

人間の動作との関係性を用いた物体認識に関する基礎検討

石井雅樹

秋田県立大学システム科学技術学部情報工学科

人間は物体を認識する際、その物体の使われ方、すなわち、人と物体とのインタラクション（相互関係）を特徴として用いていると考える。本研究では人と物体との基本的なインタラクションを検出し、物体に対する人の動作を物体の属性情報として抽出し、その属性情報を物体認識の特徴として利用する手法の確立を目的としている。本稿では RGB-D センサーを用いて抽出した物体の形状と人間の動作を、物体の属性として付与し、物体認識を行う手法を提案する。本稿では基礎検討として実験環境を固定し、対象物体を椅子に限定した状態で RGB-D センサーを用いて物体の形状および人の動作を計測した。形状認識に係る基礎実験ではクラスタリングと RANSAC を併用することで平面認識が可能であることを確認した。動作認識では「座る」と「立つ」を対象動作とし、機械学習により識別器を作成した。環境認識では人間が立つ動作を行った時、つま先が触れている平面を床面と定義した。また、人間が座る動作を行った時、臀部の触れている平面かつ床面と平行な平面を椅子の座面と定義した。その結果、椅子の認識処理では、物体に対して本来の使用用途を満たす動作が行われたとき、正しい認識が行われることを確認した。

キーワード：物体認識，動作認識，相互関係，RGB-D センサー

近年、人間の生活環境で共存することを目的としたロボットの開発が進んでいる。ロボットが人間と生活を共にするためには様々な変化に柔軟に対応し、作業を行うことが求められる。変化に柔軟に対応するための一機能として物体認識が挙げられる。従来の物体認識の手法では、入力画像に対して SIFT や SURF などを用いて特徴を抽出し、機械学習を行うことにより入力画像を識別している。また、近年では DeepLearning を用いた物体認識が行われている。DeepLearning では自動的に適切な特徴が抽出されるため、従来の物体認識よりも精度が高いことが特徴である。しかし、どちらの手法も汎化性を高めるために大量の学習データが必要となる。

人間は物体を認識する際、その物体の使われ方、すなわち、人と物体とのインタラクション（相互関係）を特徴として用いていると考える。本研究では人と物体との基本的なインタラクションを検出し、物体に対する人の動作を物体の属性情報として抽出

し、その属性情報を物体認識の特徴として利用する手法の確立を目的としている。本稿では RGB-D センサーを用いて抽出した物体の形状と人間の動作を、物体の属性として付与し、物体認識を行う手法を提案する。

提案手法

人と物体とのインタラクションとして本研究ではアフォーダンスの概念を用いた。アフォーダンスとは物体が人に情報を与え、その情報から人のどのような行動が誘発されるかといった考え方である。近年ではアフォーダンスの概念を物体認識に応用する研究が行われている（秋月, 2016 ; Koppula, 2016）。物体が人に与える情報とは、その物体の本来の使用方法であり、物体の形状によって表現されていると考える。この物体の形状を本研究では静的属性として定義した。また、その形状から実際に人が物体に

対して行った動作を動的属性として定義した。本研究では、この二つの属性をもとに物体認識を行った。なお、本稿では基礎検討として対象物体を椅子と床に限定し、実験は椅子を含む環境に限定して実施した。なお、RGB-D センサーは Kinect for Windows v2(Microsoft 社)を用いた。

形状・動作認識

本研究する椅子の認識処理の流れを図 1 に示す。

本研究では、静的属性として物体の形状に着目した。例えば、椅子であれば平面と平面が垂直に交差していることなどが挙げられる。すなわち、物体は単純な形状部分を有していることから、それらの特徴として用いることが有効と考えた。平面認識処理では、はじめに Kinect を用いて環境中の 3 次元点群情報を取得した。次に取得した点群をクラスタリングにより領域分割し、最後に領域分割した各クラスターで RANSAC (Random Sample Consensus) (Fischler, 1981) による平面検出を行い、各クラスターで平面が存在するかを判定した。

次に、本研究では、動的属性として人の物体に対する動作に着目した。ある物体に対して誘発される人の物体に対する動作は物体の形状から限定されると考えられるため、その動作を特徴として用いることが有効と考えた。本研究では認識する動作として「座る」と「立つ」の二つの動作を対象とした。動作認識のための識別器は VisualGestureBuilder を用い、学習アルゴリズムには AdaBoost (Freund, 1997) を選択した。動作の学習には 8 人分の動画像を用いた。

