティングにおいて活用することができる。しかし、防犯タグの代替として活用することは難しい。第一に、棚接近・棚交差検出率が 100%であった。第二に、物

体把持検出実験において物体の種類にかかわらず、角度が 0 度, 45 度以外の場合(物体が手に隠れている)に物体把持を認識できていた。第三に、物体把持検出実験においてペールオレンジ色の場合角度にかかわらず把持を検出することができなかった。第四に、認識率が 88.02%, 95%信頼区間が 85.94~90.06(%)であり、誤認識が存在している。

行動認識システムにおいて発生する誤認識の原因は肌色検出の正確度の低さ、深度画像に対する色画像取得タイミングの遅れである。第一に、物体把持検出部では肌色率が下がると物体を把持していることになるが、物体把持検出実験において素手での 90 度, 135 度, 180 度の把持状態が把持になっていた。第二に、行動判定実験において起こった誤認識 115 回中 114 回は物を持っていないのに持っているとして認識してしまうことに起因する誤認識であった。第三に、行動判定実験において常に物体を把持している行動③では誤認識は起こらなかった。第四に、行動判定実験において行動④の誤認識 57 回中 51 回が行動②と認識してしまう誤認識であり、肌色率の低下が棚に手を伸ばした際によく見られた。人領域は深度画像をもとに作成しているため、深度画像に対して色画像の取得タイミングが遅れていると仮定すると、棚に手を伸ばす際に人領域に人以外の背景が映り込み、肌色率の低下を招いていると考える。

行動認識システムの改善点は肌色検出手法と、深度画像・色画像のずれ、手先領域のサイズの三点である。肌色検出は現在 HSV 色空間での閾値処理のみを用いているため、より光や影の影響を受けにくい手法を導入する必要がある。深度画像・色画像のずれは、それぞれの画像の取得タイミングを計測し、各画像の遅れ分深度画像の取得タイミングを遅らせる必要がある。また、現在の手先領域のサイズでは、指先だけしか映っておらず、深度画像・色画像のずれの影響を受けやすいため、人の身長や RGB-D カメラから手までの距離に応じた適切な領域サイズを動的に決定する必要がある。

参考文献

- 1) 杉本千佳 ほか, "ウェアラブルセンサによる行動認識システムの開発", 日本時計学会報告誌. マイクロメカトロニクス 51(197), pp. 1-12(2007).
- 2) 佐藤琢磨 ほか, "機械学習を用いた摂食行動認識手法の実現と食画像ラベリング環境の構築", 情報処理学会研究報告. UBI, [ユビキタスコンピューティングシステム] 2015-UBI-46(12), pp. 1-8(2015).
- 3) 宮澤飛鳥 ほか, "調理者の手と容器の位置関係を利用した「かき混ぜる」行動の認識", 電子情報通信学会技術研究報告. DE, データ工学 112(75), pp. 25-30(2012).