Short Report

人間とのインタラクションに着目した物体認識に関する基礎検討

石井雅樹 [1]，石島樹 [2]

[1] 秋田県立大学システム科学技術学部電子情報システム学科

[2] 秋田県立大学大学院システム科学技術研究科機械知能システム学専攻

人間は物体を認識する際，その物体の使われ方，すなわち人と物体のインタラクション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相互関係(物体に対する人間の動作)を物体の属性特徴として抽出し，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する手法を提案する．本研究では人と物体とのインタラクションとして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．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，物体の形状に表現されている．この形状を本研究では静的属性として定義した．また，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．提案手法ではこの二つの属性をもとに物体認識を行った．本稿では対象物体を“椅子”と“飲み物”，対象動作を“座る”と“飲む”に限定して基礎実験を行った．その結果，どちらの処理も物体に対して本来の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, インタラクション，アフォーダンス，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．従来の物体認識手法では，入力画像に対して SIFT (Scale-Invariant Feature Transform) や SURF (Speed-Upped Robust Feature)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．しかし，主に人間が活動する環境はロボットのために整備されていることが少ないため，物体同士が重なるオクルージョンが発生することが多い．また，照明環境も一定とは限らず，見え方による幾何学的特徴も変化する．したがって，画像を用いた物体認識ではこれらのような外的要因に大きく依存することが問題である．一方，近年では Deep Learning を用いた物体認識が行われている．Deep Learning 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（Krizhevsky, 2012: Girshick, 2014）．しかし，汎化性を高めるためには大量の学習データが必要となる．

人間は物体認識を行う場合，視覚から多くの情報を獲得し，その情報の関連性や自身の経験から総合的に物体を認識している．上述の関連性とは対象物体と自身との間のインタラクション（相互関係）を意味している．インタラクションとは人間が物体に対して行った行動によって特定の効果が生じるように対話的に成り立つ作用のことである．例えば鉛筆は机上では棒である．しかし，人間が把持し，先端を紙に接したまま動かすことにより「書く」という効果が生じ，この時点で棒は鉛筆と認識される．このように人間が物体を認識する場合は，対象物体の使われ方，環境中での存在のしかたなど，対象物体と他者，周囲の物体，環境とのインタラクションが

責任著者連絡先：石井雅樹　〒015-0055 由利本荘市土谷字海老ノ口 84-4　公立大学法人秋田県立大学システム科学技術学部電子情報システム学科．E-mail: ishii@akita-pu.ac.jp

重要な情報となる．

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，インタラクションが発生した際の相互関係(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する手法を提案する．

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．アフォーダンスにおける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．また，物体は単純な形状が組み合わさって成り立っていると考える．例えば，椅子であれば座面と背面が平面であり，垂直に組み合わさっていること，ペットボトル，マグカップなどは液体を蓄える円柱形状を有することなどが挙げられる．近年ではこの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）．本研究ではこの物体の形状を特徴として利用するため，これを静的属性として定義した．また，物体に対して人が行う動作はある程度限定されていると考えた．したがって，ある形状を有する物体に対して実際に「人が物体に対して行った動作」を特徴として利用するため，これを動的属性として定義した．

提案手法の具体的な手順としては，はじめに物体が機能するための形状を認識し，静的特徴を抽出する．次にその形状から誘発され，実際に人が行った動作を認識し，動的特徴を抽出する．最終的にこの二つ属性をもとに物体認識を行う．なお，本研究では基礎検討として認識対象物体を椅子と飲み物に限定した．また，RGB-D センサーとして Kinect for Windows v2（Microsoft 社）を用いた．

形状認識

本研究では，静的属性として物体の形状に着目した．椅子については人が座るための平面形状，飲み物については液体を蓄えるための円柱形状を対象とし，以下の手順で静的属性の抽出を行った．

（1）平面認識
- ① 3 次元点群取得
- ② クラスタリング
- ③ 各クラスターで RANSAC による平面認識

（2）円柱認識
- ① 点群を XY，YZ，ZX 平面に平行投影
- ② ZX 投影点群に対してノイズ除去
- ③ ZX 投影点群から凸包座標算出
- ④ 凸包座標から最小二乗法により円の半径を算出
- ⑤ 半径の閾値判定による円柱認識