最後に、Kinect で検出した人の関節の座標と平面の接触判定を行い、物体に対する人の係わりの検出を行った。関節座標と平面の接触判定には、八分木表現による半径近傍探索を用いた。八分木表現とは立体を小さな立方体の集合で表現する手法であり、ボクセル表現のデータ量が大きくなる欠点を克服するために考案された表現法である (前川, 2001)。平面と座標間の距離を用いた場合では垂直距離のみの判定になるため、座る位置によっては判定できない。半径近傍であれば平面と腰の座標の角度に依存しないため、座る位置が極端に平面の角でも判定可能で

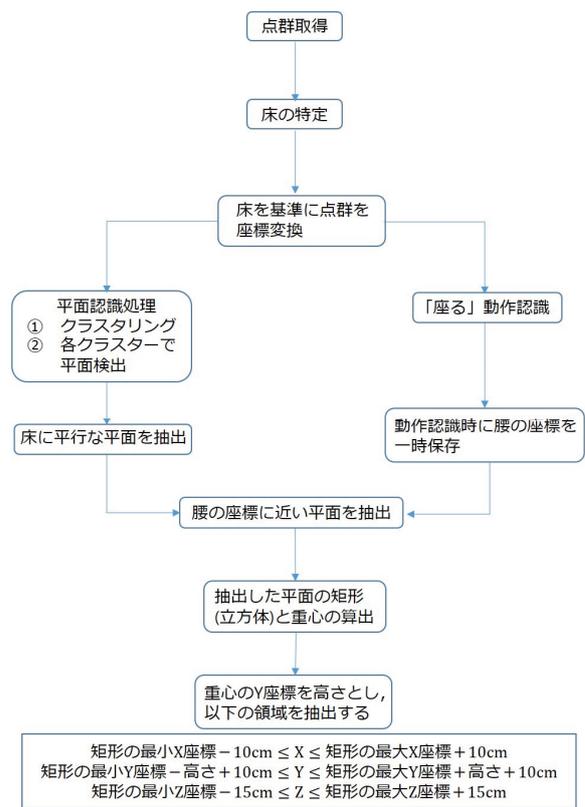


図 1 椅子認識処理手順

ある。

物体認識

前述した手法を用いて環境認識を行った結果を以下に示す。

床認識

「立つ」動作が行われた際、左足、右足のどちらかのつま先が接触している面を床として認識し、タグ付けを行った。立つ動作が複数回確認されると、つま先付近の床の点群に対して床の信頼度が上昇し、歩行可能な床であることが確定される。実験結果を図 2 に示す。同図では、立つ動作が検出された際の人領域を緑色、床とタグ付けされた箇所で歩行可能な床を青色で示している。

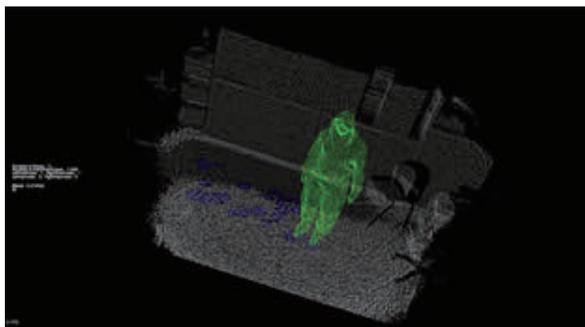


図2 起立時の動作認識

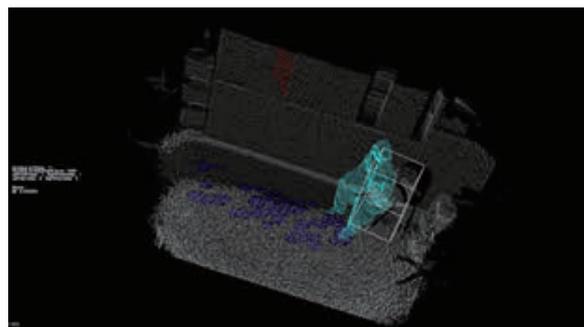


図4 着席時の動作認識

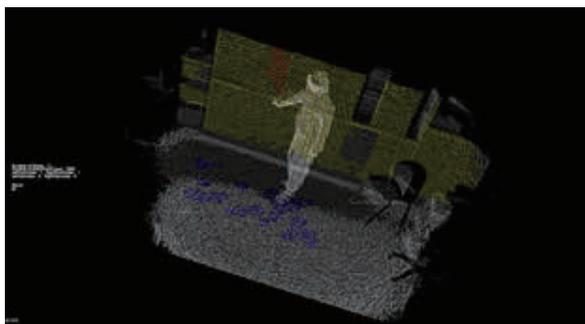


図3 手で触れた場所の認識



図5 椅子認識

手で触れた面・箇所の認識

手の関節座標が平面の一部に接触した場合、その平面全体を手で触れられた平面と認識しタグ付けを行った。また、手の関節座標から半径 70mm の領域を取得し、領域内に背景点群が存在する場合、その点群を手で直接触れた部分として認識しタグ付けを行った。実験結果を図 3 に示す。同図では、触れたと認識された平面を黄色、触れた箇所を赤色で示している。

