6ZE-04

# 距離画像を用いた人物と物体の 3次元空間でのインタラクション検出

内村 太一 †1　小西 陸斗 †2　阿部 亨 †2,†3

†1 東北大学工学部電気情報物理工学科　†2 東北大学大学院情報科学研究科
†3 東北大学サイバーサイエンスセンター

## 1　はじめに

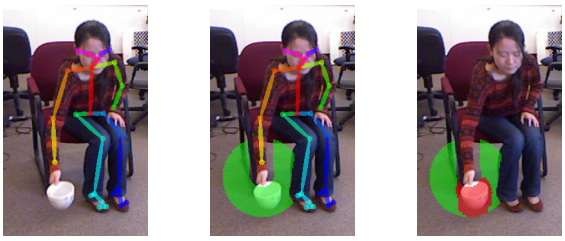
人物と物体とのインタラクションを検出する技術は防犯システムや人物の行動分析など様々な分野で利用されている [1]．特に人物が手で物体を動かす動作は人物が物体に接触していることを明確に示すため，その検出は重要である．

人物が物体を手で動かす動作については，映像から得られる 2 次元の動き情報（オプティカルフロー）に基づき，人物の前腕と手の周辺で観測される動きの関連を判定し検出を行う手法が提案されている [2,3]．この手法は，「物体領域の抽出」や「人物と物体との対応付け」が不要という利点があるものの，オプティカルフローではカメラ方向の（画像面に垂直な）動きを十分表現できず，手が動く方向によっては動作を正しく検出できない場合があった．

これに対し，本稿では，映像情報（RGB 画像）と距離情報（距離画像）から求めた 3 次元の動き情報（シーンフロー）に基づき，人物の前腕と手の周辺の動きの関連を判定することで，人物が物体を手で動かす動作をより安定に検出することが可能な手法を提案する．

## 2　関連研究

映像から得られるオプティカルフローに基づき，人物が物体を手で動かす動作を検出する既存手法 [2,3] では，まず，図 1(a) に示すように，入力画像から人物の骨格の抽出を行い，抽出された骨格に基づき，図 1(b) 中の緑の領域で示すように，手を中心とした周辺領域を設定する．次に，映像中のオプティカルフローから推定された前腕の動きで人物が物体を動かしている場合に物体上で生じるはずのオプティカルフローの予測を行う．実際に観測されたオプティカルフローと予測されたオプティカルフローを比較し，図 1(c) 中の赤の領域で示すように，観測値が予測値に類似している箇所が周辺領域内に生じているならば，人物が物体を手で動かしていると判定する．この判定は，[2] の手法では，ヒューリスティックな閾値処理により，[3] の手法では，SVM を用いた機械学習により行っている．



(a) 骨格の抽出　(b) 手の周辺領域　(c) 動きの類似性

図 1: 物体を手で動かす動作を検出する既存手法

これらの手法では，入力画像から人物と物体を抽出し，人物と物体との対応を認識する処理が不要であり，未知の物体や一部が遮蔽され認識が難しい物体の場合でも，物体を手で動かす動作の検出が可能となる．しかし，人物の前腕の動きや周辺領域内の動きがオプティカルフローにより求められているため，カメラ方向（画像面に垂直な方向）に物体を手で動かす場合，その動きを十分捉えることができず，物体を手で動かす動作の正しい検出が困難となるという問題がある．

## 3　提案手法

前章で述べた既存手法での課題を踏まえ，本稿では手の周辺領域の状態に基づいたインタラクションの検出方法をもとに，領域の決定や動作，インタラクションの有無の判定において，時系列の RGB 画像と距離画像を組み合わせた 3 次元位置情報，またその情報に基づいた動き情報を活用するインタラクション検出手法を提案する．この手法では映像中の 2 次元の動きに加えて 3 次元空間内の動き（カメラ方向の動き）も捉えることができるため，人物のカメラ方向の動き（前後の動きなど）も考慮することが可能となり，人物と物体の間のインタラクションについて効果的に分析できると考えられる．

以下に提案手法の流れを説明する．

### 3.1　シーンフローの推定

まず，RGB 画像の各画素について，対応する 3 次元位置を距離画像の深度情報から求める．次に，連続する 2 枚の RGB 画像から映像中の 2 次元の動きであるオプティカルフローを求める．RGB 画像の各画素で得られたオプティカルフローを対応する 3 次元位置に投影することで，3 次元空間中の動きであるシーンフローを求める．

---

Interaction detection between a person and an object in 3D space using range images
Taichi UCHIMURA†1, Rikuto KONISHI†2, and Toru ABE†2,†3
†1Department of Electrical, Information and Physics Engineering, Tohoku University
†2Graduate School of Information Sciences, Tohoku University
†3Cyberscience Center, Tohoku University

### 3.2 人物の骨格抽出と前腕の動きの判定

各 RGB 画像に対し，OpenPose [4] 等の手法を適用することで，画像中の人物の骨格（キーポイントの集合）を抽出する．抽出された肘と手首のキーポイントの位置を 3 次元空間の対応する位置へ投影し，投影された線を結ぶことで，3 次元空間中での前腕の骨格を決定する．