平面認識では，はじめに Kinect を用いて環境中の 3 次元点群情報を取得する．次に取得した点群をクラスタリングにより領域分割する．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，各クラスターに平面が存在するか判定する．

円柱形状の認識では，はじめに Kinect から円柱物体の 3 次元点群情報を取得する．次に円柱が平面に対して垂直に置かれていると仮定し，取得した 3 次元点群を ZX 平面に平行投影する．投影することにより円柱を真上から俯瞰した点群情報が得られる．次にノイズ除去を行い，残りの点群で凸包座標を算出する．最後に凸包座標を用いて最小二乗法による円フィッティングを行い，円の半径を算出する．その後，円の半径により閾値判定による物体が円柱形状を有しているか判定する．

動作認識

動的属性として人の物体に対する動作に着目した．本研究では認識動作として「座る」と「飲む」の二つの動作を対象とした．なお，動作識別器の作成には機械学習アルゴリズムの AdaBoost（Freund, 1997）を使用した．

「座る」動作の学習には 8 人分の動画像を用いた．また，「飲む」動作は右手に限定し，5 人分の動画像を用いた．テストデータは両者とも 1 人分の動画像を使用した．

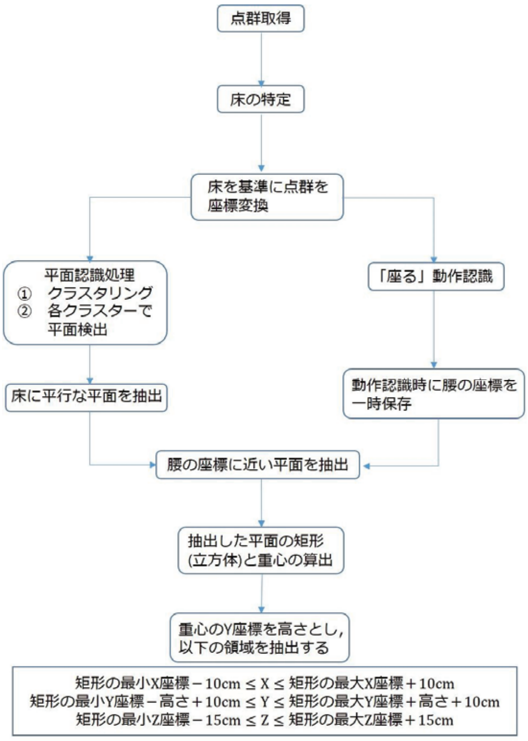

図 1　椅子認識処理手順

物体認識

形状認識による静的属性と動作認識による動的属性をもとに物体認識を行った．図 1 及び図 2 に椅子の認識処理, 飲み物の認識処理の手順を示す. また，認識対象を図 3 に示す．椅子認識実験では同図（a）に示す 8 種の椅子と 2 種の椅子以外の物体から成る計 10 種のデータセットを用いた. 飲み物認識実験では同図（b）に示す 10 種の飲み物と 2 種の飲み物以外の物体から成る計 12 種のデータセットを用いた．

椅子, 飲み物として認識された点群を図 4 に示す．椅子の認識ではデータセット内の 10 種すべてを椅子として認識した．座る動作中に対象物体を椅子と認識する過程の例を図 5 に示す．形状が異なる椅子に対しても環境中の床という情報を用いて人が座れる平面を特定でき，最終的に人が座ることで正しく椅子の認識が行われた．しかし，椅子以外の物体についても座る動作が可能な物体，すなわち，本来の用途と異なる場合に対して誤認識を認めた．

飲み物の認識では飲み物である 10 種のみを正しく認識した．また，飲み物以外の 2 種は飲み物として認識されないことを確認した．飲み物認識では人の骨格情報から手先の領域を抽出することで手によるオクルージョンを防いでいる．また，椅子の認識と同様に床の情報を用いて飲み物が置かれている平面を除去することで飲み物の点群のみを分割できる．したがって, 分割した点群に対して円柱認識を行い，最終的に人が飲む動作を行うことで飲み物の認識が