椅子認識

椅子の認識実験の結果を図 4 及び図 5 に示す。「座る」動作を行い、判別器が動作を認識した際、左右どちらかの臀部の関節座標と接触している平面をカウントし、動作を繰り返す中で多くカウントされた平面を座面としてタグ付けを行った。

図 4 は、座る動作が確認できたとき、臀部に触れている平面を認識している様子を示している。なお、同図では座る動作が認識された人の領域を水色で示している。図 5 は、人が椅子から立ち上がり、少し離れた面に移動した際、上記で認識した座面を含む

物体を椅子と認識している様子を示している（直方体領域）。

結言

本研究では人と物体との基本的な関係性を検出し、人と物体との相互関係を物体の属性情報として抽出し、その属性情報を物体認識の特徴として利用する手法について基礎検討を実施した。具体的には、認識対象を椅子、床などの環境に限定し、提案手法の有用性を検証した。

提案手法ではアフォーダンスの概念を用い、物体の形状から人がどのような動作が誘発されるかという形状と動作の関係性を物体認識に用いた。

その結果、平面認識では RANSAC とクラスタリングを併用することで平面認識が可能という結果が得られた。床認識では人間の立つ動作を検出し、その際のつま先の座標を用いることで認識を可能とした。椅子認識では人間が着席した座面で床と平行であれば椅子として認識することが可能であった。

また、人間の動作を座る、立つに限定し、物体に

対して人が誘発されて行う動作を動的属性として定義し、動作認識により属性情報の抽出を行った。座る動作及び立つ動作と平面認識により椅子と床の認識が可能であったため、物体が本来有する静的属性の他に動的属性を用いる従来手法の有効性が確認された。

今後は、対象物体及び対象動作を増やし、提案手法の有用性を検証する予定である。

本研究は、科学研究費(C)(No.22K12078)の助成により行われたことを付記する。

文献

秋月秀一, 飯塚正樹, 橋本学 (2016): 「アフォーダンスに着目した一般物体認識のための特徴量」『第 21 回知能メカトロニクスワークショップ講演論文集』, 94-96.

Koppula, H., & Saxena, A. (2016). Anticipating Human Activities using Object Affordances for Reactive Robotic Response. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 38(1), 14-29.

Fischler, M. A., & Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM*, 24(6), 381-395.

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of Online Learning and an Application to Boosting, *Journal of Computer and System Sciences*, 55(1), 119-139.

前川佳徳 (2001): 「コンピュータグラフィックス (新世代工学シリーズ)」, オーム社

〔 令和 5 年 7 月 11 日受付
平成 5 年 8 月 18 日受理 〕

Object Recognition Using Relationship with Human Movement

Masaki Ishii

Department of Information and Computer Science, Faculty of Systems Science and Technology, Akita Prefectural University

When recognizing an object, humans tend to think of the ways the object is used, that is, the interaction (mutual relationship) between the person and the object is a feature of the recognition. This study examined the basic interaction between humans and objects, extracting their mutual relationships (human movement with respect to the object) at the time in which the interaction occurs as attribute information to the object, and establishing a method of using attribute information as a feature of object recognition. In this study, a method is proposed for the recognition of an object, adding the shape of the object extracted with an RGB-D sensor and taking human movement as an attributes of the object. The experimental environment was fixed, and the shape of the object and the person's movement were measured using an RGB-D sensor, with the target object limited to the chair. The result of this experiment in relation to shape recognition confirmed that plane recognition is possible through the use of clustering and RANSAC. In motion recognition, sitting and standing were targeted, and a classifier was created that used machine learning. In environmental recognition, the surface of the floor is defined as a plane touched by the toes of a human in a standing motion. In addition, the seating surface of the chair is defined as a plane parallel to the floor that is touched by the buttocks of a human in a seated position. This confirmed that in the chair recognition process, correct recognition is performed when an action that satisfies the original intended use is performed on the object.

Keywords: Object recognition, Motion recognition, Interaction, RGB-D sensor