前腕領域として，前腕骨格の長さに応じた幅の領域を前腕骨格に沿って設定する．設定された前腕領域内の箇所を一定の数だけサンプリングし，各箇所のシーンフローの成分を要素とする特徴ベクトルを求める．得られた特徴ベクトルの状態から，前腕が動いているか否かの判定を Support Vector Machine (SVM) 等を用いた機械学習により行う．

### 3.3 手の周辺領域の決定と領域内の分析

前腕が動いていると判定された場合，前腕骨格の長さに応じ，前腕骨格を手首方向に延長した箇所を中心とした球を手の周辺領域として設定する．設定された周辺領域内の箇所を一定の数だけサンプリングし，物体が手で動かされた場合に物体上で生じるはずのシーンフロー（予測値）と実際に観測されたシーンフロー（観測値）の差を各箇所で求める．各箇所で得られた予測値と観測値の差を要素とする特徴ベクトルの状態から，手が物体を動かしているか否かの判定を SVM 等を用いた機械学習で行う．

以上のように，提案手法では，手が物体を動かしているかをシーンフローに基づき判定するため，手が動く方向によらず，人物が物体を手で動かす動作を安定に検出することが可能になると期待できる．

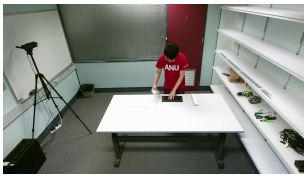
## 4 実装・実験

提案手法の一部を実装し，予備的な実験を行った．実験には，一般公開されている IKEA Assembly (IKEA ASM) データセットを用いた [5]．このデータセットは，人物が家具を組み立てている場面を撮影したものであり，同じ状況を RGB 画像（1920 × 1080 画素，25fps）と距離画像（512 × 424 画素，25fps）として撮影したものも含まれている．データセットに含まれる RGB 画像と距離画像の例を図 2(a) と (b) に各々示す．また，RGB 画像の各画素について，対応する 3 次元位置を距離画像の情報から求め，RGB 画像の画素値を 3 次元空間へ投影した例を図 2(c) に示す．

RGB 画像で求めたオプティカルフローと距離画像の情報から求めた 3 次元位置を用いてシーンフローを推定し，対応する 3 次元空間の箇所へ投影した例を図 3 に示す．図中，緑の矢印がシーンフローを表している．

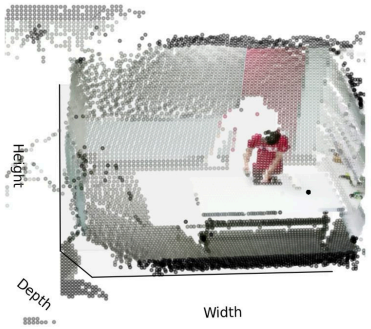
## 5 おわりに

本稿では，映像情報と距離情報から求めたシーンフローに基づき，人物の前腕と手の周辺で観測される動きの関連を判定することで，人物が物体を手で動かす動作を安定に検出する手法を提案した．今後は，提案手法の実装を進め，既存データセットおよび人物の行動を新たに撮影したデータセットを対象とした実験により，提案手法の有効性の検証を行う予定である．



(a) RGB 画像

(b) 距離画像




(c) RGB 画像の画素を 3 次元空間へ投影した例

図 2: IKEA ASM データセットの例

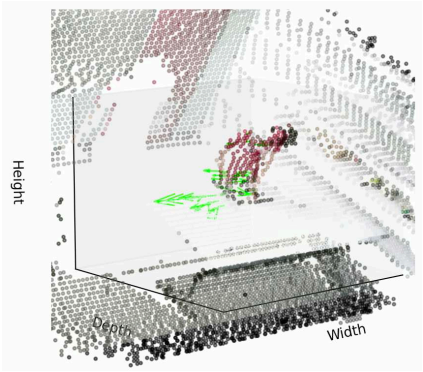



図 3: 推定されたシーンフローの例

## 参考文献

[1] Chao, Y.-W. et al.: Learning to Detect Human-Object Interactions, *Proc. IEEE Winter Conf Appl. Comput. Vision*, pp. 381–389 (2018).

[2] Tsukamoto, T. et al.: A method for detecting human-object interaction based on motion distribution around hand, *Proc. 15th Int Joint Conf. Comput. Vision, Imaging Comput. Graphics Theory Appl.*, pp. 462–469 (2020).

[3] Konishi, R. et al.: A method to detect human hands moving objects, *Proc. 12th Global Conf. Consum. Electron.*, pp. 166–167 (2023).

[4] Cao, Z. et al.: OpenPose: Realtime multi-person 2D pose estimation using part affinity fields, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 43, No. 1, pp. 172–186 (2021).

[5] Ben-Shabat, Y. et al.: The IKEA ASM Dataset: Understanding people assembling furniture through actions, objects and pose, *Proc. IEEE Winter Conf Appl. Comput. Vision*, pp. 846–858 (2021).