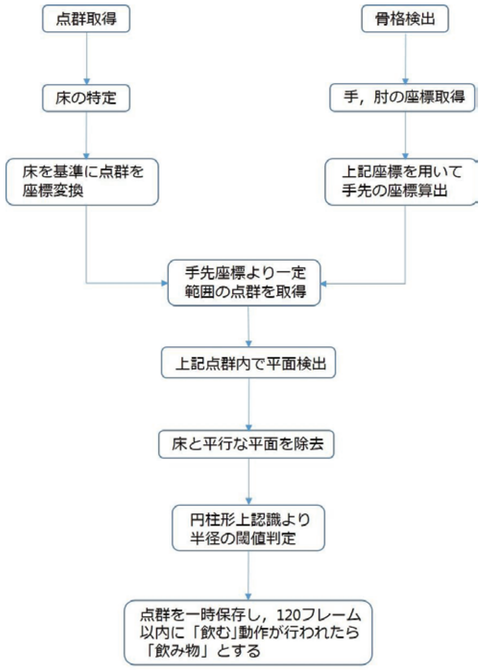

図 2　飲み物認識処理手順

(a)　椅子

(b)　飲み物

図 3　認識対象物体

(a)　椅子

(b)　飲み物

図 4　**認識結果**

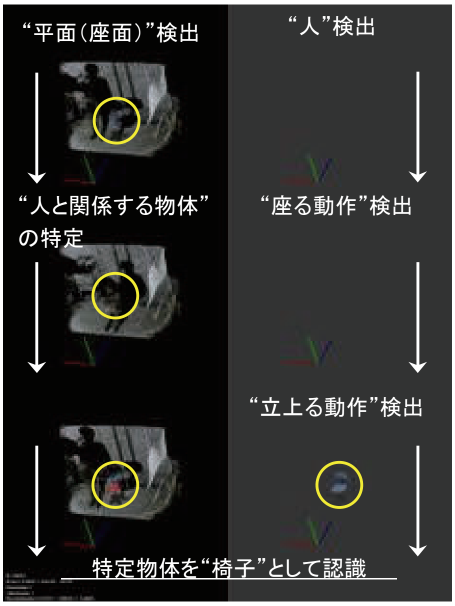

図 5　**椅子の認識の例**

正しく行われた．

結言

本研究では物体の形状部分を静的属性，人が物体に対して行った動作を動的属性として定義し，二つの属性を用いた物体認識を試みた．実際に物体の使用目的を満たす形状部分を認識し，それに対して人間が行った動作を物体に付与することで椅子と飲み物の認識が可能であることを確認した．今後は静的属性，動的属性に加え，物体そのものが有する色情報や法線情報を組み合わせた物体認識手法について検討する予定である．

文献

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 1106-1114.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proc. the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 580-587.

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.

平成 29 年 6 月 30 日受付
平成 29 年 7 月 11 日受理

Basic Study of Object Recognition with a Focus on Human Interaction

Masaki Ishii[1], Tatsuki Ishijima[2]

[1] *Department of Electronics and Information Systems, Faculty of Systems Science and Technology, Akita Prefectural University*

[2] *Course of Machine and Intelligence Systems, Graduate School of Systems Science and Technology, Akita Prefectural University*

In recognizing an object, humans are thought to rely on how that object is used; that is, on the interaction between a person and that object. This study develops a method that extracts interrelationships between people and objects (actions humans take on an object) as the object's attribute characteristic and use that attribute information as a characteristic of object recognition. This paper proposes a method that represents the shape of an object and human movements—extracted with an RGB-D sensor—as attributes of the object. In this study, the concept of affordance was used to account for the interaction between people and objects. This concept examines what kind of human action is induced from the information that an object accords to humans. The information that an object gives to a person is how the object is conventionally used, which is expressed in its shape. This shape was defined as a static attribute in this study. The movement that a person actually takes with regard to an object was defined as a dynamic attribute. These two attributes were used as bases of object recognition in the proposed method. In this paper, a basic experiment was conducted by limiting the objects to "chair" and "beverage," and the actions to "sit" and "drink." Consequently, it was confirmed that when a movement that satisfies the conventional purpose of the object is taken, the object is correctly recognized in either of the processes.

Keywords: object recognition, action recognition, interaction, affordance, RGB-D sensor

Correspondence to Masaki Ishii, Department of Electronics and Information Systems, Faculty of Systems Science and Technology, Akita Prefectural University, 84-4 Ebinokuchi, Tsuchiya, Yurihonjo 015-0055, Japan. E-mail: ishii@akita-pu.ac.jp